The Open World: Reference and Meaning on the Web

Harry Halpin

Doctor of Philosophy
Institute for Communicating and Collaborative Systems
School of Informatics
University of Edinburgh
2008
Abstract

This thesis examines the question: What does a URI mean? Or in other words, what does a URI identify and refer to? This question is of utmost importance for the creation and re-use of URIs on the Semantic Web. An philosophical and historical introduction to the Web explains the primary purpose of the Web as a universal information space for naming and accessing information. A terminology, based in widely-known philosophical distinctions, is employed to define precisely what is meant by information, language, representation, and digitality. These terms are then employed to create a foundational ontology and principles of Web architecture, which can be used to delimit the boundaries of the Web. From this perspective, the Semantic Web is then viewed as the application of the principles of Web architecture to knowledge representation. However, the classical philosophical problems of identity, reference, and meaning that traditionally caused problems for knowledge representation are inherited by the Semantic Web. Three main positions are inspected: the logicist position, as exemplified by formal semantics and a neo-Russellian descriptivist theory of names, the direct reference position, as exemplified by Putnam and Kripke’s causal theory of reference, and a neo-Wittgensteinian position that views the Semantic Web as yet another public language. After identifying the neo-Wittgensteinian viewpoint as the most promising, a solution of using people’s everyday use of search engines is proposed as a proper neo-Wittgensteinian way to determine the meaning of URIs via sense identification and disambiguation. This solution is then tested on a corpus of search terms from a real-world major search engine query log, and these terms are used to retrieve a corpus of Semantic Web URIs. Human browsing of hypertext web-pages is then used as a feature set in order to disambiguate and identify the “sense” of these URIs via a variety of machine-learning techniques. The results show that @@. Future work for the Semantic Web that follows from our argument is detailed, both of a technical and conceptual nature.
To be filled in when needed
Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Harry Halpin)
# Table of Contents

1 Introduction .......... 1
   1.1 Motivation .......... 2
   1.2 Hypothesis .......... 2
   1.3 Scope .......... 5
   1.4 Notational Conventions .......... 7
   1.5 Summary .......... 7

2 The Significance of the Web .......... 9
   2.1 The Origins of the Web .......... 11
   2.2 The Man-Machine Symbiosis Project .......... 12
   2.3 The Internet .......... 13
   2.4 The Modern World Wide Web .......... 16

3 Philosophical Prolegomenon .......... 19
   3.1 The Web and the World .......... 20
   3.2 Information, Encoding, and Content .......... 23
   3.3 Language, Interpretation, and Models .......... 29
   3.4 Purposes and Proper Functions .......... 36
   3.5 Representations .......... 39
   3.6 Digitality .......... 45
   3.7 The Extended Mind Thesis on the Web .......... 49

4 The Principles of Web Architecture .......... 53
   4.1 Foundational Terminology of the Web .......... 58
      4.1.1 Protocols .......... 58
      4.1.2 Uniform Resource Identifiers .......... 63
      4.1.3 Resources and Web Representations .......... 67
      4.1.4 Representational State Transfer .......... 74
## 4.2 The Principles of Web Architecture

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1</td>
<td>Principle of Universality</td>
<td>79</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Principle of Linking</td>
<td>86</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Principle of Self-Description</td>
<td>90</td>
</tr>
<tr>
<td>4.2.4</td>
<td>The Open World Principle</td>
<td>94</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Principle of Least Power</td>
<td>98</td>
</tr>
</tbody>
</table>

## 4.3 Conclusions

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>101</td>
</tr>
</tbody>
</table>

## 5 The Semantic Web

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>A Brief History of Knowledge Representation</td>
<td>104</td>
</tr>
<tr>
<td>5.2</td>
<td>The Resource Description Framework (RDF)</td>
<td>109</td>
</tr>
<tr>
<td>5.2.1</td>
<td>RDF and the Principle of Universality</td>
<td>109</td>
</tr>
<tr>
<td>5.2.2</td>
<td>RDF and the Principle of Linking</td>
<td>110</td>
</tr>
<tr>
<td>5.2.3</td>
<td>RDF and the Principle of Self-Description</td>
<td>113</td>
</tr>
<tr>
<td>5.2.4</td>
<td>RDF and the Open World Principle</td>
<td>115</td>
</tr>
<tr>
<td>5.2.5</td>
<td>RDF and the Principle of Least Power</td>
<td>118</td>
</tr>
</tbody>
</table>

## 5.3 The Semantic Web: Good-Old fashioned AI Redux?

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>123</td>
</tr>
</tbody>
</table>

## 6 The Identity Crisis

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>What Do URIs refer to?</td>
<td>127</td>
</tr>
<tr>
<td>6.2</td>
<td>The Logicist Position and the Descriptivist Theory of Reference</td>
<td>133</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Logical Positivism</td>
<td>133</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Tarki’s Formal Semantics</td>
<td>136</td>
</tr>
<tr>
<td>6.2.3</td>
<td>In Defense of Ambiguity</td>
<td>138</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Logicism Unbound on the Semantic Web</td>
<td>142</td>
</tr>
<tr>
<td>6.3</td>
<td>The Direct Reference Position and The Causal Theory of Reference</td>
<td>145</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Kripke’s Causal Theory of Proper Names</td>
<td>145</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Putnam’s Theory of Natural Kinds</td>
<td>146</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Direct Reference on the Web</td>
<td>147</td>
</tr>
<tr>
<td>6.3.4</td>
<td>The 303 Redirection Solution to the Identity Crisis</td>
<td>149</td>
</tr>
<tr>
<td>6.3.5</td>
<td>The Second-Generation Semantic Web</td>
<td>154</td>
</tr>
<tr>
<td>6.4</td>
<td>Critiques of the Causal and Descriptivist Theory of Reference</td>
<td>156</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Radical Translation and the Failure of the Direct Reference Position</td>
<td>156</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Radical Interpretation and the Failure of the Logicist Position</td>
<td>159</td>
</tr>
<tr>
<td>6.4.3</td>
<td>The Return of Frege’s Sense</td>
<td>164</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

To imagine a language means to imagine a form of life. Ludwig Wittgenstein (?)

The World Wide Web is without a doubt one of the most significant computational phenomena yet. Although it is impossible to tell what the future holds, it is obvious that the Web has already been massively adopted, for it is the Web more than even the word processor has made computing a necessity in everyday life. The sheer size of the Web makes empirical work difficult, although through careful sampling large amounts of progress has been made on characterizing the Web descriptively (?). Still, there are some questions that cannot be answered without having the entire Web available (?). Furthermore, there are some questions that cannot be answered without a theoretical understanding of the Web, in particular, the design of new Web standards. Although the Web is impressive as a practical success story, there has been little in the way of developing a theoretical framework to understand what, if anything, is different about the Web from the standpoint of long-standing questions of meaning and reference in philosophy. While this situation may have been tolerable so far, serving as no real barrier to the further growth of the Web, with the development of the Semantic Web, a next generation of the Web “in which information is given well-defined meaning, better enabling computers and people to work in cooperation,” these philosophical questions come to the forefront, and only a practical solution to them can help the Semantic Web repeat the success of the hypertext Web (?)

1
1.1 Motivation

Despite the hyperbole, there is little doubt that the Semantic Web faces gloomy prospects. At first inspection, the Semantic Web appears to be a close cousin both in spirit and in execution to another intellectual project, which we dub “classical artificial intelligence,” (also known as “Good-Old Fashioned AI”) an ambitious project whose progress has been relatively glacial and whose assumptions have been found to be cognitively questionable (?). The initial bet of the Semantic Web was that somehow the Web part of the Semantic Web would somehow overcome whatever problems the Semantic Web inherited from classical artificial intelligence (?). However, progress on the Semantic Web has also been relatively slow over the last near-decade, and it appears both new techniques and large amounts of data have not yet caused the Semantic Web to repeat the phenomenal success of the hypertext Web.

The first problem that is self-evident to anyone who actually attempts to deploy any ‘real-world’ data on the Semantic Web is that there is little guidance on how to name data using URIs, and what digital information to serve from these URIs (?). For long, this question was unanswered, and recently has only been cryptically answered (?). The second self-evident problem that occurs to anyone actually trying to do data integration on the Semantic Web is the different people create different URIs for the same thing. The essential bet of the Semantic Web is that decentralized agents will come to agreement on using the same URI to name a thing, including things that aren’t accessible on the Web, like people, places, and abstract concepts. Yet there is virtually no ability to even find URIs for things on the Semantic Web. Currently, each application creates its own URIs, and rarely are URIs re-used outside a particular application domain, repeating the localism of classical artificial intelligence. So many things either have no URIs or far too many. With the recent explosion of Semantic Web data on the Web given by the ‘Linked Data’ initiative, this overabundance of co-referential URIs have become even more pressing, and there is no principled way to determine when a new URI should be created (?).

1.2 Hypothesis

The scientific hypothesis of this thesis must be stated in a two-fold fashion, both to state the problem and then propose a solution. To state the problem, the Semantic Web is an extension of language, one that can defined by its conformance to the principles of
Web architecture, but nonetheless inherits the problems regarding reference and meaning from the philosophy of language. A humble solution can then be stated, that a **theory of reference and meaning suitable enough to encourage identifier re-usage on the Web can be implemented by employing relevance feedback from search engine results**. This hypothesis can be unpacked into three distinct constitutive questions, each one dependent on the previous one.

The first constitutive question is whether or not there is such a thing as the principles of Web architecture so that we can determine if there is such a thing as a language of the Web. Answering this question takes up the next four chapters of the dissertation. To preview, we phrase these five principles of Web architecture as:

- **The Principle of Universality**: Any resource can be identified by a URI (Uniform Resource Identifier, like http://www.example.org).

- **The Principle of Linking**: Any URI or Web representation (such as a web-page) can be linked to another resource identified by a URI.

- **The Principle of Self-Description**: The information an agent, like a web-browser, needs to have an interpretation of a resource should be accessible from its URI.

- **The Principle of the Open World**: The number of resources can always increase.

- **The Principle of Least Power**: A Web representation given by a resource can be described in the least powerful but adequate language.

Given the exegesis of Web architecture in Chapter 4, in Chapter 5 we propose that the Semantic Web, at least as embodied by the Resource Description Framework (RDF), is a URI-based language for data integration based on the principles of Web architecture as given above.

The second constitutive scientific question is, if the Semantic Web is a language, how can an agent tell what the meaning of a URI on the Semantic Web is? We address current theories in Chapter 6 and propose a neo-Wittgensteinian theory of meaning for the Web in Chapter ???. There are three distinct positions to this question, each corresponding to a distinct philosophical theory of reference. The first response is the **logicist position**, which states that the meaning of a URI is determined by whatever model(s) satisfy the formal semantics of the Semantic Web (?). This answer is identified
with both the formal semantics of the Semantic Web itself and the traditional Russellian theory of names and its descriptivist descendants (?). While this answer may be sufficient for machines, this answer is not sufficient for humans, as it often crucially under-determines what kinds of things the URI identifies. Furthermore, as the prevailing position in Semantic Web research, this position has born little fruit, for URIs given strictly logicist interpretations in general are not re-used. They do not in practice seem to have their meaning defined by their formal interpretation, as usually agents ignore these semantics. Another response the causal position for the Web, states that the meaning of a URI is whatever was intended by the owner. This answer is identified with the intuitive understanding of many of the original Web architects like Berners-Lee and a special case of Putnam’s ‘natural kind’ theory of meaning. This position is also a near relative to Kripke’s famous response to Russell (?). We also argue that this position is insufficient. First, despite its philosophical popularity, it is faulty in terms of philosophical argument, being undermined by Kripke himself. Second, from a sheerly practical standpoint there is no way to enforce an owner “determining” the meaning in a manner sufficient for communication. Lastly, at least as far as the Semantic Web is concerned, in practice this theory has not been a success, with efforts to embody this theory in projects failing to take off any moreso than the logicist position. Finally turn to the public language position, which states that since the Semantic Web is a form of language and as a language exists as a public form of meaning among multiple agents, then the meaning of language is the public use of the URIs, which must be established in a publicly-accessible mechanism that easily fits in the form of life of agents on the Web. This position is inspired by merging a neo-Wittgenstein understanding of language with a neo-Fregean concept of identification and the social constructivism of Searle (?). It outlines the requirements of a mechanism, a mechanism that must let agents create, find, and re-use URIs in a manner consistent with the principles of Web architecture. As vague as this position seems on first glance, we argue this analysis of meaning is the best fit to how natural language works, and it supersedes and to some extent, subsumes, the two other positions. As it works for natural language, it follows that it may be a good bet for the Semantic Web, for the Semantic Web is just a form of language, albeit an unusual one, according to this position.

The third constitutive question can then be phrased: What concrete mechanism can implement the Semantic Web as a public language game of sharing URIs? While it may be intuitively correct to endorse a neo-Wittgensteinian theory of meaning, this does little for the Web if a practical implementation of what terms like ‘language game’
and ‘form of life’ can be practically exemplified for the Semantic Web. First, one must remember that every “language game” comes with, as Wittgenstein would say, a “form of life” (?). Without a doubt, one activity that seems to be prevalent among users of the Web is searching for web-pages using natural language query terms via a search engine (?). Therefore, the obvious solution to the problem of finding out what a URI means is to take advantage of current search engines. Chapter ?? details on a high-level of abstraction a design for an implementation of determining URI meaning based on feedback from users using ordinary query terms in a hypertext-based search engines, putting the Semantic Web in a ‘virtuous cycle’ with the behavior of web users.

Our implementation is then tested with real data and real users. In Chapter ??, we describe a search engine query log from a major hypertext search engine (Microsoft Live.com), and how we derive query terms for people, places, and abstract concepts from this query log and then use those to derive a corpus of hypertext documents and Semantic Web URIs with their attendant data. Then in Chapter ?? we describe an experiment that uses human judges to disambiguate the sense of Semantic Web URIs as well as dispose of incorrect web-page results. In Chapter ??, we describe the results of using techniques from natural language processing and machine-learning to determine the meaning of Semantic Web URIs grounded in the behavior of real users. Finally, in Chapter ?? we discuss the advantages and limitations of our particular technique. The concluding Chapter ?? of the thesis then summarizes the work so far, demonstrates how we have in fact only created one component of the ‘language game” needed for the success of the Semantic Web. We present plans for future work as well as a further intriguing further philosophical questions that arise from the thesis.

Each of these chapters build upon each other to make the dissertation complete as a whole. Readers interested in particular subjects may wish to focus their attention on particular components, although they are warned that concepts and findings developed in earlier chapters are referred in later chapters. Due to nature of the project in an interdisciplinary and emergent area, there is no singular and comprehensive literature review in a separate chapter, but instead the literature is reviewed and mentioned as necessary throughout the dissertation.

1.3 Scope

This dissertation is explicitly limited in scope, concentrating only on terminology necessary to formulate, formalize, and demonstrate a limited number of principles of Web
architecture, in order to phrase a single, if broad, question: “How can we determine the meaning of a URI on the Semantic Web?” Although the dissertation is interdisciplinary, as it involves elements as diverse as the philosophy of language and machine-learning, these elements are only harnessed insofar as they are necessary to phrase our central hypothesis and present a possible solution.

Due to this constraint, this dissertation is not an attempt to develop a philosophy of computation (?), or a philosophy of information (?), or even a comprehensive “philosophy of the Web” (?). These are much larger projects outside the scope of a single dissertation, and even a single individual. These particular much larger and more important intellectual projects are also themselves still in the process of development. However, in combination with more fully-formed work in the philosophy of mind and language, we hope that at least this dissertation provides an starting point for future work in these areas. So we use notions from philosophy selectively, and then define the terms in lieu of our goal of articulating the principles of Web architecture and the Semantic Web, rather than attempting to articulate or define the terms of a truly systematic philosophy. Many of the philosophical terms in this dissertation could be explored much further, but are necessarily not explored, as to constrain the dissertation to a reasonable size. Unlike a more properly philosophical work, counter-arguments and arguments are generally not given for terminological definitions, but instead references are given to the key works that explicate these notions further.

Furthermore, this dissertation does not inspect every single possible answer to the question: What is the meaning of a URI?, but only three distinct positions. An inspection of every possible theory of meaning and reference is beyond the scope of the thesis, as is an inspection of the tremendous secondary literature that all of these theories have accrued over the years. Instead, we will focus only on theories of meaning that have been brought up explicitly in the various arguments over this question in the Web by the primary architects of the Web and the Semantic Web. Lastly, the theory of meaning that we feel best answers the question, a neo-Wittgensteinian theory of meaning that converges with much neo-Fregean work, is one of the most infamously dense yet infuriatingly obscure theories of meaning.

Lastly, while the experimental component has done its best to be realistic, it is in no way complete. Pains have been taken to ensure that the experiment, unlike much work in the Semantic Web, at least uses real data and feedback from real users. Yet our proposed solution would require for a full-scale implementation and co-operation of both a major hypertext search engine and a Semantic Web search engine. Obviously,
this is out of scope for a dissertation. Yet a real query-log from a major search engine is used to determine the terms used to retrieve real data from a large-scale Semantic Web search engine, and real human expert knowledge and browser results are used to provide the ‘gold standard’ for the implementation of our solution. While various parts of the experiment could no doubt be optimized still further, for a proof-of-concept solution to a very difficult problem, this experiment should be sufficient.

1.4 Notational Conventions

In order to aid the reader, the thesis employs a number of notational conventions. In particular, we only use “double” quotes to quote a particular author or other work. When a new word is introduced and used in an unusual manner to be clarified later, we use ‘single’ quotes. The use of ‘single’ quotes is also used when a word is supposed to be understood as the use of the word qua word, a mention of the word, rather than a use of the word. When a term is defined, the word is first labeled using **bold and italic** fonts, and then the definition is given in *italics*. Mathematical or formal terms are *italicized*, as is the use of *emphasis* in any sentence. Finally, the names of books and other works are often italicized. In general, technical terms like HTTP are often abbreviated by their capitalized initials. One of the largest problems of this whole area historically has been the eliding of levels of abstraction and use between terms, and we hope this fairly rigorous notational convention helps separate at least use, mention, definitions, and direct quotations of words. In order to make cross-referencing these terms straightforward, we provide an alphabetical list of all technically and philosophically defined terms in Appendix.

1.5 Summary

Despite its ambitious title, this dissertation is a *modest* attempt to both articulate and apply the principles of Web architecture in order to answer a question at the heart of the Semantic Web: *What does a URI mean?* We provide a solution by giving a broad overview of both philosophy of language and Web architecture, and by constructing a proof-of-concept solution. We do not claim to provide a complete or only solution, but do show our solution is better than other competing positions and solutions. Furthermore, we do not claim to have solved any of these problems regarding meaning and reference for all languages, especially natural language, and are fully confident
that philosophers will continue arguing over these issues of meaning and reference for the next century, if not more. However, what we do present is a proof-of-concept solution for these problems of meaning and reference in the special and limited case of establishing common identifiers on the Semantic Web.

Although we admit our approach of philosophically-informed engineering may seem unnecessarily dense or impractical at first glance, by taking previous work in philosophy seriously, we will demonstrate that practical solutions to difficult questions of meaning and reference on the Web are possible, although this necessitates a careful understanding of viewpoints from different disciplines. As for the evolution of the Web as a whole, as long as the all-too-human limitations of most users of the Web are kept in mind, this philosophical engineering could prove to be a crucial advantage in making the Semantic Web a reality. Ultimately it is not just pure performance that matters, but solid conceptual foundations. Furthermore, once we recognize the Web and its principles, as well as the relationship of Web to philosophy and language, we are more likely to have practical success in preserving characteristics that lead the Web to be such an astounding and universalizing information space. In the words of Italo Calvino, we must “seek and learn to recognize who and what, in the midst of the inferno, are not inferno, make them endure, give them space” (?).
Chapter 2

The Significance of the Web

If we could rid ourselves of all pride, if to determine our species we kept strictly to what historic and prehistoric periods show us to be the constant characteristic of man and of intelligence, we should not say Homo Sapiens but Homo Faber. In short, intelligence, considered in what seems to be it original feature, is the faculty of manufacturing artificial objects, especially tools for making tools. Henri Bergson (?)

The subject matter of this thesis is the nature of reference and representation on the World Wide Web, and this chapter provides the necessary background information to motivate the thesis and to make the central hypothesis of the thesis comprehensible. In this thesis, we consider the World Wide Web (from hereon referred to only as ‘the Web’) as a first-class subject matter for study. The first chapter delves into the origins of the Web so that the question of meaning and reference on the Web can be understood in its proper context.

Why the Web? Why not look at more interesting problems in a subject like artificial intelligence? In his One Hundred Billion Lines of C++, computer scientist-turned-philosopher Brian Cantwell Smith notes that the models of computing used in debates over reference and representation tend to frame the debate as if it were between “classical” logic-based symbolic reasoners and some “connectionist” and “embodied” alternative ranging from neural networks to epigenetic robotics (?). Smith then goes on to aptly state that the kinds of computational systems discussed in artificial intelligence and philosophy tend to ignore the vast majority of existing systems, for “it is impossible to make an exact estimate, but there are probably something on the order of 10, or one hundred billion lines of C++ in the world. And we are barely started. In sum: symbolic AI systems constitute approximately 0.01% of written software” (?). The same
small fraction likely holds true of “non-symbolic AI” computational systems such as robots, artificial life, and connectionist networks. While numbers by themselves hold little intellectual weight, for one could always argue that the vast majority of computational systems may have just no impact on our understanding of representation and intelligence. In this thesis we argue otherwise. The wide class of computational systems present a “middle distance” where questions of reference, representation, and intelligence come to the forefront and may even be more tractable than in the case for humans (?). One of the the most significant members to date of this wider class of computational systems is the World Wide Web, described by Tim Berners-Lee, the person widely acclaimed to be the ‘inventor’ of the Web, as “a universal information space”(?).

Michael Wheeler, a philosopher who is well-known for his Heideggerian defense of embodiment, surmises that while “the power of the Web as a technological innovation is now beyond doubt” but “what is less well appreciated is the potential power of the Web to have a conceptual impact on cognitive science” and so this thesis may provide a new “fourth way” in addition to the “three kinds of cognitive science or artificial intelligence: classical, connectionist, and (something like) embodied-embedded” (?). This taking of computational systems ‘in the wild’ like the Web seriously has been mirrored in philosophy as a move away from the “linguistic turn” to what has been termed a “informational turn,” which demonstrates that this wider class of computational systems are having an impact on the debates over reference and representation traditionally confined to the philosophy of language (?). While countless papers have been produced on the technical aspects of the Web, very little has been done explicitly on the Web qua Web as a subject matter. This does not mean there has not been interest, although the interest has come in particular more from the side of those working on developing the Web rather than those already entrenched in philosophy, linguistics, and artificial intelligence. In particular, the workshop series on Identity, Reference, and the Web has provocated many articles on these topics from prominent Web architects, although not from philosophers per se (???). In this spirit, what we will undertake in this thesis as a whole is to apply many well-known philosophical theories of reference and representation to the phenomenon of the Web.

In order to establish the relative autonomy of the Web as a subject matter, we recount its origins and so its relationship to other projects, both intellectual such as Engelbart’s Human Augmentation Project as well as more purely technical projects such as the Internet (?). It may seem odd to start out this thesis, which involves very
specific questions about meaning and reference on the Web, with a thorough history of the Web. To understand these questions we must first have an understanding of the boundaries of the Web and the normative documents that define the Web. The Web is a fuzzy and ill-defined subject matter whose precise boundaries and even definition are unclear. Unlike some subject matters like chemistry, the subject matter of the Web is not necessarily very stable, like a ‘natural kind,’ as it is a technical artifact. So we will take the advise of the philosopher of technology Gilbert Simondon, “Instead of starting from the individuality of the technical object, or even from its specificity, which is very unstable, try to define the laws of its genesis in the framework of this individuality or specificity, it is better to invert the problem: it is from the criterion of the genesis that we can define the individuality and the specificity of the technical object: the technical object is not this or that thing, given hic et nunc but that which is generated” (\textsuperscript{13}). In other words, we must first trace the creation of the Web before attempting to define it, imposing on the Web what Frederic Jameson calls “the one absolute and we may even say ‘transhistorical’ imperative,” that is: “Always historicize!” (\textsuperscript{13}). We build on the work of this chapter in Chapter 4 to delineate the precise principles of the Web.

2.1 The Origins of the Web

What is the Web, and what is its significance? At first, it appears to be a relative upstart upon the historical scene, with little connection to anything before it, an ahistorical and unprincipled ‘hack’ that came unto the world unforeseen and with dubious academic credentials. The purpose of this section is to dispel this myth.

The intellectual trajectory of the Web is a fascinating, if mostly unknown, history. Although it is well-known that the Web bears some striking similarity to Vannevar Bush’s ‘Memex’ idea from 1945, the Web is itself usually thought more of as a technological innovation rather than an intellectually rich subject matter such as artificial intelligence or cognitive science (\textsuperscript{14}). However, the Web’s heritage is just as rich as artificial intelligence and cognitive science, and can be traced back to the same roots, namely the ‘Man-Machine Symbiosis’ project of Licklider (\textsuperscript{15}). The ‘Man-Machine Symbiosis’ project gave birth to two streams of research, that of artificial intelligence and another lesser-known strand, the work on ‘human augmentation’ exemplified by the Human Augmentation Project of Engelbart (\textsuperscript{17}). Human augmentation, instead of hoping to replicate human intelligence as artificial intelligence did, only thought to enhance it. Despite similarities to artificial intelligence on the ‘Semantic Web,’ the Web
itself is a descendant of Engelbart’s vision, and this historical trajectory, which leads from Licklider to the creation of the Web, is detailed in the next sections.

### 2.2 The Man-Machine Symbiosis Project

The first precursor to the Web was glimpsed, although never implemented, by Vannevar Bush. For Bush, the primary barrier to increased productivity was the lack of an ability to easily recall and create records, and Bush saw in microfiche the basic element needed to create what he termed the “Memex,” a system that lets any information be stored, recalled, and annotated through a series of “associative trails” (\(?\)). The Memex would lead to “wholly new forms of encyclopedias with a mesh of associative trails,” a feature that became the inspiration for “linking” in hypertext (\(?\)). However, Bush could not implement his vision on the analog computers of his day.

The Web had to wait for the invention of digital computers and networks, both of which bear some debt to the work of J.C.R. Licklider, a disciple of Norbert Wiener (\(?\)). Wiener thought of feedback as an overarching principle of organization in any science, and one was equally universal among humans and machines (\(?\)). Licklider expanded this notion of feedback loops to that of the low-latency feedback between humans and digital computers. This vision of ‘Man-Machine Symbiosis’ is distinct and prior from cognitive science and artificial intelligence, both of which hypothesize that the human mind can be construed as either computational itself or even implemented on a computer. Licklider held that while the human mind itself might not itself be a computational (although Licklider cleverly remained agnostic on that particular gambit), the human mind was definitely complemented by computers. As Licklider himself put it, “The fig tree is pollinated only by the insect Blastophaga grossurun. The larva of the insect lives in the ovary of the fig tree, and there it gets its food. The tree and the insect are thus heavily interdependent: the tree cannot reproduce without the insect; the insect cannot eat without the tree; together, they constitute not only a viable but a productive and thriving partnership. This cooperative ‘living together in intimate association, or even close union, of two dissimilar organisms’ is called symbiosis. The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today” (\(?\)). The goal of ‘Man-Machine Symbiosis’ is then the enabling of reliable coupling between the humans and their ‘external’ information
as given in digital computers. To obtain this coupling, the barriers of time and space needed to be overcome so that the symbiosis can operate as a single process.

The ‘Man-Machine Symbiosis’ project was not merely an philosophical project, but an engineering project. In order to provide the funding needed to assemble what Licklider termed his “galactic network” of researchers to implement the first step of the project, Licklider became the institutional architect of the Information Processing Techniques Office at Advanced Research Projects Agency (ARPA) (?). Licklider first tackled the barrier of time. Early computers had large time lags in between the input of a program to a computer on a medium such as punch-cards and the reception of the program’s output. This lag could then be overcome via the use of time-sharing, taking advantage of the fact that the computer, despite its centralized single processor, could run multiple programs in a non-linear fashion. Instead of idling while waiting for the next program or human interaction, in moments nearly imperceptible to the human eye, a computer would share its time among multiple humans (?).

Douglas Engelbart had independently generated a proposal for a “Human Augmentation Framework” that shared the same goal as ‘Man-Machine Symbiosis’ project of Licklider, although it differed by placing the human at the center, focusing on the ability of the machine to extend to the human user, while Licklider imagined a more egalitarian partnership between humans and digital computers (?). This focus on human factors led to Engelbart to the realization that the primary reason for the high latency between the human and machine was the interface of the human user to the machine itself, as a keyboard was at best a limited channel. After extensive testing of what devices enabled the lowest latency between humans and machine, Engelbart invented the mouse (?). By employing these interfaces, the temporal latency between humans and computers was decreased even further.

### 2.3 The Internet

The second barrier to be overcome was space, so that any computer should be accessible regardless of its physical location. The Internet “came out of our frustration that there were only a limited number of large, powerful research computers in the country, and that many research investigators who should have access to them were geographically separated from them” (?). Licklider’s lieutenant Bob Taylor and his successor Larry Roberts contracted out Bolt, Beranek, and Newman (BBN) to create the Interface Message Processor, the hardware needed connect the various time-sharing computers
of Licklider’s “galactic network” that evolved into the ARPANet \textsuperscript{7}. While BBN provided the hardware for ARPANet, the software was left undetermined, so an informal group of graduate students constituted the Internet Engineering Task Force (IETF) to create software to run the Internet \textsuperscript{2}.

The IETF has historically been the main body that creates the protocols that run the Internet. It still maintains informal nature of its foundation, with no formal structure such as a board of directors, although it is officially overseen by the Internet Society. The IETF informally credits as their main organizing principle the credo “We reject kings, presidents, and voting. We believe in rough consensus and running code” \textsuperscript{5}. Decisions do not have to be ratified by consensus or even majority voting, but require only a rough measure of agreement on an idea. The most important product of these list-serv discussions and meetings are IETF RFCs (Request for Comments) which differ in their degree of reliability, from the unstable ‘Experimental’ to the most stable ‘Standards Track.’ The RFCs define Internet standards such as URIs and HTTP \textsuperscript{2}.

RFCs, while not strictly academic publications, have a normative force on the Internet and therefore on the Web, and so they will referenced considerably throughout this thesis.

Before the Internet, networks were assumed to be static and closed systems, so one either communicated with a network or not. However, early network researchers determined that there could be a “open architecture networking” where a meta-level “internetworking architecture” would allow diverse networks to connect to each other, so that “they required that one be used as a component of the other, rather than acting as a peer of the other in offering end-to-end service” \textsuperscript{3}. In the IETF, Robert Kahn and Vint Cerf devised a protocol that took into account, among other, four key factors, as cited below \textsuperscript{3}:

1. Each distinct network would have to stand on its own and no internal changes could be required to any such network to connect it to the Internet.

2. Communications would be on a best effort basis. If a packet didn’t make it to the final destination, it would shortly be retransmitted from the source.

3. Black boxes would be used to connect the networks; these would later be called gateways and routers. There would be no information retained by the gateways about the individual flows of packets passing through them, thereby keeping them simple and avoiding complicated adaptation and recovery from various failure modes.
4. There would be no global control at the operations level.

In this protocol, data is subdivided into ‘packets’ that are all treated independently by the network. Data is first divided into relatively equal size packets by TCP (Transmission Control Protocol), which then sends the packets over the network using IP (Internet Protocol). Together, these two protocols form a single protocol, TCP/IP (?). Each computer is named by an Internet Number, a four byte destination address such as 152.2.210.122, and IP routes the system through various black-boxes, like gateway and routers, that do not try to reconstruct the original data from the packet. At the recipient end, TCP collects the incoming packets and then reconstructs the data. The key assignment of IP addresses to individual machines was originally managed on behalf of the IETF by a single person, Jon Postel, but now managed by ICANN (Internet Corporation for Assigned Names and Numbers) under contract from the U.S. government. ICANN also regulates the domain name system, which is a centralized naming authority that maps names like http://www.ibiblio.org to an IP address like 152.46.7.122. This mapping is done via a hierarchical decision tree, where an agent can ask a domain name server what the mapping is for the top-level (such as .org) of the domain name, which then in turn delegates the request to other more specialized name servers. In this manner, the domain names of servers are connected to individual computers, although a single computer can host many servers or single domain name may be hosted by multiple computers.

The Internet connects computers over space, and so provides the physical layer over which the “universal information space” of the Web is implemented. However, it was a number of decades before the latency of space and time became low enough for the Web to become not only universalizing in theory, but universalizing in practice. An historical example of attempting a Web-like system before the latency was acceptable would be NLS (oNLine System) of Engelbart (?). The NLS was literally built as the second node of the Internet, the Network Information Center, the ancestor of the domain name system. NLS allowed any text to be hierarchically organized in a series of outlines, with summaries, giving the user freedom to move through various levels of information and link information together. The most innovative feature of the NLS was a journal for users to publish information in a journal for others to link and comment upon, a precursor of blogs and wikis (?).

However, Engelbart’s vision could not be realized on the slow computers of his day. Although time-sharing computers reduced temporal latency on single machines, too many users sharing a single machine made the latency unacceptably high, especially
when using an application like NLS. Furthermore, his zeal for reducing latency made NLS far too difficult to use, as it depended on obscure commands that were far too complex for the average user to master within a reasonable amount of time. It was only after the failure of the NLS that researchers at Xerox PARC developed the personal computer, which by providing each user their own computer reduced the temporal latency to an acceptable amount (?). When these computers were connected with the Internet and given easy-to-use interfaces as developed at Xerox PARC, both temporal and spatial latencies were made low enough for ordinary users to access the Internet. This convergence of technologies, the personal computer and the Internet, is what allowed the Web to be implemented successfully and enabled its wildfire growth, while previous attempts like NLS were doomed to failure as they were conceived before the technological infrastructure to support them had matured.

2.4 The Modern World Wide Web

Perhaps due to its own anarchic nature, the IETF had produced a multitude of incompatible protocols such as FTP (File Transfer Protocol) and Gopher (?). While protocols could each communicate with other computers over the Internet, there was no universal format to identify information regardless of protocol. One IETF participant, Tim Berners-Lee, had the concept of a “universal information space” which he dubbed the “World Wide Web” (?). His original proposal to his employer CERN brings his belief in universality to the forefront, “We should work toward a universal linked information system, in which generality and portability are more important than fancy graphics and complex extra facilities” (?). The practical reason for Berners-Lee’s proposal was to connect the tremendous amounts of data generated by physicists at CERN together. Later as he developed his ideas he came into direct contact with Engelbart, who encouraged him to continue his work despite his work being rejected at conferences like ACM Hypertext 1991.¹

In the IETF, Berners-Lee, Fielding, Connolly, Masinter, and others spear-headed development of URIs (Universal Resource Identifiers), HTML (HyperText Markup Language) and HTTP (HyperText Transfer Protocol). Since by being able to reference anything with equal ease due to URIs, a web of information would form based on “the few basic, common rules of ‘protocol’ that would allow one computer to talk to another, in such a way that when all computers everywhere did it, the system would

¹Personal communication with Berners-Lee.
thrive, not break down” (?). The Web is a virtual space for naming information built on top of the physical infrastructure of the Internet that could move bits around, and it was built through specifications that could be implemented by anyone, “What was often difficult for people to understand about the design was that there was nothing else beyond URIs, HTTP, and HTML. There was no central computer ‘controlling’ the Web, no single network on which these protocols worked, not even an organization anywhere that ‘ran’ the Web. The Web was not a physical ‘thing’ that existed in a certain ‘place.’ It was a ‘space’ in which information could exist” (?).

The very idea of a universal information space seemed at least ambitious, if not de facto impossible, to many. The IETF rejected Berners-Lee’s idea that any identification scheme could be universal. In order to get initiative of the Web off the ground, Berners-Lee surrendered to the IETF and renamed URIs from Universal Resource Identifiers (URIs) to Uniform Resource Locators (URLs) (?). The Web begin growing at a prodigious rate once the employer of Berners-Lee, CERN, released any intellectual property rights they had to the Web and after Mosaic, the first graphical browser, was released. However, browser vendors started adding supposed ‘new features’ that soon led to a ‘lock-in’ where certain sites could only be viewed by one particular corporate browser. These ‘browser wars’ began to fracture the rapidly growing Web into incompatible information spaces, thus nearly defeating the proposed universality of the Web (?).

Berners-Lee in particular realized it was in the long-term interest of the Web to have a new form standards body that would preserve its universality by allowing corporations and others to have a more structured contribution than possible with the IETF. With the informal position of merit Berners-Lee had as the supposed inventor of the Web (although he freely admits that the invention of the Web was a collective endeavor), he and others constituted the World Wide Web Consortium (W3C); a non-profit dedicated by “leading the Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web” (?). In the W3C, membership was open to any organization, commercial or non-profit organization. Unlike the IETF, W3C membership came at a considerable membership fee. The W3C is organized as a strict representative democracy, with each member organization sending one member to the Advisory Committee of the W3C, although decisions technically are always made by the Director. By opening up a “vendor neutral” space, companies who previously were interested primarily in advancing the technology for their own benefit could be brought to the table. The primary product of the World Wide Web Consortium is a W3C Recommendation, a standard for the Web that is explicitly voted on
and endorsed by the W3C membership. W3C Recommendations can be considered to
similar to IETF RFCs, with normative force due to the degree of formal verification
given via voting by the W3C Membership. A number of W3C Recommendations have
become very well known technologies, ranging from the vendor-neutral later versions
of HTML (?), which stopped the fracture of the universal information space, to XML,
which has become a prominent transfer syntax for many types of data (?). This thesis
will cite W3C Recommendations when appropriate, as these are one of the main norm-
ative documents that define the Web. With IETF RFCs, these normative standards
collectively define the foundations of the Web. It by agreement on these standards that
the Web functions as a whole. However, the rough-and-ready process of the IETF and
even W3C has led to a terminological confusion that must be sorted in order to inspect
the problem of how URIs can identify things outside the Web itself.
Chapter 3
Philosophical Prolegomenon

The commonwealth of learning is not at this time without master-builders, whose mighty designs in advancing the sciences, will leave lasting monuments to the admiration of posterity; but every one must not hope to be a Boyle, or a Sydenham; and in an age that produces such masters as the great Huygenius, and the incomparable Mr. Newton, with some others of that strain; it is ambition enough to be employed as an under-labourer in clearing the ground a little, and removing some of the rubbish that lies in the way to knowledge. John Locke (?)

A major focus of this dissertation is to use terminology from philosophy of computation, language, and the mind to produce a finite set of moderately clearly-defined terms that we can use these terms to express the question; what does a URI mean? Afterwards, we use these terms to determine what the boundaries of the Web are in Chapter 4 and to clarify the Semantic Web in Chapter 5.

For the sake of brevity we will not in this chapter explore all the nuances and consequences arising from our admittedly broad-sweeping terminology. This is unfortunate, as there is just not enough space to address, much less defuse, all possible counter-arguments. In this manner, this chapter will be decidedly non-philosophical, although we will attempt to mitigate this problem by at least providing references to well-known philosophers from whom we have adopted our terminology, although often in a slightly-modified form so that the terminology may fit the problem at hand. However, the theoretical framework and terminological definitions given in this chapter provides the foundation for the entire thesis, coming to a head in our proposed solution to the issues of reference and representation on the Semantic Web in Chapter ??.

While this chapter may in this chapter appear to be wide-ranging, the single
golden thread that necessitates this cleaning of the conceptual basement will become more apparent as the thesis continues. Again, we claim neither that our historical and philosophical foundations of Web architecture are complete and systematic, but just systematic and complete enough to pose and solve our hypothesis, without either the question or our solution using vacuous terminology. Otherwise, like many other discussions on this topic in Web circles, the result will be terminological confusion, causing any reader to fall into an intentional swamp of undefined and fuzzy terms like meaning, reference, and representation. In this regard, this entire chapter is in the spirit of John Locke’s famous under-labourer conception of philosophy, where the central task philosophy is viewed as the ability to clarify some of the dusty corners of some phenomenon, like the Web (?). We tackle this task first exploring the notion of ‘information’ at the heart of Berners-Lee’s definition of the Web as a ‘universal information space’ and then rebuild a notion of ‘representation’ and ‘reference’ on top of our notion of information, concluding with a foray into understanding the notion of ‘digitality,’ since the Web is composed not just representations, but digital representations.

3.1 The Web and the World

The central thesis of the Web is that multiple agents can use the same Uniform Resource Identifier (URI), such as http://www.example.org, on the Web to access the same information and refer to the same thing. Since the Web is an engineered artifact, unlike more obvious direct products of evolution like natural language, one would imagine that the solution to the problem of naming is of an engineering variety. However, no engineering solution seems to be trivially forthcoming. This is because identification, naming and its attendant issues are part and parcel of the “intentional sciences: sciences dealing with symbols, meaning, reference, interpretation, truth” (?). The intentional sciences, which expand beyond their traditional arenas in cognitive science and the philosophy of the mind to the more afar fields such as computer science are still very much a landscape of sciences in development. Informatics, the study of “the representation, processing, and communication of information in natural and artificial systems” can be thought of as the science of these intentional systems in their most all-encompassing sense (?). It is the task of this chapter of the thesis then to make a first initial foray into creating some common terminology to understand the intentional terminology of the Web, all while remaining true to a robustly common-sense view of the world based on the natural sciences.
On the surface an intentional term like ‘representation’ seem to be what Brian Cantwell Smith calls “physically spooky,” since a representation can refer to something with which it is not in physical contact (?). This spookiness is a consequent of a violation of common-sense physics, since representations allow us to have some sort of what appears to be a non-physical relationship with things that are far away in time and space. This relationship of ‘aboutness’ or intentionality is often called ‘reference.’

While it would be premature to define ‘reference,’ a few examples will illustrate its usage: someone can think about the Eiffel Tower in Paris without being in Paris, or even having ever set foot in France; a human can imagine what the Eiffel Tower would look like if it were painted blue, and one can even think of a situation where the Eiffel Tower wasn’t called the Eiffel Tower. Furthermore, a human can dream about the Eiffel Tower, make a plan to visit it, all while being distant from the Eiffel Tower. Reference also works temporally as well as distally, for one can talk about someone who is no longer living such as Gustave Eiffel. Despite appearances, reference is not epiphenomenal, for reference has real effects on the behavior of agents. Specifically, one can remember what one had for dinner yesterday, and this may impact what one wants for dinner today, and one can book a plane ticket to visit the Eiffel Tower after making a plan to visit it.

Can we get to the heart of this mystery at the heart of representation and other intentional terminology? The trick would be to define what precisely our common-sense notion of reference is, and to do this requires some terminological ground work while avoiding delving into amateur quantum physics. The terminology here is supposed to reconstruct rather carefully some common-sense demarcations in a uncontroversial yet broad manner so that these terms can deal with a suitably broad range of phenomena, including the Web. To pin the supposed ‘spookiness’ of reference down, we will introduce a few terms. A process is a general-purpose term used to denote events, objects, and proto-objects in a “patch of metaphysical flux,” where the process can be defined by having some regularity - a lack of difference - in time and space, that can distinguish it from other possible processes (?). However, since the word ‘process’ is often rather awkward, we shall often use term thing interchangeably with process. We will also sometimes use the term system when we are emphasizing the fact that one thing can be also, on a different level of abstraction, given as multiple things. All things and processes are the world. This can be considered a mere change of focus, for the term ‘thing’ emphasizes the everyday, solid, and static nature of the “metaphysical flux,” while the term ‘process’ refers more to its dynamic aspect (?). There are generally
two kinds of separations possible in processes in a relativistically invariant theory, a physical theory that obeys the rules of special relativity so that the theory looks the same for any constant velocity observer, since processes may be separated in time or space. Processes that are separated by time and space are distal while those processes that are not separated by time and space are proximal. As synonyms for distal and proximal, we will use non-local and local, or just disconnected and connected. Although this may seem to be an excess of adjectives to describe a simple distinction, this aforementioned distinction will underpin our notions of representation and reference. In figures, physical local relationships will be marked with a $\psi$ symbol, while distal (and so referential) relationships are marked with the $\phi$ symbol.

While a discussion about counterfactuals and causation is far beyond our scope, we will rely on the common-sense intuition that if one process is connected with another process and a change in the former process is followed by a change in the latter process, that former process may have caused the change in a former process. In other words, the first process is effective, and the other processes that may be effected by a particular process are within its effective or causal reach. Anything that appears to violate these common-sense intuitions about physics and causation is spooky, while anything that does not is non-spooky. A property of the distal is that it is beyond effective reach; as Smith puts it, “distance is where no action is at” (?). For example, a tourist hitting their toe on the Eiffel Tower has no effect on someone in Edinburgh.

Reference allows an agent to use local processes (neural states, speaking words, using identifiers in a computer program) to track a non-local process. These local phenomena can then be used to bring a particular process in and out of connection, such as an illucationary speech act: the utterance “Ralph, please come here!” in the right context could make Ralph come over from the next room, from being distal to being local. This force of reference can also be generalized, such as when the act of someone asking an travel agent to book a trip to the Eiffel Tower causes the purchase of a plane ticket with the attendant complex mixture of fiscal, legal, and computational effects. This eventual bringing of the previously distal Eiffel Tower itself into effective reach exemplifies the cycle of connection and disconnection that seems crucial to understanding the interplay between reference and representation.

The impact of the Web on this cycle of connection and disconnection should not be underestimated. The various hardware and software technologies underlying and constituting the Web, ranging from high-speed fiber optic cables to its universalizing identification scheme, allow the effective reach of agents to be spread further than
ever before, making the boundaries of space and time increasingly irrelevant. While someone in Edinburgh cannot view the Eiffel Tower directly, they could use the Web to retrieve web-pages and other facts about the Eiffel Tower, or even open up a web-cam with a view of the Eiffel Tower itself. Although the intuitive impact of the Web on these issues of reference and effective reach are obvious, what reference and representation actually are and how they work is far from trivial. While a complete analysis is outside the scope of this thesis, we will sketch some definitions to serve as the conceptual foundation for further work on the issues of representation and reference on the Web, mostly by appeal to philosophers whose works can be referenced for a more complete definition.

3.2 Information, Encoding, and Content

The Web has been defined as a “universal information space” by Berners-Lee, and we will take this definition seriously and attempt to unpack it, in hope that it will provide clues on how we can define both ‘representation’ and ‘reference’ in a manner that can do justice to the Web. The strategy to be employed is to inspect Berners-Lee’s evocative notion of the Web in order to provide a less complex notion of information that can serve as the foundation for building the more complex notion of representation. The first question to be answered then is the perennial question: What is information? Although we cannot comprehensively answer this question in full, we can sketch some crucial distinctions.

We will rely on the notion of information, central to Shannon’s communication theory, that information is whatever is in common between two processes, a source and a receiver (\textsuperscript{1}). To have something in common means to share the same regularities, e.g. parcels of time and space that cannot be distinguished at a given level of abstraction.\textsuperscript{1} Whether or not the information is preserved over time or space is due to the properties of a physical substrate known as the channel. This correlates with information being the inverse of the amount of “noise” or randomness in a system, and the amount of information being equivalent to a reduction in uncertainty. This preservation or failure to preserve information can be thought of as sending of a message between the source and the receiver over a channel. The message is the physical process that can be

\textsuperscript{1}The notion of the ‘level of abstraction’ is not trivial in of itself either. Although in general this idea has been inadequately theorized, we explore the notion more detail in Section 3.6, and for a more fully-worked out, albeit more limited, theory see (\textsuperscript{2}).
thought of as bearing information due its local physical properties, so we will often call some particular message with some particular information an ‘information-bearing message.’

Shannon’s theory deals with finding the optimal encoding and size of channel so that the message can get from the sender to the receiver (?). The best attempt we have to give an definition of the notion of encoding is given by Nelson Goodman (?). Goodman defines what we would call an encoding as a series of marks, where mark is a physical characteristic, such as one used in an encoding of information. Goodman uses this notion of encoding to encompass everything from utterances to alphabetic characters, which we can easily extend to cover bits and series of voltage patterns (?). To be reliable in conveying information, an encoding should be physically “differentiable” and thus maintain what Goodman calls “character indifference” so that (at least within some context) each character (or characteristic) can not be mistaken for another character. So, an encoding is the precise regularities realized by the information-bearing message. Encodings are usually given these regularities in virtue of being in a language, which is explicated in Section.

Every particular information-bearing message must have an encoding. To use an example, Daniel in Paris (the source) is trying to send a message to Amy (the receiver), a secretary in Boston, that one of her fellow workers, Ralph, has won a trip to the Eiffel Tower. Daniel can send this message as a signal encoded in a variety of ways: a secret binary code in e-mail, a carrier pigeon carrying the note in English, or taps on a telegraph. The encoding itself is just the precise physical regularities at a level of abstraction, which is a way of recognizing certain physical differences and regularities. For example, given a hand-written letter in English, one can focus on the the low-level of abstraction, such as the details of the various pen-strokes and the texture of the paper, or progressively higher levels of abstraction, such as recognizing letters in a alphabet, words, or sentences. The encoding can then be thought of as certain physical characteristics that, at a level of abstraction, can convey the message. Therefore, the same encoding in our example is the note in English saying that ‘Ralph has won a ticket to Paris,’ which can then be realized by various physical things, such as e-mails, postcards, billboards, and so on. To say a thing has a certain encoding is to say that it is a exemplification or realization of the encoding, that is the physical thing that locally carries the information.

It is useful to think of the relationship between an abstract encoding and the things that realize that information as similar, although in a subtle manner different, to the
classical distinction between type and tokens. Even encodings exist on a level of abstraction. The text of *Moby Dick* can be thought of as information, and the text of *Moby Dick in English* is an encoding of the abstract information of *Moby Dick*, with precise regularities given by the *very letters* of the language. The abstract information of the novel *Moby Dick* could be expressed in a different language, like French, and the precise regularities that convey *the same information at a level of abstraction* could be given by *different* letters. However, even the text of *Moby Dick* in a particular language like English exists at a level of abstraction, as it could be realized in multiple manners, as a copy in English of *Moby Dick* could be realized by a web-page or a book, and even two different physical books, such as a web-page and a book. So *encodings also exist at a level of abstraction*. Furthermore, realizations of encodings can equally exist at levels of abstraction, as a web-page can realize the information of an English version of *Moby Dick*, but at different moments in space-time different, or even the same, bits can be sent down the wire that realize this web-page. Likewise, the same book in *Moby Dick* in English can have two different copies that both realize the same information.

Shannon’s genius was to show that messages over a channel can be encoded into a series of bits, with each bit representing a binary decision between two possibilities. Information can then be quantitatively measured, with each bit quantifying *how much* information there is. Let’s assume there’s eight possible co-workers of Amy in Boston. To encode the reduction of possibilities from eight to one requires three bits, as trivially $\log_2(8) = 3$. Encodings and their realizations then may also be layered on each other, on different levels of abstraction, eventually grounding out in the physical world of process of things in space-time (?). In the typical e-mail message, the lowest level of the message’s physical marks that can be physically differentiated are a series of voltages. From these bits can be interpreted other encodings, such as a format for e-mail messages like SMTP (Simple Mail Transfer Protocol), which in turn may have as its payload alphabetic characters in UTF (?). These alphabetic characters can then in turn encode words, which in turn encode sentences, which can then be part of a discourse. Encoding schemes can be realized on top of each other, one inside of another, with each level of abstraction having its own unique interpretation, all being realized on the same physical things in space-time. In this manner, only at the very ‘bottom’ level does the realization relationship ground out in some physical thing in space-time. However, even by grasping some particular thing or process in space-time,

\(^2\)“Unicode Transformation Format” is one of many possible character encoding schemes for text that encodes a given character in one through four octets.
one is grasping some regularities, and so every thing, even things we normally consider tokens, exist only on some level of abstraction.

However, Shannon’s theory does not explain the notion of information fully, since giving someone the number of bits that a message contains does not tell the receiver what information is encoded. Shannon explicitly recognizes this by stating “The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem” (\textit{?}). In fact, as Weaver states in his introduction to Shannon’s work, “The word information, in this theory, is used in a special sense that must not be confused with its ordinary usage...in fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, regards information” (\textit{?}). Although Shannon’s theory attempts to quantify how much information the message contains, many intuitions about the notion of information have to deal with not only how the information is encoded or how to encode it, but what a particular message is about, the content of an information-bearing message. ‘Content’ is a term we adopt from Israel and Perry (\textit{?}), as opposed to the more confusing term ‘semantic information’ as employed by Floridi and Dretske (\textit{??}), as we wish to reserve the term ‘semantics’ for a later related use. The content of the information-bearing message is not causally ineffective, but important. Just showing that determining a single employee out of eight won the lottery requires three bits does not tell Amy which employee in particular won the lottery, but only measures how many bits are needed to tell Amy precisely who won. After all, the false message that another office-mate Sandro won a trip to Paris is also three bits. One cannot tell by quantitative information alone what information is being conveyed by a message. The content is not independent of the encoding in practice, for it is conveyed by the precise number of bits in a particular message. as the content of “001” could be Ralph while “010” could be Sandro. Content is neither in theory orthogonal from Shannon’s theory of information encoding, for the amount of information given by a channel imposes constraints on what content can be sent (\textit{?}). If there are only two possible bits of information and all eight employees need one unique encoding, Daniel cannot send a message specifying what friend got the trip since there aren’t enough options in the encodings to go around so that each employee gets a unique encoding. If “01” has the content that “either Sandro or Ralph won the ticket,” the message has
not been successfully transferred from Daniel if the function of the message is to tell someone precisely which employee won the ticket.

One of the first attempts to formulate a theory of content was due to Carnap and Bar-Hillel (?). Their theory attempted to bind a theory of content closely to first-order predicate logic, and so while their “theory lies explicitly and wholly within semantics” they explicitly do not address “the information which the sender intended to convey by transmitting a certain message nor about the information a receiver obtained with a certain message,” since they believed these notions could eventually be derived from their formal apparatus (?). However, their overly restrictive notion of the content of information as logic did not gain widespread traction, and neither did other attempts to develop alternative theories of information such as that of Donald McKay (?). In contrast, Dretske’s semantic theory of information defines the notion of content to be compatible with Shannon information theory, and his notions have gained some traction within the philosophical community (?). Dretske, while accepting Shannon’s theory as a theory of encoding, notes that the content of a message and the amount of information in terms of uncertainty are different. For example, “saying ‘There is a gnu in my backyard’ does not have more content than the utterance ‘There is a dog in my backyard’ since the former is, statistically, less probable” (?). While information that is less frequent may require a larger number of bits in encoding, the content of information should be viewed as separable from this notion of reduction of uncertainty, since otherwise one is led to the “absurd view that among competent speakers of language, gibberish has more meaning than semantic discourse because it is much more less frequent” (?).

Is there a way to precisely define the non-quantitative content of a message? Dretske defines the content of information: “a signal $r$ carries the information that $s$ is $F$ when the conditional probability of $s$’s being $F$, given $r$ (and $k$) is 1 (but, given $k$ alone, less than 1). $k$ is the knowledge of the receiver” (?). This definition forms the heart of what content is conveyed by a signal, or message. The content of any information is whatever is held in common between the source and the receiver as a result of the conveyance of a particular information-bearing message. While this is similar to our definition of information itself, it is different, as information measures the total in common between a source and receiver. For example, two humans can share quite a lot in common, and so share information, despite never having conveyed messages between each other. The content is whatever is shared in common as a result of a particular message, such as the conveyance of notion that ‘Ralph won a ticket to the Eiffel
Tower.’ To unpack this definition and relate it to Dretske’s definition, the content of a message is called the “facts” by Dretske, \((F)\). This content is conveyed from the source \((s)\) successfully to the receiver \((r)\) when the content about the source can be used by the receiver with certainty. Daniel can only successfully convey the content that Ralph won a trip to Paris if before receiving the message Amy does not know Ralph won the trip to Paris and after receiving the message Amy does know that Ralph won the ticket. To communicate a fact successfully, both the source and receiver have to be using the same encoding scheme (bits, English, etc.) and the source has to encode the content into the message relative to what the receiver already knows or capacities the receiver possesses. Thus, if Amy does not know who is specified by the term “Ralph” given by the encoding scheme, but only knows him as “the guy with the black beard,” Daniel needs to explain in his message the additional fact that the “fellow with the black beard at your office is Ralph.” Furthermore, Daniel and Amy must share the same relationships between the content and the encoding scheme, so that Amy knows “001” has the content “Ralph won the ticket” as opposed to “Sandro won the ticket.”

Millikan rightfully notes that Dretske states his definition too strongly by stating that the conditional probability must be 1, and then notes that this “probability of 1” is just an approximation of a statistically “good bet” indexed to some finite domain where the information was learned to be recognized \(\Omega\). For example, lightening carries the content that “a thunderstorm is nearby” in rainy climes but in certain other domains, such as an arid prairie, lightening can convey a dust-storm.

These terms all illustrated in Figure 3.1. A source (Amy) is sending a receiver (Daniel) a message. The information-bearing message realizes some particular encoding such as a few sentences in English and a picture of the Eiffel Tower, and the content of the message can be interpreted to be about the Eiffel Tower.

![Figure 3.1: Information, Encoding, Content](image)

Dretkse notes that content is always finite, since information “does not mean that
a signal must tell us everything about a source to tell us something,” it just has to tell enough so that the receiver is now certain about the content within the domain (\(\_\_\_\_\_\_\_\)). Despite its use of humanistic terms like “certainty” and “knowledge,” one can also rephrase this definition of content in ways that apply to technology such as computers whose epistemic properties are at best uncertain. For a message being successfully conveyed with certainty requires that the entire regularity is preserved over the channel so that the message is capable of performing its function. What Dretske calls “knowledge” are the regularities already present in the system that may or may not contribute to the regularity being preserved in the act of information transfer. One could easily replace the message about a free trip to Paris being a message to book an airplane ticket for Ralph from one dumb server to another over the Internet. For this to be successful, the servers must share the same encoding schemes so that the contents of the message can be decoded, and the message should convey the same content from the first server of the airplane booking agency to the second server of the airplane company.

### 3.3 Language, Interpretation, and Models

What does it mean to share an encoding scheme? The sharing of an encoding scheme means that both encodings have the same interpretation, or relationship to content. When an encoding has a relationship to its content, what is taking place is an **interpretation** relationship between an encoding to content. The interpretation ‘fills’ in the necessary background and maps the encoding to some content. In our previous example, a mapping could be made between the encoding 001 to the content of Ralph while the encoding 010 could be mapped to content of Sandro. The word ‘interpretation’ is probably one of the most embattled words in the intentional sciences, and an in-depth study of its usage far exceeds the scope of this thesis.\(^3\) However, is Dretske’s definition really adequate to form the basis of the mysterious ‘content’ that is caused by the

---

\(^3\)Somewhat unusually, our usage of the term “interpretation” is as a relationship between processes, not a first-order process itself. Usual definitions of “interpretation” tend to conflate these issues. In formal semantics, the word “interpretation” often can be used either in the sense of “an interpretation structure, which is a ‘possible world’ considered as something independent of any particular vocabulary” or “an interpretation mapping from a vocabulary into the structure” or as a shorthand for both (\(\_\_\_\_\_\_\_\)). The difference in use of the term seems somewhat divided by fields. For example, computational linguistics often uses “interpretation” to mean the what Hayes called the “interpretation structure.” In contrast, we use the term ‘interpretation’ to mean what Hayes called the “interpretation mapping.” reserving the word ‘content’ for the “interpretation structure.” Also, this quick aside into matters of interpretation does not explicitly take on a formal definition of interpretation as done in model theory, although our general definition has been designed to be compatible with model-theoretic and other formal approaches to interpretation.
interpretation, and has not the physical spookiness merely been displaced now from the content to the relationship of interpretation itself?

Taking on the latter inquiry first, an interpretation requires an interpreter, a thing that is capable of having an interpretation from a particular encoding and a content. To uphold our requirement for physical non-spookiness, the interpreter and the encoding must be in some form of physical local connection, such as a mathematician looking at symbols or a humans looking at bytes. An interpretation can then also cause physical effects upon the interpreter. The content of a particular message depends very much on the encoding scheme used by the interpreter. For example, a mathematician can interpret the “11” as either the number eleven in decimal notation, or the number three in binary notation. This is not to say that humans are some sort of magical interpretation machine, as non-human processes can interpret messages. In our previous example involving two servers communicating a message about booking Ralph a plane ticket, both machines encode and decode bytes describing Paris, planes, and Ralph. They may not interpret the content of the message in the same manner that a human does, by knowing that Ralph is a human, but they interpret the message nonetheless, since the message has some physical effect upon the machine, causing it to send other messages to other machines that eventually result in a plane ticket being printed for Ralph when he gets to the airport. However, the content is not arbitrary, just as an interpretation is not arbitrary, if the plane ticket comes out for Sandro instead of Ralph, or sends Ralph to Berlin, something has gone amiss. Furthermore, in the tradition of Goodman, we consider the realization or exemplification can be the reverse of interpretation. So, if a realization is interpreted to be about some content, that realization realizes the content. The information on a level of abstraction that is contained in the text of Moby Dick is exemplified or realized by a physical copy of the book.

Interestingly enough, our definition of interpretation then opens the field to another definition, which is the possibility of at some point an encoding not having an interpretation, due to either an inability of an interpreter to recognize it as information, the lack of an necessary capacity (the $k$ of Dretske) of an interpreter to decode the encoding scheme. This would be the standard definition of data, which is information without an interpretation. Our definition works well with other “textbook” definitions of data and information, such as that of Davis and Olson, which states that “information is data that

---

4Again, this may not necessarily be a change in the interpreter, for the content could simply cause the receiver to resist, say, change of decay over time, as is the norm in digital systems.
has been processed into a form that is meaningful to the recipient” (7). This does not mean that the information does not possibly have an interpretation, but at some given moment it cannot be interpreted. One example would be if the message from Daniel that Ralph had won the plane ticket had been delivered via e-mail in French. While Amy could have been aware of some very limited aspects of the e-mail (such as the time sent and the sender), she would lack the necessary knowledge of French to decode the information’s content and so have an interpretation of the message. In this manner, the e-mail from Daniel, while having a definite interpretation for people French speakers, would lack an interpretation for Amy. To Amy, the message would just be data. Of course, Amy could learn French and eventually read the message, or send it to a machine-translation program, or ask a French speaker to translate the message for her, and so could eventually transform the encoding from data to information. One can also imagine cognitive constraints leading to a lack of an interpretation. For example, the amount of digital data gathered by modern telescopes is absolutely enormous, so large that much of it lies as uninterpreted reams of data rather than information, as it is beyond a single human to interpret this data, and even groups of humans trying to interpret it in a distributed manner are still struggling to catch up with the amount of data produced by the telescope.

Unlike many others, including Dretske, who have we relied on the help define information, we shall make no claims about the nature of information and truth, in particular if what appears to be ‘false’ information is really misinformation or pseudo-information. By remaining studiously neutral on this long-standing debate and not employing heavy-duty notions like ‘truth,’ our definition of information is suitably vague enough to stand for all sorts of messages. For example, if Daniel was sending the message to Amy that Ralph had a free plane ticket to Paris as some sort of jest or lie, Amy could still decode and interpret the message, and by filling in normal background assumptions (as Dretske put it, the “channel assumptions”) she might assume that the message was true. Amy would still have a interpretation of the content of the message, it would just be different than Daniel’s interpretation. However, we should notice that even if interpretations are different (or, in the case of data, the interpretation is just missing), information obeys the famous maxim of Hegel, that there is no matter without form and no form without content, where we substitute the term ‘encoding’ for ‘form’ (7). In other words, information always may have an encoding and content, as content is always carried by some physical encoding, and information can always be interpreted by some interpreter in some, although not necessarily the same, manner.
Information-bearing messages do not have to be things with a singular decoding onto some sort of simple propositional content. Instead, information often is layered upon other information, and so the content is also layered upon other content. When this happens, one is no longer dealing with a simple message, but some form of information-bearing language. A language can be defined as a system in which information-bearing messages are related to each other in a manner that can change their interpretation. Unless they are truly unique, encodings are therefore in languages. The relationships between encodings and content are usually taken to be based on some form of (not necessarily formalizable or even understood) rules. In other words, every information-bearing message must have an encoding, and these encodings are themselves in languages. If one is referring to a system in which the encoding of information-bearing messages are related to each other, then one is talking about the syntax of a language. If one is referring to a system in which the content of information-bearing messages are related to each other, then one is referring to the semantics of the language. The lower-level of a language can be terms, regularities in marks, that do or do not have their own interpretation, such as the alphabet or words. Any combination of terms that is valid according the language’s syntax is a sentence or expression in the language, and any combination of terms that has an interpretation according to the language’s semantics is a statement in the language. In this way, marks form a language. The relationship between semantics and syntax can be straightforward or only vaguely known, depending on the language in question. For example, formal languages almost always have an explicitly humanly-defined syntax and even model-theoretic semantics, while the semantics of English seem to escape easy definition. In natural language, one hypothesis, attributed to Frege, is the principle of compositionality, where the content of a statement is related systematically to terms in the which it is composed. The content of the sentence such as ‘Ralph has a plane ticket to Paris so he should go to the airport!’ can then be derived the more elementary content of the sub-statements, such as ‘Ralph has a plane ticket’ which in turn can be derived from the content of words such as ‘Paris’ and ‘ticket.’ The distinction between sentences and statements is due to the fact that sentences that are accepted by the syntax of a language, such as Chomsky’s famous “Colorless green ideas sleep furiously” may have no obvious interpretation outside of the pragmatics of Chomsky’s particular exposition (?), while sentence fragments like Shakespeare’s ‘to die, to sleep; to sleep, perchance to dream’ may have an interpretation only in terms of a dialogue or discourse, and otherwise appear senseless. There are obviously different types of languages.
Also, we do not restrict our use of the word ‘language’ to primarily linguistic forms, but use the term ‘language’ for anything where there is a systematic relationship between syntax and (even an informal) semantics, and this may include non-linguistic languages. On such investigation into this matter is Nelson Goodman’s *Languages of Art* (?). Although our examples so far have been in natural language, our definition of language is purposefully neutral as regards languages for humans (or even possibly languages for other animals) and “formal” languages for machines such as programming languages for computers. There are *iconic languages* based on images and *natural languages* based on human linguistic expressions, as well as *formal languages* with an explicitly defined syntax and possibly model-theoretic semantics, and so these formal languages can be interpreted regularly by computers. Indeed, while the debate is still out if human languages are truly compositional (?), programming languages almost always are compositional. Many computer languages not considered to be programming languages are also languages insofar as they have some normative or even informal interpretation, such as HTML. Furthermore, due to some biases against computer languages, we will use the term *format* as a synonym for computer-based language. *Linguistic expressions of an natural or formal language* are called *descriptions* while *the expressions of a iconic language* is called *depictions*. Lastly, just as encodings and content may be embedded in each other to form a language, languages themselves may be embedded in each other to form new languages. A *language embedded as a subset of another language* is a *dialect* of the language. Many machine languages like XML have as their primary purpose the expression of other dialects ?.

The relationship between information and realization also exists in language, as shown by the relationship between the encoding of some particular information in a sentence and its physical utterance by a speaker at a given point in space-time.

An interpretation is usually thought of as a mapping from some expressions in a language to a state-of-affairs in a world. This world is often thought to be the everyday world of concrete trees, houses, and landscapes that humans inhabit. We will not engage in any metaphysical speculation as regards the nature of the world besides our previous minimal definitions of physically connected or disconnected things and processes, so allowing for others to debate the existence of possible worlds or the metaphysical status of the past and future. Regardless, informally an interpretation can be considered to be a mapping where the physical world itself, a mapping rather appropriately labeled ‘God Forthcoming’ (?). However, often we do not have access to the world or state-of-affairs or can not communicate such access, and even if we
could, it is unclear if such as simple definition such as “the truth of a sentence consists in its agreement with (or correspondence to) reality” makes any sense, for “all these formulations can lead to various misunderstandings, for none of them is sufficiently precise and clear”? In an attempt to define a formal notion of truth, Tarksi ends up defining the interpretation of a language (which he terms the “object” language in terms of a “meta-language” (>). If both the language and the meta-language are suitably formalized, the interpretation of the language can then be expressed in terms of a satisfaction of a mathematical model, where satisfaction can be defined as an interpretation that defines whether or not every statement in the language can be given a content in the world, which in the tradition of Frege is usually thought of as a ‘truth’ value. The model “stands-in” for the vague and fuzzy world or some portion thereof. While Tarksi originally applied this only to suitably formal languages, others such as Montague have tried to apply this approach, with varying degrees of success and failure, to natural language. Regardless, the approach where an interpretation of a language is to a mathematical model is a model-theoretic semantics. The model is a mathematical representation of the world or the language itself. The relationship is summarized below in Figure 3.2, where the relationship between the model and the world is thought to be distal (such that the model represents the world), and the distal relationship between the content of the language given by the model and the world. The relationship between the language, the model, and the world is often thought to be distal, although not always, as in the case of code that interacts ‘directly’ with the world. The model may be thought of as picking up relevant features in the world.

The adequacy of models is usually judged by whether or not they fulfill the purposes to which the language is design, or whether or not their behavior adequately serves as a model of some portion of the world. Given a model-theoretic semantics, a model-theoretic interpretation can be given as “a minimal formal description of those aspects of a world which is just sufficient to establish the truth or falsity of any expression” in the language (>). One distinction that comes in useful is between that of intension and extension. While again the history and debate over these terms is outside the scope of this thesis, in general the original notion, as pioneered by Carnap (?), is that a certain kind of thing may only be described, and so given an intension, while the things that satisfy this description (which may be more than one thing) are extensions. A statement or statements are consistent if it can be satisfied, it is inconsistent if otherwise. Lastly, note that an entailment a semantic relationship where an interpretation of one sentence always satisfies the interpretation of another sentence , i.e.
the first statement entails the second. When an interpretation \( (I) \) can satisfy a sentence \( s \), one says that it \textit{entails} the sentence, or \( I \models s \). In contrast, an \textit{inference} is a syntactic relationship where one sentence can be used to construct another sentence in a language. For one sentence \( s_1 \) that infers another sentence \( s_2 \), we can write \( s_1 \vdash s_2 \). An inference that ‘follow in lock-step’ with an entailment are considered valid. In detail, as shown in Figure 3.2, the syntactic inference mechanisms over time produces more valid inferences, and because these “line up” with entailments in the interpretation, and since since model lines up with the world, these inferences therefore describe the world. Models can be captured in various ways, of which we have primarily described a denotational semantics, but often an axiomatic and operational semantics are equally powerful formalisms. Inference can usually be accomplished by some local inference procedure, like a computer program. The inference procedure of a language is \textit{sound} if every inferred sentence can be satisfied (i.e. the inference mechanism preserves ‘truth’), and it is \textit{complete} if every satisfied sentence can be shown to be entailed (i.e. all ‘true’ statements can be proven). This is necessarily a quick overview of the large field of formal semantics, and these issues are discussed more in depth in Chapter 6. This illustrated below as the contrast between the causal (\( \psi \)) relationships between the syntactic sentences of an traditional ‘knowledge representation’ inference system, which by churning syntax, manages to still produce semantic statements that \textit{accurately} refer to the world, and this is done by having the distal relationship between the sentences and the model, as well as the world itself, maintain itself as ‘true’ via any soundly inferred sentence.
The notion of interpretation does not necessarily mean any type of relationship must be established at all between the information-bearing message and the interpreter, besides the interpreter itself providing whatever mechanism is needed for interpretation. The notion of interpretation is between an information-bearing message and its content (or a model). When Daniel sends Amy the e-mail to tell her Ralph had a plane ticket to Paris, Amy interpreted the message by filling in various background information, and so determining that Ralph at her office, and not Ralph at the hot-dog stand down the street, has a plane ticket to the Paris. Amy has successfully interpreted the message. However, if Amy merely sat at her desk, content in her knowledge, but did not tell Ralph, then something would have gone awry from Daniel’s standpoint. The preservation of content can be, but does not have to be, detected because the message has the effect on the receiver, and the exact effect determines whether or not the message has successfully completed its function. However, what do we mean by *function*?

Information often has a ‘purpose’ that is beyond its interpretation to some content in a strict sense. For example, the same message “Police!” might always have the same interpretation (i.e. to a nearby policeman) but it would be radically different in purpose if it was muttered by a thief who had just managed to pick-pocket a tourist, than if the exact same expression was used by the tourist who had just been pick-pocketed. A single identical message can have a single interpretation, but two entirely different purposes or functions. The notions a function or purpose in this context is inherently *normative*, something should, but does not necessarily have to, fulfill its function. This normativity can be defined in terms on whether or not the function fulfills what Millikan calls a “proper function” (?). While the notion of a proper functions is too large subject to delve into detail here, Millikan summarizes her more extended presentation in *Language, Thought, and Other Biological Categories* (?) by saying “A thing’s proper functions are effects which, in the past, have accounted for selection of its ancestors for reproduction, or accounted for selection of things from which it has been copied, or for selection of ancestors of the mechanisms” (?). She later extends this definition to deal not just with natural selection by genetic selection, but mimetic selection, where imitation counts as a form of reproduction (?). To summarize, the

---

This does not necessarily mean that the receiver has changed, instead, the successful receipt of the message may mean that the function of the message on the receiver was to stay the same. This would be exemplified when measuring the degradation of information on a hard-drive, where the amount of information preserved from the selfsame hard-drive at one moment in time to another is considered the message.
The notion of proper function, instead of being a teleological notion of function, is a historical notion, one that grounds norms out in selection and so fulfills Dennett’s condition that “all normativity does ride of Darwin’s coat-tails” (9). The idea of proper functions makes special sense in the case of dealing with biological organisms, such as Dretske’s famous bacteria which possesses a magnetosome, which provides the bacteria with the information regarding which direction magnetic north is (9). Let’s also say the bacteria needs oxygen-free water in order to find food and reproduce. Somehow, the bacteria interprets whatever signals the magnetosome sends to orient itself towards magnetic north, and the bacteria goes towards magnetic north in order to find the oxygenless water. According to a Millikan-inspired understanding, the proper function of the magnetosome is to detect oxygenless water, since oxygenless water is precisely what the bacteria needs in order to increase the chances of its genes passing on. When the bacteria is put in an artificial environment where a magnet disturbs its magnetosome and points the poor bacteria towards oxygen-rich bacteria, the proper function of the magnetosome is being violated, since it is no longer in relation to the “normal” context in which it evolved.

Does this account help us with the Web? Since technological artifacts rarely themselves reproduce or appear to be directly involved in natural selection, we will expand the word “reproduction” to include both the production and reproduction of the environment that produced the process. So the proper function of a thing is either enabling its own reproduction, like the reproduction of the genes, or the reproduction of whatever produced the thing, such as the successful use of everything from flint tools to e-mails providing an evolutionary advantage to the humans that created them. Although it seems odd, e-mail messages in the large have the proper function of delivering information from one agent to another, since the communication of messages over the Internet has allowed agents to co-ordinate behavior and provide some measure of an evolutionary advantage. If the message is lost or garbled, it has not fulfilled its proper function. The only caveat is that determining the connection of a function of a thing in relation to any evolutionary story is difficult due to the remoteness of the connection and the limits of our own knowledge, and so even judging the success of any particular function is often outside the grasp of any observer. However, in most of the examples we are dealing with, our notion of success is straightforward even if its connection to an evolution is unclear: the message to Amy that Ralph won the plane ticket is suc-
cressful if Amy receives the content of the message, and this can be detected by Amy acting appropriately, such as when she tells Ralph that he has a plane ticket to Paris. Without the ability to accurately receive and transmit messages, one would assume that the species would be less likely to survive, and technology such as sending e-mail and the Web as a whole is successful insofar as it provides an evolutionary benefit to its users. E-mail does provide an advantage over carrier pigeons in speed and accuracy of delivering messages, so as Andy Clark puts, it “by seeing tools as entities with their own selective histories” we can understand what Terrence Deacon calls the “flurry of adaption...going on outside the brain” (??).

While a proper function is “a natural purpose,” many technological artifacts have an unnatural purpose, particularly those designed in research and not yet released “into the wild” to suffer the travails of selection either by nature or the market. This unnatural purpose is the purpose for which an artifact has been designed, which it may or may not succeed. Imagine a new kind of genetically engineered bacteria that was “designed” to detect oxygen changes in water, and was so thought be its scientists to have the purpose of “seeking oxygen” more directly than the magnetosome-possessing bacteria by detecting minute changes in oxygen content and following those changes, in a “hill-climbing” fashion. Now imagine this bacteria being put in water where its tactic works wonders, and as it is not confused by errant magnets. Since it didn’t get it’s oxygen-seeking abilities from natural selection, would it count as a proper function? Now imagine that most of the time the bacteria that followed magnetic north in general had a better chance of getting to oxygen-less water, for the purely local presence of a bit more oxygen here and there was not a reliable indicator for oxygenless water. Or simply that the “direct oxygen detection” capabilities of the genetically engineered bacteria just didn’t work that well. Either way, despite its purpose of detecting oxygen, it didn’t get its purpose from its selection history, since its purpose failed to help it reproduce in anyway.

The same general rule holds even more true for technological artifacts than biological organism. For every successful technological artifact, there are countless failures that never replicated, or at least beyond their initial order. For example, many hyper-text systems like HyperG were designed with the purpose to provide better access to documents, yet they failed and the Web succeeded (?). On a smaller scale, the precise words “Ralph has won a free ticket to Paris!” may never have been employed by Daniel before, or heard by Amy, and yet Daniel had a purpose in mind when he sent them to Amy in an e-mail. While it reasonable to hold that natural language itself is
the product of natural and mimetic selection, is there a proper function of a sentence that has never been uttered before? Indeed, it seems unreasonable, but they do have a purpose. So, what is a purpose? One distinction is that something can only have a proper function after reproduction or imitation. Therefore, the first, and original usage of a thing is not its proper function. So, the first magnetosome evolved by the bacteria just happened to accidentally detect magnetic north, but that was not its proper function. It seems absurd to give the accidental creation of a magnetosome by evolution the “purpose” of doing anything. However, there is a difference between natural selection and technological innovation, since technological innovation can try to explicitly fulfill goals, not just through mutation. A purpose is the intended use of a thing as given by its creator, regardless of its history of natural selection or imitation. While these “unnatural” purposes parasite off of the capabilities provided by natural selection, they are nonetheless very real. Also, we should notice that many things spread, especially by imitation, regardless of any proper function. As Millikan notices, “Many conventions seem to have no functions. They seem to proliferate only because people are creatures of habit, or unthinking conformists, or because they venerate tradition, and so forth” (?). From this we can get a definition of convention, such as choosing to drive on the right side of the road as opposed to the left, as the use of a thing based purely on previous history, without regard to imitation or natural selection.

3.5 Representations

By claiming to be a “universal space of information,” the Web is asserting to be a space where any encoding can be transferred about any content (?). However, there are some distinct differences between kinds of content, for some content can be distal and other content can be local. In a message between two computers, if the content is a set of commands to “display these bytes on the screen” then the server can translate these bytes to the screen directly without any care of what those bytes represent to a human user. However, the content of the message distal, such as the string “Ralph won a ticket to the Eiffel Tower in Paris,” which refers to many things outside of the computer. Differences between interpretations and receivers allow the self-same content of a message to be both distal and local. The message to “display these bytes on the screen” could cause a rendering of a depiction on the Eiffel Tower to be displayed on the screen, so the self-same message causes not only a computer to display some bytes but also causes a human agent to receive information about what the Eiffel Tower
looks like. In this regard, the content of the message is dependent both on a receiver and an interpretation of the content.

The kinds of information-bearing messages having distal content are called representations. Representations are then a subset of information-bearing messages, and inherit the characteristics outlined of all information, such as having an encoding scheme, a realization, and often a purpose or proper function. In this scheme, to have some relationship to a thing that one is disconnected from is to be about something else. Generally, the relationship of an thing to another thing to which one is immediately causally disconnected is a relationship of reference to a referent, the distal thing referred to by a representation. The thing which refers to the referent we can generally call a representation, and take this to be equivalent to being a symbol. To refer to something is taken to be synonymous with to denote something, so the content of a representation is its denotation. In the tradition of Bretano, the reference relation is considered intentional due to its apparent physical spookiness. Just as the reverse of interpretation is given by realization, realization can be thought of as the reverse of reference. Both the web-page about the Eiffel Tower and a map about the Eiffel Tower realize some information about the distal Eiffel Tower. A simple explanation of the difference between representation and realization given by Brian Cantwell Smith in the form of a test: If \( X \) represents \( Y \), destroying \( X \) has no effect on \( Y \), but if \( X \) realizes \( Y \), destroying \( X \) also destroys \( Y \) (?). In this matter, a web-page about the Eiffel Tower is a digital encoding whose interpretation can be the ‘Eiffel Tower’ itself, as removing that webpage does not count as blowing up the Eiffel Tower. However, this relationship is due to representations having distal content. Sometimes information may not have distal content, a referent per se. For example, imagine one could have the last copy of *Moby Dick* in the entire world. If that book was blown up, since no other thing realized the information of *Moby Dick*, the content would be gone, along with the encoding, when the realization was destroyed.

The very idea of representation is usually left under-defined as a “standing-in” intuition, that a representation is a representation by virtue of “standing-in” for its referent (?). The classic definition of a symbol from the Physical Symbol Systems Hypothesis is the genesis of this intuition regarding representations (?): “An entity \( X \) designates an entity \( Y \) relative to a process \( P \), if, when \( P \) takes \( X \) as input, its behavior depends on \( Y \).”

There are two subtleties to Newell’s definition. First, the concept of a representation is grounded in the behavior of a process. So, what precisely counts as a represen-
tation is never context-free, but dependent upon the process completing some function with the representation. Second, the representation simulates action at a distance on what is being represented, and so the representation must be local while the referent may be non-local: “This is the symbolic aspect, that having \( X \) (the symbol) is tantamount to having \( Y \) (the thing designated) for the purposes of process \( P \)” (?) We will call \( X \) a representation, \( Y \) the object and referent of the representation, a process \( P \) the subject. This definition does not seem to help us in our goal of avoiding physical spookiness, since it pre-supposes a strangely Cartesian dichotomy between both subject and object and object and referent. To the extent that this distinction is held a priori, then it is physically spooky, as it seems to require the referent and representation to somehow magically line up in order for the representation to serve as a suitable “stand-in” for its missing referent.

The only way to escape this trap is to give a description of how representations arise without the a priori subject-object dichotomy. Brian Cantwell Smith tackles this challenge by developing a theory of representations that explains how they arise temporally and does not assume distinct subjects and objects (?). Imagine Ralph finally gets to Paris and is trying to get to the Eiffel Tower. In the distance, Ralph sees the Eiffel Tower. At that very moment, Ralph and the Eiffel Tower are both physically connected via light-rays. At the moment of tracking, connected as they are by light, Ralph, its light cone, and the Eiffel Tower are a system, not distinct individuals. An alien visitor might even think they were a single individual, a “Ralph-Eiffel Tower” system. While walking towards the Eiffel Tower, when the Eiffel Tower disappears from view (such as from being too close to it and having the view blocked by other buildings), Ralph keeps staring into the horizon, focused not on the point the Eiffel Tower was at before it went out of view, but the point where he thinks the Eiffel Tower would be, given his own walking towards it. Only when parts of the physical world, Ralph and the Eiffel Tower, are now physically separated can they be considered suitable for becoming subjects and objects. The subject can only then use a representation, such as the case of Ralph using an internal “mental image” of the Eiffel Tower to direct his walking towards it, even though he cannot see it. The subject is distinguished from its object by virtue of not only its physical disconnection but by the subject’s attempt to track its object, “a long-distance coupling against all the laws of physics” (?) After disconnection, and possibly more cycles of disconnection and re-connection, the proto-subject can stabilize as a full-blown individual subject and its object as an individual object, with considerable work on the subject’s side to “track” its object.
local physical processes used to track the object by the subject are the representation.

This notion of representation is independent of the representation being either internal or external to the particular subject, regardless of how one defines these boundaries. Imagine that Ralph had been to the Eiffel Tower once before. He could have marked its location on a piece of paper by scribbling a small map. Then, the marking on the map could help guide him back as the Eiffel Tower disappears behind other buildings in the distance as well as, if not better than, any mental image. This characteristic of the definition of representation being capable of including “external” representations is especially important for any definition of a representation to be suitable for the Web, since the Web is composed of information that is considered to be external to its human users.

However fuzzy the details of Smith’s story about representations may be, what is clear is that instead of positing the subject, object, and representation a priori, they are introduced as products of a temporal process. This process is at least theoretically “non-spooky” since the entire process is capable of being grounded out in physics without any “spooky” action at a distance. To be grounded out in physics, all changes must be given in terms of connection in space and time, or in other words, via effective reach. Representations are “a way of exploiting local freedom or slop in order to establish coordination with what is beyond effective reach” (?).

In order to clarify Smith’s story and improve the definition of the Physical Symbol Systems Hypothesis, we consider Smith’s theory of the “origin of the objects” to be a representational cycle with distinct stages (?):

- **Presentation**: Process $S$ is in connected with process $O$. $S$ is the proto-subject that evolves into the subject, while $O$ is the proto–object that evolves into the object.

- **Input**: The process $S$ is connected with $R$. An input procedure of $S$ puts $R$ in some causal relationship with process $O$. This is entirely non-spooky since $S$ and $O$ are both connected with $R$. $R$ eventually becomes the representation.

- **Separation**: Processes $O$ and $S$ change in such a way that the processes are disconnected.

- **Output**: Due to some local change in process $S$, $S$ uses its connection with $R$ to initiate local behavior that is in part caused by $R$.

---

6The defining of “external” and “internal” boundaries is actually non-trivial, as shown in (?).

7In terms of Newell’s earlier definition, $0$ is $X$ while $S$ is $P$ and $R$ is $Y$.
In the “input” stage, the referent is the cause of the content of the representation, what the representation refers to.\(^8\) The relationship of reference is the relationship between the representation and the referent, as given by the interpretation of the content at the “Output” stage. So we have constructed an ability to talk about representations and reference while not presupposing that behavior depends on internal representations or that representations, subjects, and objects exist a priori at all. Representations are only needed when the relevant intelligent behavior requires some sort of distal coordination with a disconnected or “decoupled” object. Representations are physically non-spooky, since they are not a metaphysical or psychological assumption, but exist as part of a rich temporal dynamic that gives rise to dichotomies between subjects, objects, and representations.

The previously defined notions of encoding, content, interpretation, and realization help flesh out the notion of representation. A representation is just a particular kind of information-bearing message whose content is distal, and so the interpretation of a representation is simply an interpretation that is dependent on distal content. In this manner, the act of reference can then be defined as the interpretation of a representation. This would make our notion of representation susceptible to what is known as a correspondence theory of truth (?), where a representation refers by some sort of structural correspondence to some referent. However, our notion of representation is

\(^8\)The referent that causes the content of a representation may change, as a painting of Gustave Eiffel may still represent Gustave Eiffel at some point in the past, even if Gustave Eiffel is no longer existing at the present time.
much weaker, requiring only a causation between the representation - and not just any causal relationship, but one that entails an affordance for the interpreter - as opposed to some tighter notion of correspondence such as an ‘isomorphism’ between a representation and its “target,” the term used by Cummins to describe what we have called the “referent” of a representation (?). So an interpretation should therefore not be viewed as mapping to referents, or even just truth-values. Assertoric content is just one kind of behavior. Furthermore, our definition of distal is merely non-local, and so the referent may exist in the past, the future, or not at all. One can make statements about unicorns and other imaginary objects, as well as the future such as ‘See you next year’ or definite statements whose exact referent such as ‘the longest hair on your head on your next birthday.’ The distal referent of a representation does not imply the existence of the referent or its direct knowledge by the user of a representation, only non-local content.

However, this seems to contradict our ‘input’ stage in the representational cycle, which implies that part of our definition of representation is historical: for every representation there must be a presentation, an encounter with the thing presented. By these conditions, the famous example of Putnam’s example of an ant tracing a picture of Winston Churchill by sheer accident in the sand would not count as a representation by our definition (?). If Ralph didn’t know where the Eiffel Tower was, but walked down streets in Paris at random until he found it, or navigated the streets of Paris and found the Eiffel Tower by reference to a tracing of a Kandinsky painting in his notebook, then Ralph would not then be engaged in any representational behavior, since the Kandinsky painting lacks the initial presentation with the Eiffel Tower. However, the presentation does not have to be done by the subject that encountered the object directly. Representations are usually indirect. A representation that is created by one system in the presence of a referent can be used by another system as a ‘stand-in’ for that referent if the second system shares much of the context, like the ability to interpret the same encoding. It is unlikely that Ralph has been to the Eiffel Tower before, so instead of relying on his own vision or a map he made on his last trip, he buys a map and so relies on the ‘second-order’ representation of the map-maker, who has some historical connection to someone who actually traveled the streets of Paris and figured out where the Eiffel Tower was. In this regard, our definition of representation is very much historical, and the original presentation of the object can be far back in evolutionary time, and the representational content preserved by natural selection, in line with evolutionary accounts as given by Millikan (?). To return to Millikan’s disagreement
with Dretske’s infamous example of the magnetosome-possessing bacteria, according to Millikan the magnetosome detects and sends the bacteria towards magnetic north, so you could say that the distal content of the magnetosome was oxygen-less water, not magnetic north (contra Dretske)?, since in evolutionary terms the magnetosome was selected due to its ability to represent oxygen-less water.

Imaginary referents and referents constructed by definition - and the former is always distal and the latter usually is distal - then also allow some sort of “presence,” it is just that “presentation” of the referent is created via the initial description or imagining of the imaginary referent. Due to “noise” and other factors in the sending of messages, it is always possible that referents does not match to anything existing in the world, however, that does not stop the content of the representation from being both distal and having an effect on the interpreter. For example, exchanging representations of “ghosts” - even if they do not quite identify a coherent class of referents - can govern the behavior of ghost-hunters. Indeed, it is the power and flexibility of representations of these sorts that provide humans the capability to escape the causal prison of their local environment, to plan and imagine the future. Insofar as the information on the Web can be representational, it provides an even more tremendous advantage to its users.

It is not just any sort of representation, but a particular type that the Web provides: digital representation. Thinking back to our representational cycle as defined earlier, for a subject to form a representation, it must have some capability of putting part of itself in some causal relationship with an object. This capability can be viewed as a capacity of the subject to have an imprint of the object in some thing best described as a sort of memory, although the complexities of how precisely this relationship forms may vary tremendously from case to case. There are are some sorts of memory that seem to have special properties that allow the imprint of the object to be flawless and reliably retrieved; this sort of memory is called ‘digital.’ It is a crucial component of the Web, one that differentiates Bush’s Memex from the World Wide Web, is that the information, including representations, that can be transferred over the Web are digital. To define this notion is the central question of the next section.

### 3.6 Digitality

One of the defining characteristics of information on the Web is that this information is digital, bits and bytes being shipped around by various protocols that then get interpreted in various ways. Yet there is no clear notion of what ‘being’ digital consists of.
To get to the heart of this question, a working notion of digitality is necessary. Much like the Web itself, we can know something digital when we spot it, and we can build digital devices, but an encompassing notion of digital is difficult.

One philosophical essay that comes close to defining a notion of digitality is Nelson Goodman’s *Languages of Art*: Given some physically distinguishable marks, which could compose an encoding, (7) defined marks as "finitely differentiable" when it is possible to determine for any given mark whether it is identical to another mark or marks. This can be considered equivalent to how, despite the variation in handwriting, a person recognizes hand-written letters from a finite alphabet. So, equivalence classes of marks can be thought of an application of the philosophical notion of types. This seems close to ‘digital,’ so that given a number of types in a system, a system is digital if any mark of the encoding is interpreted to a discrete type of content. Therefore, in between any two types there can not be an infinite number of other types. Digital systems are the opposite of the famous informal definition of information: Being digital is simply having a difference that does not make difference (7). This is not to say there are characteristics of a mark which do not reflect its assignment in a type. These are precisely its analog characteristics which are lost in digital systems. So in an analog system, every difference in a realization makes a difference, since between any two types there is another type that subsumes a unique characteristic of the token. In this manner, the prototypical digital system is the discrete distribution of integers, while the continuous numbers are the analog system par excellence, since between any real number there is another real number.

Lewis took aim at Goodman’s interpretation of digitality in terms of determinism by arguing that digitality was actually a way to represent possibly continuous systems using the combinatorics of discrete digital states (7). To take a less literal example, discrete mathematics can represent continuous subject matters. This insight caused Haugeland to point out that digital systems are always abstractions built on top of analog systems (7). The reason we build these abstractions is because digital systems allow perfect reliability, so that once a system is in a digital type (also called a ‘digital state’), it does not change unless it is explicitly made to change, allowing both flawless copying and perfect reliability. Haugeland reveals the purpose of digitality to be “a mundane engineering notion, root and branch. It only makes sense as a a practical means to cope with the vagaries and vicissitudes, the noise and drift, of earthy existence” (7). Yet Haugeland does not tell us what digitality actually is, although he tells us what it does, and so it is unclear why certain systems like computers have been
wildly successful due to their digitally (as in the success of analog computers was not so widespread), while others like ‘integer personality ratings’ have not been as successful. Without a coherent definition of digitality, it is impossible to even in principle answer questions like whether or not digitality is purely subjective, a mere matter of opinion (?).

In contrast, it seems sensible to state that certain physical processes have the potential to be digital objectively. Different interpreters can interpret the same physical encoding as “digital” in different ways. The marks “11” can mean eleven in decimal and three in binary notation. So there are multiple ways one can state a system is digital since digitality is a convergence between an abstract mode of interpretation and an objective system that physically implements a correspondence between the possible states of the system and discrete types of content in the interpretation. An interpretation is a **discrete interpretation** when it is a relationship from an encoding to some content where the content itself is also finitely differentiable, so that any given content of a type can be distinguished from any other content of another type. In order to distinguish types, there must be some physical regularity that serves as a boundary. Due to this, digitality then allows some finitely differentiable encoding to map via an interpretation to some finitely differentiable content. When reading letters in a book, we concentrate on the letters, not any minor variations in the quality of the paper - these analog details are left out of our interpretation. It is a convergence between an discrete interpretation, our discrete encoding of the alphabet (and language in general), and a realization of this information in a particular book somewhere in space-time. If we attempt to use an analog system as if it were digital and it does not have the proper physical characteristics, such as writing letters in water, then digitality seems to elude us. Furthermore, all information exists on a level of abstraction from the analog realization of the information. Since all information exists at a level of abstraction from the analog world, on a very deep level no mere concept can do justice to the complexity of the world, where every difference makes a difference.

There are some things that are analog that are not continuous. For example, let us take the phenomenon of being late or early to a meeting, and even ‘on time.’ One cannot be 12 and two-third percent late, early, or ‘on time.’ Fundamentally, there are three discrete types of content, so these types are not analog-as-continuous, for it seems they resist measurement on a continuous scale. With ‘being on time’ and ‘being late’ there is often no difference that makes a difference, so these types could be considered discrete types of an interpretation. One could argue this is a subjective judgment.
Yet there are ‘boundaries’ between the types, just fuzzy boundaries. As Brian Smith has done, one often could make a similar argument around less contentious temporal expressions like ‘day’ and ‘night’ or spatial expressions like ‘in the desert’ and ‘out of the desert.’ Where is the boundary of a desert or the precise boundary between day and night? Despite having a lack of precise boundaries, there are definitely still distinct interpretations of these terms.

Any thing is digital is when a discrete boundary in a particular interpretation of some content converges with a regularity in a physical system. This includes the case if the regularity is, so to speak, fuzzy, and some philosophers like Brian Cantwell Smith hold this “slop” or “fuzziness” to be a fundamental property of the world (?). Conversely, any thing is analog when it is not digital. To implement digitality, there must be a small chance that it can be considered to be in a state that is not part of the discrete types given by the interpretation. The regularities that compose the physical boundary allows within a margin of error this discrete boundary decision to be made in the interpretation. So a system is digital if that buffer created by the margin of error has an infinitesimal chance at any given time of being in a state that is not part of a discrete state. For example, the hands on a clock can be on the precise boundary between markings on the clock, just not for very long. It does requires a continuous amount of time to go from one digital state to another, but this time must be so infinitesimal that it can be ignored by the interpretation. This is why digital systems can have a temporal component such as a double clock. While there are differences that make differences in physics, on the level of the digital those differences don’t make a difference. The boundary between day and night doesn’t take very long, and one definitely knows pretty quickly when one is in the desert and out of the desert.

In a given level of abstraction, the margin of error does not propagate upwards to other levels of abstraction that supervene on the first level of abstraction. This former level of abstraction is ‘first-order’ digital, and other latter levels can be ‘higher-order’ digital. First-order digital is created from analog physics, as we have outlined earlier. There can be higher-order analog on top of lower-order digital states. For example, in order to capture apparently analog music stored in a digital format, one should sample the wavelength twice as often as the highest frequency of the waveform, and this leads the human to have an analog experience of the music. Digital systems interact with and are based on analog systems. This is because information, which on some level of abstraction, can be digital, as it must in the final instance, no matter how many

---

9 An example taken from unpublished lecture notes by Brian Cantwell Smith.
layers of realization and encoding are built into each other, be realized in very concrete and therefore analog realization. Digital things, which like any other information, are realized by the analog world.

Some realizations of information are better than others. Since we can create physical systems through engineering, we can create physical substrata that have low probabilities of being in states that do not interpret to discrete types at a given level of abstraction. As Mueller puts it, “a transistor can be in a voltage state that is clearly of type ‘on’ or ‘off,’ but it can also be on the borderline between the two - it just so happens that our computing machines are made with systems that do not usually get stuck in intermediate states” (?)..

The success of the Web lies in no small part on the vast proliferation of digital computers that allow users to create, store, and retrieve information, and use the Web as a naming space to share this information with others. While, according to Hayles, “the world as we sense it on the human scale is basically analog,” and the Web is yet another development in a long-line of biological modifications and technological prostheses to impose digitalization on analog realizations (?). The vast proliferation of digital technologies is possible because there are physical substrata, some moreso than other, which support the implementation of the digital and give us the advantages that Haugeland rightfully points out: flawless copying and perfect reliability in a flawed and imperfect world (?)

3.7 The Extended Mind Thesis on the Web

With both the history of the Web and the definitions of information, encoding, content, representation, and digitality in hand, the earlier definition of the Web as a universal for digital information should be more clear. The Extended Mind thesis sets the framework for our understanding of the utility of these digital representations on the Web (?). To explain the Extended Mind thesis, Clark introduce us to Otto, a man with an impaired memory who navigates about his life via the use of his notebook, in particular to the Museum of Modern Art (?) . We will rephrase this example in the more familiar terms of Ralph’s visit the Eiffel Tower. Let us assume Ralph has a serious memory impairment. Ralph is trying to navigate to the Eiffel Tower from the airport, and uses his notebook as a surrogate memory in order to discover the location. Ralph has a map in his notebook to the Eiffel Tower made for the precise purpose of navigating individuals to the monument. Ralph can get to the museum with the map, but without the map he
would be lost. In this regard, the map qualifies as an ‘external’ representation that can drive the cognitive processes of an agent in a similar fashion to the way that classical artificial intelligence assumed internal representations did. Interestingly enough, Clark point out that if external factors are driving the process, then they deserve some of the credit: “If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process” (?). The map and other external representations have been dubbed by “cognitive technology” (?).

When computers are mentioned in concrete examples that Clark and Chalmers use in the Extended Mind Thesis, they use a lone person sitting in front of a computer screen. To press upon the Extended Mind thesis in the context of the Web, imagine the world to be inhabited by multiple agents that can access the same information. The obvious example would be two people using the Web to both share a single representation. One could imagine Ralph trying to find his way to the Eiffel Tower and instead of a notebook having a mobile telephone with access to a webpage that contains a map. One could also imagine Ralph’s friend Carine does not have a memory impairment but nonetheless cannot remember her way to the Eiffel Tower unless she has access to the same map via a Web browser on her personal digital assistant. Since they are sharing the exact same representation via digitality and their behavior is successful by virtue of being dependent on the same representation, Carine and Ralph can be said to partially share the same cognitive state. Clark and Chalmers agree that cognition can be socially extended: “What about socially extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle” (?). This socially extended cognition is accomplished via shared external representations.

To push the Extended Mind thesis even further, imagine not only that Ralph and Carine are co-ordinating their behavior via a webpage with a map, but a webpage that gives users the ability to add annotations to the map. The webpage is updated with the annotations viewable by every user in near real-time due to the low latency of Web access. Carine realizes that the main bridge over the Seine to the Eiffel Tower is closed temporarily due to construction and so Ralph should go over the Seine using a bridge a few blocks over, and she adds this annotation to the map. Luckily, with this change to the map Ralph can now find the entrance to the museum, while without it he would have been hopelessly lost. This active manipulation of a representation lets Ralph and Carine partially share a dynamic cognitive state and collaborate for their
greater collective success. Their shared cognitive process is functioning not via telepathy but via shared external representations that are universally accessible over the Web. These Web-based external representations are increasingly important components of intelligence. The Web provides both standardized encodings for information and a universalizing naming scheme needed to access this information.

One of the requirements employed by Clark and Chalmers for something to count as part of the Extended Mind is “reliable coupling” (9). One of the obvious requirements for any process to be part of the Extended Mind is that it accessible when needed to solve some problem. The obvious requirement is that the “external” representation needed by the subject be within its effective reach, not separated from the subject in space or time. So if Ralph’s notebook with the map to the Eiffel Tower has been left at home in Boston when he is in Paris, the notebook can not count as part of his Extended Mind. Furthermore, if his notebook exists only in the past, such that it was destroyed in a fire before Ralph could use it, then the notebook easily can not count as part of Ralph’s Extended Mind. The point here is that at least a minimal condition for anything to be cognitive technology is that be accessible over the bounds of space and time when needed with a reasonable latency - in other words, have “reliable coupling,” (9). The technical trajectory of Licklider’s “Man-Machine Symbiosis” project, which could be considered the engineering twin of the philosophical Extended Mind thesis, is precisely to overcome the barriers of time and space that separate processes in order to enable this reliable coupling required by cognitive technology. The development of the time-sharing computer and then the personal computer allowed temporal divisions to be overcome, and the development of the Internet allowed the division of space to be overcome. Further technological developments have made the latency lower and lower, enabling further reliable coupling with digital information.

The Web can then be considered the realization of the vision of Bush, Licklider, and Engelbart: that built on top of an infrastructure of low-latency digital computing, all information should be accessible to anyone using a single universe of names, and this information should be capable of being linked to by anyone. From this derives the Web’s claim to be universal. Since there will always be some information that is not on the Web, the Web is more accurately not a universal space per se, but a universalizing space, capable of in theory - increasingly in practice - of containing evermore information. At the current moment, the Web is primarily a space of hypertext documents. A more precise explanation of how the Semantic Web employs the Web as a universalizing space for things beyond hypertext documents, and the problems that
the next chapter and the rest of the thesis.

The motivation for this thesis is the presumption that some sort of collective agreement on names on the Web, URIs, is necessary to enable what Engelbart called “collective intelligence” (?). For example, if Ralph and Carine could not access the same web-page because they could not find it’s URI, then they would be unable to solve their problem. If they can agree on whatever names are being used, then they can use these names to co-ordinate their behavior and achieve their goals. The Extended Mind thesis makes this robustly engineering project philosophically justified (?). Yet unlike the examples brought up by Clark and Chalmers in their original thesis, the key to collective intelligence using the Web requires some form of agreement between multiple agents on the meaning of information on the Web given by URIs (?). Determining how agreement can be made and what this entails is far from trivial, and the theoretical and practical ramifications of how agents can reach agreement comprise the central question of the thesis. Before going any further, the thesis will first deploy the intentional terminology defined in this terminology on the architecture of the Web itself, in order to understand how, in the context of the Web, agents could theoretically reach some agreement on URIs before presenting a number of practical proposals on how this could be achieved.
Chapter 4

The Principles of Web Architecture

*You have abandoned the old domain, the old concepts. Here you are in a new domain, for which new concepts will give you the knowledge. The sign that a real change in locus and problematic has occurred, and that a new adventure is beginning, the adventure of science in development.* Louis Althusser (?)

While the significance and history of the Web have been explained, the task remains to show that the Web is a well-defined system with a unique combination of properties, and then how these properties. In Chapter 5 we will demonstrate how these principles can in turn be applied to the Semantic Web.

Can the various technologies that go under the rubric of the World Wide Web be found to have common principles? This question would at first seem to be shallow, for one could say that any technology that is described by its creators, or even the public at large, to be considered trivially ‘part of the Web.’ This shallow answer merely begs the question. To further complicate the matter, the terms like the ‘Web’ and the ‘Internet’ are elided together in common parlance, and so are often deployed as synonyms. In a single broad stroke, we can distinguish the Web and the Internet. The Internet is a type of packet-switching network as defined by its use of the TCP/IP protocol. The purpose of the Internet is to get bits from one computer to another. In contrast, the Web is a space of names defined by its usage of URIs. So, the purpose of the Web is the use of URIs for accessing and referring to information. The Web and the Internet are then strictly separable, for the Web, as a space of URIs, could be realized on top of other types of networks that move bits around, much as the same virtual machine can be realized on top of differing physical computers. For example, one could imagine the Web being built on top of a network built on principles different than TCP/IP, such as OSI, an early competitor to the TCP/IP stack of networking protocols (?). Likewise,
before the Web, there were a number of different protocols with their own naming schemes built upon the Internet like Gopher (?

The problem of defining the Web then can then be divided into two distinct tasks. The first problem is to define the components of the Web that distinguish the Web from the Internet, such as URIs. The definition of these components can then serve as a foundational ontology of Web architecture. In order to be clear and not self-referential, these terms should be grounded in the philosophical terminology from Chapter 3. The second task is to define the various normative principles that govern the relationships between these components. These relationships are the ‘Principles’ of Web architecture. These principles should be normative, as in any program should be capable of being judged as either compliant or not. In order for such a judgment to take place, the principles should be finitely enumerable. Lastly, both the ontology and the principles should be parsimonious, and so kept to a minimum. Lastly, these principles should be compositional insofar as some principles result from the application of other principles, and all operate from a basic terminology. The desiderata of our principles of Web architecture are that these principles are normative, enumerable, parsimonious, and compositional. This combination of principles should ideally defines whatever properties of the Web that distinguish the Web from other computational phenomenon, and also possibly the reasons that led the original Web to experience such phenomenal growth (?).

Is it not presumptuous of us to even hope that such an unruly phenomenon such as the Web even has guiding principles? In this way, again we must appeal to the fact that unlike natural language or chemistry, the Web is like other engineered artifact, created by particular individuals with a purpose, and designed with this purpose in mind. Unlike the case of the proper function of natural language, where natural selection itself will forever remain silent to our questions, the principal designers of the Web are still alive to be questioned in person, and their design rationale is overtly written down on various notes, often scribbled on some of the earliest web-pages of the Web itself. It is generally thought of that the core of the Web consists of the following standards, given in their earliest incarnation, HTTP (?), URI (?), and HTML (?). So the basic protocols and data formats that proved to be successful were the creation of a fairly small number of people, such as Tim Berners-Lee, Roy Fielding, and Dan Connolly. It essence, this chapter then will provide an exegesis, clarification, and systematization of their work, more than any original contribution per se. And while the Web may have evolved, so that whatever purpose that these original architects of the Web had may no longer be
relevant per se, it seems to be the place to beginning is an analysis of the normative documents and informal notes of these original Web architects.

The primary source for our principles of Web architecture is a document entitled *The Architecture of the World Wide Web* (AWWW), a W3C Recommendation edited by Ian Jacobs and Norm Walsh to “describes the properties we desire of the Web and the design choices that have been made to achieve them” (??). The AWWW is an attempt to systematize the thinking that went into the design of the Web by some its primary architects, and as such is both close to our project and an inspiration. In particular, this document is an exegesis of Tim Berners-Lee’s notes on “Design Issues: Architectural and philosophical points”\(^1\) and Roy Fielding’s dissertation “Architectural Styles and the Design of Network-based Software Architectures” (??). The rationale for the creation of such a document of principles developed organically over the existence of the W3C, as new proposed technologies were sometimes considered to be either informally compliant or non-compliant with Web architecture. When the proponents of some technology were told that their particular technology was not compliant with Web architecture, they would often demand that somewhere there be a description of this elusive Web architecture. The W3C in response set up the Technical Architecture Group (TAG) to “document and build consensus” upon “the underlying principles that should be adhered to by all Web components, whether developed inside or outside W3C,” as stated in its charter.\(^2\) Although the membership of the TAG may change, at the time of its inception in 2004 it consisted of Tim Berners-Lee (the Director of the W3C and co-editor of most of the original Web specifications, often acclaimed as the “inventor of the Web”) along with several other key Web specification editors, such as Dan Connolly (co-editor of the original IETF HTML 1.0 specification) and Roy Fielding (co-editor of the IETF HTTP specifications), along with elected representatives, often from companies such as SUN Microsystems and Google. The TAG also maintains a numbered list of problems (although the numbers are in no way sequential) that attempts to resolve issues in Web architecture by consensus, with the results released as notes called ‘W3C TAG findings,’ which are also referred to in this discussion. The TAG’s only Recommendation at the time of writing is the aforementioned *Architecture of the Web: Volume 1* but it is reasonable to assume that more volumes of *Architecture of the Web* may be produced to make that after enough findings have

---

\(^1\)There exist as a collection of unordered personal notes available at: [http://www.w3.org/DesignIssues/](http://www.w3.org/DesignIssues/), which we also refer directly to in the course of this chapter.

\(^2\)Quoted from their charter, available on the Web at: [http://www.w3.org/2001/07/19-tag](http://www.w3.org/2001/07/19-tag) (last accessed April 20th, 2007).
On the Internet, the closest parallel to this attempt of the TAG to create a codified architecture of the Web is the *Architectural Principles of the Internet* of the IETF (?:). It is a list of principles in natural language, including the famous “be strict when sending and tolerant when receiving” principle (?:). One maxim that Carpenter gives is that “we would be foolish to imagine that the principles listed below are more than a snapshot of our current understanding” because “principles that seemed inviolable a few years ago are deprecated today... principles that seem sacred today will be deprecated tomorrow,” ending on the note that “the principle of constant change is perhaps the only principle of the Internet that should survive indefinitely.” (?). This is trivially correct as regards any sort of statement of principles, yet it is empirically incorrect to claim that the Internet is a sort of Heracleitan flux, such that after any change it would still be the ‘Internet.’ If the Internet disposes of the very use of IP addresses, it would not still be the Internet, at least in any recognizable technical form. If *anything* could be part of the Internet, this would include absurdities such as my dog that has never exchanged a single bit with the rest of the Internet. While constant change may describe some underlying metaphysical fact about reality, there must be some collection of relatively stable properties that allow ordinary people to talk about the Web or Internet without being perpetually mired in a cloud of perpetual unknowing. Certain principles such as IP addresses, while incrementally evolving, obey fundamentally the same principles as they did at the foundation of the Internet: Each computer or network should have a distinct IP address, as given by the principle stated as “a single naming structure should be used” and “addresses must be unambiguous” (?). While the *Architectural Principles of the Internet* is of interest, ultimately it is an informal collection of aphorisms in natural language that contains many gems but does not hold up to our all of desiderata for Principles of Web Architecture, although they are at least enumerable (?).

In contrast to the more homely *Architectural Principles of the Internet*, the W3C TAG’s AWWW is a blend of common-sense and sometimes surprising conclusions about Web architecture that attempts to unify diverse web technologies with a finite set of core design principles, constraints, and good practices (?). In a nutshell, the Web is composed of a set of resources that are identified by URIs, which agents can interact with using standardized protocols, usually retrieving representations of the resource via standardized languages. In more concrete terms, if an agent wants to learn about the resource known as the Eiffel Tower in Paris, the agent can access using HTTP a web-page in HTML via the URI http://www.tour-eiffel.fr/. While each of
components is examined later, the W3C TAG does provide at least a standardization of the terminology used in the Web. The AWWW does set out five principles: (?

- Orthogonal abstractions benefit from orthogonal specifications.
- Agents that recover from error by making a choice without the user’s consent are not acting on the user’s behalf.
- Agents do not incur obligations by retrieving a representation.
- Global naming leads to global network effects.
- An application developer or specification author should not require networked retrieval of representations each time they are referenced.

The AWWW also contains ‘Good practices’ that are guidelines set to be less strictly enforced than principles, which range from the detailed “a representation provider should not assign Internet media types beginning with text/ to XML representations” to the broad “a URI owner should provide representations of the resource it identifies” (?). There is no principled way to distinguish constraints, such as “assign distinct URIs to distinct resources,” from the TAG’s principles and good practices, and no systematization of the derivation of either the TAG’s constraints or practices from its principles (?). Furthermore, the principles themselves are rather vague and broad, ranging from error-recovery to the benefits of global naming, and so it is equally unclear if these principles are parsimonious or compositional. Considerable work must be done to make the AWWW match our desiderata for principles of Web architecture.

Besides innumerable works about trends on the Web, the other well-known attempt to formulate the principles of Web architecture is Small Pieces Loosely Joined: A Unified Theory of the Web by the neo-Heideggerian philosopher David Weinberger, but the work fails to deliver on its title (?). His theory of the Web is a series of informal vignettes that give an exposition of his “small pieces” in everyday examples of the effect of the Web on our everyday life (?). Weinberger’s work proceeds without an even coherent definition of terms, much less principles or even argumentation (?). As such, it cannot be used as the foundation for any systematic approach to Web architecture. Another work, entitled Laws of the Web by physicist Bernardo Huberman includes summaries of his ground-breaking studies of the structure and growth of the Web (?). What Huberman invariably found in his experiments was that the Web empirically follows a ‘power-law’ distribution (as generated by function $y = cx^a$) in its growth and
structure, a now well-known finding that the Web is composed of a few hubs with many links (like Yahoo!) and a much larger amount of sites with only a few. This finding also applies to the length of “click-throughs” and the occurrence of spikes in traffic on the Internet. The discovery of power-laws should not be a surprising result and is not unique to the Web. Power-law distributions are a well-known distribution that results from many empirical studies of complex systems where each individual agent is establishing connections in a decentralized manner. This is precisely how the Web works, as users may follow links and add links to sites without centralized control. Second, a power-law distribution or even a generative model for such a distribution is not a normative requirement, but only an empirically observed phenomenon. Huberman does not give any explanation on how architectural design decisions on the Web enabled this sort of distribution to arise. While empirically important, Huberman’s observations cannot function as principles of Web architecture as they are empirical rather than normative, and not unique to the Web.

4.1 Foundational Terminology of the Web

To begin our reconstruction of Web architecture, the first task is the definition of terms, as otherwise the technical terminology of the Web can lead to as much misunderstanding as understanding. To cite an extreme example, people coming from communities like the artificial intelligence community use terms like ‘representation’ in a way that is different from those involved in Web architecture. We begin with the terms commonly associated with a typical exemplary Web interaction. For an agent to learn about the resource known as the Eiffel Tower in Paris, a person can access its representation using its Uniform Resource Identifier (URI) http://www.tour-eiffel.fr/ and retrieve a webpage in the HTML language using the HTTP protocol.

4.1.1 Protocols

An agent is any thing capable of having an interpretation, a broad definition that includes everything from humans to web-browsers and is synonymous with our earlier definition of an interpreter. A protocol is a convention for transmitting information between two or more agents, an equally broad definition that encompasses everything from computer protocols like TCP/IP to conventions in natural language like those employed in diplomacy. A payload is the information transmitted by a protocol. Gal-
loway notes that protocols are “the principle of organization native to computers in distributed networks” and that agreement on protocols are necessary for any sort of network to succeed in the acts of communication (?). The paradigmatic case of a protocol is TCP/IP, where the payload transmitted is just bits in the body of the message, with the header being used by TCP to ensure the lossless delivery of the bytes. TCP/IP transmits strictly an encoding of data as bits and does not force any particular interpretation on the bits; the payload could be a picture of the Eiffel Tower, web-pages about the Eiffel Tower, or just meaningless random bits. All TCP/IP does is move some particular bits from one individual computer to another, and any language that is built on top of the bit-level are strictly outside the bounds of TCP/IP. Since these bits are usually communication with some purpose, the payload of the protocol is almost always information with an interpretation to some content above and beyond that of the raw bits themselves.

Protocols are necessary and not unique to computers. Imagine a network of computers without protocols sending bytes to each other. In this strange network without protocols, the receiver would be receiving messages without even knowing when a single message began and ended. There would be no way for a computer to determine if it received the payload, or if it should pass the payload to another computer, and so coordination of any type would be impossible. The same example applies to non-digital systems such as natural language, where the raw encodings are the analog waveforms emitted by mouths, which can then interpreted by our biological hearing organs as phonemes, words, sentences, and so on. Without a protocol, these signals would just be senseless wave-length of pure sound without beginning or end, sent for reasons unknown. Considering natural language to be a type of protocol, these waves-lengths become a form of communication and co-ordination directed at particular individuals or groups.

In the argument of Lewis, natural languages rely on convention based on previous history to solve a co-ordination problem (?). Millikan notes that conventions survive insofar as they are historically successful in solving some problem, and so have a proper function (?). Yet for an artificial language composed of bits, in order to interpret the bits our machines must be designed or programmed with a convention, created not necessarily out of prior history but through explicit formulation of a design to solve some problem thrown by the world to humans. In fact, a convention for the encoding and interpretation of information is precisely the definition of a standard. Such a standard ensures that a computer can determine when a series of bits arrived, and
whether or not then to decode interpret the bits in some fashion outside their existence *qua* bits, such as using these bits to display a picture of the Eiffel Tower to a user. In this manner, a protocol is a shared convention for communication, necessary for any communication to exist. Since protocols are necessarily accepted over more than one party in order to be used, the creation of protocols like HTTP is the main work of organizations such as the IETF and the W3C, whose diverse members use the organization body as a way to propose and ratify standards.

The Web is based on a **client-server architecture**, meaning that that **protocols take the form of a request for information and a response with information**. The **client** defined as the agent that is requesting information and the **server** is defined as the agent that is responding to the request. In a protocol, an **endpoint** is any process that either requests or responds to a protocol, and so includes both client and servers. The client is often called a **user-agent** since is the user of the Web. A user-agent may be anything from a web-browser to some sort of automated reasoning engine that is working on behalf of another agent, often the specifically human user. The main protocol in this exposition will be the **HyperText Transfer Protocol** (HTTP), as most recently defined by IETF RFC 2616 (?). HTTP is a protocol originally purposed for the transfer of hypertext documents, although its now ubiquitous nature often lets it be used for the transfer of almost any encoding over the Web, such as its use to transfer XML-based SOAP (originally the *Simple Object Access Protocol*) messages in Web Services (?). HTTP consists of sending a **method**, a request for a certain type of information from a user-agent to the server, including information that may change the state of the server. These methods have a list of **headers** that specify some information that may be of used by the server to determine the response. The **request** is the method used by the agent and the headers, along with a blank line and an optional message body.

The methods in HTTP are HEAD, GET, POST, PUT, DELETE, TRACE, OPTIONS, and CONNECT. We will only be concerned with the most frequently used HTTP method, GET. GET is informally considered ‘commitment-free,’ which means that the method has no side effects for either the user-agent or the server, besides the receiving of the response (?). Practically, GET should not cause an exchange of any data of the user-agent with the server, and so a GET method should not be used to change the state of a user-agent, such as charging someone for buying a plane ticket to Paris. To change the state of the information on the server or the user-agent, either PUT (for uploading data directly to the server) or POST (for transferring data to the server that will require additional processing, such as when one fills in a HTML form)
should be used. A sample request to http://www.example.org from a Web browser user-agent is given in Figure 4.1.

The first part of a response from the server then consists of an HTTP status code which is *one of a finite number of codes gives the user-agent information about the server response itself*. The two most known status codes are HTTP 200, which means that the request was successful, or 404, which means the user-agent asked for data that was not found on the server. The first digit of the status code indicates what general class of response it is. For example, the two hundred series (2xx) response codes mean a successful request, although 206 means partial success. The 4xx codes indicate that user-agent asked for a request that the server could not fulfill, while 1xx is informational, 3xx is redirectional, and 5xx means server error. After the status codes there is an entity which is “*the information transferred as the payload of a request or response*” (?). An entity consists of “entity-header fields and... an entity-body” (?). An HTTP response consists of *the combination of the status code and the entity*. These responses from the server can include an additional header, which specifies the date and last modified date as well as optional information that can determine if the desired representation is in the cache and the content-type of the representation. A sample response to the previous example request, excluding the entity-body, is given below in Figure 4.2.

In the response, an entity body is returned. The encoding of the entity body without any reference to its protocol is given by the entity header fields that specify its Content-type and Content-language. These are both considered different languages, as a single webpage can be composed in multiple languages, such the text being given in English with various formatting given in HTML. Every entity body should have its particular encoding specified by the Content-type. *The formal languages that can be explicitly given in a response or request in HTTP are called content types*. In the example response, based on the header that the content type is text/html.
a user-agent can interpret ('display as a web-page') the encoding of the entity body as HTML. Since the same encoding – such as a series of some bits – can theoretically represent many different languages besides HTML, so even if one is given a single payload, a user-agent can only know definitely how to process a message through using the content type, or if no content type is provided, through guessing the content type through various heuristics including looking at the bytes themselves, a process informally called *sniffing*. A user-agent can specify what media types they can prefer, so that a web-server that can only present JPEG images can specify this by also asking for the content type *image/jpeg* in the request.

Content-types in HTTP were later generalized as ‘Internet Media Types’ so they could be applied with any Internet protocol, not just HTTP and MIME (*Multimedia Internet Message Extensions*, an e-mail protocol) (>). A media type consists of a two-part scheme that separates the type and a subtype of an encoding, with a ‘/’ indicating the distinction. Internet media types are centrally registered with IANA at [http://www.iana.org/assignments/media-types/](http://www.iana.org/assignments/media-types/), although certain ‘experimental’ media types (those beginning with “x-“) can be created in a decentralized manner (>). A central registry of media types guarantees the interoperability of the Web, although increasingly new media-types are dependent on extensions to specific applications (plug-ins) in order to run. Support for everything from new markup languages to programming languages such as Javascript can be declared via support of its media type.

To move from concrete bits to abstract definitions, a protocol can be defined and implemented in many different types of ways. In the early ARPANet, the first wide-area network and foundation of the Internet, the protocol was “hard-wired” in the hardware of the Interface Message Processor (IMP), a separate “computer” attached
to computers wishing to be attached to ARPANet (?). As more and more networks multiplied, these heterogeneous networks began using different protocols. While the invention of TCP/IP let these heterogeneous networks communicate, TCP/IP does not interpret messages beyond bits. Further protocols built on top of TCP/IP, such as FTP (File Transfer Protocol) for the retrieval of files (?), Gopher for the retrieval of documents (?), and SMTP (Simple Mail Transfer Protocol) for the transfer of mail (?).

Since one computer might hold many different kinds of information, IP addresses were not enough as they only identified where a particular device was on the network. Thus each protocol created its own naming scheme to allow it to identify and access things on a more fine-grained level than IP addresses. Furthermore, each of these protocols was often associated (via registration with a governing body like IANA, the Internet Assigned Numbers Authority) with particular ports, such that port 25 was used by SMTP and port 70 by Gopher. With this explosion of protocols and naming schemes, each Internet application was its own ‘walled garden.’ Names created using a particular protocols were incapable of being used outside the original protocol, until the advent of the naming scheme of the Web (?).

### 4.1.2 Uniform Resource Identifiers

The World Wide Web is defined as by the AWWW as “an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI)” (?). A naming scheme, not any particular language like HTML, is the primary identifying characteristic of the Web. This naming scheme arose from a need to organize the “many protocols and systems for document search and retrieval” that were in use on the Internet, especially considering that “many more protocols or refinements of existing protocols are to be expected in a field whose expansion is explosive” (?). Despite the “plethora of protocols and data formats,” if any system was “to achieve global search and readership of documents across differing computing platforms,” gateways that can “allow global access” should “remain possible” (?). The obvious answer was to consider all data on the Internet to be a single space of names with global scope.

URIs accomplish their universality over protocols by moving all the information used by the protocol within the name itself. The information needed by to identify any protocol-specific information are all specified in the name itself: the name of the protocol, the port used by the protocol, any queries the protocol is responding to, and
the hierarchical structure used by the protocol. The Web is then first and foremost a naming initiative “to encode the names and addresses of objects on the Internet” rather than anything to do with hypertext (?). The notion of a URI can be viewed as a “meta-name,” a name which takes the existing Internet protocols and addresses and wrapped them in the name itself, a process analogous to reflection in programming languages (?). Instead of limiting itself to only existing protocols, the URI scheme also abstracts away from any particular set of protocols, so that even protocols in the future or non-Internet protocols can be given a URI; “the web is considered to include objects accessed using an extendable number of protocols, existing, invented for the web itself, or to be invented in the future” (?).

One could question why one would want to name information outside the context of a particular protocol. The benefit is that the use of URIs “allows different types of resource identifiers to be used in the same context, even when the mechanisms used to access those resources may differ” (?). This is an advantage precisely because it “allows the identifiers to be reused in many different contexts, thus permitting new applications or protocols to leverage a pre-existing, large, and widely used set of resource identifiers” (?). This ability to access with a single naming convention the immense amount of data on the entire Internet gives an application such as the ubiquitous Web browser a vast advantage over an application that can only consume application-specific information.

Although the full syntax in Backus-Naur form is given in IETF RFC 3986 (?), a URI in a simplified manner a URI can be given as the regular expression $\text{URI} = [\text{scheme} ":"] [\text{hierarchical component}]^* [ "?" \text{query }]? [ "#" \text{fragment}]?$. First, a scheme is the name of the protocol or other naming convention. Unlike protocols, a scheme does not have to be capable of transmitting information. A hierarchical component is the left to right dominant component of the URI that syntactically identifies the resource. URIs are federated, insofar as each scheme identifies the syntax of its hierarchical component. For example, with HTTP the hierarchical component is given by [authority] ["/" ["":" port]? ["/" path component]]*. The authority is a name that may, but does not have to be translatable by the domain name system into an IP address, and is often the name of the server. However, in URI schemes like tel for telephone numbers, there is no notion of an authority for telephone numbers in the scheme. The hierarchical component contains special reserved characters are in HTTP characters such as the backslash for locations as in a file system. For absolute URIs, there must a single scheme and the
scheme and the hierarchical component must together identify a resource such as http://www.example.com:80/main/architect/Gustave_Eiffel in HTTP, which signals port 80 of the authority http://www.example.com with the path component /main/architect/Gustave_Eiffel. The port authority is usually left out, and assumed to be 80 by HTTP-enabled clients. Interestingly enough there are also relative URIs in some schemes like HTTP, where the path component itself is enough to identify a resource within certain contexts, like that of a web-page. This is because the scheme and authority itself may have substituted some special characters that serve as indexical expressions, such as ‘.’ for the current place in the path component and ‘..’ as the previous level in the path component. So, ..Gustave_Eiffel is a perfectly acceptable relative URI. Relative URIs have a straightforward translation into absolute URIs, and it is trivial to compare absolute URIs for equality (≠).

The ‘hash’ (#) and ‘question mark’ (?) are special characters at the end of URI. The question marks to denote query terms, so that a query for Gustave Eiffel’s birthday in URI form could be http://www.example.com/Gustave_Eiffel?birthday=date. The ‘hash’ traditionally declares a fragment identifier, which identifies fragment of the document but according to the TAG, it can also identify a “secondary resource,” which is defined as “some portion or subset of the primary resource, some view on representations of the primary resource, or some other resource defined or described by those representations” where the “primary resource” is the resource identified by the URI without reference to either a hash or question mark (?). The fragment identifier (specified by a ‘hash’ followed by some string of characters) is stripped off for the request to the server, and handled on the client side. Often the fragment identifier causes the local client to go to a particular part of the accessed entity. If there was a web-page about Gustave Eiffel, a sentence containing his birthday date could be identified with the URI http://www.example.com/Gustave_Eiffel#birthday. Figure 4.3 examines a sample URI, http://www.example.org/Gustave_Eiffel#birthday:

The first feature of URIs, the most noticeable in comparison to IP addresses, is that they can be human-readable, although they do not have to be. As an idiom goes, “written on the side of a bus.” In other words, URIs are human-readable and easily memorable, unlike IP addresses. URIs can then in of themselves have a natural language interpretation due to their use of terms from natural language, such as www.whitehouse.gov referring to the White House or the entire executive branch of the United States government. Yet it would be unwise for any human to depend on whatever information they can glean from the URI itself, since to a machine the natu-
ral language terms used by the URI have no interpretation. For an agent, all URIs are opaque, with each URI being just a string of characters that can be used to either refer to or access information, and so syntactically it can only be checked for equality with other URIs and nothing more. This is captured well by the good practice of **URI opacity**, which states that “agents making use of URIs should not attempt to infer properties of the referenced resource” (\(\text{?}\)). To rephrase, we could state that a *URI should never itself have an interpretation, only the information referred to or accessed by that URI should have an interpretation*. This point becomes crucial in trying to determine ‘what a URI identifies’ as inspected in detail in Chapter 6.

Second, a URI has an owner. The **owner** which is *the agent that is accountable for originally determining what the URI identifies*. Usually for URIs schemes such as HTTP, where the hierarchical component begins with an authority, the owner of the URI is simply whoever controls that authority. In HTTP, since URIs can delegate their relative components to other users, the owner can also be considered the agent that have the ability to create and alter the Web representation accessible from the URI, not just the owner of the authority. Each scheme should in theory specify what ownership of a URI means in context of the particular scheme.

There are a few points of contention about URIs, in particular the notion that URIs are ‘context-free.’ If they are context-free, a URI should identify the same resource within any context on the Web. This was such a large point of debate that within the IETF that “Universal Resource Identifiers” had to be transformed into “Uniform Resource Identifiers” (\(\text{?}\)). URIs are close to being context-free in the sense that their meaning should always be the same regardless of context. For URIs, the normal context of identification of information on the Internet (via protocol, location, port number) has been encapsulated and subsumed by its URI itself. However, a URI is not entirely context-free since the actual behavior of the application usage of a URI may
be dependent on its local context. This is admitted by the specification itself, for while URIs are “global in scope” and “interpreted consistently regardless of context...the result of that interpretation may be in relation to the end-user's context,” i.e. their implementation of the protocol, the user's needs, and the capabilities of the machine they are using (?). In this manner, the meaning of a URI can depend on deixis, such as the use of the deitic http://localhost:127 or http://127.0.0.1 to access a web server is on port 127 of the local machine, the machine the user is using themselves when they use that URI.

4.1.3 Resources and Web Representations

While we explained what a URI does in terms of the Internet, we have yet to define what a URI is. To inspect the acronym itself, a Uniform Resource Identifier (URI) is an identifier for a “resource.” Yet this does not solve any terminological woes, for the term “resource” is undefined in the earliest specification for “Universal Resource Identifiers” (?). Berners-Lee has remarked that one of the best things about resources is that for so long he never had to define them (?). Eventually Berners-Lee attempted to define a resource as “anything that has identity” (?). Other specifications were slightly more detailed, with Roy Fielding, one of the editors of HTTP, defining (apparently without the notice of Berners-Lee) a resource as “a network data object or service” (?). However, at some later point Berners-Lee decided to generalize this notion, and in some of his later works on defining this slippery notion of ‘resource,’ Berners-Lee was careful not to define a resource only as information that is accessible via the Web, since not only may resources be “electronic documents” and “images” but also “not all resources are network retrievable; e.g., human beings, corporations, and bound books in a library” (?). This generalization of the term ‘resource’ to refer to things that are not the network-accessible digital things of earlier specifications seems to have first officially in RFC 2396, due to URIs being used in an initiative called the Semantic Web, as explored later in Chapter 5. Also, these resources do not have to be singular but can be a “collection of other resources” (?).

Resources are not only concrete realization or sets of possible realizations at a given temporal junction, but a looser category that includes individuals changing over time, as “resources are further carefully defined to be information that may change over time, such as a service for today’s weather report for Los Angeles”(?). Obviously, a web-page with “today’s weather report” is going to change in content over
time, so what is it that unites the notion of a resource over time? The URI specification defines this tentatively as a ‘conceptual mapping’ (presumably located in the head of an individual creating the representations for the resource) such that “the resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content – the entities to which it currently corresponds – changes over time, provided that the conceptual mapping is not changed in the process” (?). This obviously begs an important question: If resources are identified as conceptual mappings in the head of an individual(s), then how does an agent know, given a URI, what the resource is? Is it our conceptual mapping, or the conceptual mapping of the owner, or some consensus conceptual mapping? This question and further questions of identity come to center stage in Chapter 6. The latest version of the URI specification deletes the confusing jargon of “conceptual mappings” and instead re-iterates that URIs can also be things above and beyond concrete individuals, for resources can be “abstract concepts can be resources, such as the operators and operands of a mathematical equation” (?). After providing a few telling examples of precisely how wide the notion of a resource is, the URI specification finally ties the notion of resource directly to the act of identification given by a URI, for “this specification does not limit the scope of what might be a resource; rather, the term ‘resource’ is used in a general sense for whatever might be identified by a URI” (?). Although this definition seems at best tautological, the intent should be clear. A resource is any thing capable of having identity. The only constraint is upon a resource is that under a given interpretation, an agent could tell whether or not one resource could be differentiated from another. In other words, the agent must have as termed by Gareth Evans a “discriminating concept” of the resource (?). Since information can possibly have any content and so identify any process, in practice within certain protocols that allow access to information, a resource is typically not a particular encoding of the information but the content of the information that can be given by many encodings. To rephrase, the URI identifies the content of any information about the resource, not the encoding of the information. So, a URI identifies a resource usually not like a particular webpage of the Eiffel Tower in HTML, but the notion of a web-page for the Eiffel Tower, even if the web-page were in a language other than HTML. However, while this is best practice on the Web, there is nothing to forbid someone from identifying a particular encoding of information with its own URI and resource. For example, one could also have a distinct URI for a webpage of the Eiffel Tower in English, or a webpage of the
Eiffel Tower in English in HTML. In other words, a resource can identify anything at a level of abstraction, and the same thing, such as a web-page, can be given multiple URIs, each corresponding to a different level of abstraction.

We illustrate these distinctions in a typical HTTP interaction in Figure 4.4, where the content of the information is an image of the Eiffel Tower that a user agent like Ralph’s web browser wants to access via the URI http://www.example.org/EiffelTower/image. While on a level of abstraction a protocol allows an user-agent to identify some content, what the user-agent usually accesses concretely is some realization of that content in a particular encoding, such as a webpage in HTML or a picture in the JPEG language (?). In our example, the URI is resolved using the domain name system to an IP address of a concrete server, which then transmits to the user-agent an entity (which we call a ‘Web representation,’ as explained in the next paragraph) that realizes the resource, i.e. the content identified by the URI. In this example, all the interactions are local, since the webpage realizes the information given by the resource. This entity can then be interpreted by a browser as a rendering on the screen of Ralph’s browser. Note this is a simplified example, as some status codes like 307 may cause a redirection to yet another URI and so another server, and so on possibly multiple times, until an entity may finally be accessed.

Figure 4.4: An user agent accessing a resource

One of the most confusing issues of the Web is that a URI does not necessarily identify a single entity, but can identify multiple entities. This leads to a surprising and little-known aspect of Web architecture known as content negotiation. Content Negotiation is a mechanism defined in a protocol that makes it possible to respond to a
request with different Web representations of the same resource depending on the preference of the user-agent. A “representation” on the Web is then just “an entity that is subject to content negotiation” (?). Historically, the term “representation” on the Web was originally defined in HTML as “the encoding of information for interchange” (?). A later definition given by the W3C did not mention content negotiation explicitly, defining a representation on the Web as just “data that encodes information about resource state” (?). To descend further into a conceptual swamp, “representation” is one of the most confusing terms in Web architecture, as the term “representation” is used differently across philosophy. In order to distinguish the technical use of the term “representation” within Web architecture from the standard philosophical use of the term “representation,” we shall use the term “Web representation” to distinguish it from the ordinary use of the term “representation” as given earlier in Section 3.5. A Web representation is the encoding of the content given by a resource given in response to a request, which must include then the entity headers that specify an interpretation, such as character encoding and media type. A Web representation can be considered to have two distinct components, and the media type that lets us interpret the encoding. Notice that Web representations are realized by bits on actual computers, the servers of the client-server architecture. URIs themselves only give a space of identifiers, but do not in of themselves communicate information due to the opacity of URIs. Instead, any information about a resource with a URI must be communicated by Web representations. This is noted by the TAG, “the Web is designed so that agents communicate resource information state through representations, not identifiers” (?).

Our typical Web transaction, as given earlier in Figure 4.4, can becomes more complex due to this possible separation between content and encoding on the Web. Different kinds of Web representations can be specified by user-agent as preferred or acceptable, based on the preferences of its users or its capabilities, as has been explained in Section 4.1.1. The owner of a web-page about the Eiffel Tower decides to host a resource for images of the Eiffel Tower. The owner creates a URI for this resource, http://www.eiffeltower.example.org/image. While a single URI is used, the content that is encoded in either SVG or JPEG is the same, namely that of an image of the Eiffel Tower, and all hosted Web representations should realize this information, just using different languages. There are two distinct encodings of the image of the Eiffel Tower available on a server in two different iconic languages, one in a vector graphic language known as SVG and one in a bitmap language known as JPEG (?). These depictions are rendered identically on the screen for the user. If
a web-browser only accepted JPEG images and not SVG image, the browser could request a JPEG by sending a request for `Accept: image/jpeg` in the headers. Ideally, the server would then return the web-page with the entity header `Content-Type: image/jpeg`. Had the browser wished to accept the SVG picture as well, it could have put `Accept: image/jpeg, image/svg+xml` and received the SVG version. In Figure 4.5, the user agent specifies its preferred media type as `image/jpeg`. So, both the SVG and JPEG images are Web representations of the same resource, an image of the Eiffel Tower, since both the SVG and JPEG information realize the same information, albeit using different languages for encoding. Since while a single resource is identified by the same URI `http://www.example.org/EiffelTower/image`, different user-agents can get a Web representation of the resource in a language they can interpret, even if they can not all interpret the same language.

![Figure 4.5: An user agent accessing a resource using content negotiation](image)

In Web architecture, content negotiation can also be deployed over not only differing formal languages, but differing natural languages, as the same content can be encoded in different natural languages such as French and English. An agent could request the web-page of the Eiffel Tower from its URI `http://www.example.org/EiffelTower` and set the preferred media type to `'Accept-Language: fr'` so that they receive a French version of the webpage as opposed to an English version. Or they could set their preferred language as English but by using `'Accept-Language: en.'` The preferences specified in the headers are not mandatory for the server to follow, the
server may only have a French version of the resource available, and so send the agent a French version of the web-page regardless of their preference. This extension of content negotiation to operate over different natural languages can be considered controversial. Different natural languages may not be able to encode the same content. Is it really true that two different languages can, even on a high level of abstraction, encode the same information? In some cases, this seems reasonable. Yet it is well-known there are some words in French that are difficult if not impossible to translate into English, such as ‘frileusement.’ Indeed, saying that one natural language encodes the same content as another natural language is akin to hubris in the general case. If this is the case, then it is perfectly reasonable to establish different resources and so URIs for the French and English language encodings of the resource, such as http://www.eiffeltower.example.org/francais and http://www.eiffeltower.example.org/anglais.

In fact, if one believes the same image can not be truly expressed by both SVG and JPEG image formats, one could give them distinct URIs as well. Regardless, what Figure 4.5 shows is that the Web representations are distinct from the resource, even if the Web representations are bound together by realizing the same information given by a resource, since accessing a resource via a single URI can return different Web representations depending on content negotiation.

The only architectural constraint that connects Web representations to resources is that they are retrieved by the same URI. So one could imagine a resource with a URI called http://www.example.org/Moon, that upon accessing using English as the preferred language would provide a web-page with a picture of the moon, and upon accessing with English as the preferred language would provide a picture of blue cheese. While this seems odd, this situation is definitely possible. What binds Web representations to a resource? Is a resource really just a random bag of Web representations? Remember that the answer is that the Web representations should have the same content regardless of their particular encoding if it is accessible from the same URI. This notion depends on our notion of informational content as given in Section 3.2, which we define by an appeal to Dretske’s semantic theory of information (?). To recall, Dretske’s definition of semantic information, “a signal \( r \) carries the information that \( s \) is \( F \) when the conditional probability of \( s \)’s being \( F \), given \( r \) (and \( k \)) is 1 (but, given \( k \) alone, less than 1). \( k \) is the knowledge of the receiver” (?). We can then consider the signal \( r \) to be a Web representation, with \( s \) being a resource and the receiver being the user-agent. However, instead of some fact \( F \) about the resource, we want an interpretation of the Web representation by different the user-agent to be
to the same content. Of course, one can not control the interpretations of yet unknown agents, so all sorts of absurdities are possible in theory. As the interpretation of the same encoding can differ among agents, there is a possibility that the owner of the URI http://www.example.org/Moon really thinks that for French speakers a picture of blue cheese has the same content as a picture of the Moon has for English speakers, even if users of the resource disagree. However, it should be remembered that the Web is a space of communication, and that for communication to be successful over the Web using URIs, it is in the interest of the owner of the resource to deploy Web representations that they believe the users will share their interpretation of. So content negotiation between a picture of blue cheese and picture of the moon for a resource that depicts the Moon is, under normal circumstances, the Web equivalent of insanity at worse or bad manners at best.

The key content negotiation is that the owner of a URI never knows what the capabilities of the user-agent are, what natural and formal languages are supported by it. This is analogous to what Dretske as the “knowledge” or $k$ of the receiver ($?$. The responsibility of the owner of a URI should be, in order to share their resource by as many user-agents as possible, to provide as many Web representations in a variety of formats as they believe are reasonably necessary. So, the owner of the URI for the Eiffel Tower may wish to have a number of Web representations in a wide variety of languages and formats. By failing to provide a Web representation in Spanish, they prevent speakers of only Spanish from accessing their resource. Likewise, by supporting only the newest image format, the owner could prevent users of older web browsers from receiving the same interpretation. Since the maintainer of a resource can not reasonably be expected to predict the capabilities of all possible user-agents, the maintainer of the resource should do their best to communicate their interpretation within their finite means.

The reason URIs identify resources, and not individual Web representations, is that Web representations are too ephemeral to want to identify in of themselves, being by definition the response of server to a particular response and request for information. While one could imagine wanting to access a particular Web representation, in reality what is usually wanted by the user-agent is the content, which may be present in a wide variety of languages. What is important is the content gets transferred and interpreted by the user agent, not the individual bytes of a particular encoding in a particular language at a particular time. With this insight in mind, an attempted definitional unification can be made between Web representations (and, technically speaking, entities in
HTTP responses in general) and the philosophical notion of representation. For a Web representation to count as a philosophical representation, it would have to be effective in notion of having some distal content. The content of a Web representation could be thought to be the resource itself. However, sometimes, like in the case of the information given by an image of the Eiffel Tower, but the Web representations directly realize this information. So some times the Web representation may have distal content and be a philosophical representation, and some times the resource itself may be capable of being realized by Web representations on the Web. We’ll return to these issues in depth in Chapter 6. From a purely practical standpoint, this distinction between Web representations and resources holds particularly true in modern technology, as typical ‘Web 2.0’ transitions involve access of a single URI and so one resource, but use the dynamic update of a Web representations in response to user behavior to create rich Web applications far beyond a single Web representation, like a web-page, being sent over the wire as one big chunk of information.

With this insight in hand, some clarification on the relationship between representations, resources, and URIs should be given. First, a URI may identify only a single resource, as otherwise multiple resources would have both the same URI and the same set of Web representations (with the same content), and so the URI would be indistinguishable. The opposite of this is when the same resource has multiple URIs, which is called URI collision (?). However, a single URI may not be identified only with its currently accessible Web representations, since those representations may change in the future as the resource changes. A resource for the weather in Paris will have to change in order to remain accurate. Likewise, two sets of otherwise identical Web representations may be for different resources. These Web representations may only be represented the same but diverge in the future. A resource for pictures of the tallest monument in Paris would at the time of writing be the same as a resource of the Eiffel Tower but if an even larger monument was built in Paris, then the resources for the capital of France and pictures of Paris should be given distinct URIs.

### 4.1.4 Representational State Transfer

What precise method of interaction allows the Web, unlike previous hypertext systems, to scale globally? The most prominent story is the **Representational State Transfer** (REST) architectural style, as given by Roy Fielding in his doctoral dissertation (?). REST can be defined as an architectural style in which all state where the information
state of the interaction between the server and client is stored on the client. While never officially formalized or given as a standard, REST builds off on the relevant IETF specifications, such as HTTP, which Fielding co-authored (?), and has been a large influence on the W3C’s AWWW (?).

Fielding states that the primary reason the Web scaled globally was that it stored all relevant state on the client, not the server, “All REST interactions are stateless. That is, each request contains all of the information necessary for a connector to understand the request, independent of any requests that may have preceded it” so “each request from client to server must contain all of the information necessary to understand the request, and cannot take advantage of any stored context on the server” (?). This means that unlike object-oriented systems, the resource’s state is not encapsulated in a “black box” (?). On the Web, the entire state of a resource is public and available directly via protocol access mechanisms to a client. Due to this treating all clients the same and the default of storing state publicly on the server, a single server can treat all clients exactly uniformly. Given a HTTP method (such as HTTP GET) and a URI, without any special client-side state, the server always returns the same result to the user. Therefore, the server can perform caching and parallelization much more efficiently, since all the clients can be considered the same and all requests are idempotent.

This phenomenon of caching and proxies is what in turn allows the Web to scale. A URI does not just identify a single set of Web representations, or even multiple sets of a Web representations, on a particular server, for it is a virtual space of names whose underlying relationship to Web representations is flexible. First, multiple servers can host the same resource in parallel and each provide the same set of Web representations to user-agents. The URI http://www.example.org may have multiple servers that host it, with user-agent requests being directed to particular servers based on a criteria such as their spatial location and so optimizing latency. A cache is when a user-agent has a local copy of a Web representation that it accesses in response to a request rather than getting a Web representation from the server itself. For idempotent HTTP methods, this allows the user-agent to press the “Back” button and, assuming REST, receive the exact same Web representation they would have if they had requested it from the server with a very low latency. A proxy is a cache that is not stored on the user-agent itself, but shared among multiple user-agents by a server or group of servers. The proxy can then have a lower latency for some user-agents. Both for user-agents and proxies, caches are kept co-ordinated with the Web representations they are caching by the use of the ETag or Last-Modified headers that specify how long a Web particular representation
may be cached without an update. A proxy or local cache serves a Web representation instantaneously if it still within the specified caching time as given by the client. For example, when a URI such as http://www.example.org/EiffelTower is accessed. Instead of accessing the server(s) that host http://www.example.org/EiffelTower, the user-agent first checks locally and then with a proxy cache to see if a cache with the appropriate representation is available. Due to caching and proxies, to identify a server on the Web with a single computer is a mistake, as a ‘server’ may be spread among multiple computers (proxies) or even the client itself (the cache).

The key insight of REST is that the Web can be considered a distributed finite state machine for information, where a URI specifies precisely one state of the information. In this machine, each URI represents a state, and the state transitions among them to be evoked by moving from URI to URI. Links given by a Web representation to other URIs define the space of possible transitions between states, with an ε transition to any state is transversed whenever a user-agent accesses a URI directly without following a link, such as by typing a URI directly in their browser. REST considers “hypermedia as the engine of application state” (?). Fielding supports this interpretation, albeit not as abstractly, since he notes the “user progresses by selecting links (state transitions), resulting in the next page (representing the next state of the application) being transferred to the user and rendered for their use”(?).

This idea of URIs holding the entire state of the transaction is the end-to-end argument of the Internet architecture in another guise. The end-to-end argument holds that “information about the state of the end-to-end communication” inside the Internet “should be maintained only in the endpoints” such that “the state can only be destroyed when the endpoint itself breaks” (?). So, both the server and client can be considered an endpoint in TCP/IP. This argument was originally phrased in such a way to encourage a network to consist of “smart terminals” and “dumb networks” since for any sort of interoperability between diverse machines, the interpretation of a particular encoding to content should be done not by the terminal, not the network (?). The purposeful restriction of the capacities of the network to just shipping bytes makes sense, since it both simplifies the network architecture, allowing protocols like TCP/IP to focus only on moving bits, while developers of a particular application on the Internet can then focus without worry on what bits have been received at the end-point: “the function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system. Therefore, providing that questioned function as a feature of the communication system itself
is not possible.” (9). The revolutionary move of REST is take the end-to-end argument and move all state to only one end of connection, the client. The other end-point, the server, should always maintain the same state at each URIs it hosts.

This gives REST-compliant systems a level of scalability that distributed object systems lack, since in a distributed object system each object much keep track of its own internal state in regard to its users, which would lead to an exponentially increase in the number of states having to be tracked as the amount of users increase. Historically, the earlier competitor to the Web was the distributed object system CORBA and recently distributed object models have returned in peer-to-peer systems (9). Another aspect of REST that distinguishes it from classic agent-based systems is that information is therefore always moved to processing on the user’s local client, unlike agent-based or mobile code systems that move the code itself over the network to the client (9), although “the benefits of the mobile object style are approximated by sending a representation that consists of instructions in the standard data format” capable of giving the client an interpretation (9).

The REST model by itself is not sufficient for any principles of Web architecture. REST is both too broad and narrow to categorize the Web as a whole. REST is too broad for the Web, since it can be considered not specifically a model of the Web per se, but a software architectural style in general that the Web - for the most part - conforms to. There have been claims that protocols like MODBUS for programmable logic circuits conform to REST (9). Some even claim that a large part of the success of the cut-copy-paste paradigm is due to its unconscious conformance to REST, since a successful cut-copy-paste implementation should use a distinct interface with a finite number of interactions, such that cut, copy, and paste can be considered operations over an abstract interface in the same way that REST considers HTTP PUT and GET operations an abstract interface, and that the operations can only be successful if the object of the ‘cut and paste’ operation is self-described to such an extent that it can be transferred in between applications. Furthermore, it seems REST is also too narrow for the Web. REST is a model of a “hypermedia engine serving application state” while we have earlier defined the Web as a universal information space of names, where hypertext is itself is just one type of information the Web traffics can identify (9).

The Web in practice does not conform normatively to the REST model. Fielding recognizes that “REST is not intended to capture all possible uses of the Web protocol standards” yet he believes any movement away from the REST model is a fundamental error that could not be “optimized for the common case.” (9). As the Web has evolved
over the last decade, it has evolved away from rather than towards REST. Increasingly, servers maintain multiple states in a single resource via features of HTTP like cookies. This violates REST, since “a cookie is defined as being attached to any future requests” for a URI “rather than being associated with the particular application state” (1). The conclusion is the well-known problem that since “the Back button is subsequently used to back-up to a view prior to that reflected by the cookie, the browser’s application state no longer matches the stored state represented within the cookie” (1). In this manner, Fielding’s REST model is surely the correct way to maintain state. Yet almost all modern Web sites use cookies, and although the ‘Back Button’ problem is commonly agreed upon to actually be a problem. Fielding maintains that functions of cookies “should have been accomplished via anonymous authentication and true client-side state,” yet it is unclear how to actually do this (1). While REST described the Web up to the mid-1990s well, it should come as no surprise that, while it is superior to many ad-hoc uses of the Web, REST itself is need of updating. It is beyond the scope of this thesis to provide an update to the architectural style of REST, since at the time of writing of this thesis the Web is too inherently unstable to be characterized by single architectural style. Instead, we focus on the less grandiose task of loosening REST’s constraints when articulating more general Principles of Web Architecture.

4.2 The Principles of Web Architecture

In light of having both the philosophical terminology defined in Chapter 3 and the terminology of the Web defined Section 4.1, it now is possible to show how the various Web terms are related to each other in a more systematic way. These relationships are phrased as five finite principles that serve as the normative Principles of Web architecture: The Principles of Universality, Linking, Self-Description, the Open World, and Least Power. In practice that many applications violate these principles, and by virtue of their use of URIs and the HTTP protocol, many of these applications would be in some sense ‘on the Web.’ However, these Principles are normative insofar as they define what could be considered compliance with Web architecture, and so an application that embodies them is compliant with Web architecture.
4.2.1 Principle of Universality

The **Principle of Universality** can be stated that *any resource can be identified by a URI*. The notion of both a resource and a URI was from their onset universal in its ambition, as Berners-Lee said, “a common feature of almost all the data models of past and proposed systems is something which can be mapped onto a concept of ‘object’ and some kind of name, address, or identifier for that object. One can therefore define a set of name spaces in which these objects can be said to exist. In order to abstract the idea of a generic object, the web needs the concepts of the universal set of objects, and of the universal set of names or addresses of objects” (?). The more informal notes of Berners-Lee are even more startling in their claims for universality, stating that the first ‘axiom’ of Web architecture is “Universality” where “by universal’ I mean that the Web is declared to be able to contain in principle every bit of information accessible by networks” (?). Although it appears he may be constraining himself to only talk about digital ‘objects’ that are accessible over the Internet in this early IETF RFCs, in later IETF RFCs the principle quickly ran amok, as the Semantic Web wanted to use URIs to refer to “human beings, corporations, and bound books in a library” (?).

There seems to be a certain way that web-pages are ‘on the Web’ in a way that human beings, corporations, unicorns, and the Eiffel Tower are not. Accessing a web-page in a browser means to receive some bits, while one cannot easily imagine what accessing the concept of unicorns or the Eiffel Tower in a browser even means. This property of being ‘on the Web’ is a common-sense distinction that separates things like a web-page about the Eiffel Tower from things like the Eiffel Tower itself. This distinction is a matter of between the use of URIs to access and reference, between the local and the distal. The early notes of Berners-Lee that pre-date the notion of URIs itself address this distinction between access and reference, phrasing it as a distinction between locations and names. As Berners-Lee states, “conventionally, a ‘name’ has tended to mean a logical way of referring to an object in some abstract name space, while the term ‘address’ has been used for something which specifies the physical location” (?). So, a **location** is *an term that can be used to access the thing*, while a **name** is *an term that can be used to refer to a thing*. Under this understanding, of course names can server as identifiers by distal things, and so be representations in the sense defined in Section 3.5. However, Berners-Lee immediately puts forward the hypothesis that “with wide-area distributed systems, this distinction blurs” so that “things which at first look like physical addresses...cease to give the actual location of
the object. At the same time, a logical name...must contain some information which allows the name server to know where to start looking” (?). He posits a third neutral term, “identifier” that was “generally referred to a name which was guaranteed to be unique but had little significance as regards the logical name or physical address” (?). In other words, an **identifier** is a term that can be used to either access or refer, or both access and refer to, a thing. The problem at hand for Berners-Lee was how to provide a name for his distributed hypertext system that could get “over the problem of documents being physically moved” (?). Using simple IP addresses or any scheme that was tied to a single server would be a mistake, as the thing that was identified on the Web should be able to move from server to server without having to change identifier.

The way to overcome this problem was to provide a translation mechanism for the Web that could provide a methodology for transforming “unique identifiers into addresses” (?). Mechanisms for translating unique identifiers into addresses already existed in form of the domain name system that was instituted by the IETF in the early days of the expansion of ARPANet (?). Before the advent of the domain name system, the ARPANet contained one large mapping of identifiers to IP addresses that was accessed through the Network Information Center, created and maintained by Engelbart (?). However, this centralized table of identifier-to-address mappings became too unwieldy for a single machine as ARPANet grew, so a decentralized version was conceived based on **domain names**, where each domain name is a specification for a tree structured name space, where each component of the domain name (part of the name separated by a period) could direct the user-agent to more specific “domain name server” until the translation from an identifier to the name to IP address was complete. An example of this would be the translation of the domain name http://www.example.org. Starting from the right and moving left, first the user-agent would look up the domain-name server for org to provide an IP address for mapping to the org domain name server which would provide a mapping for example, which at last provide the actual IP address, such as 152.46.7.122 for http://www.example.org. URIs, by relying on the pre-existing domain name system, had all that was needed to provide mappings for its identifiers to concrete computers.

While the original draft of the URI notion was adamant about the universality of resources, in order to be ratified by the IETF process a distinction was made between URIs which eventually allowed users to access a digital encoding over the Internet and those that did not (?). Many participants in the IETF felt like the blurring of this
distinction that Berners-Lee made was incorrect, so URIs were bifurcated into two distinct specifications. A scheme for locations that allowed user-agents to via an Internet protocol access a realization of information were called **URLs** (Uniform Resource Locations) (?) while a scheme whose names that could refer to things outside of the causal reach of the Internet were called **URNs** (Uniform Resource Names) was created (?). All analog things naturally had to be given URNs, so only digital things can be transmitted over the Internet. Interestingly enough, URNs count as only as a scheme, as opposed to a protocol like HTTP, because they cannot access locally any information. While one could imagine a particular Web-accessible realization, like a web-page, disappearing from the Web, it was felt that identifiers for things that were not accessible over the Web should “be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name” (?). However, while URLs took off explosively with the advent of the Web, despite their cosmic pretensions, URN schemes have only really ever been used to map already centralized names such as “isbn” and “mpeg” and have never had the explosive growth traditionally associated with URLs despite their advantage of being both globally unique and persistent. This is possibly because very few organizations did due to costs in time and effort to register a URN scheme through the IETF, as well as the obscurity of IETF process, in comparison to the straightforward registering of a domain name that could be used as URL. Once a URN scheme was registered, the body that registered the scheme had to guarantee any use of the URN would be persistent and unique and the registering body had complete and total control of the URN scheme. In stark contrast, while there is a centralized managing authority in the form of the top-level domain name system once one has registered a domain name for URLs, the amount of URLs that one can produce under that URL is infinite and not necessarily centrally governed by the owner of the top-level URL. This could be one key to the explosive growth of URLs and the failure of URNs.

It is more likely the over-riding reason for the success of URLs is they provided concrete access to information. URLs and URNs simply reflect a often-ignored philosophical distinction between the use of an identifier for either access or reference. **Reference** is the use of an identifier for a thing to which one is immediately causally disconnected. In other words, the exact same relationship a name has with its distal content is the same relationship a representation has with a referent, one of reference, as discussed in Section 3.5. **Access** is the use of a identifier to create immediately a
causal connection to the thing identified (7). In general, names are thought to be used primarily for reference, while location are used primarily for access. In other words, the use of an identifier as a names is representational, while the use of an identifier to access is not representational, but realizes some local casual relationship between the agent using the name and the thing identified by the name.

Examples abound both in and outside of the Web, especially in natural language. When Ralph is in Boston and says ‘I would like to visit the Eiffel Tower,’ he is using the identifier ‘Eiffel Tower’ in reference, as he is physically at the time of utterance disconnected from the Eiffel Tower. Since reference may be distal in time as well as space, having Ralph refer to ‘Gustave Eiffel’ is also reference. On the Web, URNs are identifiers used for reference and although there were experimental methods for translating URNs to URLs, these never took off (7). One should also note that identifiers used for reference are representational in the sense defined in Section 3.5. Using locations for access is fairly straightforward. A piece of mail has an address such as ‘Henry Thompson at 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland, UK’ in order to be delivered to a particular location and come into contact with the identified Henry Thompson. URLs are obvious locations used for access, such how the URL http://www.tour-eiffel.fr/ is used primarily to access a webpage, as through the domain name system this URL can be translated into a IP address “193.108.167.119” for a server. Interestingly enough, while most names in natural language are used for reference, in contrast most of the use of URIs on the Web are used for access, and so these URIs are URLs. This makes sense, as the class of things that are outside of immediate causal access is much larger than those things that are accessible, and most of natural language is concerned with things outside those things that are immediately accessible.

Just because an identifier can be used for reference does not mean that it can also not be used for access, and vice versa. The use of natural language names for access can be found in imperative sentences such as having Ralph utter or e-mail ‘Amy, come to my office!’ that uses the name ‘Amy’ to bring her into some sort of causal connection. This example also shows identifiers can also be used to refer to states of affairs in the future, and attempt to be made to create that future state. So the identifier ‘Amy’ can be used to both refer to a person and bring her into access. Second, even a location can be used referentially, as stating that “The Eiffel Tower is in Paris” when in Boston. Identifiers for locations on the Internet can also be used referentially, such as by stating that “193.108.167.119 is a computer running a web server” or that “The website
The relationship between reference and access is not exactly symmetrical. While using an identifier for access implies some sort of reference is possible, using an identifier for reference usually does not imply any sort of access. One can refer to things in the past like “the Battle of the Bulge,” imaginary things like “unicorns,” theoretical constructs whose existence is a matter of debate like the “Higgs Boson,” without having to access any of them. For the exact path on how deitic identifiers used primary for egocentric and indexical access can evolve into vaguely context-free identifiers used for reference, one is best to consult the evolutionary argument put forward by Millikan (?), with the most detailed inspection of natural language identification being that of Gareth Evans as put forward in his *Varieties of Reference* (?).

![Figure 4.6: A Venn Diagram describing the relationships between URIs, URNs, and URLs](image)

Building on this observations about the “blurring of identifiers,” the notion of URIs implodes the distinction between identifiers used only for access (URLs) and the identifiers used for reference (URNs). A **Uniform Resource Identifier** is an unique identifier whose syntax is given in (?), that may be used to either or both refer to or access a resource. URIs are a super-set of both URIs and URNs, as shown in Figure 4.6. Berners-Lee and others were only able to push this standard through IETF process years after the take-off of the Web. Anything that may be referred to, and so identified, by an agent may have a URI. However, just because something may be referred to does not mean it already has a URI. There are many concepts that are simply too indexical to be referred to be communicated, such as qualia like the feeling of being a bat (?). Indeed, the space of acts of identification that defines the resources that an agent may wish to give a URI to is biased towards things that can either be accessed (such as web-pages), or relatively uncontroversial things on a low-level of abstraction that
may be referred to, like people, places, and commonly-shared concepts, as we explore in Chapter ???. These things are non-controversial because they are not too indexical, such that they do have a “discriminating concept” that can be used by other agents and so are capable of being re-identified by multiple agents, even in the absence of their presence (?). URIs exist to identify things that one may wish want to communicate about, and a subset of these things that one may wish to communicate about can be directly accessed over the Internet.

So, as shown in Figure 4.7, a URI such as http:///www.example.org/EiffelTower could be used to refer to the Eiffel Tower itself, and the URI http:///www.example.org/EiffelTower/image could allow a user-agent to access an image about the Eiffel Tower. So the retrieved Web representation refers to the Eiffel Tower, and so the URI refers to a distal resource, the Eiffel Tower. The server itself can not of course realize the analog Eiffel Tower itself, but only digital representations of it. Precisely how the TAG believes a URI used to refer to something that is not Web-accessible should itself access a Web representation is addressed in Chapter 6.

![Figure 4.7: URIs for reference](image)

However, should we use HTTP URIs, or some other scheme like URNs, to refer to things? If the gap between reference and access was an unbridgeable chasm, then there would be good reason to invest in separate URIs schemes to separate access and reference. In this manner, URNs could be considered only the first try. Other schemes, such as Masinter’s t.db (“Thing Denoted By”) URI scheme, have been proposed, and entire non-URI systems have also been proposed like the Kahn’s Digital Object Identifier scheme (?). These schemes are not wild-eyed attempts to split the Web into
two over the access and reference divide, as Masinter was one of the co-authors of the original URI specification and Kahn one of the co-inventors of TCP/IP. Both of them assumed that somehow the distinction between reference and access needed to be made in the identifier itself. Yet both these schemes have never taken off, primarily because it seems that since the same identifier can be used for both access and reference, there is a distinct advantage in using HTTP URIs, with their large deployed base of users, as identifiers for both access and reference. While DOIs provided a “handle” system to convert DOIs into HTTP URIs, it just shows that DOIs provide no concrete benefit about HTTP URLs outside of the questionable benefit of a centralized controlling authority, much like URNs (\footnote{\ref{URN}}).

While it may seem strange that HTTP URIs in particular should be used for both accessing digital information and referring to almost anything, it should be stated that the Web is without a doubt the most successful identification system that has been ever been created in the world. There have been many other identification schemes with widespread success, such as telephone numbers, person names, and mailing addresses, yet all of these can easily be incorporated as URIs. The tel scheme makes phone numbers into URIs, such as tel:+1-919-859-9586. If there were some greater standardization of addresses and person names URI schemes could be created for them as well. In particular, HTTP URIs exhibit two features that any distributed identification system should possess: distributed creation and access to information.

From the standpoint of access, any scheme like HTTP must have some kind of global lookup mechanism that can access information to establish the context of use of the identifier. Both URNs and DOIs lack this feature of access, and it is hard to find a better deployed base for access outside HTTP. In the case of URIs, this is the pre-existing Internet domain name system, and the context-of-use can be provided by whatever Web representation can be accessed via the URI. In principle, this access mechanism can be used to resolve debates around reference, as explored in detail in Chapter 6. On a secondary note, such a system should have the ability to structure the identifier, which URIs do with their hierarchical component. One can violate URI opacity and tell from \url{http://www.example.org/Gustave_Eiffel} quite a bit about the relationship between Gustave and the Eiffel Tower. From the standpoint of being distributed, it must since the hierarchical component of a URI is unbounded and so URIs are scalable since there is no upper bound on what can be identified. For a counter-example, IP addresses limits the total amount of addresses, and so versions of the IP address scheme itself must be perpetually updated. Lastly, as the domain name
system itself is distributed, there is no single centralized control, but just enough to make the naming scheme work through the domain name servers.

The notion of that all information accessible on the Internet should share a single naming convention follows straightforwardly the maxim of Internet design that “if there are several ways of doing the same thing, choose one” (?). However, the Internet had previously held up that while “a single naming structure should be used” in any protocol, for the Internet qua Internet, this naming structure was the IP address. IP addresses was found by almost all protocols to be too limiting, as its primary purpose was to ensure the accurate delivery of bit-level messages over the Internet, not for use by particular applications. As detailed earlier, with each protocol inventing its own naming conventions, the Internet did risk fragmenting into mutually incommensurable information spaces, and thus the need for URIs became accepted slowly. The idea of a single global space of names is not new.

Historically, URIs are a return to a pre-Internet principle given by Engelbart known as “Every Object Addressable,” that “in principle, every object that someone might validly want/need to cite should have an unambiguous address (capable of being portrayed in a manner as to be human readable and interpretable)” (?). Engelbart was the first to explicitly propose some sort of global naming scheme for digital networks, although his NLS (oNLine System) system never reached widespread use. Yet components of NLS, such as the use of the ‘hash’ (#) to identify fragments of documents were directly used by the URI scheme. The general vision of some sort of global naming scheme pre-dates both Engelbart and Berners-Lee, going far back in historic imagination to legends of the ‘true names’ of everything being a unique name, perhaps as given in biblical legends about naming. Early proposals for universal names, ranging from Raymond Lull to Leibniz, all missed the crucial advantage of the Web; while classically names in natural language are used for reference, on the Web names can be used to access information (?).

### 4.2.2 Principle of Linking

The **Principle of Linking** states that any URI or Web representation can be linked to another resource identified by a URI. No resource is an island, and the relationships between resources are captured by the linking, transforming lone resources into a Web. A **link** is a connection between resources. The **resource that the link is directed from** is called its **starting resource** while the **resource a link is directed to** is the **ending**
What are links for? Just as URIs may be used for both access and reference, links may be used for access, reference, or both. In particular, in HTML the purpose of links is for access to additional hypertext documents, and so they are sometimes called hyperlinks. This access is often called following the link, a transversal from one Web representation to another, that results in access to Web representations of the ending resource. The ability to concretely access a representation via a link determines that a unidirectional link between only two resources is the predominant kind of link in hypertext, since if a single link could connect multiple ending resources in hypertext, it would be unclear which ending resource an agent following the link would access in a traditional browser. Furthermore, access by linking is transitive, for if an user-agent can access a Web representation of the ending resource from the starting resource, then it can access any links present in the Web representation, and thereby access a Web representation of an ending resource. This following of links to find a particular resource is sometimes called a trail of links. It is precisely this ability to transitively access documents by following links that led the original Web to be a seamless Web of hypertext.

There are several ways in which hyperlinking in HTML, although by far the most well-known usage of linking, is itself just a particular exemplar of the more general concept of linking. Since Web representations are realizations of a resource and a URI identifies a resource, the link is still to and from a resource. To be specific, hyperlinks in HTML are from a particular location within the Web representation, as determined by whatever is spanned by the \texttt{a} element, to another ending resource whose URI is given by the \texttt{href} value. As an aside, since content negotiation demands that the same content be the same to some extent in all available encodings of a resource from a server, every Web representation should in theory maintain either the same links. This is because on a level of abstraction, it is not the Web representations per se that are linked to resources, but the resource that the Web representations encode that are linked to other resources. While particular non-local Web representations are too ephemeral to be given a stable identifier and thus their own URIs, the main motivation for using URIs as the ending resource of a link is to prevent broken links, where a user-agent follows a link to a resource that no longer there, due to the Web representation itself changing. As put by the TAG, “Resource state may evolve over time. Requiring a URI owner to publish a new URI for each change in resource state would lead to a significant number of broken references. For robustness, Web architecture promotes
independence between an identifier and the state of the identified resource” (?).

However, one of the distinguishing features of the Web is that links may be broken by having any access to a Web representation disappear, due to just the lack of hosting a Web representation, loss of ownership of the domain name, or some other reason. These reasons are given in HTTP status codes, such as the infamous 404 Not Found that signals that while there is communication with a server, the server does not host the resource. Further kinds of broken links are possible, such as 01 Moved Permanently or a 5xx server error, or an inability to even connect with the server leading to a time-out error. This ability of links to be ‘broken’ contrast to previous hypertext systems. Links were not invented by the Web, but by the hypertext research community. Constructs similar to links were enshrined in the earliest of pre-Web systems, such as Engelbart’s oNLine System (NLS) (?), and were given as part of early hypertext work by Theodor Nelson (?). Engelbart is specific in his requirement for such “back links” for “when reading a hyperdocument online, a worker can utilize information about links from other objects within this or other hyperdocuments that point to this hyperdocument – or to designated objects or passages of interest in this hyperdocument” and that hypertext systems should guarantee access so “hyperdocuments can be submitted to a library-like service that catalogs them and guarantees access when referenced” (?). The plethora of pre-Web hypertext systems were systematized into the Dexter Reference Model (?). According to the Dexter Reference Model, the Web would not even qualify as hypertext, but as “proto-hypertext,” since the Web did not fulfill the criteria of “consistency,” which requires “in creating a link, we must ensure that all of its component specifiers resolve to existing components” (?).

To insure a link must resolve and therefore not be broken, this mechanism requires a centralized link index that could maintain the state of each resource and not allow links to be created to non-existent or non-accessible resources. Many early competitors to the Web like HyperG had a centralized link index (?). As an interesting historical aside, it appears that the violation of this principle of maintaining a centralized link index was the main reason why the World Wide Web was rejected from its first academic conference, ACM Hypertext 1991, although Engelbart did encourage Berners-Lee and Connolly to pursue the Web further. While a centralized link index would have the benefit of not allowing a link to be broken, the lack of a centralized link index removes a bottleneck to growth by allowing the owners of resources to link to other resources without updating any index besides their own Web representations. This was doubt-

---

3 Personal communication with Tim Berners-Lee
less important in enabling the explosive growth of linking. The lack of any centralized link index, and index of Web representations, is also precisely what search engines like Google create post-hoc through spidering, in order to have an index of links and webpages that enable their keyword search and page ranking algorithms. As put by Dan Connolly in response to Engelbart, “the design of the Web trades link consistency guarantees for global scalability” (?). So, broken links and 404 Not Found status codes are purposeful features, not defects, of the Web.

Links in HTML are just one special case of a wider ontology of links. The most complete reference to possible kinds of links is given in the XLink specification, even if that specification deals only with the XML encoding (?). XLink itself built out of years of work exploring the full range of possible types of links, as described in more detail in the Hytime hypertext standard (?). First, links may be bidirectional, although this case is so rare that we will in general assume, as we have till now, that links are unidirectional. Second, a link is a connection between an arbitrary number of resources, not just two. While the traditional case is a unidirectional connection between a single starting resource and a single ending resource, in theory link can be between multiple starting resources with the link being a ‘dangling’ link going to no resource, or vice versa. However, again this is rarely encountered in actuality. While links in HTML are used primarily for access, there is no reason why they can also be used for reference. For example, there may be an ‘architect of’ relationship between Gustave Eiffel and the Eiffel Tower, and Gustave Eiffel, the Eiffel Tower, and the abstract concept ‘architect of’ that the particular link exemplifies can all be considered non-Web accessible resources. In fact, this usage of links ends up being crucial to the Semantic Web, as detailed in Section 5.2.2.

Since links may be themselves be resources, they may be given identifiers such as URIs. For links that are used for access, these URIs could possibly label the link itself, or other information about the link itself. This is rarely seen. More likely is that the URI of a link describes the class of link, such as the level of accessibility which a link involves. The URI of a link that provides information about the class of the link itself is called an arc role. Another word for link is ‘arc’ and the link describes the ‘role’ of the arc. Both the link for ‘architect of’ and a kind of link that determines level of accessibility would be arc-roles. A URI that can be attached to a link to provide information about the ending resource is called a role. Links that are inserted into Web representations directly and go beyond the local Web representation to an distal ending resources are outbound links. One could have the opposite, where the ending
resource is a local Web representation and the distal starting resource is given by an identifier, and so have inbound links (?). One can also have links from one part of a local Web representation to another local part, as is done in HTML with fragment identifiers. Finally the links can be represented outside of any Web representation of the starting or ending resource, in what is called a linkbase, as pioneered by pre-Web systems like Microcosm (?). This form of link is especially important as it serves as one of the foundations of the Semantic Web, as detailed in Section 5.2.2.

### 4.2.3 Principle of Self-Description

One of the goals of the Web is for resources to be ‘self-describing,’ currently defined as “individual documents become self-describing, in the sense that only widely available information is necessary for understanding them” (?). While it is unclear what “widely-available” means, one way for information to be widely-available is for it to be linked to from the Web representation itself. The **Principle of Self Description** states that the information an agent needs to have an interpretation of a resource should be accessible from its URI.

How many and what sort of links are necessary to adequately describe a resource? A resource is successfully described if an interpretation is a possible, although the definition of a ‘successful’ interpretation is a difficult one, discussed in Chapter ???. Any representation can have links to other resources which in turn can determine valid interpretations for the original resource. Remembering that ‘interpretation’ is defined as broadly in Section 3.3, ranging from a formal interpretation that maps the Web representation onto a mathematical model to an informal interpretation by a human that maps the representation to ‘real-world’ referents. Where the owner of the representation wishes to communicate this interpretation to other agents, one way is to apply the Principle of Linking to connect the Web representation to other resources that ‘fill’ in the necessary background needed for the intended interpretation. These ‘links’ do not have to explicit in the Web representation in a single-manner, such as hypertext links. These ‘links’ may be implicit, depending on the language, and the HTTP protocol defines a number of implicit ‘linking’ mechanisms like media types. This process of following whatever data is linked in order to determine the interpretation of a URI is informally called ‘following your nose’ in Web architecture.

The **Follow-Your-Nose algorithm** states that if a user-agent encounters a representation in a language that the user-agent can not interpret, the user-agent should, in
1. **Dispose of Fragment Identifiers:** As mandated by the URI specification (?), user-agents can dispose of the fragment identifier in order to retrieve whatever Web representations are available from the racine (the URI without fragment identifier). For example, in HTML the fragment identifier of the URI is stripped off when retrieving the webpage, and then when the browser retrieves a Web representation, the fragment identifier can be used to locate a particular place within the Web representation.

2. **Inspect the Media Type:** The media type of a Web representation provides a normative declaration of how to interpret a Web representation. Since the number of IETF media-types is finite and controlled by the IETF, a user-agent should be able to interpret these media types.  

3. **Follow any Namespace Declarations:** Many Web representations use a generic format like XML to in turn specify a customized dialect. In this case, a language or dialect is itself given a URI, called a **namespace URI**, a *URI that identifies that particular dialect*. A namespace URI then in turn allow access to a **namespace document**, a *Web representation that provides more information about the dialect*. In a Web representation using this dialect, a **namespace declaration** then specifies the namespace URI. In this case, the user-agent may follow these namespace declarations in order to get more information needed to interpret the Web representation. As a single Web representation may be encoded in multiple languages, it may have multiple namespace URIs to follow.

4. **Follow any links:** The user-agent can follow any links. There are some links in particular languages that may be preferred, such as the ending resource of a link header in HTML or in RDF Schema links such as rdfs:isDefinedBy links, or links like OWL by the owl:imports. If links are typed in some fashion, each language may define or recommend links that have the normative status, and normative links should be preferred. However, for many kinds of links, their normative status is unclear, so the user-agent may have to follow any sort of link as a last resort.

---

4 The finite list is available at [http://www.iana.org/assignments/media-types/](http://www.iana.org/assignments/media-types/), and a mapping from media types to URIs has been proposed at [http://www.w3.org/2001/tag/2002/01-uriMediaType-9](http://www.w3.org/2001/tag/2002/01-uriMediaType-9).
Using this algorithm, the user-agent can begin searching for some information that allows it to interpret the Web representation. It can follow the first three guidelines and then follow the fourth, applying the above guidelines recursively. Eventually, this recursive search should bottom out either in a program that allows an interpretation of the Web representation (such as a rendering of a web-page or inferences given by a Semantic Web language) or specifications given by the IETF in plain, human-readable text, the natural bottoming point of self-description. This final fact brings up the point that the information that gets one an interpretation is not necessarily a program, but could be a human-readable specification that requires a human to make the mapping from the names to the intended interpretation.

An example should suffice to explain the Principle of Self-Description and the follow-your-nose algorithm. If no Web representation is at http://www.example.org/EiffelTower#loc, it would not be warranted to assume that URI is about the Eiffel Tower due to URI opacity. The browser may drop the fragment identifier loc. If the web-page returned from a HTTP GET on the URI is XHTML (a XML-based version of HTML with some additional capacities (?)), the user-agent may be able to be discover the media type of the language as application/xhtml+xml and process appropriately. If the browser is unable to understand XHTML, the user-agent cannot find a processor, it could retrieve the normative IETF specification for the media-type in the baseline human-readable media-type of text/plain and present that to a human if possible. In an effort to find a machine-readable interpretation of the Web representation, the user-agent would follow the namespace declaration in the XHTML Web representation to find the namespace URI, and so retrieve the Web representation at http://www.w3.org/1999/xhtml. At this namespace, it could retrieve a namespace document, and this namespace document can in turn access the relevant W3C Recommendations or browser plug-ins via linking.

Namespace documents have long been a space of activity within the communities that design Web language. Since each customized dialect is a resource, it follows that each dialect should be given a URI and so have an accessible Web representation. However, what concretely is that Web representation for dialects? It would seem that at least a list of definitions of terms (or ‘names’) in the languages would be useful. In XML, a ‘name’ in an XML dialect is given by adding the namespace URI in front of it with a colon, and so creating a QName or qualified name. However, there is no standardized mapping from QNames to complete URIs (such as from xhtml:li to http://www.w3.org/1999/xhtml#li). Furthermore, there are two readings of how
terms in a language relate to the namespace document. The minimalist reading of namespaces states that anyone can mint a new name by just adding a name and qualifying it by adding the namespace URI in front of it in a document they produce (\(\text{?}\)). The power of defining the “names in a namespace” is in the hands of the user, not the owner of the namespace URI. As noted by Henry Thompson, “The minimalist reading is the only one consistent with actual usage – people mint new namespaces by simply using them in an expanded name or namespace declaration, without thereby incurring any obligation to define the boundaries of some set” (\(\text{?}\)). A maximalist reading of namespaces would state that there is some finite number of names in a language, with the number and usage of names in a namespace is defined by the owner of the namespace URI. The TAG has at least decided that for at least for a namespace it governed for XML, as “namespace consisted all possible local names and that only a finite (but flexible) number of them are defined at any given point in time” (\(\text{?}\)). However, “specifications that define namespaces should explicitly state their policy with respect to changes in the names defined in that namespace” (\(\text{?}\)). So the namespace document should at least tell user-agents if it prefers a maximalist, minimalist, or some other interpretation of names in the namespace. This issue is explored at length elsewhere, but it is key for the Principle of Self-Description that media types and namespace documents exist (\(\text{?}\)).

For the user-agent in our earlier example, just finding a human-readable link to a normative specification is not enough. In order for the Principle of Self-Description to work without human intervention, one would want full-fledged computer programs that can interpret the language of a Web representation to be accessible from the Web representation itself, furthering the seamless space of information the Principles of Universality and Linking are supposed to engender. These interpreters could be attached to a namespace document or from the URI accessible from the media type’s URI, or via a normative link. These should be for a variety of programming languages and platforms, ideally with some that can be installed either “on the fly” with permission (in the manner of browser plug-ins) or available as Web services. While this does present security concerns, these could be addressed via authentication, local policy, and human intervention just as any program installation currently does. To prevent the overloading of certain URIs, caching can be implemented.

The main technical problem with the final guideline in the follow-your-nose algorithm and thus the Principle of Self-Description itself is that a browser needs to be able to know which links to follow in order to find an interpreter for a language it can not
interpret, or at least some automated way to discover whether or not a given Web representation confirms to the language. An informal standard for namespace documents that provides the types of links needed, such as those to applications and normative references, already exists in the form of the RDDL (Resource Directory Description Language) standard (?). RDDL gives standardized links, using both arc-roles and roles in XLink links, for both human and machine consumption (?). Therefore, by combining the accessibility given by the Principle of Universality with the Principle of Linking we can now describe how the Principle of Self-Description can solve decentralized deployment of not only resources and their Web representations, but the languages that given them their interpretation.

4.2.4 The Open World Principle

The Open World Principle states that the number of resources on the Web can always increase. There can always be new acts of identification, carving out a new resource from the world and identifying it with a URI. At any given moment, a new webpage may appear on the Web, and it may or may not be linked to. This is a consequence of the relatively decentralized creation of URIs for resources given by the Principle of Universality and the decentralized creation of links by the Principle of Linking. Without any centralized link index, there is no central repository of the state of the entire Web. While approximations of the state of the entire Web are created by indexing and caching web-pages by search engines like Google, due to the Open World Principle, none of these alternatives will necessarily ever be guaranteed to be complete. Imagine a web-spider updating a search engine index. At any given moment, a new resource could be added to the Web that the web-spider may not have not crawled. So to assume that any collection of resources of the Web can be a complete picture of the whole Web is at best impudent.

The ramifications of the Open World Principle are surprising, and most clear in terms of judging whether a statement is true or false. This repercussions transform the Open World Principle into its logical counterpart, the Open World Assumption, which logically states that statements that cannot be proven to be true can not be assumed to be false. Intuitively, this means that the world can not be bound. On the Web, the Open World Principle holds that since the Web can always be made larger, with any given set of statements that allows an inference, a new statement relevant to that inference may be found. So any agent’s knowledge of the Web is always partial and incomplete,
and thus the Open World Assumption is a safe bet for agents on the Web.

The Open World Principle is one of the most influential yet challenging principles of the Web, the one that arguably separates the Web from traditional research in artificial intelligence and databases in practice. In these fields, systems tend to make the opposite of the Open World Assumption, Closed World Assumption. The **Closed World Assumption** states that logically statements that cannot be proven to be true can be assumed to be false. Intuitively, this means that somehow the world can be bounded. The Closed World Assumption has been formalized on a number of different occasions, with the first formalization being due to Reiter (?)). This assumption has often been phrased as an appeal to the Law of the Excluded Middle (\(\forall p. p \lor \neg p\)) in classical logic (?). Negation as failure is an implementation of the Closed World assumption in both logic programming and databases, where failure for the program to show a statement is true implies the statement is false (?)..

The Closed World Assumption has its history as a solution in temporal reasoning to the **Frame Problem** (?). Originally a problem in reasoning with fluents, the Frame Problem can be more broadly construed: How can one infer the state of the future world from a model of the present world (?)? Philosophically, if a prediction required the entire past world to be formalized into a model, then the process to reasoning about the future would be impossible, since by the nature of abstraction, any given model always excludes some possibly relevant part of the world (?). Computationally, how can a realistic model of the world be computationally tractable, as the amount of facts needed to model the world seems infinite. The obvious engineering answer to both these questions is to assume the world is bounded and finite, which is the Closed World Assumption. As introduced originally in the Stanford Research Institute Problem Solver (STRIPS) system in robotics, the Closed World Assumption holds that anything that is not provable in the model is false (?).

The original Frame Problem may seem distant to the Open World Principle, but they are deeply intertwined. While the Frame Problem is a problem concerning inference between one set of statements, the ‘past,’ and another, the ‘future,’ where the number of statements in the ‘past’ are assumed to be smaller than the number of statements in the ‘future’ and at least some of ‘the future’ can be inferred from the past. Yet, one could randomly select statements from the set of ‘the future’ and combine these with statements from ‘the past’ and commit inferences over this set to a new set. While it is normally assumed that there are less statements about the ‘past’ than the ‘future,’ this is not necessary. The question that both the Open and Closed World assumptions
are trying to address is thus not one about the ‘future’ or ‘past’ but of the relationship between two sets, where it is often convenient to think of the smaller set as the “past” and the larger set as ‘the future,’ where the larger set consists includes inferences from the past. So the relationship between these two sets is arbitrary in the Frame Problem. The problem that the Open and Closed World Assumption then is tackling is the same as the Frame Problem: When is it valid to say a statement is inference is false?

There are a few common misunderstandings of the differences between the Closed World and Open World Assumption. First, the Closed World Assumption does not mean that every statement not present in a particular store of data, like a database, is false. It means it is false if it can not be proven by means of inference. Yet since many relational database query facilities do not provide any inference, data that is not explicit in the database can be assumed to be false. This leads to the notion that programs with a Closed World Assumption do not allow one to express data unless it fulfills the explicit constraints of the program, i.e. in database terminology, any new data must fit in a place where there are rows and columns already for it. In the Closed World, features like integrity constraints that prevent ‘false’ values from being asserted within a model make sense. Yet, under the Open World Assumption statements can still be false, but they just aren’t considered false only when they can’t be proven to be true, the ‘assumption’ in the Open World Assumption. A statement can be proven to be false in a system that implements the Open World Assumption. For example, a statement may include a contradiction, so that the interpretation of that statement may entail everything trivially, and so be considered false. Likewise, an explicit third-value besides true or false, such as “unknown” can be added to systems without violating either the Closed or Open World Assumptions. It is just that with the Open World Assumption, the value of “unknown” is implicit in the assumption the system is making, since a certain statement may be true in some models, and not true in others, and under the Open World Assumption a system cannot assume a particular model is the one and only correct interpretation.

These assumptions can lead to two distinct kinds of reasoning. The Open World Assumption can be stated inferentially that for any set of statements, any inference made from that set should still hold even if a new statement is added to the set. In other words, all reasoning under the Open World Assumption should be monotonic, such that an inference relationship \( \vdash \) is monotonic if and only if for all sets of statements \( s_1 \) and \( s_2 \), and all inferred statements \( s_3 \), if \( s_1 \vdash s_3 \) and \( s_2 \supset s_3 \) then \( s_2 \vdash s_3 \). Otherwise, the agent would face the prospect of having to exhaustively recompute all
inferences, potentially a very expensive task, every time a new statement was added. In other words, the Open World Principle tackles the Frame Problem by holding that any inference from the “past” should not have to be retracted when in the face of the “future.” When monotonicity does hold for a system capable of inference, the system is non-monotonic. In non-monotonic inference, as new elements are added to the set, these additional statements may change the status of statements, such as when statements previously assumed by the Closed World Assumption to be false are added to a system with the Closed World Assumption. This type of reasoning is often used in “common-sense” human reasoning. For example, if one does not have any reason to believe that the Eiffel Tower is in Texas, one usually assumes that statement is false. Under the Open World Assumption, one would just not know that the Eiffel Tower is in Texas. Although, the relationship between monotonic and non-monotonic inference does not strictly parallel the Open and Closed World Assumptions. If the Closed World assumption is stated explicitly and some provenance for this assertion is added in the conclusion (such as in a system that provides belief revision), then inferences under the Closed World can be monotonic. Again, it is the implicit nature of these assumptions that make them assumptions to begin with.

One effect of monotonicity is that one is not allowed to remove assumptions. Can one really assume though that while the Web is always getting larger, the already existing Web is not changing? Indeed, Berners-Lee seems to think this is true in his infamous memo, Cool URIs Don’t Change (?). Berner-Lee notes that “URIs change when there is some information in them which changes” but when “when you change a URI on your server, you can never completely tell who will have links to the old URI” (?). Berners-Lee has two reasons making the case against changing URIs. First, the Web succeeds by ignoring by time. A mere Web representation is considered something too transient to give a URI, since one does not want to identify a resource about the ‘The Eiffel Tower in August for 2006,’ rather one wants to identify a resource about The Eiffel Tower simpliciter, as that is less likely to change and so break the links. The changing of resources can break links, as when an owner fails to host Web representations or loses control of their domain, such as when their ownership expires. This breaking of links cannot be detected to the lack of a centralized link index given by the Principle of Linking. The solution is for the owner of URIs to make what information a URI identifies as stable as possible and maintain all URIs (?). If the information given by the Web representations in a resource are expected to change over time and a user-agent may want to link to this time-specific information, then resource should be given
a separate URI. On the Web history should not disappear, but just given new URIs. Second, on the Web, if a resource disappears, statements that partake of this resource may change, and this violates the ideal monotonicity of inference on the Web. Although URIs are due to their binding to fallible humans and the domain-name system non-persistent, normatively URIs should be made as persistent as possible. Attempts to explicitly index the time of URIs (in dated URI scheme (??)) and create new persistent schemes like URNs have failed in the past. Due to its widespread accessibility mechanism and decentralized creation, it is better to use HTTP ‘as if’ it were a persistent identification scheme, and create new technologies that help manage the lost of URIs, such as the PURL initiative (Persistent Uniform Resource Locator or the ARK system (?)). Ultimately, persistence is a social, not a technical, issue.

4.2.5 Principle of Least Power

The Principle of Least Power states that a Web representation given by a resource should be described in the least powerful but adequate language. This principle is also normative, for if there are multiple possible Web representations for a resource, the owner should chose the Web representation that is given in the ‘least powerful’ language. The Principle of Least Power seems odd, but it is motivated by Berners-Lee’s observation that “we have to appreciate the reasons for picking not the most powerful solution but the least powerful language” (?). The reasons for this principle are rather subtle. The receiver of the information accessible from a URI has to be able to decode the language that the information is encoded so the receiver can an interpretation of the content. Furthermore, an agent may be able to decode multiple languages, but the owner of the URI does not know what languages an agent wanting to access their URI may possess. Also, the same agent may be able to interpret multiple languages that can express the same content. So, the question always facing any agent trying to communicate is what language to use? In closed and centralized systems, this is ordinarily not a problem, since each agent can be guaranteed to use the same language. In an open system like the Web, where one may wish to communicate a resource to an unknown number of agents, each of which may have different language capabilities, the question of which language to deploy becomes nearly insurmountable. Obviously, if an agent is trying to convey some content, then it should minimally choose a language to encode that content which is capable of conveying that content. Yet as languages can usually convey whole ranges of content, and the same content can be conveyed by different
languages, what language to choose?

The Principle of Least-Power is a common-sense engineering solution to this problem of language choice. The solution is just to build first a common core language that fulfills the minimal requirements for use of the content, and then extend this core language. Using HTML as an example, one builds first a common core of useful features such as the ability to have text be bold and have images inserted in general areas of the text, and then as the technology matures, slowly add features such as the precise positioning of images and the ability to specify font size. The Principle of Least Power allows a straightforward story about compatibility to be built to honor “be strict when sending and tolerant when receiving” maxim of the Internet, since it makes the design of a new versions an exercise in strictly extending the previous version of the language (?). When a single resource is to be described using a language, a version of the language that sufficiently expresses content of the resource, but does not include unnecessary features, can then be used. Since the ‘power’ of the core should be minimal, as it allows agents to easily be created that understand the minimal language without implementing the more complex features of ‘more powerful’ languages. Furthermore, if the languages are built from a ‘less powerful’ and a ‘more powerful’ language is deployed to encode a resource, then if some part of the ‘more powerful’ language is not supported by the receiving agent, then at least the information transferred by the common core can be given an interpretation by the agents. This feature is often called graceful degradation.

Although the various versions of HTML do not strictly follow the Principle of Least Power, the Principle of Least Power does determine how processing a HTML document on the Web works. When a web-page given the media type of text/html is received and there is mark-up for terms in another language that the browser does not understand, the browser ignores these terms but still interprets as many of the terms as possible. For example, if a browser capable of understanding only HTML 3.2 encountered a web-page encoded in HTML 4.0, it would not completely fail and reject the entire page. Instead, the browser would interpret, which in this case means rendering to the screen of the client the web-page, as many of the terms as it could, and since many of the terms are overlapping with HTML 3.2, and so a good portion of the web-page should be interpretable. The Principle of Least Power both compliments and provides an alternative for content-negotiation. In content negotiation, usually mutually incommensurable Web representations of a single resource can be hosted, such as a bit-map and vector graphic version of a picture of the Eiffel Tower, or an English and
French version of the same text. In contrast, the Principle of Least Power is a design strategy for choosing the language to encode a single Web representation in, as well as a maxim for the design of new languages. So, if the creator of the Web representation has a choice between a ‘less powerful’ language such as HTML to convey textual information, or convey this information using a more high-powered solution such as a programming language like Java, the Principle of Least Power dictates that HTML should be used.

The purpose of the Principle of Least Power is ultimately to facilitate information re-use. Berners-Lee goes onto claim that “the less powerful the language, the more you can do with data stored in that language,” where he states that the reason he “chose HTML not be a programming language” was “because I wanted different programs to do different things with it” (?). This brief explanation is viewed by many as rather confusing and unintuitive, so the W3C TAG clarified to state the “Rule of Least Power” more concisely: “powerful languages inhibit information reuse” (?). Take for example the normal grounding out of HTML and XML in text as opposed to a binary encoding. While this could be considered inefficient, the ability of many Web languages to be serialized as plain text allows for the famous ‘view source’ capabilities, and that combined with the Principle of the Open World, allow for the ‘cut and pasting’ of information on the Web. In this way, by grounding out all languages on their ‘least powerful’ variant, the Web allows rapid adoption by virtue of low barriers to entry, and easily debugging in the least powerful language. Furthermore, and perhaps most importantly, the least powerful mechanism is suppose to foster information re-use. Ideally, more agents can then interpret the information, create their own extensions of the language, and combine this information with information in other languages.

A gaping hole in the middle of the Principle of Least Power is no consistent definition of the concept of ‘power.’ The W3C TAG, building off of Berner-Lee’s more informal notion of re-use, seems to conflate power with the Chomsky Hierarchy, “Computer languages range from the plainly descriptive (such as Dublin Core metadata, or the content of most databases, or HTML), through logical languages with limited propositional logic (such as access control lists, conneg content negotiation or regular expressions), to the nearly Turing-complete (PDF or some versions of SQL), through those that are in fact Turing-complete though one is led not to use them that way (XSLT, more powerful versions of SQL), through those that are functional and Turing-complete (Haskell), to those that are unashamedly imperative and Turing-complete (Java, Javascript/ECMAScript or C)” (?). While the Chomsky Hierarchy provides one
way of understanding the notion of “power”, it is unclear how it applies to many common cases on the Web, such as HTML, that simply do no in-of-themselves compute, although their interpreter may. This appeal to Turing-complexity has further trouble when considering comparing languages like HTML 4 with HTML 3.2, which adds a \texttt{button} element that allows richer presentation than the \texttt{input} element of HTML 3.2. Also, one Turing-complete programming languages differ widely from each other, and some may offer ‘more features’ than another language. Indeed, although the work needed to formalize a notion of ‘power’ of languages that makes finer-grained distinctions than allowed in the Chomsky hierarchy is an open research problem, there are domains of languages, such as logic, that have clear notions of power. We return this in Section 5.2.5.

4.3 Conclusions

The Web, while to a large extent being an undisciplined and poorly-defined space, does contain a set of defining terms and principles. While previously these terms and principles have been scattered throughout various informal notes, IETF RFCs, and W3C Recommendations, in this chapter we have systematized both the terminology and the principles in a way that reveals how they internally build of of each other. The next question is how can these principles be applied to domains outside the hypertext Web, and this will be the topic of Chapter 5, as we apply these principles to the notion of a knowledge representation language for the Web.
Chapter 5

The Semantic Web

*All the important revolutions that leap into view must be preceded in the spirit of the era by a secret revolution that is not visible to everyone, and still less observable by contemporaries, and that is as difficult to express in words as it is understand.*

G.W.F. Hegel (?)

The Web is a universal information space, but so far it has been one composed entirely of hypertext documents. As said by Berners-Lee at the WWW conference in 1994, “to a computer, then, the web is a flat, boring world devoid of meaning...this is a pity, as in fact documents on the web describe real objects and imaginary concepts, and give particular relationships between them” (?). The heart of this particular insight is the realization that it is content of the information, not its encoding in hypertext, that is of central importance to the Web. The purpose of the architecture of the Web is to connect information of any kind in a decentralized manner, and this architecture can be applied beyond the hypertext of its initial incarnation.

The next step in Berners-Lee’s programme to expand the Web beyond hypertext is called the *Semantic Web*, a term first used by Foucault in *The Order of Things* (?). The most cited definition of the Semantic Web is given by Berners-Lee et. al. as “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation” (?). How can information be added to the Web without encoding it in hypertext? The answer is to find a language capable of representing the information about “real objects and imaginary concepts.” This requires a *knowledge representation language*, a language whose primary purpose is the representation of non-digital content in a digital formal language.
As the previous exposition of Web architecture explained in detail, resources on the Web are given by a URI that identifies the same content on the Web across different encodings. What drives the Semantic Web is the realization that at least some of the information on the Web is representational, i.e. information about distal content. Then instead of the hypertext language, which is mainly concerned with the presentation and linking of natural language for humans, the Web needs a knowledge representation language that describes the represented content as fully as possible without regard to presentation for humans. The mixture of information about some content with information for human presentation forces web-spiders to “scrape” valuable content out of hypertext. In theory, encoding information directly in a knowledge representation language gives a spider more reliable and direct access to the information. As Berners-Lee puts it, “most information on the Web is designed for human consumption, and even if it was derived from a database with well defined meanings (in at least some terms) for its columns, that the structure of the data is not evident to a robot browsing the web” (\cite{Berners-Lee2004}). This has led him to consider the Semantic Web to a Web “for expressing information in a machine processable form” and so making the Web “machine-understandable” (\cite{Berners-Lee2004}). This leads to the contrast between the Semantic Web as a ‘web of data’ as opposed to the hypertext ‘web of documents.’ W3C standards such as XML were originally created, albeit rarely used, precisely in order to separate content and presentation (\cite{Berners-Lee2004}).

Furthermore, the purpose of the Semantic Web is to expand the scope of the Web itself. Most of the world’s digital information is not natively stored in hypertext. Instead, it is stored in databases and other non-hypertext documents and spreadsheets. While this information is slowly but surely migrating towards the Web, as more and more of this information is being exposed to the Web via scripts that automatically and dynamically converts data from databases into HTML, the Semantic Web imagines that by having a common knowledge representation language across the entire Web, all sorts of information that previously were not on the Web can become part of the Web. This makes the Semantic Web not a different and parallel Web to the hypertext Web, but an extension of the current Web, where hypertext serves as just one possible language.

### 5.1 A Brief History of Knowledge Representation

The creation of the Semantic Web then depends on the creation of a knowledge representation language for the Web, and so the Semantic Web inherits both the successes
and failures of previous efforts to create knowledge representation languages in artificial intelligence. The earliest work in digital knowledge representations was spearheaded by John McCarthy’s attempts to formalize elements of human knowledge in first-order predicate logic, where the primary vehicle of intelligence was to be considered some form of inference (?). These efforts reached their apex in Hayes’s “Naive Physics Manifesto,” which called for parts of human understanding to be formalized as first-order logic - since although actual physics was best understood using mathematical techniques such as differential equations, Hayes conjectured that most of the human knowledge of physics, such as “water must be in a container for it not to spill” could be conceptualized better in first-order logic (?). Hayes took formalization as a grand long-term challenge for the entire AI community to pursue, “We are never going to get an adequate formalization of common sense by making short forays into small areas, no matter how many of them we make” (?). While many researchers took up the grand challenge of Hayes in various domains, soon a large number of insidious problems were encountered, primarily in terms of the expressivity of first-order logic and its undecidability of inference. In particular, first-order logic formalizations were viewed as not expressive enough, being unable to cope with temporal reasoning as shown by the Frame Problem, and so had to extended with fluents and other techniques (?). Since the goal of artificial intelligence was to create an autonomous human-level intelligence, another central concern was that predicate calculus did not match very well with how humans actually reasoned. For example, humans often use default reasoning, and various amendments must be made predicate calculus to support this (?). Further efforts were made to improve first-order logic with temporal reasoning to overcome the Frame Problem, as well as the use of fuzzy and probabilistic logics to overcome issues brought up by default reasoning and the uncertain nature of some knowledge (?).

Under increasing criticism from its own former champions like McDermott, first-order predicate calculus was increasingly abandoned by those in the field of knowledge representation (?). McDermott pointed out that formalizing knowledge in logic requires that all knowledge be formalized as a set of axioms and that “it must be the case that a significant portion of the inferences we want...are deductions, or it will simply be irrelevant how many theorems follow deductively from a given axiom set” (?). McDermott found that in practice neither can all knowledge be formalized and that even given some fragment of formalized knowledge, the inferences drawn are usually trivial or irrelevant (?). Moving away from first-order logic, the debate focused on what was
the most appropriate manner for AI to model human intelligence. Some researches championed a *procedural* view of intelligence that regarded the representation as itself irrelevant if the program could successfully solve some task given some input and output. This contrasted so heavily contrasted with earlier attempts to formalize human knowledge that it was called as the *declarative versus procedural* debate. Procedural semanticist Terry Winograd stated that “the operations on symbol structures in a procedural semantics need not correspond to valid logical inferences about the entities they represent” since “the symbol manipulation processes themselves are primary, and the rules of logic and mathematics are seen as an abstraction from a limited set of them” ( harms). While the procedural view of semantics first delivered impressive results through programs like SHRDLU ( harms), since the ‘semantics’ were ad-hoc and task-dependent, so that they could not be used outside the limited domain in which they were created. Furthermore, there became a series of intense debates on whether these programs often purported to do what they wanted even within their domain, as Dreyfus critiqued that it was ridiculous that just because a program was labeled ‘understand’ that it did actually in any way understand ( harms). Interestingly enough, the debate between declarative and procedural semantics is, under the right formal conditions, a red herring since the Curry-Howard Isomorphism states that given the right programming language, there is a tight coupling between logical proofs and programs so that the simplification of proofs can be equivalent to steps of computation ( harms).

Within AI, research began into other forms of declarative knowledge representation languages besides first-order logic that were supposed to be in greater concordance with human intelligence and that could serve as more stable substrates for procedural knowledge-based systems. Most prominent among these alternatives was **semantic networks**, a “a graphic notation for representing knowledge in patterns of interconnected nodes and arcs” ( harms). Semantic networks are as old as classical logic, dating back to Porphyry’s explanation of Aristotelian categories ( harms), although their first self-described usage was as a common knowledge-representation system for machine-translation systems by Masterman ( harms). Motivated by a correspondence with natural language, semantic networks were used by many systems in natural language processing, such as the work of Wilks in resolving ambiguities using preference semantics and the work of Schank using conceptual dependency graphs to discover identical sentences regardless of their syntactic form ( harms). Soon semantic networks were being used to represent everything from human memory to first-order logic itself ( harms). The approach of semantic networks was given some credence by the fact that often when
attempting to make diagrams of ‘knowledge,’ humans often start by drawing circles connected by lines, with each component labeled with some human-readable description. A semantic network about ‘The architect of Eiffel Tower was Gustave Eiffel’ is given in Figure 5.1. Note that it refers declaratively to things in the world, but uses ‘natural-language-like’ labels on its nodes and edges.

When researchers attempted to communicate or combine their knowledge representation schemes, no-one really knew what the natural language description ‘meant’ except the author, even when semantic networks were used as a formal language. The ‘link’ in semantic networks was interpreted in at least three different ways (?) and no widespread agreement existed on the most common sort-of link, the IS–A link, which could represent both subclassing, instantiation, close similarity, and more. This led to an assault on semantic networks by champions of first-order logic like Hayes, who believed that by providing a formal semantics that defined ‘meaning’ first-order logic at least allowed knowledge representations to be transportable across domains, and that many alternative knowledge representations could be re-expressed in first order-logic (?). In response, the field of knowledge representation bifurcated into separate disciplines. Some researchers removed any natural language labels from their semantic networks, leading to the evolution of neural networks and machine-learning (?). Many of the former champions of logic currently do not believe that human intelligence can be construed as logical inference, but researchers still actively pursue the field as it is of crucial importance to many systems such as mathematical proof-proving and it still used in many less ambitious knowledge-reasoning systems such as ISO Common Logic (?).

The classical artificial intelligence programme, while fixated on finding a formal language capable of expressing human knowledge, had ignored the problem of infer-
ence. This problem came to abrupt attention when KRL, one of the most flexible of knowledge representation languages pioneered by Winograd was found to have intractable inference even on simple problems of cryptarithmetic, despite its representational richness. Furthermore, while highly optimized inference mechanisms existed for first-order logic, even first-order predicate logic is undecidable. These disadvantages of alternative representational formats and first-order logic led many researchers, particularly those interested in an alternative “slot and value” knowledge representation language known as frames to begin researching the decidability of their inference mechanisms (?). This research into frames then evolved into research on description logics, where the trade-offs between the tractability and expressivity where carefully studied (?). The goal of the field was to produce a logic with decidable inference while maintaining maximum expressivity. Although the first description-logic system, KL-ONE, was proven to have undecidable inference for even subsumption, later research produced a vast proliferation of description logics with carefully categorized decidability and features (??).

Ultimately, the project of artificial intelligence to design a single knowledge representation system suitable for creating human-level intelligence has not yet succeeded and progress seems glacial at best. With no unifying framework, the field of artificial intelligence itself fragmented into many different diverse communities, each with its own family of languages and techniques. Researchers into natural language embraced statistical techniques and went back to practical language processing tasks, while logicians have produced an astounding variety of different knowledge representation languages, and cognitive scientists moved their interests towards dynamical systems and specialized biologically-inspired simulations. The lone hold-out seemed to be the Cyc project, which continued to pursue the task of formalizing all 'commonsense' knowledge in a single knowledge representation language (?). In one critique of Cyc, Smith asked instead what lessons knowledge representation languages could learn from hypertext, “Forget intelligence completely, in other words; take the project as one of constructing the world’s largest hypertext system, with Cyc functioning as a radically improved (and active) counterpart for the Dewey decimal system. Such a system might facilitate what numerous projects are struggling to implement: reliable, content-based searching and indexing schemes for massive textual databases,” a statement that strangely prefigures not only search engines, but the revitalization of knowledge representation languages due to the Semantic Web (?).
5.2 The Resource Description Framework (RDF)

What makes knowledge representation language on the Web different from classical knowledge representation? Berners-Lee’s early thoughts, as given at the first World Wide Web Conference in Geneva in 1994, were that “so adding semantics to the Web involves two things: allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values” (?). Having information in “machine-readable forms” requires a knowledge representation language that has as some sort of relatively content-neutral language for encoding (?). The parallel to knowledge representation in artificial intelligence is striking, as it also sought to find one universal encoding, albeit encode human-intelligence. The second point, of “allowing links,” means that the basic model of the Semantic Web will be a reflection of the Web itself: The Semantic Web consists of connecting resources by links. The Semantic Web is then easily construed as a descendant of semantic networks from classical artificial intelligence, where nodes are resources and arcs are links. Under the aegis of the W3C, the first knowledge representation language for the Semantic Web, the Resource Description Language (RDF) was made a W3C Recommendation, and it is clearly inspired by work in AI on semantic networks. This should come as no surprise, for RDF was heavily inspired by the work of R.V. Guha on the Meta-Content Framework (?). Before working on MCF, Guha was chief lieutenant of the Cyc project, the last-ditch Manhattan project of classical artificial intelligence (?). There are nonetheless some key differences between semantic networks and RDF, as RDF was built in accordance with the Principles of Web Architecture as given in Chapter 4, as detailed in the next subsections.

5.2.1 RDF and the Principle of Universality

Semantic networks were hobbled with because of their use of ambiguous natural language terms to identify their nodes and arcs, which became a problem when they were transported between domains and different users, a problem that would be fatal in the decentralized and multi-lingual environment of the Web ?. According to the Principle of Universality, since a resource can be anything, then a component of the knowledge representation language should be considered a resource, and thus can be given a URI. Instead of labeling the arcs and nodes with natural language terms, in RDF all the arcs and nodes can be labeled with URIs. Although few applications had ever taken advantage of the fact before RDF, due to the Principle of Universality, URIs could be minted
for things like the Eiffel Tower *qua* Eiffel-Tower, an absolute necessity for knowledge representation. Since the content of statements in knowledge representation is usually about content in the world outside the Web, this means that Semantic Web crucially depends on the rather strange fact that URIs can be minted for things outside the Web.

This does not restrict the knowledge-representation language to merely refer to things that normally we would consider outside of the Web, since normal web-pages use URIs as well, and so the Semantic Web can easily be used to refer to normal web-pages. This has some advantages, as it allows RDF to be used to model the relationships between web-accessible resources, and even mix kinds of relationships. This sort of “meta-data” is exemplified by the relationship between a web-pages and its human author, which in with RDF would both be denoted by URIs. Lastly, this ability to describe everything with URIs leads to some unusual features, for RDF can then model its own language constructs using URIs, and make statements about its own core language constructs. However, just as all components of RDF may be considered resources, just as all resources may not have URIs, all components of RDF may not have URIs. For example, a string of text or a number may be a component of RDF, and these are called *literals* by RDF. In RDF specified anonymous resources are not given a URI, and these are called *blank nodes*. Furthermore, all sorts of information may be accessible via a URI in RDF. It would be premature declare that the deployment of URIs in RDF signal a major improvement over the natural language labels, for URIs can be just as ambiguous as natural language labels. A further analysis of the scope of this problem is in Chapter 6.

### 5.2.2 RDF and the Principle of Linking

The second step in Berners-Lee’s vision for the Semantic Web, “allowing links to be created with relationship values,” follows straightforwardly from the application of the Principle of Universality to knowledge representation. Since RDF is composed of resources, and any resource may link to another resource, then any term in RDF may be linked to another term. This linking forms the heart of RDF, as it allows disparate URIs to be linked together in order for statements in RDF to be made. The precise form of a statement in RDF is a *triple*, which consists of two resources connected by a link, as shown in Figure 5.2. While the W3C standardized encoding of RDF is in a verbose XML format called “RDF/XML”, a simple encoding called *Turtle* for triples is much more brief. In Turtle, a triple is three space-delimited terms (the subject,
There are some restrictions to linking on the Semantic Web. As opposed to the vast numbers and kinds of links possible in XLink, linking on the Semantic Web is directed, like hyperlinks (?). The resource which is linked from in the triple is called the subject, while the link itself is called the predicate, and the resource that is linked to in the triple is the object. The predicate usually is a role as opposed to an arc role. The major restriction on the Semantic Web is that the subject must be a URI or a blank node, and the predicate must also be a URI. The subject, on the other hand, is given the most flexibility, as it may be either a URI, a blank node, or a literal. This predicate-argument structure is a well-known and familiar structure from logic, linguistics, and cognitive science. Triples resemble the binary predicates in propositional logic needed to express facts, relationships, and the properties of individuals. Furthermore, triples seem similar to simple natural language sentences, where the subject and objects are nouns and the predicate is a verb. In cognitive science, there has even been speculation that such a subject-predicate-object structure might be even present in our neural structure (?).

From the perspective of the traditional Web, the main feature of RDF is that links in RDF themselves have a required role URI. It is through this role that URIs are given to relationships outside the Web in RDF. For example, the relationship of ‘is architect of’ between Gustave Eiffel and the Eiffel Tower could be formalized as a link (as shown in Figure 5.2), as could the relationship between Tim Berners-Lee and the
creation of his web-page. Since the relationships are abstract, these URIs then refer to these relationships, the URIs may not be accessible, and RDF predicates are unlike links in traditional hypertext systems. A set of RDF triples is essentially a linkbase, such as those pioneered in earlier hypertext systems like Microcosm (\(^2\)). Similarly, a triple by itself can only state a simple assertion, but webs of links may be made between triples to explain. A set of triples that share resources is called a graph, as illustrated in Figure 5.3 by two triples having the same subject, namely that ‘The Eiffel Tower’s in Paris has as an architect Gustave Eiffel.’

The inspiration of RDF to Berners-Lee comes more from relational databases than linguistics (\(^2\)). Triples can be thought of as the most elementary information that can be structured in a relational databases, where rows are subjects, columns are properties, and values are objects. URIs then can be seen to serve as a space of globally unambiguous identifiers, analogous to the use of keys in relational databases. In databases the foreign key is usually local to the database, so one uses a number of constraints whose sum serves as a foreign key. The Semantic Web hopes to bypass this work by using a URI to be a ‘globally unique key’ for integration, and so allowing data from different sources to combined without any further effort beyond re-using or creating new URIs for the information described in the sources.

With the ability to make separate statements using URIs, the main purpose of RDF is revealed to be information integration. Due to their reliance on URIs, RDF graphs can graph merge, when two formerly separate graphs combine with each other when they use any of the same URIs. The central purpose of URIs is to serve to allows independent agents to make statements about the same referent. With a common language of URIs, agents can merge information about the referents of the URIs in a decentralized manner. This is one of the most important applications of the Semantic Web, and it will be further explored in Chapter 6.
5.2.3 RDF and the Principle of Self-Description

Once the Principle of Universality and the Principle of Linking are obeyed, the Principle of Self-Description naturally follows, and RDF is no exception. Self-description is a crucial advantage of RDF in decentralized environments, since an agent by following links can discover the context of a triple needed for its interpretation. As witnessed by the Brachman and Smith survey of knowledge representation systems, a bugbear of semantic networks was their inability to be used by transferred outside of the closed domain and centralized research group that designed them (?). The crucial context for usage of a particular semantic network was always lost in transfer, so that what precise “IS-A” means could vary immensely between contexts, such as the difference between a sub-class relationship or individual identity (?). By providing self-description, RDF triples can be transported from one context to another, at least in an ideal world where normal conditions, such as when the URIs in the triple can be used to access a webpage describing its content, and correct media types are used. Furthermore, as RDF is imagined to be used as a basic meta-language for other dialects, these dialects can also have their intended interpretation discovered by the follow-your-nose algorithm.

The hypertext Web, when every resource is linked together, provides a seamless space of linked documents. For example, the W3C tries to deploy its own internal infrastructure in a manner compatible with the principles of Web architecture. Its e-mail lists are archived to the Web, and each e-mail is given a URI, so an agent may follow links seamlessly from one e-mail message to another, and by following links launch applications to send e-mail, discover more about the group, and in new e-mails reference previous topics. Likewise, an initiative called “Linked Data” attempts to deploy massive public data-sets as RDF, and its main tenet is to follows the Principle of Self Description (?). The hope is that the Semantic Web can be thought of as a seamless web of linked data, so that an agent can discover the interpretation of Semantic Web data by just following links. These links will then go to more data which may host formal definitions or informal natural language descriptions and multimedia depictions. For example, if one finds an RDF triple such as URI ex:EiffelTower ex:hasArchitect ex:Gustave_Eiffel and discover more information about the Eiffel Tower, like a picture of it or the fact that construction was finished in 1889 by accessing http://www.example.org/EiffelTower (Although the devil is in the details, especially when eliding the Eiffel Tower and some digital information about it, as explored in Chapter 6.
Since it is often too much work to provide accessible information from every URI in a RDF statement, RDF statements are usually grouped together using namespaces. For example, DBpedia hosts also URIs for representational web-pages and referent itself in its namespace http://dbpedia.org (?). It also uses predicates from a taxonomy created by a merger of the Wikipedia categories and Wordnet terms called YAGO (?). So one could write a triple similar to our example by using two namespaces, such as http://dbpedia.org/resource/Eiffel_Tower yago:architect http://dbpedia.org/resource/Gustave_Eiffel. If one then wants to retrieve a namespace document from the namespace URI, one can often retrieve an RDF Schema, a document in natural language. Often, since separate URIs for different resources can be given by the use of fragment identifiers, the Semantic Web does the often rather unorthodox maneuver - at least in the world of hypertext, which normally uses fragment identifiers to identify parts of a single document - of using fragment identifiers with the same namespace URI to group together a set of related things. This is done, for example, by well-known “Geo” vocabulary for geospatial data, where a ‘spatial thing’ is http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing so RDF describing the entire vocabulary can be accessed from the namespace document, http://www.w3.org/2003/01/geo/wgs84. The namespace document which gives a definition of “spatial thing” as “Anything with spatial extent, i.e. size, shape, or position. e.g. people, places, bowling balls, as well as abstract areas like cubes”(?).

Since RDF is supposed to an all-purpose knowledge representation system for the Web, RDF statements themselves can also be described using RDF. RDF itself has a namespace document at http://www.w3.org/1999/02/22-rdf-syntax-ns#, which provides a description of RDF in RDF itself. In other words, RDF can be meta-modeled using RDF itself, in a similar manner to the use of reflection in knowledge representation and programming languages (?). For example, the notion of a RDF predicate is http://www.w3.org/1999/02/22-rdf-syntax-ns#predicate, and is defined there as “the predicate of the subject RDF statement.” The same holds for almost all RDF constructs, and a conformant RDF processor can derive from any RDF triple a set of axiomatic triples that define RDF itself, such as rdf:predicate rdf:type rdf:Property (all RDF predicates are of the type property). For any RDF statement like ex:EiffelTower ex:hasArchitect ex:Gustave_Eiffel, an RDF-aware agent can then infer that ex:hasArchitect rdf:type rdf:predicate, which states in RDF that an architect relationship is a predicate in a RDF triple. However, usually RDF is not hosted according to the Principle of Self-Description. Use of
the media type application/rdf+xml is not consistent usually, and the namespaces URI of specifications like the RDF Syntax namespace just allow access of to some RDF triples, which is useless to a machine incapable of understanding RDF in the first place, instead of a more useful RDDL document?. A version of RDDL in RDF (?) with an associated GRDDL transform exists in order to make it even easier for Semantic Web agents to follow namespace documents to associated resources (?). There is also a very large open question, as explored in Chapter 6, of how to link authoritative information to a URI that is supposed to about a non-accessible thing.

5.2.4 RDF and the Open World Principle

The Principle of the Open World is the fundamental principle of inference on the Semantic Web. A relatively simple language for declaring sub-classes and sub-properties, RDF Schema, abbreviated as RDF(S), was from the beginning part of the vision of the Semantic Web and developed simultaneously with RDF. Yet determining how to specify exactly what other triples may be inferred from a given RDF triple is a non-trivial design problem, since it required adding an inference mechanism to a semantic network, which historically in AI featured little or no inference. Those that do not remember the history of artificial intelligence are bound to repeat it, and the process of specifying inference in RDF led to an almost complete repeat of the 'procedural versus declarative' semantics debate. As originally as defined, the original RDF specification defined its inference procedure by natural language and examples. Yet differing interpretations of the original RDF specification led to decided different inference results, and so incompatible RDF processors. This being unacceptable for a Web standards organization, the original defender of formal semantics in artificial intelligence, Pat Hayes, oversaw the creation of a declarative, formal semantics for RDF and RDF(S) in order to give them a principled inference mechanism.

Indeed, the Open World principle was considered to be a consequent of the lack of centralized knowledge implied by the decentralized creation of URIs and links as given by the Principles of Universality and Linking. The parallel to the removal of centralized link indexes is that on the Semantic Web, “we remove the centralized concepts of absolute truth, total knowledge, and total provability, and see what we can do with limited knowledge” (?). Hayes argued, in a similar fashion as he had argued in the original ‘procedural versus declarative’ semantics debate in AI, that the Semantic Web should just use standard first-order predicate logic. Yet while Berners-Lee accepted the
need for a logic-based semantics, he argued against Hayes for the Principle of Open World and monotonicity, and the formal semantics of RDF was designed to obey the Open World Assumption (\textsuperscript{2}). The reason for maintaining the Open World Assumption was that adding triples in a graph merge never should change the meaning of a graph so one could never retract information may simply adding more triples, or invalidate previously-made conclusions. This monotonicity is considered key, since otherwise every time a RDF triple was merged into a graph the interpretation of the graph could change and so the entire graph might have to be re-interpreted, a potentially computationally expensive operation. By enforcing monotonicity, RDF allows interpretations to be changed incrementally, to scales well in the sea of partial information of the Web. Hayes himself eventually agreed, saying that reasoning on the Semantic Web “needs to always take place in a potentially open-ended situation: there is always the possibility that new information might arise from some other source, so one is never justified in assuming that one has ’all’ the facts about some topic’” (\textsuperscript{3}).

RDF Schema is on the surface a very simple modeling and inference language (\textsuperscript{4}). RDF Schemas can provide additional information about properties and classes that describe a resource. Due to the Open-World assumption, unlike schemas in relational databases or XML Schemas, RDF Schemas are not prescriptive, but merely descriptive, and so an agent cannot validate RDF triples as being either consistent or inconsistent with an RDF Schema (\textsuperscript{5}). They cannot make the information given by a triple itself change, but only enrich the description of an existing triple. RDF Schema adds two main features to RDF. First, it allows there to be classes and sub-classes, as well as properties to have sub-properties. Second, it allows properties to have types for domains and ranges, such that in for a triple the subject is the domain and the object is the range of a property. Imagine that the property ex:hasArchitect has the range ex:people and domain ex:building. Note that RDF Schemas are not automatically applied to triples even if they are mentioned in a triple, such that for a statement like ex:Eiffel_Tower ex:hasArchitect ex:Gustave_Eiffel, the fact that the domain of ex:hasArchitect is buildings and the range is people, is not known unless the RDF Schema is automatically imported and so merged with the triple itself. An RDF(S)-aware agent that has retrieved the RDF Schema can deduce from the triple that ex:Gustave_Eiffel rdf:type yago:Person or that Gustave Eiffel is indeed a person. This sort of simple reasoning is again encoded as a set of axiomatic triples and rules for inference and semantic conditions for applying these axioms to infer more triples. See the RDF Formal Semantics for full details (\textsuperscript{5}). From here on out, the
acronym ‘RDF’ refers to both RDF and RDF(S), whose formal semantics are given together (?).

In practice, the Principle of the Open World has surprising results. One of the ramifications in RDF is that there is no proper notion of false, but only the notion that something is either inferred or not, and if it is not inferred, it may simply be undefined. Although it seems straightforward, in practice this leads to surprising results. Take the following example: “Gustave is the father of Valentine,” which in RDF is `ex:Gustave ex:fatherOf ex:Valentine` . Is George also the father of Valentine (`ex:George ex:fatherOf ex:Valentine`)? Operating under the closed world assumption, the answer would be no. Yet operating under Open World Principle, that statement would be possible, for there is no restriction that the there someone can only have only a single father, and in RDF(S) stating such a restriction is impossible.

OWL (Web Ontology Language), is an open-world extension of the Semantic Web, allows restrictions, such as cardinality, to be placed on predicates. However, even if one set the cardinality of the `ex:fatherOf` predicate to one (so that one could have at most one father), the results will be surprising: the reasoner will conclude that `ex:George` and `ex:Gustave` refer to the same individual. In contrast to the expected behavior of many other inference engines, including people, there is no Unique Name Assumption, the assumption that each unique name refers to a unique individual, due to the Open World Principle. The Unique Name Assumption, while very useful for counting, makes an implicit assumption about each name refers to only one individual, and if an individual can not be found that satisfies the name then that individual must not exist. This further reinforces the tendency of URIs on the Semantic Web, despite their global scope, to be ambiguous, a point we shall return to in the Chapter 6.

All is not well with the Principle of Self-Description and the Open World on the Semantic Web, for ultimately even on the Semantic Web, URI persistence is a problem. If some of the URIs in an RDF graph are no longer maintained by their owner, or even temporarily go down, problems could ensue if any application attempts to use the Principle of Self-Description in order to find an interpretation of that RDF graph. Furthermore, the problem can become even worse if the owner of URI changes. Imagine if the owner changes the Web representations to represent another resource entirely! This sort of Semantic Web spam and disrepair is dangerous. If an agent followed its nose to `http://www.example.org/EiffelTower` and an RDF Web representation was no longer hosted there, how could the agent make an authoritative decision on whether or
not that URI represented the Eiffel Tower in France or in Texas, especially if there was some disagreement over the meaning of the URI. Even worse, assume in some ‘semantic land-grab’ that the town of Paris, Texas cleverly stole the URI of the resource that at a point in the past identified the Eiffel Tower in France, and changed the geographical co-ordinates attached to the URI. What should our agent do? While this event would break the Open World Principle, as the agent would either have to maintain its previous interpretation of the graph, or update its new one. In this manner, the lack of ‘Cool URIs’ (i.e. persistent URIs) breaks the Open World Assumption for monotonic reasoning, as it requires all inferences to be updated if there has been a change. The only alternative is for the agent to abandon the follow-your-nose algorithm, but then where else would it find more data?

What this problem brings up is the strange fact that the Principle of Self-Description, and the Open World Principle that relies upon it, essentially require bodies that are accountable for the Web representations hosted at the URI. While this may apply to the W3C, it may not apply to your average URI owner on the street, who is ultimately unaccountable. Does this mean that the Principles of the Web, or even the Web itself, is broken? Ultimately not. First, the Principles of the Web are normative, but not descriptive, so they describe an ideal situation to which the reality of the Web is measured. Second, even in messy reality, due to the various advantages that following the Principles of the Web may incur, there may be incentives that even self-interested individuals may follow for maintaining URIs? We inspect these in more detail in Chapter 6.

5.2.5 RDF and the Principle of Least Power

Insofar as it is applied to the Semantic Web, the Principle of Least Power is strangely counter-intuitive: Traditionally knowledge representation languages were always striving for greater power, yet the Semantic Web begins with RDF, a language purposefully designed to be the least powerful language. The true bet of the Semantic Web is then on triples as the most basic language upon which other languages can be based. The challenge for the Principle of Least Power is how to build the rest of the Semantic Web by expanding on the language of triples.

Inspired by the Principle of Least Power, he envisaged that each language would extend and build upon lower-level languages. On top of RDF, Berners-Lee envisaged a whole stack of more expressive languages being constructed. Although the vagaries
Figure 5.4: The Semantic Web Stack

of the standardization process have caused various changes in ‘Semantic Web stack’ and numerous conflicting versions exist, the original and most popular version of the Semantic Web stack is given in Figure 5.4 (Figure 5.4). The W3C has commenced standardization efforts in a number of these areas, and research in almost all levels of the stack has begun. The majority of the research has focused on extending the Semantic Web with “ontologies” based on description logic, known as the Web Ontology Language (somewhat strangely abbreviated as “OWL”) (Figure 5.4). As should be suspected given their heritage in artificial intelligence, most of the work in description logic applied to OWL has focused on determining the most expressive possible language that preserves decidable inference. OWL itself works will with the Open World Principle, since it only makes an inference by adding inferred statements and classifications, and so remains monotonic. While almost any possible triple is acceptable RDF, OWL allows users to design ontologies can even add constraints, such as cardinality and data-typing, that can make some RDF triples inconsistent with a given OWL ontology. Another part of the Semantic Web, originally unforeseen, is the query language SPARQL, a query language for RDF similar to the popular database query language SQL (Figure 5.4). Current work is focused on Rule Interchange Format (RIF), a rule-language similar to Prolog for both serializing normal rules and operating over RDF data (Figure 5.4). Other higher-levels on the Semantic Web stack such as ‘Unifying Logic’ remain mysterious, if poetic and evocative.

The relationship of building a more powerful language for the Web on top of a ‘less’ powerful one is difficult, as the relationship between RDF and OWL exemplify.
While OWL maintains the Open World Principle by not having a strict definition of false, a version of OWL, OWL-DL (Description Logic), can decidably prove whether ontologies or either inconsistent or consistent (?). Yet RDF is not precisely “less” powerful than OWL-DL. While RDF on first glance may appear may appear to be just a propositional logic restricted to binary arguments, this is misleading. Due to the Principle of Linking, features of higher-order logic are allowed in RDF, such as making statements about predicates, classes serves serving as individuals (and vice versa), and even the reification of triples. In contrast, OWL-DL gains this expressivity by virtue of maintaining a strict separation between classes and individuals and so forbidding the ability to add constraints to basic constructors of the language. Following the Principle of Linking, RDF(S) allows any element to be linked to any other element, and so cycles of class inheritance are allowed, while this is forbidden by OWL-DL. A version of OWL, OWL Full, allows any feature of OWL to be added to any legal RDF graph, and while its decidability is lost, it maintains (with a few caveats (?)) internal consistency and full compatibility with RDF, so any legal RDF graph is an OWL Full graph and vice versa, but this is not true for OWL-DL. So, OWL Full builds a ‘more powerful’ language on top of RDF, while OWL-DL abandoned the Principle of Linking and so strict compatibility with RDF in order to maintain decidability, although both are compliant with the Open World Principle. Further up the the stack, fractures are developing in the Semantic Web community (?). The primary designers of RIF have expressed that Closed World reasoning is more suitable for first-order logic (?). Furthermore, RIF does not even have a syntactic RDF serialization of their rules, although it should be able to handle RDF data (?). More widely-used than RIF is N3 Logic, a RDF-native rule language originally created by Berners-Lee, whose formal properties are unknown (?). The second version of OWL, OWL2, is going to abandon any connection with RDF besides its use possibly as a serialization syntax (?). Regardless, it appears for the future development of the Semantic Web, the Principle of Least Power is in peril.

One way to phrase the Principle of Least Power is the Weak Syntactic Version of the Principle of Least Power, which states that some of the sentences acceptable in the lower-level language are also accepted in a higher-level language. For example, the same predicate, like rdf:type, can be used by both OWL and RDF. One can also phrase this principle more strongly, Strong Syntactic Version of the Principle of Least Power, which states that all of the sentences accepted in the lower-level language are also acceptable in a higher-level language. In this way, the higher-
level language can be considered strictly backwards-compatible with the lower-level language, so that every sentence in the lower-level language is acceptable in the higher-level language. In this manner, every RDF sentence is a valid sentence in OWL Full, and so OWL Full obeys the Strong Syntactic Principle of Least Power. In practice most versions of higher-level languages share some sentences with lower-level versions, but not all (like OWL-DL and RDF), and so obey the Weakly Syntactic version of the Principle of Least Power. This means that the terms in the ‘higher-level’ and ‘lower-level’ are more a matter of perspective, as both may just share a some set of terms with each other. The various problems of versioning and compatibility are beyond this particular document, and are detailed in length elsewhere (?). 

However, there is a semantic alternative to phrasing the Principle of Least Power. After all, phrasing it in terms of only syntax would mean that RDF would be at best yet another ‘transfer syntax’ like XML. The notion of ‘power’ is easily formalized in logic in terms of model-theoretic interpretation and expressivity, a much better interpretation of ‘power’ than the Chomsky Hierarchy put forward by the W3C and more suitable than just syntactic acceptability (?). To review, an interpretation \( I \) of any group of sentences \( s \) in a language \( L \) can be formalized as a mapping to a statement in model \( m \) that satisfies the sentence \( s \), i.e. the statement is ‘true’ (?). As there are many possible models that satisfy any given number of sentences, there is a possible infinite number of models that satisfy \( s \). Given two different sets of sentences \( s_1 \) and \( s_2 \) from two different languages \( L_1 \) and \( L_2 \) respectively, if every \( s_1 \) in \( L_1 \) can be satisfied by \( m \) and every sentence \( s_2 \) in \( L_2 \) can also be satisfied by \( m \), then \( L_1 \) is reducible to \( L_2 \), so \( L_2 \) is at least as expressive as \( L_1 \). Furthermore, if there are sentences in \( L_2 \) that can not be satisfied by any model of \( L_1 \), then \( L_2 \) is more expressive than \( L_1 \).

The disparate languages of the Semantic Web can be judged on a continuum of power as regards expressivity. As the Semantic Web consists of a stack of languages, one way to enforce the Principle of Least Power is the Strongly Semantic Version of the Principle of Least Power, which can be phrased that for for any higher-level language and lower-level language that share the same sentences, all the entailments provided by these sentences in the lower-level language would be entailed by the higher-level language. So the entailments of the lower-level language of these representations should be monotonic with regard to the entailments of the higher-level language, so that going ‘up a level’ in a language just adds entailments to, but does not subtract from, any entailments given by the lower-level language. This can be considered a multi-language version of the monotonicity given by the Open World Principle. For
OWL Full, every RDF entailment, such as the axiomatic triples of RDF, still holds. However, as the Semantic Web grows, it becomes unclear if this is possible at higher-levels. For example, it is not possible in the layering of a rule language since universal quantification of variables is impossible to embed in RDF semantics, although it may be given in RDF syntax (?). This problem was not unnoticed in the early days of the design of RDF, and Hayes attempted to convince Berners-Lee and others that it would be more sensible to reverse the Principle of Least Power for the Semantic Web, and so create the most powerful language first and build the Semantic Web by creating lower-level fragments of the language. Not surprisingly, Hayes and Guha suggested that a version of first-order predicate logic, which they called LBase, would be suitable for the task (?). However, this was not endorsed by the W3C, which proceeded to standardize RDF, and the general complaint has been that it is impossible to build new languages on top of “RDF straw” while maintaining the strongly semantic Principle of Least Power (?). It seems that the Semantic Web has managed to repeat the mistake of classical AI of KRL: in attempting to create a knowledge representation language first (RDF) and then attach inference mechanisms (OWL, rule-languages) on top of it post-hoc, fundamental problems with the knowledge representation itself have been revealed (?).

As it may be difficult, if not impossible, to maintain the strongly semantic Principle of Least Power, one can weaken the semantic requirement of languages so new languages with different and possibly more powerful semantics to be built on top of the Semantic Web stack without fracturing the Semantic Web into mutually semantically incommensurable languages. The Weakly Semantic Version of the Principle of Least Power which can be phrased that for any higher-level language and lower-level language that share the same sentences, some of the entailments provided by these sentences in the lower-level language would be entailed by the higher-level language. In this version, the interpretation of the higher-level language does not always have to possess the same entailments that the lower-level interpretation. Some sentences may have the same entailments, just as how certain syntactic triples in RDF have the entailments as triples in OWL-DL. In this manner, new languages can add features like quantification without violating the Principle of Least Power. Let us imagine an example of a graph that contains a mixture of OWL-DL and RDF. A Semantic Web-aware agent that had only the lower-level language could then scrape an interpretation of this graph by either ignoring OWL-DL constructs altogether or giving these constructs a standard RDF interpretation, while an agent that had OWL-DL capabilities could likewise ig-
nore RDF triples that violated its semantic constraints while giving an interpretation to triples that it could. Inspired by how agents interpret HTML documents, the Principle of Least Power can be applied successfully to the Semantic Web.

5.3 The Semantic Web: Good-Old fashioned AI Redux?

To many, it has seemed that the Semantic Web was nothing but a second coming of classical artificial intelligence. As put by Yorick Wilks, “the initial presentation of the Semantic Web by Berners-Lee, Hendler and Lassila to be a restatement of the Good-Old-Fashioned AI agenda in new and fashionable World Wide Web terms” (?). So why would the Semantic Web succeed where classical knowledge representations failed? The first reason would be a difference in the underlying intellectual project. A second reason would be a difference in technology.

The difference of the project is one both of scope and goal. The Semantic Web is, at first glance at least, a more modest project than artificial intelligence. To review the claims of artificial intelligence in order to clarify their relation to the Semantic Web, we are best served by remembering the goal of AI as stated by John McCarthy at the 1956 Dartmouth Conference, “The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it” (?). However, ‘intelligence’ itself is not even vaguely defined. The proposal put forward by McCarthy gave a central role to “common-sense,” so that “a program has common sense if it automatically deduces for itself a sufficient wide class of immediate consequences of anything it is told and what it already knows” (?).

In contrast, the Semantic Web does not seek to replicate human intelligence and encode all common-sense knowledge in some universal representational scheme. The Semantic Web instead leaves “aside the artificial intelligence problem of training machines to behave like people” but instead tries to develop a representation language that can complement human intelligence, for “the Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help” (?). Despite appearances, the Semantic Web is in the spirit of Licklider and Engelbart rather than McCarthy, Minsky, and even latter-day proponents of AI like Brooks. Berners-Lee is explicit that the project of encoding human intelligence is not part of the problem, as the Semantic Web “does not imply some magical artificial intelligence which allows machines to compre-
hend human mumblings. It only indicates a machine’s ability to solve a well-defined problem by performing well-defined operations on existing well-defined data” (9). Instead, the Semantic Web is an intellectual project whose goal is philosophically the opposite of artificial intelligence, the creation of new forms of collective intelligence. As phrased by Licklider, this would be a “man-machine symbiosis,” in which in “the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions” (9).

While the goals of the Semantic Web are different, it does still employ the same fundamental technology as classical artificial intelligence: knowledge representations languages. As put by Berners-Lee, “The Semantic Web is what we will get if we perform the same globalization process to knowledge representation that the Web initially did to hypertext” (9). Yet there is a question about whether or not knowledge representation itself might be the problem, not just scale. As put by Karen Spärck Jones, one of the founders of information retrieval, “there are serious problems about the core [Semantic Web] idea of combining substantive formal description with world-wide reach, i.e. having your cake and eating it, even if the cake is only envisaged as more like a modest sponge cake than the rich fruit cake that AI would like to have” (9). What we have argued so far is that the best hope of the Semantic Web lies not in knowledge representation per se, but in the application of the Principles of Web Architecture to knowledge representation, putting the emphasis on ‘Web’ as opposed to ‘semantic’ in the Semantic Web. Indeed, as the Web has proven to be extraordinarily successful, so any knowledge representation language which is based on the same principles as the Web may fare better than its ancestors in artificial intelligence.

This was accomplished by first explaining what precisely the fundamental components of the Web consist of and the Principles that define the Web in Chapter 4. These terms were defined using the philosophically-grounded terminology given earlier in Chapter 3. Second, we showed how in detail a particular knowledge representation system, the Resource Description Framework (RDF), obeys these Principles of Web Architecture in this chapter. This reliance on Web architecture furthermore explains many of the non-traditional oddities of RDF from a knowledge representation viewpoint.

Will Web architecture save the Semantic Web from the fate of classical artificial intelligence? The hypothesis put forward in the next chapter is that the problems of the Semantic Web is actually much more general; the Semantic Web is itself a form
of language, and so the Web signals a return to the very same questions of reference and meaning and naming that have troubled philosophers of language for so long. ‘Meaning’ itself is a thorny word; does ‘meaning’ mean “machine-readable” or “has a relation to a formal model?” Or could it mean “easily understood by humans,” or “somehow connected to the world in a robust manner?” The original W3C Recommendation defining RDF did not even define ‘meaning’ informally, much less formally (?). Later versions of the specification did define a formal notion of meaning in terms of model theory, but Hayes notes that “exactly what is considered to be the ‘meaning’ of an assertion in RDF or RDFS in some broad sense may depend on many factors, including social conventions, comments in natural language” so unfortunately “much of this meaning will be inaccessible to machine processing” such that a “a full analysis of meaning” is “a large research topic” (?). However, as the entire Semantic Web is built on top of the notion of URIs having some sort of sharable ‘meaning.’ for the success of the Semantic Web, there is no choice but to engage questions of reference and meaning, questions that we turn in the next chapter.
Chapter 6

The Identity Crisis

Meaning is what essence becomes when it is divorced from the object of reference and wedded to the word. W.V.O. Quine (?)

6.1 What Do URIs refer to?

For multiple agents to exchange knowledge representations on the Semantic Web, they must share the meaning of a URI. What precise resource does a URI identify? The question of how an agent can determine what a URI refers to lies in the heart of Web architecture itself, although it only becomes noticeable on the Semantic Web, since on the hypertext Web are thought in general to trivially URIs identify the hypertext documents that those URI allow access to, although content negotiation does complicate even that simple story. However, on the hypertext Web this question could be in general ignored as an obscure edge-case, yet on Semantic Web it is absolutely central, since the information identified by Semantic Web URIs should be shared universally in a decentralized manner. In a nutshell, the problem is that URIs identify not only hypertext documents and other digital information, but analog things that have no causal connection to the Web. How can an agent determine the meaning of a URI? How can a Semantic Web URI for the Eiffel Tower be used to refer to the Eiffel Tower in Paris itself? Should the Eiffel Tower itself have a URI? If so, should that URI allow access to any Web representations? This cluster of questions has been dubbed the Identity Crisis of the Semantic Web.

There are at least two distinct positions on this question. The first position, the direct reference position, is that on the Web, the meaning of a URI is whatever was
The owner of the URI should be able to unambiguously declare and communicate the meaning of any URI, including a Semantic Web URI. In this position, the referent is generally considered to be some individual unambiguous single thing, like the Eiffel Tower or the concept of unicorns. This viewpoint is the one generally held by many Web architects, like Berners-Lee, who imagine it holds not just for the Semantic Web, but the entire Web.

The second position, the logicist position, is that for the Semantic Web, the meaning of a URI is given by whatever model(s) satisfy the formal semantics of the Semantic Web. Adherents of this position hold that the referent of a URI is ambiguous, as many different things can satisfy whatever model is given by the interpretation of some set of sentences using the URI. There are a few minor variations on this theme, with some people believing a URI has no meaning in of itself, but only in the context of its use in other triples, while others hold that one should be able to access logical descriptions from the URI itself. This position is generally held by logicians, who claim that the Semantic Web is entirely distinct from the previous, hypertext, Web.

In general, the positions were subterranean in the development of the Semantic Web, until a critical point was reached in an argument between Pat Hayes, the AI researcher primarily responsible for the formal semantics of the Semantic Web, and Berners-Lee. This argument was provoked by an issue was called ‘Social Meaning and RDF’ brought about by the following draft statement in the RDF Concepts and Abstract Syntax Recommendation, “The meaning of an RDF document includes the social meaning, the formal meaning, and the social meaning of the formal entailments” so that “when an RDF graph is asserted in the Web, its publisher is saying something about their view of the world” and “such an assertion should be understood to carry the same social import and responsibilities as an assertion in any other format” (?). During the period of comments for the RDF Working Drafts, Bijan Parsia commented that the above-mentioned sentences do not “really specify anything and thus can be ignored” or are “dangerously underthought and underspecified” and so should be removed (?). While at first these sentences about the meaning of RDF seemed to be a rather harmless and in concordance with common-sense, the repercussions on the actual implementation of the Semantic Web are surprisingly large, since “an RDF graph may contain ‘defining information’ that is opaque to logical reasoners. This information may be used by human interpreters of RDF information, or programmers writing software to perform specialized forms of deduction in the Semantic Web” (?). In other words, a special type of non-logical reasoning can therefore be used by the Semantic Web.
An example of this extra-logical reasoning engendered by the fact that URIs identify ‘one thing’ is as follows. Assume that a human agent has found a URI for the Eiffel Tower from DBpedia, and so by accessing the URI a Semantic Web agent can discover a number of facts about the Eiffel Tower, such that it is in Paris and that its architect is Gustave Eiffel, and these statements are accessed as an RDF graph. However, a human can have considerable background knowledge about the Eiffel Tower, such as a vague belief that at some point in time it was the tallest building in the world. This information is confirmed by the human agent employing the follow-your-nose algorithm, where by following the subject of any triple, the human would be redirected to the hypertext Wikipedia article about the Eiffel Tower, where the agent discovers via a human-readable description that the Eiffel Tower was in fact the tallest building till 1930, when it was superseded in height by New York City’s Chrysler building. This information is not explicitly in the RDF graphs provided. It is furthermore difficult to even phrase this sort of temporal information in RDF. Furthermore, the human agent discovers another URI for the Eiffel Tower, a RDF version of Wordnet in the file synset-Eiffel_Tower-noun-1.rdf. When the human agent access this URI, there is little information in the RDF graph except that this URI is used for a noun. However, in the human-readable gloss property explains that the referent of this URI is ‘a wrought iron tower 300 meters high that was constructed in Paris in 1889; for many years it was the tallest man-made structure.’ Therefore, the human agent believes that there is indeed a singular entity called the ‘Eiffel Tower’ in Paris, and that this entity was in fact at some point the tallest building in the world, and so the two URIs are equivalent in some sense, although the URIs do not formally match. What the ‘Social Meaning’ clause was trying to state is that the human should be able to non-logically can infer that both URIs refer to the Eiffel Tower in Paris, and they use this information to merge the RDF graphs, resulting in perhaps some improved inferences in the future.

This use-case was put forward primarily by Berners-Lee, and the W3C RDF Working Group decided that deciding on the relationship between the social and formal meaning of RDF was beyond the scope of the RDF Working Group to decide, so the RDF Working Group appealed to the W3C TAG for a decision. As TAG member Connolly noticed, they “didn’t see a way to specify how this works for RDF without specifying how it works for the rest of the Web at the same time” (W3C). In particular, Berners-Lee then put forward his own viewpoint that “a single meaning is given to each URI,” which is summarized by the slogan that a URI “identifies one thing.” (W3C).
In response, Hayes said that “it is simply untenable to claim that all names identify one thing” (?). Furthermore, he goes on to state that this is one of the basic results of the knowledge representation community and 20th century linguistic semantics, and so that the W3C can not by fiat render the judgment that a URI identifies one thing. Berners-Lee rejects Hayes’s claim that the Semantic Web must somehow build upon the results of logic and natural language, instead claiming that “this system is different from natural language: we design it such that each URI identifies one and only one concrete thing in the real world or one and only one globally shared concept” (?). At this point, in exasperation, Hayes retorts that “I’m not saying that the ‘unique identification’ condition is an unattainable ideal: I’m saying that it doesn’t make sense, that it isn’t true, and that it could not possibly be true. I’m saying that it is crazy” (?). Since Hayes and Berners-Lee deadlocked on this issue, in order to move RDF from a Working Draft to a Recommendation, the W3C RDF Working Group removed all references to social meaning from the RDF documents.

One should be worried when two as prominent researchers such as Berners-Lee and Hayes have such a titanic disagreement, where no sort of consensus agreement seems forthcoming. Yet who is right? At first, the argument would clearly seem to won by Hayes, for although he is not explicit in his references to academic literature, he seems to have the general weight of academic work in logic and linguistics behind him. However, there is reason to pause to consider the possibility that Berners-Lee is correct. First, while his notion may seem counter to academic ‘common-sense,’ it should be remember that as far as practical results are concerned, the project of logic-based modeling of common-sense knowledge in classical artificial intelligence inaugurated by Hayes earlier is commonly viewed to be a failure by current researchers in AI and cognitive science (?). In contrast, despite the earlier and eerily similar argument that Berners-Lee had with original hypertext academic researchers about broken links and with the IETF about the impossibility of a single naming scheme for the entire Internet, the Web is without a doubt an unparalleled success. While in general the intuitions of Berners-Lee may seem to be wrong, history has proven him right in the past. Therefore, one should take his seemingly odd and even incoherent pronouncements seriously.

Berners-Lee’s background is in the Internet standardization bodies like the IETF, and it is primarily his intuitions behind Web architecture as given in Chapter 4. Hayes, whose background in logic jumpstarted the field of knowledge representation in artificial intelligence, comes from a different and sometimes competing tradition from
the Web, as discussed in Chapter 2. The Identity Crisis is not just a conflict between merely two differing individual opinions, but a conflict between two entire disciplines: the nascent discipline of ‘Web Science’ as given by the principles of Web architecture, and that of knowledge representation in AI and logic (?).

This disagreement would not nearly as worrisome were not the Semantic Web itself not in such a state of perpetual disrepair. In a manner disturbingly similar to classical artificial intelligence, the Semantic Web is always thought of as soon-to-be arriving. The reason given by Semantic Web advocates is that the Semantic Web is suffering from simple engineering problems, such as a lack of some new standard, some easily-accessible list of vocabularies, or a dearth of Semantic Web-enabled programs. However, given the Semantic Web has not yet experienced the dizzying growth of the original hypertext Web, after an even longer period of gestation, might point to some deeper problems in the entire intellectual enterprise. It is very possible that this profound disagreement over the nature of URIs and identity, which forms the basis of all of Web architecture and differentiates the Semantic Web from previous knowledge representation systems, is the root of the problem for the lack of the growth of the Semantic Web.

Far from being a mandarin metaphysical pursuit, this philosophical issue is the very first practical problem one encounters as soon as one wants to actually use the Semantic Web. If an agent receives a graph in RDF, then the agent should be able to determine an interpretation of these triples. The inference procedure itself may help this problem, but it may instead make it worse, simply producing more uninterpretable RDF statements. The agent could employ the follow-your-nose algorithm, but what information, if any, should be accessible at these Semantic Web-enabled URIs? If a user wants to add some information to the Semantic Web, how many URIs should they create? One for the representation, and another for the referent the representation is about? Should the same URI for the Eiffel Tower itself be the one that is used to access a web-page about the Eiffel Tower?

What is then be necessary to explain these vast differences over such a basic issue would be a more full explanation of the differing background assumptions between Berner-Lee’s direct reference position and Hayes’s logicist position. Understanding these positions belongs primarily to the domain of the philosophy, even if Hayes and especially Berners-Lee do not articulate their positions with the relevant academic citations. Indeed, the Identity Crisis is just the infamous Symbol-Grounding Problem in new Web guise, “How can the semantic interpretation of a formal symbol system be
made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?” (?). However, the key difference is that on the Web, the symbols that one is concerned with, URIs and Web representations, are in fact external and explicit rather than internal and implicit sort of symbols systems of interest to cognitive science. Furthermore, while being parasitic on meanings “in the head” is unacceptable to cognitive science and AI, on the Web this is acceptable, as long as we can tell a story about URIs and meaning that allows decentralized agents to share the meaning of a URI. In this manner, the precise domain of philosophy that the Identity Crisis falls under is the philosophy of language. Indeed, this argument over reference on the Web is nothing more than the untimely return of blatantly philosophical issues in reference and meaning first brought up by Frege, issues that heralded the birth of analytic philosophy (?).

URIs on the Semantic Web can be thought of as analogous to natural language names, as names in natural language can be used to refer as well. Therefore, what needs to be done is to distinguish within analytic philosophy the various theories on naming and reference in general, and then see how these various theories either do or do not apply on the Semantic Web. What is truly remarkable is that the position of Hayes, the logicist position, corresponds to a well-known theory of meaning and reference, the ‘descriptivist theory of reference’ attributed to early Wittgenstein, Carnap, Russell, and turned into its pure logicist form by Tarski (?). However, it is common currency in philosophical circles that the descriptivist theory of reference was overthrown by the ‘causal theory of reference’ championed by Kripke and extended by Putnam (?). It is precisely this causal theory of reference that Berners-Lee justifies in his direct reference position. Thus, the curious coincidence is that both opposing positions on the Semantic Web correspond to equally opposing positions in philosophy. The purpose of the rest of this chapter is then the full explication of these two theories of reference in philosophy of language, and then inspect their practical success (or lack thereof) in the context of the Semantic Web, while at the end offering a critique of both, paving the way for a third theory of meaning.
6.2 The Logicist Position and the Descriptivist Theory of Reference

The origin of the logicist position is the descriptivist theory of reference. In the **descriptivist theory of reference**, the referent of a name is given by whatever satisfies the descriptions associated with the name. Usually, the descriptions are thought to be logical, so a name is actually a disguised logical description. The referent of the name is then equivalent to the set of possible things, given normally by a mathematical model, such that all statements containing the name are satisfied.

### 6.2.1 Logical Positivism

The roots of the descriptivist theory of reference lay with the confluence of philosophers who are known as *logical positivists* and *logical atomists*, whose most systematic proponents were Rudolf Carnap and Bertrand Russell respectively. Although logical positivism is a vast school of thought that has proven tremendously influential, even in its current discredited state, for our purposes we will only concern ourselves with one particular doctrine common to both logical positivism and logical atomism, the problem of how natural language terms relate to the logic descriptions, and logical descriptions to the world. The difference between the two schools is mainly one of focus, for the logical positivists hoped to rid the world of metaphysical and epistemological statements through the use of logic and empiricism, while logical atomists thought that the metaphysics and even our epistemology should be phrased in terms of logic.

The logical positivists and Bertrand Russell were inspired by Wittgenstein’s early philosophical work in the [*Tractatus Logico-Philosophicus*](https://en.wikipedia.org/wiki/Tractatus_Logico-Philosophicus). In it, Wittgenstein strongly argues for *logical atomism*, that logic is the true language of the world; “Logic is not a body of doctrine, but a mirror image of the world” for “the facts in logical space” are the world (?). So logical statements are “laid against reality like a measure” (?). This is possible because the world is metaphysically determinate at its base, being composed of “simple” and “unalterable” objects that “make up the substance of the world” so that “the configuration of objects produces states of affairs” where “the totality of existing states of affairs is the world” (?). In other words, there is no – as Brian Cantwell Smith would put it – “flex” or “slop” in this picture, no underlying “metaphysical flux” that somehow resists easily being constrained into these fully determinate “objects” (?). Although the nature of the world consists *true* logical facts, humans, since they
picture facts” to themselves, can nonetheless make false logical statements, since these pictures merely “model reality”? Contrary to his own logical atomist teacher Russell, Wittgenstein thought that the primary job of the logician is then to state true facts, and “what we cannot speak about” in the form of true logical statements “we must pass over in silence,” a phrase he believed was consistently misinterpreted by logical positivism (?). Note that unlike the more mature standpoint of Hayes, the early logical atomism of Wittgenstein allowed logical statements to directly refer to single things in the world, he and the logical positivists reified the logical model to be the world itself.

According to Carnap, in his The Logical Structure of the World, is that all statements (at least, “scientific” statements with “cognitive content”) can be reduced to logical statements, where the content of this logical language is given by sensory experiences (?). These “elementary experiences” can not be directly described, as they are irreducible, but only described by a network of logical predicates that treats these experiences as logical constants (?). While Carnap’s ultimate goal was to use this logical empiricism to render any scientific hypothesis either verifiable by sense experience or not; their general position was since natural language is the part of the world, the structure of language too must be logical, and range over these elementary sense experiences. In this regard, names are given their referents by concordance with a logical structure ranging over these elementary sensory experiences.

It is precisely the logical structure of names as representations of the world that Bertrand Russell explored. Russell begins his investigation in On Denoting with a deceptively simple claim: “Is the King of France bald?” (?). To what referent does the description “the King of France” refer to ()? Since in Russell’s time there was no King of France, it could not refer to any of the elementary sense data of Carnap. In this regard, Russell makes a crucial distinction. Carnap’s elementary sensory experiences are known through acquaintance, in which we have some sort of direct “presentation of” the thing. Yet knowledge of a thing can be based on description, which are those “things we only reach by means of denoting phrases” (?). Russell believed, much like Carnap, that “all thinking has to start from acquaintance, but it succeeds in thinking about many things with which we have no acquaintance” via the use of description (?). Russell was most interested in whether those things that which we have direct acquaintance can be considered true or false, or whether a more mysterious third category such as ‘nonsense’ needed. Russell opts to reject creating imaginary but true ‘things’ as well as any third category, but instead holds that such statements such as “the bald King of
France” are false, since “it is false that there is an entity which is now King of France and is bald” (?). This solution then raises the alarming possibility that “the King of France is not bald” also may come out false, which would seem to violate the Law of the Excluded Middle. So, Russell counters this move by introducing the fact that the “King of France is bald” is actually a complex logical statement involving scope and quantification, namely \((\exists x. F(x) \land G(x)) \land (\forall y. F(y) \rightarrow x = y)\), where \(F\) is “being the King of France” and \(G\) is “being bald” (?). According to the analysis, the “King of France” is merely a disguised complex logical statement. Furthermore, this treatment can be extended to proper names such as ‘Sir Walter Scott,’ who can be identified with ‘the author of Waverly,’ so that instead of being a tautology, even a proper name of a person, even if known through acquaintance, is sort of shorthand for a large cluster of logical statements. So to use our previous example, the ‘Eiffel Tower’ can be thought of as a short-hand for not only the ‘there exists an entity known as the Eiffel Tower’ but also logical statement was ‘the aforementioned entity had Gustave Eiffel as its architect.’ If someone did not know that ‘the aforementioned entity was also the tallest building in the world up till 1930,’ once could then make a statement such as ‘The Eiffel Tower is identical with the tallest building in the world up till 1930’ without merely stating a tautology, and such a statement would add true and consistent knowledge to a hearer who was not aware of the statement.

As sensible as Russell’s programme appeared, there are difficulties in building all names on of maintaining, as Quine put it, such a “slender basis” as elementary sense data and logic (?). The crux of the problem for any descriptive theory of names comes for the use of names of any “kind of abstract entities like properties, classes, relations, numbers, propositions,” for such entities could not have an interpretation to any content using such a simple sensory epistemology (?). Carnap’s *Empiricism, Semantics, and Ontology* made an argument for basing such entities purely on linguistic form itself. Carnap believed that, despite the difficulty of determining the interpretation of names for abstract entities, “such a language does not imply embracing a Platonic ontology but is perfectly compatible with empiricism” (?). His position was that while “if someone wishes to speak in his language about a new kind of entities, he has to introduce a system of new ways of speaking, subject to new rule,” which Carnap calls the “construction of linguistic framework for the new entities in question.” From within a linguistic framework, Carnap believed to commit to any statement about the “existence or reality of the total system of the new entities” was to make a “pseudo-statement without cognitive content”(?). Although this particular position of Carnap’s
was devastated by Quine’s argument against analyticity in *The Two Dogmas of Empiricism*, Carnap made an important advance in the idea of a name of even abstract things being defined by linguistic descriptions, the problems brought up by Quine forced later logicians to abandon the notion of the logic ranging over “elementary sense data” (?).

### 6.2.2 Tarski’s Formal Semantics

Tarski abandoned the quaint epistemology of Russell and Carnap and defined reference purely in terms of logic in his *The Concept of Truth in Formal Languages*. Reference was just defined as a consequence of the truth *only* in terms of satisfaction of a formal language (?). To set up his exposition, Tarski defines two languages, the first being the syntactic *object language* \( L \) and the second being the *meta-language* \( M \). The *meta-language* should be *more expressive* (in the sense given in Section 5.2.5) such that it can describe every sentence in the object language, and furthermore, that it contain axioms that allow the truth of every sentence in the object language to be defined. In his first move, Tarski defines the *formal conception of truth* as ‘Convention T,’ namely that for a given sentence \( s \) in \( L \) and a statement \( p \) that is a theorem defining the truth of \( s \) via a translation of \( s \) into \( M \) (?). Tarski then later shows that truth can be formally defined as “\( s \) is true if and only if \( p \)” (?). For example, if the object language is exemplified by a sentence uttered by some speaker of English and the meta-language was an English description of the real world, “‘The Eiffel Tower is in Paris’ is true if and only if the Eiffel Tower is in Paris.” The sentence ‘The Eiffel Tower is in Paris’ must be satisfied by the Eiffel Tower *actually being* in Paris. While this would at first seem circular, it’s non-circularity is better seen through when the object language is not English, but another language such as German. In this case, “‘Der Eiffelturm ist in Paris’ is true if and only if the Eiffel Tower is in Paris.” However, Tarski was not interested in informal languages such as English, but in determining the meaning of formal languages. In order to establish theories of semantics for formal languages, If one were defining a formal semantics for some fragment of a knowledge representation language like RDF, one had a statement such as http://www.eiffeltower.example.org ex:location dbpedia:Paris is true if and only if \( \exists a b . R(a,b) \) where \( R, a, \) and \( b \) are given in something like first-order predicate logic, which is more expressive than that particular RDF statement.

This straightforward approach to formal semantics runs into a difficulty, as shown in the above example; if one is defining a formal Tarski-style semantics for a lan-
language, what should one do when one encounters complex statements, such as ‘The Eiffel Tower is in Paris and had as an architect Gustave Eiffel.’ The answer is the heart of Tarski’s project, namely that the second component of Tarski’s formal semantics is to use of the principle of compositionality so that any complex sentence can have its truth conditions derived from the truth conditions of its constituents. To do this, the meta-language has to have finitely many axioms, and each of the truth-defining theorems produced by the meta-language has to be generated from the axioms (\textbf{?}). So, the aforementioned complex sentence is true if and only if $\exists ab.R(a,b) \land Q(a,c)$, where $Q$ can be the architect of relationship, $c$ can be Gustave Eiffel and $a$ the Eiffel Tower. Tarski’s theory as explained so far only deals with ‘closed’ sentences, i.e. sentences containing no variables or quantification. The third, and final component of Tarski’s formal semantics is to use the notion of satisfaction via extension (\textbf{?}). For a sentence such as “All monuments have a location,” we can translate the sentence to $\forall a.m(a) \rightarrow l(a)$ which is true if and only if there is an extension $x$ from the world that satisfies the logical statements made about $a$. In particular, Tarski has as his preferred extensions infinite ordered pairs, where the ordered set could be anything (\textbf{?}). For formal languages, as explained in Section 3.3, a model-theoretic semantics with a model composed by set theory was standard. For example, the ordered pairs in some model of $(\text{EiffelTower}, \text{Paris})$ would satisfy our example statement, as would $(\text{ScottMonument}, \text{Edinburgh})$ but not $(\text{Paris}, \text{EiffelTower})$. However, there is no reason why these models could not be “God Forthcoming,” things in the the real world itself, albeit given in set-theoretic terms that would violate the “metaphysical flux” of world (\textbf{?}). Henceforth we will assume all extensions used by Tarski-style semantics are models. To summarize Tarski’s remarkably successful programme, model-theoretic semantics can produce a theory of truth that defines the semantics of a sentence in terms of the use of a translation of the sentence into some formal language with a finite number of axioms, then using compositionality to define the truth of complex sentences in terms of basic sentences, and finally determining the truth of those basic sentences in terms of what things in a model satisfy the extensions of the basic sentences as given by the axioms. This work marks the high-point of the logicist programme, as all questions of meaning are reduced to questions about giving the interpretation of a sentence in terms of a formal notion of truth, and this notion of truth is not restricted by the logical atomist’s quaint epistemology of elementary sense data, but instead can range over any possible formal language and any possible worlds.
6.2.3 In Defense of Ambiguity

The descriptivist theory of reference, taken to its conclusion, results in the logicist position on the Semantic Web. While this work in the descriptivist theory of reference seems distant from the Identity Crisis of the Web, it is in fact central to the position of Hayes and the Semantic Web as whole. This is primarily because Hayes’s background was in the logicist tradition, with his particular specialty being the creation of Tarski-style semantics for knowledge representations languages. What Hayes calls the “basic results in 20th century linguistic semantics” that Berners-Lee’s dictum that “URIs identify one thing” violates is the interpretation of URIs in a Tarski-style formal semantics (?). For the logicist position, the semantics in the Semantic Web derive from the Tarski-style formal semantics Hayes created for the Semantic Web (?).

Before delving into the RDF Formal Semantics, it should be noticed that these semantics are done in a extensionally, far more so than even standard formal semantics as Hayes has completed for first-order logic in KIF (?). The reason for this is the Principle of Linking, in particular, the unusual features of RDF that “a property may be applied to itself” and that classes “may contain themselves” (?). This is done by distinguishing the class qua class and property qua property in RDF from whatever their extensions are, so while a class and property in RDF may or may not be satisfied by some model or world, the extension of the class or property are not considered to have the same identity as the property or class.

A simple example should suffice. What is the formal semantics of ex:EiffelTower ex:architect ex:Gustave_Eiffel? To simplify slightly, Hayes defines the formal semantics of set theory, where there is a set of resources that compose the model of the language, a set of properties, and a set of URIs that can refer to resources. The interpretation of any RDF statement is then given as an extensional mapping from the set of properties to the powerset of resources, the set of set of pairs of resources. So, given a set-theoretic model consisting of elements (given by italics) Gustave Eiffel and The Eiffel Tower and being the architect of, then ex:EiffelTower |= The Eiffel Tower, ex:Gustave_Eiffel |= Gustave Eiffel and ex:architect |= being the architect of. Someone using common-sense human intuitions will likely believe that this interpretation maps to our common-sense content of ex:EiffelTower ex:architect ex:Gustave_EiffelTower, and using the axiomatic triples defined in the RDF formal semantics, a few new triples can be inferred, such as ex:architect rdf:type rdf:Property. However, the inherent pluralism of the Tarski’s approach to mod-
els also means that another equally valid interpretation would be the inverse, i.e. the mapping of \texttt{ex:EiffelTower} to \textit{Gustave Eiffel} and \texttt{ex:Gustave_Eiffel} to \textit{The Eiffel Tower}. Due to the unconstrained nature of RDF, \texttt{ex:architect} has no ‘natural’ relationship to anything in particular, but could easily assigned either \textit{The Eiffel Tower} or \textit{Gustave Eiffel} just as easily as \textit{being the architect of}. Furthermore, the model could just as easily be given by something as abstract as the integers 1 and 2, and an equally valid mapping would be for \texttt{ex:EiffelTower} \models 1 and \texttt{ex:Gustave_Eiffel} \models 2, and \texttt{ex:Gustave_Eiffel} \models 2. Indeed, the extreme pluralism of a Tarski-style semantics shows that, at least if all one has is a single lone triple statement, that triple can be satisfied by any model. As the number of triples increased, the amount of possible models that satisfy the model are thought to decrease, but in such a loose language as RDF, as mandated by the Principle of Linking, Hayes notes that it “usually impossible to assert enough in any language to completely constrain the interpretations to a single possible world, so there is no such thing as ‘the’ unique interpretation” (\cite{Hayes}). This descriptivist theory of reference, where descriptions are logical statements in RDF, is illustrated in Figure 6.1.

The essay \textit{In Defense of Ambiguity} lays out a vigorous case is laid out against the Berners-Lee’s position that a “URI denotes one thing” (\cite{Hayes}). What is at stake is Principle of Universality, namely that anything can be identified by a resource. Hayes puts forward the thesis that the word ‘identify’ is simply incoherent, as it has two distinct readings, as explored earlier in Section 4.2.1, that of \textit{access} and \textit{reference}. While Hayes makes no claim that access to some Web representations via HTTP is not possible, he claims that such access to Web representations is orthogonal to the
question of what a URI could refer to, since “the architecture of the Web determines access, but has no direct influence on reference” (>). Furthermore, he claims that reference to resources is completely independent of whatever Web representations can be accessed, even if those contain logical expressions. Hayes subscribes to the logical atomist epistemology of Russell, as he says that “reference can either be established by either reference or ostentation” with ostentation being defined as the use of Russelian demonstrative (like ‘that’ or ‘this’) identifying a particular “patch of sense data” via a statement such as that ‘That is the Eiffel Tower’ (?). Since most of the things referred to by names are not accessible, reference can only be determined by description, and these descriptions are inherently ambiguous (?).

The argument over the ambiguity of description is exemplified in natural languages. If a person is trying to identify the Eiffel Tower to a friend, then the person may attempt to communicate their thought about the Eiffel Tower by uttering a description such as “the monument in Paris.” Yet even the friend may think they are talking about the Arc de Triomphe without further information. If the person tries to give more descriptions, such as “the steel tower,” then the hearer might think of the Eiffel Tower, but there are no guarantees. The hearer may also think of the steel dome of Galeries Lafayette. Even if the person said, “the structure made by Gustave Eiffel,” the hearer may think of a lesser-known structure like La Ruche. One can imagine that with enough descriptions a person could uniquely pick out the referent for the hearer. Even with an infinite amount of descriptions this may be impossible, since it involves the large presumption that the hearer shares our same metaphysical or perceptual ontology of things in the world. The hearer may simply have no conception that the Eiffel Tower even exists, and so may be unable to grasp the referent regardless of the number of descriptions given.

Even what appears to be stable reference by description can be easily disrupted by new information. Hayes illustrates this by referring to a famous example about whether “a fitted carpet was ‘in’ an office or ‘part of’ the office in which “two competent, intelligent adult native speakers of English each discovered, to their mutual amazement, that the other would believe what they thought was an obviously false claim” but that “over an hour of discussion it gradually emerged, by a process of induction from many examples, that they understood the meaning of ‘office’ differently” (?). For one person ‘office’ referred to “roughly, an inhabitable place” while for the other it referred to “something like a volume of space defined by the architectural walls” (?). These two people had shared the same office for years, and only upon the appearance of a carpet,
that they had different mental meanings for ‘office’ and more generally, for ‘room.’

On the Semantic Web, the negative effects of adding new information also hold. Often simple formal ontologies are more stable, as “if all one wants to say about persons is that they have mailboxes and friends, then one can treat ‘person’ as a simple category” (?). Even when a stable situation of mutual reference has been reached in some simple formal ontology, it can be upset by the addition of new ontological distinctions, as can be made by a so-called “upper ontologies” such as DOLCE (?). For example, DOLCE claims that the identity of a person continues over time, while other upper-level ontologies do not (?). So, does the Semantic Web to distinguish “Tim Berners-Lee the continuant from Tim Berners-Lee the four dimensional history?” (?). For purposes of inference, such a minor distinction can really matter. If one is not careful with one’s upper-level ontology, one can produce “immediate logical contradictions, such as inferring that Berners-Lee is both 52 years old and 7 years old” (?).

The situation with descriptions in real life, with the possibility of multiple underlying ontologies and differing interpretations, is thought by Hayes and others to be modeled by the radical model-theoretic pluralism of Tarski-style formal semantics, i.e. for any language “sufficient to express arithmetic” to have many different ‘non-standard’ models (?). As our example showed, RDF in general says so little inferentially that many different models to satisfy almost any given RDF statement. Therefore, Hayes considers it essential to ditch the vague word ‘identify’ as used in URIs, and distinguish between the ability of URIs to access and refer. While access is constrained by Web architecture, according to Hayes, reference is absolutely unconstrained except by formal semantics, and so “the relationship between access and reference is essentially arbitrary” (?). From this philosophical position, the Identity Crisis dissolves into a pseudo-problem, for the same URI can indeed access a web-page and refer to a person unproblematically, as they no longer have to obey the dictum to identify one thing. Hayes compares this situation to that of overloading, using a single name to refer to multiple referents, and it is instead of a problem, since “it is a way of using names efficiently” and not a problem for communication, as “natural language is rife with lexical ambiguity which does not hinder normal communication,” as these ambiguities can almost always be resolved by sufficient context (?). Overall, the argument of Hayes against Berners-Lee in the Identity Crisis is the position of keeping the formal semantics of reference separate from the Web as given by the Principles of Web architecture.
While the logicist position may seem relatively sensible, the logicist position would also hold that the Semantic Web is more or less unremarkable, since “the Semantic Web languages would operate exactly unchanged if the identifiers in them were not URIs at all, and if the Web did not exist” (?). In this manner, we should be worried, for then the Semantic Web would be no different than the traditional project of knowledge representation in classical artificial intelligence. Indeed, the first generation of the Semantic Web was built upon this logicist vision, with a focus on inference, exemplified by the creation of inference programs and hosts of academic papers detailing how description logics could efficiently implement Open World reasoning (??). Given the emphasis on inference, not surprisingly almost all work in producing information for the Semantic Web became focused on the creation of formal ontologies, and while some of the simple ones such as FOAF (Friend-Of-A-Friend) survived, most of these ontologies languish unused (?). This complete disregard for the Principles of Web architecture make sense from the logicist perspective, as the referential mechanism of RDF and other Semantic Web languages should have absolutely no relationship with the accessibility of Web representations. While this first generation of the Semantic Web was an academic success story, the Semantic Web nonetheless did not have the tremendous growth of the original hypertext Web. Indeed, its success seems to be confined primarily to becoming a de-facto standard among the knowledge representation community in AI, rather than the more universal vision of Berners-Lee.

Practically, there was never a consensus on the first generation Semantic Web about how one delimits what logical descriptions determine, even ambiguously, the referents of a URI. One implicit viewpoint dominant on the first-generation Semantic Web is a localist reading of the scope of URIs; the URIs refer to whatever model could be interpreted to in context of just the current RDF graph given by some Web representation. Yet this makes it difficult, if not impossible, for the Semantic Web to be used for its primary purpose of data integration. One proposal on this point was to assume the localist reading any Semantic Web statement unless other URIs were explicitly imported via owl:imports statements (?). However, this would put the responsibility for data integration on the server-side hosting of Web representations, not data integration ‘on-the-fly’ by a user-agent. The second option, the holist reading, is that as names has a URI refer to whatever model can satisfy all uses of the URI on the Semantic Web. Yet this option makes little sense, for as given by the Principle of the Open World, it is
impossible to gather all uses of a URI in Semantic Web statements spread throughout the entire Web.

One possibility in combining the Principles of Web architecture with a logicist theory of reference would be to have a URI refer to whatever satisfied all logical descriptions which are accessible from the URI itself, a viewpoint championed by David Booth under the title *URI declarations* (?). This particular possibility of using URIs as names would be an almost perfect analogy to Russell’s definition of names as a cluster of logical descriptions (?). URI declarations have a number of advantages over both the localist and holist logicist readings of URIs. First, URI declarations allows the URI to access “a set of core assertions that are intended to characterize the resource” that can then be determined by the owner of the URI (?). This allows if an agent encounters a previously unseen URI in a Semantic Web statement and the interpretation of the statement itself is not satisfactory, perhaps because the interpretation is too ambiguous, the agent can use the Principle of Self-Description to discover some core assertions. However, the creation of other statements using this URI are not banned, for “different URI users will necessarily wish to make” possibly “mutually incompatible” and so “different sets of assertions involving the URI” (?). According to Booth, these “mandatory core assertions permits the meaning of a URI to be anchored, to prevent it from drifting, and this in turn increases the likelihood that independent assertions made using the URI can be successfully joined” (?).

While this standpoint makes sense, it is also very limiting for agents and may not encourage re-use, since “if you do not want to accept the core assertions specified by the URI declaration, then you should not use that URI to make statements about its denoted resource” (?). If one doesn’t agree with the interpretation of the core assertions in the URI declaration, then one should go mint a new URI. In turn, this violates the strict separation of reference and access that Hayes puts forward as central to the formal semantics of RDF, even though URI Declarations still maintains a belief in the primacy of logic (?). Furthermore, it is unclear where the follow-your-nose algorithm should stop in its quest for accessing logical statements. Should an agent follow a HTTP Link header, or the Link elements in HTML? Should the agent follow HTTP redirect headers, and if so, which ones? These questions are unanswered by the follow-your-nose algorithm. While Rees has developed a more formally specified an algorithm called the *URI Documentation Protocol*, there is no W3C standardized follow-your-nose algorithm for logical descriptions associated with a URI, and many other possibilities, such as *Concise Bounded Descriptions* (?). For at least these reasons, URI declara-
The inability of a purely descriptivist theory of reference to reach standardization, or even ad-hoc conventional usage, has led the initial first-generation Semantic Web applications to fail. Most of these first generation OWL or RDF(S) ontologies, such as DOLCE, did not in any way re-use URIs and did not let any Web representations be dereferencable from the original URIs (?). OWL ontologies were stored as one large inaccessible file, difficult to index by search engines and virtually impossible to find by anyone except the creator of the file. This lack of URI re-usage and inability communicate about the referents of Semantic URIs have led for the actual possible referents of many Semantic Web URIs to be so drastically underdetermined as to make the URI itself unusable. Strictly speaking, it was impossible to determine reference except via the relatively weak inference mechanisms of OWL and RDF, which usually did not infer much of interest as predicted by McDermott earlier ?. In an attempt to ameliorate the situation, natural language strings were added to describe Semantic Web URIs using properties like rdfs:label, but it was left unknown how this information effected the formal semantics. Since an agent could never be clear about the referential status of a Semantic Web URI, rather than trust already-existing Semantic Web URIs, everyone simply created new URIs rather than re-using them. This dire situation has led the first-generation of the Semantic Web to be more of scattered semantic islands rather than vast inter-linked semantic continents, a ghostly web of logical reference separate from the hypertext Web. Yet the failure of this first-generation of the Semantic Web should not be surprising, for it is not a test of the Semantic Web hypothesis as a knowledge representation language built according to the principles of Web architecture. The first-generation of the Semantic Web has almost nothing to do with the Principles of Web architecture besides the Open World Principle, and so is only a decentralized version of knowledge representation as done in classical artificial intelligence with a single unified logic-based monotonic semantic network language. As such, its failure is more of a test of decentralized knowledge representation of the unified logic-based AI defended by Hayes’s *In Defense of Logic* rather than with the Semantic Web, and this failure should be depressingly familiar (?).
6.3 The Direct Reference Position and The Causal Theory of Reference

The alternative slogan of Berners-Lee, that “URIs identify one thing,” may not be completely untenable after all (?). It appears to even be intuitive, for when one says ‘I went to visit the Eiffel Tower,’ one believes one is talking about a very particular thing in the real world called the ‘Eiffel Tower,’ not a cluster of descriptions or model of the world. The direct theory of reference of Berners-Lee has a parallel in philosophy, namely Saul Kripke’s ‘causal theory of reference,’ the classic devastating argument against the descriptivist theory of reference, and so the logicist position of Hayes (?).

In contrast to the descriptivist theory of reference, where the content of any name is determined by ambiguous interpretation of logical descriptions, in the causal theory of reference any name refers via some causal chain directly to a referent (?).

6.3.1 Kripke’s Causal Theory of Proper Names

The causal theory of reference was meant to be an attack on the descriptivist theory of reference attributed to Russell, and its effect in philosophy has been to discredit any neo-Russellian descriptivist theory of reference (?). Surprisingly, the causal theory of reference also has its origin in logic, since Kripke as a modal logician felt a theory of reference was needed that could make logical statements about things in different logically possible worlds (?). However, Kripke did not directly confront the related position of Tarski, his argument does nonetheless attempts to undermine the ambiguity inherent in Tarski’s model-theoretic semantics, although a Tarki-style semantics can merely ‘flatten’ models of possible worlds into a singular mode (?). Still, as a response in philosophy of language, it is accepted as a classical refutation of the descriptivist theory of reference.

In Kripke’s Naming and Necessity, an agent fixes a name to a referent by a process called baptism, in which the referent, known through direct acquaintance is associated with a name via some local and causally effective action by the agent (?). Afterwards, a historical and causal chain between a current user of the name and past users allows the referent of a name to be transmitted unambiguously through time, even in other possible worlds. For example, the person known as ‘Gustave Eiffel’ was named via a rather literal baptism, and the name ‘Gustave Eiffel’ would still refer to that baptized person, even if he had not been the architect of the Eiffel Tower, and so failed to satisfy
that definite description. Later, the causal chain of people talking about ‘Gustave Eiffel’ would identify that very person, even after Gustave Eiffel was dead and gone. In this regard, a name functions much like a representation as given by our representation cycle in Section 3.5, where some baptismal ‘input stage’ between a name and a thing is necessary to assign the name directly on the referent. Descriptions aren’t entirely out of the picture on Kripke’s account – they are necessary for disambiguation when the context of use allows more than one interpretation of a name, and they figure in the process by which things actually get their names, if the thing cannot be directly identified. However, this use of descriptions are a mere afterthought with no causal bearing on determining the referent of the name itself, for as Kripke puts it, “let us suppose that we do fix the reference of a name by a description. Even if we do so, we do not then make the name synonymous with the description, but instead we use the name rigidly to refer to the object so named, even in talking about counterfactual situations where the thing named would satisfy the description in question” (?) So what is crucial is not satisfying any description, but the act of baptism and the causal transmission of the name.

6.3.2 Putnam’s Theory of Natural Kinds

Kripke’s examples of the causal theory of reference used proper names, such as ‘Cicero’ or ‘Aristotle,’ and he did not extend his analysis to the whole of language in a principled manner. However, Hilary Putnam, in his *The Meaning of “Meaning,”* extends Kripke’s analysis to all sorts of names outside traditional proper names, and in particular Putnam uses for his examples the names of natural kinds. Putnam was motivated by an attempt to defeat what he believes is the false distinctions between intension and extension. The set of logical descriptions, which Putnam identifies with a “psychological state,” that some thing must satisfy to be given a name is the intension, while those things in a given interpretation that actually satisfy these descriptions, is the extension. Putnam notices that while a single extension can have multiple intensions it satisfies, such as the Eiffel Tower both being “in Paris” and “a monument,” a single intension is supposed to have the same extension in a given interpretation. If two people are looking for a “monument in Paris,” the Eiffel Tower should satisfy them both, even though the Eiffel Tower can also have many other possible descriptions.

Putnam’s analysis can be summarized as follows: Imagine that there is a world “very much like Earth” called ‘Twin Earth.’ On Twin Earth there “the liquid called
‘water’ is not $H_2O$ but a different liquid” whose chemical formula is abbreviated as $XYZ$, and that this $XYZ$ is “indistinguishable from water at normal temperatures and pressures”, since it “tastes like water and quenches thirst like water” (?). A person from Earth would incorrectly identify $XYZ$ for their normal referent of water, as it would satisfy all their descriptions. In this regard, this shows that meaning’s “ain’t in the head” but are in fact determined, not by individual language use or descriptions, but by some indexical relationship to “stuff that is like water around here” normally. That “stuff” should get its name and meaning from experts, since “probably every adult speaker even knows the necessary and sufficient condition ‘water is $H_2O$, but only a few adult speakers could distinguish water from liquids which superficially resembled water...in case of doubt, other speakers would rely on the judgment of these ‘expert’ speakers” who would ideally test $XYZ$ and determine that it was indeed, not water (?). Indeed, less outlandish examples, such as the difference between “beech trees” and “elm trees” are trotted out by Putnam to show that a large amount of our names for things, perhaps even extending beyond natural kinds, are actually determined by expert knowledge (?). In this way, Kripke’s baptism can extend to almost all of language, and scientists can be considered a special sort of naming authority capable of baptizing all sorts of things with a greater authority than everyone else. As even Putnam explicitly acknowledges “that Kripke’s doctrine that natural-kind words are rigid designators and our doctrine that they are indexical are but two ways of making the same point” (?).

6.3.3 Direct Reference on the Web

This expert-ruled causal theory of reference is naturally close to the direct reference position of Berners-Lee, whose background is in expert-created databases. He naturally assumes the causal theory of reference is uncontroversial, for in database schemas, what a term refers to is a matter best left to the expert designer of the database. So Kripke and Putnam’s account of unambiguous names can then be transposed to the Web with a few minor variations in order to obey Berner-Lee’s ‘crazy’ dictum that “URIs identify one thing” regardless of interpretation or even accessible Web representations (?). While it may be a surprise to find Berners-Lee to be a closet Kripkean, Berners-Lee says as much, “that the Web is not the final arbiter of meaning, because URI ownership is primary, and the look-up system of HTTP is...secondary” (?). There is also an element of Grice in the direct theory of reference, for the intended interpretation of the owner is the one that really matters to Berners-Lee, not any publicly
accessible particular Web representation (?). However, ultimately Berners-Lee has far more in common with the causal theory of reference, since although the URI owner’s intention determines the referent, after the minting of the new URI for the resource, the intended interpretation is somehow never supposed to vary (?).

To apply the causal theory of reference as applied to URIs, baptism is given by the registration of the domain name, which gives a legally binding owner to a URI. The referent of a URI is established by fiat by the owner, and then optionally can be communicated to others in a causal chain in the form of publishing Web representations accessible from the URI or by creating Semantic Web statements about the URI. This causal theory of reference for URIs is illustrated in Figure 6.2.

In this manner, the owner of the URI can thereby determine the referent of the URI and communicate it to others, but ultimately the act of baptism and so the determination of the referent are in the hands of the owner of the URI, the self-professed ‘expert’ in the new vocabulary term introduced to the Semantic Web by his URI, and he has no real responsibility to host any Web representations at the URI. Since the owner can causally establish a name for a non-Web accessible thing via simply minting a new URI, under the causal theory of reference the Semantic Web can be treated as having as a giant translation manual mapping URIs directly to referents, where the URIs refer directly to objects in the world outside of the Web. In this manual, one could look up the URI http://www.example.org/Gustave_Eiffel and get back Gustave Eiffel himself. From the direct reference position, if an agent got a URI like http://www.example.org/Gustave_Eiffel and one wanted to know what the URI referred to, one could use a service such as whois to look up the owner of the URI, and
then call them over the telephone to ask them what the URI referred to if there was any
doubt of the matter. Since obviously such URIs cannot access things outside the Web,
what kinds of Web representations, if any, could this giant Semantic Web dictionary
return? If it returns no Web representation, how can a user-agent to distinguish for a
referent outside the Web from that of a URI for some Web-accessible resource?

6.3.4 The 303 Redirection Solution to the Identity Crisis

It was in this decidedly Kripkean spirit that the W3C TAG took on the Identity Crisis,
calling it the httpRange-14 issue, which was phrased as the question: What is the
range of the HTTP dereference function? (?) In Kripke’s causal analysis, since there
is some metaphysical distinction between a web-page that represents a thing and the
thing itself, this distinction should be upheld in Web architecture. In their solution, the
TAG defined a class of resources on the Web called an information resource, which is
as a resource “whose essential characteristics can be conveyed in a message” (?). This
implies there is some converse set, which we dub non-information resources, whose
characteristics can’t be conveyed in a message. It appears that this distinction is trying
to get at the heart of the distinction between a resource being a web-page about the
Eiffel Tower and a resource for the Eiffel Tower itself.

However, does the TAG’s creation of a term called ‘information resource’ help or
hurt? The definition of an information resource is a matter of great and ongoing con-
troversy. First, at least according to our parlance, all resources contain information. A
web-page may contain information that ‘The Eiffel Tower is in Paris,’ but the Eiffel
Tower itself carries that information by virtue of being in Paris itself. There are two
main readings of the TAG’s notion of an information resource. One reading is that
the term ‘information resource’ is groping towards the distinction that some things are
accessible on the Web and others are not, such that an information resource is any-
thing that can be sent as a message sent using HTTP. In this case, a web-page is an
information resource, but the Eiffel Tower itself is not, and neither is the text of Moby
Dick. One could presumably hook-up an web-cam to transmit messages about the Eif-
fel Tower, but the resource then would be an HTTP endpoint that sent representations
of the Eiffel Tower to clients, not the Eiffel Tower itself.

This distinction at first seems cut-and-dry, but falls apart upon closer inspection.
The question is not whether something is accessible on the Web, but whether some-
thing could be accessible on the Web. For example, imagine a possible world where
the Eiffel Tower does not have a web-page. In this world, it would seem counter-intuitive to claim that the webpage of the Eiffel Tower is then not an information resource just because it happens to not exist at this moment. This is not as implausible as it sounds, for imagine if the Eiffel Tower’s web server went down, so that http://www.tour-eiffel.fr returned a 404 status code. A more intuitive case is that of the text of *Moby Dick*. Is the text of *Moby Dick* an information resource? If the complete text of Moby Dick isn’t on the Web, one day it might be. However, a particular collector’s edition of Moby Dick’ could not be an information resource, since the part of that resource isn’t the text, but the physical book itself. One one reading, an information resource includes the text of *Moby Dick*, but on the other it doesn’t, so maybe the notion of an information resource isn’t a straightforward distinction. Do people have to have remarkably scholastic discussions about whether or not something is essentially information before creating a Semantic Web URI?

What is the precise distinction that the TAG is looking after? Our previous terminology as defined in Chapter 3 comes to the rescue. Both a web-page about the Eiffel Tower and the text of *Moby Dick* are, on some level of abstraction, carrying information about some content in some encoding. So, if any information resource is any resource that is carrying information in the manner we defined it in Chapter 3, then both the text of *Moby Dick* and a web-page about the Eiffel Tower are information resources, even if they are not currently ‘on the Web.’ Furthermore, these kinds of information resources can be transmitted via digital encodings, and so can in theory be on the Web by being realized as Web representations, even if the resource does not allow access to Web representations at a given time. Lastly, a particular edition of Moby Dick, or Moby Dick in French, or even some RDF triples about *Moby Dick*, are all in in this reading information resources, with various encodings specified at certain levels of abstraction.

Still, the mysterious non-information resource eludes any grasp. The most intuitive definition is that it a resource which can not be realized as a digital encoding. In this category would fit things like the Eiffel Tower itself and people like Tim Berners-Lee himself. Despite however we cut them up In this matter, the crucial distinction is that some things, because they are not realized as a digital encoding, can only be represented in the philosophical sense as given in Section 3.5 via a digital encoding on the Web. This important distinction even covers bizarre edge-cases, such as if Berners-Lee, ala the cybernetician Kevin Warwick, put an HTTP endpoint in a chip connected to the Internet via wireless under his skin. To recall the distinction made
by Smith between representation and realization in Section 3.5, if the chip stopped working, one wouldn’t claim that Tim Berners-Lee was himself destroyed, at least upon first guess. It appears that the best story we have to tell about non-information resources is that they are things that are analog and so resist direct digital encoding, but can only be indirectly encoded via representations of the thing in a suitable language. Since this ability for URIs to represent resources that are not digital encodings occurs primarily on the Semantic Web, we shall replace the awkward terminology of ‘non-information resource’ with the more fitting term **Semantic Web resource**, which is *(some resource that is analog).* A **Semantic Web URI** is then just a **URI for a Semantic Web resource**. Likewise, an **information resource** is *some resource that is information with the possibility of a digital encoding*, even if that resource is not currently digitally encoded, or only can be coded on some level of abstraction. So the **Semantic Web** itself can be thought of as **the use of the Web as a language to represent analog things**.

Then what is the kind of resource that can be accessed over the Web, the class to which a web-page about the Eiffel Tower belongs but the text of *Moby Dick* does not? Obviously it is some subset of information resources, which we call a ‘Web resource.’ A **Web resource** is *an information resource that has accessible Web representations that realize its information.* A Web resource can then be thought of as a mapping from time of request to a series of Web representation responses, where the information realized by those Web representations are the Web resource. Of course, this definition has problems, but it close in spirit to the original pre-Semantic Web thinking behind resources in IETF 1630, as well as in IETF RFC 2616 where it is defined as “a network data object or service” and coherent with Engelbart’s original use of the term ‘resource’ (??).

The precise relationship of a Web resource to its Web representations is complex. The information of a Web resource is *realized* by a Web representation. The same information can be transmitted via many different Web representations in different languages, as shown by content negotiation. Yet Web representations also exist on a level of abstraction, for these Web representations themselves are further realized by some concrete series of ‘bits sent down the wire’ at some moment in space-time in the analog world. So, Web representations should not be considered to equivalent to tokens of some type of Web resource. In fact, the relationship between a Web resource and its Web representations can vary, which Berners-Lee phrases as “a relationship of genericity between the generic and the relatively specific resource” (??). To think of resources as on levels of specificity, take the examples of the book *Moby Dick*, the more
specific book *Moby Dick* in English, and the even more specific particular rendering of *Moby Dick* in English as ASCII characters. On the most specific level, a **specific resource** is a Web resource that “corresponds to only one possible bit stream” and so is always, at least in an ideal ‘Cool URI’ situation, equivalent to a particular realization to a Web representation that should not change. This Web representation should not vary over time, and always should be encoded in the same formal language (i.e. media type) and with the same natural language. More **generic resources** may exist that are *Web resources that vary over time, media type, and natural language.*

Furthermore, the distinction between information resources and Semantic Web resources has real effects. When the average hacker on the streets wants to add some information to the Semantic Web, the first task is to mint a new URI for the resource at hand, and the second task is to make some of this new information available as a Web representation. However, should a Web representation be accessible from a URI for a Semantic Web resource? If not, should Web representations can be accessed from such a Semantic Web resource, as it might confuse the Semantic Web resource itself with a Web resource that merely represents that resource. Yet how else would fulfilling the Principle of Self-Description for Semantic Web resources be possible? To refuse to allow access to any Web representations would make the Semantic Web – as Hayes pronounced earlier – completely separate from the Principles of Web Architecture. It seems Semantic Web resources need **documentation resources**, resources that have as their primary purpose the representation, however incomplete, of some Semantic Web resource. In other words, documentation resources are classical examples of metadata. Since according to the Kripkean analysis, the documentation resource is a separate thing from the Semantic Web resource it represents, the documentation resource should be given a separate URI. This would fulfill the common-sense requirement that the URI for a thing itself on the Semantic Web should be separate from the URI from a web-page about the thing. The TAG officially resolved *httpRange-14* by saying that disambiguation between these two types of resources should be done through the [303 See Other](http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html#sec10) HTTP header. The official resolution to Identity Crisis by the TAG is given below as:

- If an HTTP resource responds to a GET request with a `2xx` response, then the resource identified by that URI is an information resource;

- If an HTTP resource responds to a GET request with a `303` (See Other) response, then the resource identified by that URI could be any resource;
If an HTTP resource responds to a GET request with a 4xx (error) response, then the nature of the resource is unknown.

To give an example, let’s say an agent is trying to access a Semantic Web URI that refers to the Eiffel Tower itself, http://www.example.org/EiffelTower. Upon attempting to access that resource with a HTTP GET request, since the Eiffel Tower itself is not an information resource, no Web representations are directly available. Instead, the agent gets a 303 See Other that in turn redirects them to a documentation resource that hosts Web representations about the Eiffel Tower, such as the information resource http://www.tour-eiffel.fr/. When this URI returns the 200 status code in response to an HTTP GET request, the agent can infer that http://www.tour-eiffel.fr/ is actually an information resource. The Semantic Web URI used to refer to the Eiffel Tower itself, http://www.example.org/EiffelTower, could be any kind of resource, and so could be a Semantic Web resource. This 303 redirection then allows the Semantic Web resource given by http://www.example.org/EiffelTower to comply with the Principle of Self-Description. This example is illustrated in Figure 6.3.

An alternative to the obtuse 303 redirection is the hash convention, which one uses the fragment identifier of a URI to get redirection for free. If one wanted a Semantic Web URI that referred to the Eiffel Tower itself without the hassle of a 303 redirection, one would use the URI http://www.tour-eiffel.fr/# to refer to the Eiffel Tower itself. Since browsers, following the follow-your-nose algorithm, either dispose of it or treat the fragment identifier as a fragment of a document or some other Web rep-
presentation, if an agent tries to access via HTTP GET a Semantic Web URI that uses the hash convention, the server will not return a 404 Not Found status code, but instead resolve to the URI before the hash, http://www.tour-eiffel, which can then be treated as a documentation resource. In this way, Semantic Web inference engines can keep the Semantic Web URI that refers to the Eiffel Tower and a documentation resource about the Eiffel Tower separate by taking advantage of some predefined behavior in web browsers.

6.3.5 The Second-Generation Semantic Web

While at first these distinctions between Semantic Web resources and information resources seems ludicrously fine-grained, clarifying them and pronouncing an official W3C policy on them had an immense impact on the Semantic Web, since once there was definite guidelines on how to publish information on the Semantic Web, users could start creating Semantic Web URIs and connecting them to relevant documentation resources. The TAG’s decision on redirection was made part of a tutorial for publishing Semantic Web information called How to Publish Linked Data on the Web (?). It recommends that HTTP URIs should be used for everything, and that for any Semantic Web resource, one is likely to have to mint not two, but three URIs, “an identifier for the resource” as well as two documentation resources. One should be an human-readable HTML-based documentation resource while another distinct documentation resource should allow access to RDF (?). Furthermore, the tutorial encouraged the deployment of the Principles of Web architecture, in particular the Principle of Linking by encouraging interlinking between data-sets, as well re-use of URI for already identified resources for data integration. A few large data-sets, such as a transformation of Wikipedia to RDF called DBpedia, and well as geographical data in Geonames and biomedical knowledge to RDF in the Bio2RDF project were released (?). This “Linked Data” initiative is the second-generation of the Semantic Web. Unlike the first-generation approach, it followed the Principles of Web architecture and implemented a Kripkean distinction between Semantic Web resources and mere representations of these Semantic resources. It followed the Principle of Least Power by focusing primarily on RDF, and only using terms from “higher-level” languages when deemed absolutely necessary. In marked contrast, the second-generation Semantic Web ignored almost all inference, and focused on producing as much Semantic Web information as possible, even if the published data was inconsistent. The growth of the
Linked Data has so far been astounding, as from it grew from a few million to over a 100 million reusable RDF documents (?).

While the Linked Data initiative created URIs for many things, such as those things referred to by Wikipedia URIs, it has not created URIs for everything, such as the local pub and the proper names of people not famous enough to be on Wikipedia. For any real-world Semantic Web application, it is precisely these types of URIs that are necessary for data integration over something as simple as a Semantic Web-enabled review aggregation site like Revyu (?). Where are these URIs to come from, especially if the relevant things or owners of things aren’t going to mint the URI themselves? However, if every single application creates these not-so-well-known URIs themselves, then each application will create its own distinct URIs, so that these URIs can not be used for graph merger or any other sort of information integration.

One outcome of the Kripkean analysis of creating URIs for Semantic Web resources is the OKKAM project, which declares as its motto the famous principle of Occam’s Razor, namely rephrasing the famous maxim to “entity identifiers should not be multiplied beyond necessity” (?). The goal of this ambitious project is to provide HTTP URIs for every conceivable ‘entity,’ where an entity is taken to be some concrete ‘thing’ such as “electronic documents to bound books, from people to cars, from conferences to unicorns” as opposed to a more ‘abstract concept’ such as “predicates, relations, assertions” (?). Roughly speaking, the distinction is equivalent to the distinction in description logics between ‘entities’ as individuals in an ABox and ‘concepts’ in a Tbox which assertions can use), so that an OWL reasoner can use the formal ontology (or terminology) of the TBox to classify and make assertions about the entities (?). Following Hayes’s insight that high-level ontological distinctions are more likely to produce ambiguity, OKKAM puts forward the thesis that “while any attempt at ‘forcing’ the use of the same URIs for ‘logical resources’ [abstract concepts] is likely to fail (as every application context has its own peculiarities, and people tend to have different views even about the same domain), the same does not hold for entities” ?. Everyone is likely to disagree about the concept of justice or even person hood but OKKAM supposes there is unlikely to be disagreement about concrete entities like Gustave Eiffel or the Eiffel Tower. However, in a decidedly Kripkean move, instead of building a huge database that contains logical descriptions of the entities, OKKAM merely will construct an enormous and open-ended list of Semantic Web URIs to serve as names for referent. OKKAM can be thought of as the reverse of URI declarations, the only documentation resources to be attached to these OKKAM Semantic Web URIs
will be non-logical: collections of pictures, text from other web-pages which mentions the same referent, and the like. OKKAM stores “untyped data for the reason that typing an entity’s attributes would require us to classify the entity” because any logical description could lead to disagreement and thus harm re-use of the URIs (?). OKKAM so hopes to concretely realize the dream of the Semantic Web as a giant manual that can translate URIs to referents, but without logical descriptions at all.

### 6.4 Critiques of the Causal and Descriptivist Theory of Reference

While as compelling as a Kripkean vision of the Semantic Web as a giant translation manual from URIs to referents is, it is immediately beset by Quine’s famous thesis on the *indeterminacy of translation* (?). Furthermore, simply retreating and accepting the logicist theory of reference also fails, for the translation problem present in some magical Kripkean translation manual is also present in the realm of logical interpretation. This does not bode well, for just as the logicist theory of reference has clearly failed for the Semantic Web, already cracks are beginning to appear in the more Kripkean Linked Data initiative and the OKKAM project. Therefore, the only way forward seems to move from the primacy of reference over meaning to the primacy of some more all-encompassing notion of meaning over reference, which Frege’s ‘sense’ may provide an outline of. From these moves, a theory of meaning for the Semantic Web can be rehabilitated, one that does justice to the purpose of the Semantic Web.

#### 6.4.1 Radical Translation and the Failure of the Direct Reference Position

The argument of Quine for the indeterminacy of translation was originally phrased as an argument about the creation of syntactic translation manual from one language to another. Imagine a linguist, trying to construct a translation manual for the language of a “hitherto untouched people” (?). The problem of how a linguist should proceed is considered the problem of *radical translation*. Imagine that one of the natives pointed to what appeared to the linguist to be a rabbit and uttered the word “gavagai” (?). The natural response of the average linguist would be to identify the occasion of the rabbit with the name “gavagai,” yet Quine argues that there are actually “an infinity of log-
ically possible meanings for gavagai” (¿). The rabbit could have just passing by, and ‘gavagai’ could be the native’s way of saying ‘Lo, let’s go hunting!’ However, one would assume after enough back and forth observation and questioning, a more forgiving philosopher than Quine would assume that ‘gavagai’ really does mean ‘rabbit.’ However, the native could have really meant ‘undetached rabbit-part’ or ‘a temporal stage of a rabbit’ by ‘gavagai,’ and our linguist could not tell the difference. Quine points out that even our verbal disposition to baptize names for referent is necessarily underdetermined by any physical stimuli. This has serious consequences for Kripkean baptism, as it makes problematic how to causally then connect a word with a distinct referent.

The argument of radical translation easily applies to the notion of the Semantic Web as a translation manual, where URIs are causally attached to referents by an act of baptism. Given Quine’s observation that even the most detailed of documentation resources would undermine any translation, how can a URI translate to a referent unambiguously, even through OKKAM-style non-logical information? It appears all translation from one language is to another is indeterminate, including translation from URIs to whatever referents a user-agent wants to use on the Semantic Web. First, http://www.example.org/EiffelTower can not give us the Eiffel Tower itself. Violating URI opacity, an agent could guess that the URI referred to the ‘Eiffel Tower,’ by inspecting the URI but it could just as easily refer to ‘part of the Eiffel Tower’ or copies of the monument like the one in Paris, Texas? Even being redirected via a 303 redirection or the hash convention to a documentation resource with a definition in natural language like the English words “The Eiffel Tower in Paris” with some pictures would not escape indeterminacy. Even Web representations that contain natural language terms or sentences are not too much help if one was translating a language one does not already know, such as a hitherto unused form of computing language like RDF. One would hope at least one could detect a Semantic Web resource when one found it, but there is no clear mark of being a Semantic Web resource, no special URI scheme, no graphic, nothing whatsoever.

Practically, the 303 redirection of the W3C TAG and the hash convention, on purely technical grounds, are severely indeterminate. Imagine the redirection from http://www.example.org/EiffelTower via 303 to http://www.tour-eiffel.fr. Assume that an agent goes through this redirection process; the only thing the agent knows after the 303 redirection is that http://www.tour-eiffel.fr/ is an information resource can be used to access webpages. We know nothing about the original
URI, http://www.example.org/EiffelTower, since the 303 status code can not possibly tell an agent that the resource redirected from was a Semantic Web resource, arises because the 303 status code was specified as ‘See Other’ before the Identity Crisis was even noticed or the Semantic Web even existed.\(^1\) As an HTTP status code, there is no reason why it can’t be used – and usually is used – to redirect from one information resource to another information resource. For humans, since the attendant Web representations of the documentation resource bears no trace of the redirection, no disambiguation is achieved for human user-agents, at least the vast majority of humans that do not monitor HTTP status codes.

The same general argument applies against the hash convention, since fragment identifiers were originally meant to be a fragment of a document or some other representation. The URI specification says “the semantics of a fragment identifier are defined by the set of representations that might result from a retrieval action on the primary resource. The fragment’s format and resolution is therefore dependent on the media type of a potentially retrieved representation, even though such a retrieval is only performed if the URI is dereferenced” (\(\text{?}\)). Only “if no such representation exists, then the semantics of the fragment are considered unknown and are effectively unconstrained” (\(\text{?}\)). Only if an agent get a 404 status code from http://www.tour-eiffel.fr/ can http://www.tour-eiffel.fr/# mean a non-accessible Semantic Web resource. If a Web representation with the text/html media type is returned by accessing the URI without a hash, then according to the HTML specification, “the fragment identifier designates the correspondingly named element” (\(\text{?}\)). From the point of view of current standards fragment identifiers could just as well be used for named elements in the document instead of a shortcut for distinguishing Semantic Web resources and their associated description resource. In the case where the application/rdf+xml media type is returned by the accessible URI, things are different. “In RDF, the thing identified by a URI with fragment identifier does not necessarily bear any particular relationship to the thing identified by the URI alone” so the hash convention can legitimately identify anything, including non-accessible resources (\(\text{?}\)). This seems to defeat the

\(^{1}\) Arguably, from a purely technical standards viewpoint, this use of 303 is also an abuse, since as put by the specification, “this method exists primarily to allow the output of a POST-activated script to redirect the user agent to a selected resource,” not to distinguish between Semantic Web resources and information resources like documentation resources (\(\text{?}\)). Valid only in HTTP 1.1, the 303 status code was invented not to distinguish between Semantic Web resources and documentation resources, but due to the over-use of the HTTP 1.0 302 status code to redirect both temporarily and permanently, so the 307 status code was created for temporary redirection while the HTTP 303 status code dealt with permanent redirection (\(\text{?}\)).
point of returning representations, since unlike rendered HTML, RDF is much more easily used by machines than humans. If people accessed http://www.tour-eiffel.fr/ and received RDF/XML most would have no idea what to do with it. Furthermore, one is in deeply treacherous and indeterminate waters if the Web representation is given by content negotiation, as one could apply either the RDF or HTML interpretation. So, using hash URIs has the exact same problem as 303 redirection, since it doesn’t normatively define any sort of relationship between the two URIs, much less distinguish between Semantic Web resources and information resources.

6.4.2 Radical Interpretation and the Failure of the Logicist Position

Regardless of its relatively poor implementation, Quine’s argument has demolished very theoretical advantages behind the direct reference position’s claim that unambiguous reference is possible. Furthermore, the argument for the indeterminacy can even be applied to the logicist theory of reference, to question the very idea of truth conditions even ambiguously providing some determination of reference. This translation of the syntactic radical translation of Quine to the level of semantics is called the radical interpretation by Davidson (?). Davidson notes that radical translation “deals with the wrong topic, the relation between two languages, where what is wanted is an interpretation of one” (?). Donald Davidson’s primary intellectual project was to apply Tarski’s logicist conception of truth to the analysis of meaning in natural language as to reduce the notion of meaning in language to truth over satisfaction conditions. Davidson safely escapes the problems radical translation causes the direct reference position, since he views terms in language as ambiguous, and radical translation does not deal with construction of the interpretation. Yet the argument can return. Imagine trying to understand the language of a visiting alien coming to earth? Forget making a dictionary from the language to English, in radical interpretation one wants to construct a theory of meaning for the alien’s sentences, even if no direct one-to-one translation of words to either other words or singular referents is possible.

For Davidson, the construction of these interpretations takes the form of assign Tarski-style truth conditions to sentences. Just as in Tarski’s theory, at some point after decomposing a statement into its components, an agent bottom out in sentences that have meaning even if the referents of each term therein is ambiguous. Just as in logic, Davidson claims that in natural language it is impossible for a single name by itself to mean anything except in its relationship to other names. It seems correct that
a word finds its meaning in other words, but Davidson adopts a logicist position, so the meaning does not lie in words themselves, but in the network of logical statements containing that term and their truth conditions.

Exactly what precise cluster of statements are necessary for determining the interpretation of a single name is not clearly defined. Davidson himself seems to believe in holism, stating that “a sentence (and therefore a word) has meaning only in the context of a (whole) language” (?). This defining meaning in terms of the truth conditions of an entire language leads to a curious paradox Davidson has noted (?). In the style of Tarski, the sentence ‘The Eiffel Tower is a monument in Paris’ is true if and only if the Eiffel Tower is a monument in Paris. However, ‘The Eiffel Tower is a monument in Paris’ is true if and only if a monument can occupy physical space. Also, something like a grass can occupy physical space. Furthermore, grass is green. So, ‘The Eiffel Tower is a monument in Paris’ is true if and only if grass is green. Indeed, it seems absurd to state that the meaning of ‘The Eiffel Tower is a monument in Paris’ has anything to do with the fact that grass is green.

The only way out of this impasse is to employ what Davidson entitles the Principle of Charity (a version of which is also found in Quine), such that if we were interpreting the language of some visiting alien, we must assign “truth conditions to alien sentences that make native speakers right when plausibly possible, according, of course, to our own view of the right. What justifies the procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement” (?). Does this homely dictum somehow save the logical holism from radical interpretation, and so save the logicist position on the Semantic Web, at least on the level of theory?

As sensible as it appears, the Principle of Charity fails to actually provide shared referents, since vastly different and even opposing interpretations are given by humans from different cultures (?). Take for example the litmus test given by Kripke to distinguish the causal theory of reference from the descriptivist theory of reference, “Suppose that Gödel was not in fact the author of [Gödel’s theorem. A man called ‘Schmidt’ ... actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the [descriptivist] view in question, then, when our ordinary man uses the name ‘Gödel’, he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description ‘the man who discovered the incompleteness of arithmetic’... but it seems we are not. We are simply not” (?). Despite Kripke’s convincing argument that seems to show that the
The descriptivist theory of reference violate what appears to be common-sense, multiple empirical studies have determined that, at least insofar as this particular example is concerned, people from Western cultures significantly favor the causal theory of reference and so believe that Gödel is the correct reference, while people from Chinese cultures favor the descriptivist theory, and so think that Schmidt is the correct reference (\textendash\).

The Principle of Charity runs into even further trouble on the Semantic Web since the interpretation involves the inference machinery of formal languages like OWL and RDF used by machines, not natural languages spoken by humans. A human agent can not reliably use the Principle of Charity and apply it to an inference engine, for humans are well-known at being unable to predict accurately the results of inference (\textendash\). For example, even most competent human mathematicians cannot single-handedly interpret a complex machine-aided proof like the proof of the ‘Four-Color’ theorem (\textendash\).

However, a Semantic Web inference engine could make a Principle of Charity for other inference engines, since all computational Semantic Web-enabled agents should have implemented the same formal semantics for OWL and RDF. However, as mentioned in Section 6.2.4, this is difficult as there is no strict way to determine what, given a URI, exactly what statements one computational agent should assume another agent to use in its interpretation. If we wish to consider \textit{symbiotic complexes of humans and machines}, both the radical indeterminacy of translation effects any natural language text found at the URI, and ambiguity is still rife in the formal interpretation of RDF and OWL by an inference engine. Combining both natural and formal languages two can make matters \textit{worse} rather than \textit{better} in an attempt to find the referents, such as if the information in the two languages contradict each other, or if they just seem orthogonal. Reasonably, they may help each other, but then both combined will still fall victim to radical translation and radical interpretation.

The second argument against radical interpretation involves the objection to logical holism. We will call Davidson’s theory ‘logical’ holism rather than ‘linguistic’ holism as Davidson is putting forward the thesis that understanding reference in a language requires understanding the Tarski-style truth conditions of every statement in the language (\textendash\). Obviously, one objection to holism is that in practice, it violates the \textit{learnability} of language. If one can only understand a language if one understands the truth conditions of \textit{all} statements in a language, then it would be impossible to learn any new language piecemeal, which is obviously seems possible in at least human practice. As Dummett puts it, “For holism, language is not a many-storeyed structure,
but rather, a vast single-storyed complex; its difficulties in accounting for our piecemeal acquisition of language result from the fact that it can make no sense of the idea of knowing part of a language” (?). Second, it is even unclear if it even possible for a single human agent to claim to understand even its native natural language tongue, since the truth conditions for many statements, such as “Every man is mortal” require in a model require quantification ranging over infinite sequences, and this seems at first glance cognitively impossible (?).

The most potent objection is that the very idea of the truth-conditions for an entire language is incoherent. To illustrate, Dummett invokes an extension of the problems arising from Wittgenstein’s observations about the name ‘Moses,’ noting that “quite a few things are assumed to be true about the referent of ‘Moses,’ such as “he was brought up in a royal palace, that he led his people out of slavery” and so on (?). Given the possible descriptions of ‘Moses,’ does any possible referent have to involve satisfying all descriptions about Moses, or are some descriptions more important than others? Instead of becoming less problematic, this becomes more problematic as new names are added to a language, such as as the sentence ‘Moses and Aaron were brothers.’ The holistic interpretation of a language could be construed as consisting of the any interpretation that makes most of the names refer to things in a model that successfully satisfy their descriptions.

While the satisfaction of ‘most of the truth’ conditions may seem plausible if a language is considered to be spoken only by a single speaker, an idiolect which is “an individual’s partition, and often in part incorrect, understanding of his language,” there are substantial problems with the “personal theory of truth” (?). Namely, a language is “something essentially social” the truth-conditions of statements in an idiolect “need to be explained in terms of the notion of a shared language, and not conversely” (?). If more than one agent is using a language in order to share an interpretation, then holism runs into trouble. What if one person believes ‘Moses’ is ‘Aaron’s brother’ and the other person is not familiar with ‘Aaron’ or even thinks that the two were not brothers? If there is disagreement or lack of knowledge, does the shared holistic interpretation compromise “just those sentences which all speakers accept as true, or at least, which many accept as true and none reject as false, and therefore, as including only sentences that have no significant indexical feature” (?). This, of course, would eliminate logical holism from applying to most of natural language. Even with non-indexical statements in formal computer languages, determining what this core of statements must involve somehow already knowing or even guessing the interpretations of all other agents us-
ing that language in order to determine what precisely are the truth-conditions majority of sentences, so the theory loses all practical value in determining what the interpretation of an individual statement. Furthermore, “no one speaker comes anywhere near having a grasp of the theory of meaning for the entire language, since the vast majority of the judgments will be unknown to him” (?). If there is no actual procedure for determining the truth conditions for an entire language even in theory, then holism must be abandoned.

The primary alternative to holism, some form of atomism as put forward by Russell, is already discredited. The only alternative to holism is some version of molecularism, in that the meaning of sentences or term in a language is found in the meaning of a finite number of sentences or terms, a subset of all possible sentences in a language. Of course, the question then is how to determine what precisely a molecule is. On the Semantic Web, this resonates well the very practical difficulties, detailed in Section 6.2.4, of determining what precise set of statements on the Semantic Web should be used to determine the truth-conditions of the interpretation. Due to the Principle of the Open World, it would be impossible to state that the interpretation should satisfy all possible uses of a URI on the Semantic Web, since the Web is constantly growing and the URI may appear elsewhere on the Web, unbeknown to the agent. The Principle of Self-Description seems to be in favor of an approach like URI Declarations, but because of the somewhat open-ended nature of the follow-your-nose algorithm, it is unclear how to determine the precise bounds of the URI Declaration (?). Seeing as the question of how precisely to determine the boundaries of these molecules both in natural language and formal languages is far from agreed upon, seeming to be a mystery, perhaps it is better to throw out the entire logicist position of trying to ground out meaning in reference not only due to its failure in practical terms as detailed in Section 6.2.4, but on the level of theory as well. Is there a way that molecularism can be saved without recourse to purely and only logical interpretation? Yet, if both Berners-Lee and Kripke’s causal theory of reference as well as the logicist theory of reference championed by Hayes and Davidson are implausible, what new third position be staked out? What is the definition of meaning, if not the interpretation of terms in a language to referents?
Despite their best efforts and intentions, both the logicist and causal theory of reference have been unable to provide a satisfactory account of the reference of URIs for the Semantic Web, either on the level of philosophical theory or on the level of actual practice. Both these theories of reference sought to reduce the concept of the meaning to that of reference, with the causal theory of reference hoping to reduce meaning to direct unambiguous reference, and the logicist theory of reference hoping to reduce meaning to the usually ambiguous satisfaction of truth conditions. In brief, despite their apparent distinctions, both the causal and logicist account are essentially want to build meaning out of reference. The causal account could be summarized as the slogan *the priority of reference over meaning*, where reference is established via baptism or scientific knowledge, and the logicist account as the *composition of meaning via reference*, where the conditions of reference are given by truth-conditions of logical descriptions of the referents. Yet, meaning could somehow be a larger, more expansive concept than any simple notion of reference or even truth.

To discover what we are this more difficult notion of meaning could be, we cant return to Frege’s original controversial theory of sense and reference as given in *Sinn und Bedeutung* (\(^2\)). This work whose distinction between ‘sense’ and ‘reference’ provoked both Russell and Kripke’s intellectual projects to build an entire theory of meaning on top of only reference, since the notion of ‘sense’ was thought of by both Russell and Kripke, as well as philosophers like Davidson and Quine, as vague and unnecessary. This is likely because Frege himself was quite cryptic with regards to any definition of ‘sense.’

The key idea lies in Frege’s contention that the meaning of any term including names, is determined by what Frege calls the “sense” of the sentences that use the term, rather than any direct reference of the term (\(^2\)). According to Frege, two sentences could be the same only if they shared the same sense. Take for example the two sentences “Hesperus is the Evening Star” and “Phosphorus is the Morning Star.” Since the ancient Greeks did not know that ‘The Morning Star is the same as the Evening Star,’ they did not know that the names ‘Hesperus’ and ‘Phosphorus’ share

---

\(^2\)The ambiguous translation of this work from original German has been a source of great philosophical confusion. While the word ‘Sinn’ has almost always been translated into ‘sense,’ the word ‘Bedeutung’ has been translated into either ‘reference’ or ‘meaning,’ depending on the translator. While ‘Bedeutung’ is most usually translated into the fuzzy English word ‘meaning’ by most German speakers, the use to which Frege puts it is much more in line with how the word ‘reference’ is used in philosophy. So in the tradition of Michael Dummett, we will translate Frege’s ‘Bedeutung’ into ‘reference’?.
the same referent when they baptized the same star, the planet Venus, with two different names (??). Therefore, Frege says that these two sentences have distinct ‘senses’ even if they share the same referent. Furthermore, Frege pointed out that, far from being meaningless, statements of identity that would be mere tautologies in a theory of reference are actually meaningful, and that one can understand a statement like ‘The Morning Star is the Evening Star’ without knowing that both refer to Venus. Frege defines ‘sense’ in terms of the mysterious *mode of presentation*, for “to think of there being connected with a sign (name, combination of words, letter), besides that to which the sign refers, which may be called the reference of the sign, also what I should like to call the sense of the sign, wherein the mode of presentation is contained” (??).

Regardless of how mysterious ‘sense’ is, Frege believed that the notion of sense is what allows an agent understand sentences that may not have a referent, for “the words ‘the celestial body most distant from Earth’ has a sense, but it is very doubtful is there is also a thing they refer to...in grasping a sense, one certainly is not assured of referring to anything’ (??). Frege did believe, as was developed further by Tarski, that complex sentences could be built from simple constituent sentences and phrases. The phrase “the King of France”, which under a Russelian logicist theory of reference would be false or under a Kripkean causal theory of reference would somehow be indexed to a particular person, perhaps the last king of France or all kings of France. Instead of having each name have a single (or range of things, or nothing) thing it refer to, a name without a reference have their meaning grounded out in sense rather than reference. This leads to the rather curious finding that a phrase such as ‘the King of France is bald’ – unless it is given in some further group of sentences that specify further information, such as ‘the King of France is bald in 1700’ – has no truth value and no referent. This is not to deny the role of reference whatsoever, since “to say that reference is not an ingredient in meaning is not deny that reference is a consequence of meaning..it is only to stay that understanding which a speaker of a language has of a word in that language...can never consist merely in his associating a certain thing with it as its referent; there must be some particular *means* by which this association is effected, the knowledge of which constitutes his grasp of its sense” (??). In other words, the sense of a sentence would naturally for many sentences give an agent the correct determination of reference for names in the language, the

Frege himself was a mathematician and the inventor of most of the foundations of modern logic, and so his theory of language was merely a stepping-stone in his purpose to create a new language capable of expressing *true* mathematical thoughts. Therefore,
in his use of the theory and reference, he endorsed the notion that the sense of sentence could refer to the ‘true’ or the ‘false’ and that the purpose of a formal language like logic was to eliminate the ambiguity of natural languages that were capable of creating sentences without any truth or falsity. It is not surprising then that Russell, Tarski, Hayes, and others have tried to ground out Frege’s theory of sense in notions of logical truth, taking the sense to be some sort of linguistic shortcut to a cluster of logical descriptions. While some truth-maintaining inference is necessary, the Principle of the Open World maintains that the distinction between truth and falsity is not the primary concern, but moreso the ability to name a non-Web accessible thing on the Semantic using the Principle of Universality and share the name and its representations with others via the Principle of Linking and Self-Description. So, what would be a theory of sense for the Semantic Web be if not based on logical truth and falsity?

The reason the notion of sense was thought of as so objectionable by many like Russell and Kripke was that it was viewed as a private, individual notion, much like the Lockean notion of an idea. Frege himself strictly separates the notion of a sense from an individual subjective idea of a referent, for Frege believed that sense was inherently objective, “the reference of a proper name is the object itself which we designate by using it; the idea which we have in that case is wholly subjective, in between lies the sense, which is indeed no long subjective like the idea, but is yet not the object itself” (?). A sense is objective insofar as it a shared part of an inherently public language, since a sense is the “common property of many people, and so is not a part of a mode of the individual mind. For one can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another” (?). While the exact nature of a sense is still unclear, it’s main characteristic is that it should be whatever is objectively shared between agents as regards their use of names in a language. This positing of language, not as a shorthand for logical statements or some gift from experts who “really” know meaning, but as a set of public practices, leads to what Dummett notices is a strange coincidence, that “Frege’s thesis that sense is objective is thus implicitly an anticipation of Wittgenstein’s doctrine that meaning is use” (?). Therefore, given the failure of the logicist and causal solutions to the Identity Crisis, in the next chapter we demonstrate how a neo-Wittgensteinian theory of meaning can solve to the Identity Crisis on the level of theory, and then even solve it in practice in the Chapter ??.

Appendix A

Glossary of Terms

The task of classifying all the words of language, or what’s the same thing, all the ideas that seek expression, is the most stupendous of logical tasks. Anybody but the most accomplished logician must break down in it utterly; and even for the strongest man, it is the severest possible tax on the logical equipment and faculty.

Charles Sanders Peirce, letter to editor B. E. Smith of the Century Dictionary

This appendix presents the terminology used in this thesis. Some of this terminology is further formalized and presented as a Semantic Web ontology in Chapter ??.

- **absolute URI**: a URI in which there must a single scheme and the scheme must identify a name of a resource.

- **access**: the use of a identifier to create immediately a causal connection to the process identified.

- **agent**: Any thing capable of having an interpretation.

- **analog** anything that is not digital.

- **arc role**: the URI of a link that provides information about what kind of link the link itself belongs to.

- **authority** is a name that may, but does not have to be translatable by the domain name system into an IP address. Often the name of a server.
• **A WWW (Architecture of the World Wide Web)**: A Recommendation produced by the W3C to describe the defining characteristics of the Web, available at http://www.w3.org/TR/webarch/.

• **cache**: when a user-agent has a local copy of a Web representation that it accesses in response to a request rather than getting a Web representation from the server itself.

• **causal theory of reference**: any name refers via some causal chain directly to a referent

• **channel**: The physical substrate that determines whether or not the information is preserved over time or space.

• **client-server architecture**: protocols that take the form of a request for information and a response with information.

• **client**: The agent that is requesting information. In the context of the Web, called a user-agent.

• **complete**: The inference procedure of a language if every satisfied statement can be shown to be entailed

• **content**: whatever is held in common between the source and the receiver as a result of the conveyance of a particular information-bearing message

• **consistent**: A statement or statements that can not be satisfied.

• **content negotiation** is a mechanism defined in a protocol that makes it possible to respond to a request with different Web representations of the same resource depending on the preference of the user-agent

• **content types**: The types of formal languages that can be explicitly given in a response or request in HTTP.

• **convention**: the use of a thing based purely on previous history, without regard to imitation or natural selection.

• **depictions**: The expressions of a iconic language.

• **descriptions**: linguistic expressions of an natural or formal language.
• **descriptivist theory of reference**: the referent of a name is given by whatever satisfies the descriptions associated with the name.

• **dialect**: A language embedded as a subset of another language.

• **digitality**: a discrete boundary in a particular interpretation of some content converges with a regularity in a physical system. See *discrete interpretation*.

• **direct reference position**: On the Web, the meaning of a URI is whatever was intended by the owner.

• **discrete interpretation**: a relationship from an encoding to some content where the content itself is also finitely differentiable, so that any given content of a type can be distinguished from any other content of another type. In order to distinguish types, there must be some physical regularity that serves as a *boundary*.

• **domain names**: a specification for a tree structured name space, where each component of the domain name (part of the name separated by a period) could direct the user-agent to more specific “domain name server” until the translation from an identifier to the name to IP address was complete.

• **encoding**: the precise regularities physically and locally realized by the information-bearing message.

• **ending resource**: The resource a link is directed to.

• **endpoint**: Any process that either requests or responds to a protocol.

• **entailment**: A semantic relationship where an interpretation of one sentence always satisfies the interpretation of another sentence.

• **entity**: the information transferred as the payload of a request or response, which may include an entity body and optional headers. In HTTP, the entity consists of optional headers and an entity body.

• **entity body**: the information transferred as the payload of a request or response excluding any optional headers. Sometimes also called the content, although we use that term in a different sense, see *content* for our use.

• **exemplification**: the physical thing that locally carries the information.

• **extension**: Things that satisfy a description.
• **generic resource**: Web resources that vary over time, media type, and natural language.

• **graph merge**: when two formerly separate graphs combine with each other when they use any of the same URI.

• **finitely differentiable**: when it is possible to determine for any given mark whether it is identical to another mark or marks.

• **follow-your-nose algorithm**: In summary, an agent can follow the following steps in to help interpret a resource identified by a URI: dispose of any fragment identifier, inspect the media type of the retrieved Web representation, follow any namespace declarations, and follow any links. Available in full in Section 4.2.3.

• **formal languages**: a language with an explicitly defined syntax and possibly model-theoretic semantics, so suitable for regular interpretation by computers.

• **format**: synonym for *language*, particularly for on computer-based language.

• **headers**: In HTTP, the part of the method that specify some information that may be of used by the server to determine the response or that specifies to the client information about the response.

• **hierarchical component** the left to right dominant component of the URI that syntactically identifies the resource.

• **holism**: a sentence has meaning only in the context of a whole language.

• **HTTP (HyperText Transfer Protocol)**: A protocol originally purposed for the transfer of hypertext documents, although its now ubiquitous nature often lets it be used for the transfer of almost any encoding over the Web.

• **iconic languages**: a language based on images.

  *identifier* is a *term that can be used to either access or refer, or both access and refer to, a thing.*

• **inbound links**: where the ending resource is a local Web representation and the distal starting resource is given by an identifier.

• **inconsistent**: A statement or statements that can not be satisfied.
• **intension**: kind of thing may only be described.

  **identifier**: a term that can be used to either access, refer to, or access and refer to a thing.

• **information**: whatever in common between two processes, where one process is called the *sender* and the other is called the *receiver*. To have something in common means to share the same regularities, e.g. parcels of time and space that cannot be distinguished at a given level of abstraction.

• **information resource**: some resource that is information with the possibility of a digital encoding.

• **interpretation**: the relationship between an encoding and its content. Formally, used in two distinct but related ways, an *interpretation mapping* that denotes the relationship between a language and a model, and the *interpretation structure*, which is a model that satisfies a particular interpretation mapping.

• **interpreter**: a thing that is capable of having an interpretation from a particular encoding and a content.

• **knowledge representation language**: a language whose primary purpose is the representation of non-digital content in a digital formal language.

• **level of abstraction**: a way of recognizing certain physical differences and regularities.

• **link**: a connection between resources.

• **linkbase**: where the links can be represented outside of any Web representation of the starting or ending resource.

• **location** is an identifier that can be used to access a thing.

• **logicist position**: For the Semantic Web, the meaning of a URI is given by whatever model(s) satisfy the formal semantics of the Semantic Web.

• **mark**: is a physical characteristic.

• **media type**: A generalization of content types to any Internet protocol. It consists of a two-part scheme (separated by the `\/'`) that separates the type and a subtype of an encoding.
• **message**: The physical process that can be thought of as bearing information due its local properties. In HTTP, messages are also things that have headers and optional entity bodies.

• **method**: A request for a certain type of information from a user-agent.

  *molecularism*: the meaning of sentence or term in a language is found in the meaning of a finite number of sentences or terms.

• **model**: a mathematical representation of the world or the language itself.

• **model-theoretic semantics**: When an interpretation of a language is to a mathematical model

• **monotonic**: in a system capable of inference, when the inference relationship ⊢ is monotonic if and only if for all sets of statements $s_1$ and $s_2$, and all inferred statements $s_3$, if $s_1 ⊢ s_3$ and $s_2 ⊃ s_3$ then $s_2 ⊢ s_3$.

• **language**: a system in which information-bearing messages are related to each other in a manner that can change their interpretation.

• **name**: an identifier that can be used to refer to a thing.

• **namespace declaration**: Within a given Web representation in a particular dialect, the information that specifies the namespace URI of the dialect.

• **namespace document**: a Web representation that provides more information about the dialect.

• **namespace URI**: a URI that identifies that particular dialect.

• **natural languages**: a language based on human linguistic expressions.

• **non-monotonic**: when monotonicity does hold for a system capable of inference.

• **non-spooky**: anything that does not appear to violate these common-sense intuitions about physics and causation.

• **payload**: The information transmitted by a protocol.

• **path component**: a number of text strings delimited by special reserved characters that identify a resource.
• **Principle of Least Power**: A Web representation given by a resource can be described in the least powerful but adequate language.

• **Principle of Linking**: Any URI or Web representation can be linked to another resource identified by a URI.

• **Principle of the Open World**: the number of resources on the Web can always increase.

• **Principle of Self-Description**: The information an agent needs to have an interpretation of a resource should be accessible from its URI. See the *Follow-Your-Nose* algorithm.

• **Principle of Universality**: any resource can be identified by a URI.

• **proper function**: whatever characteristics which a thing has in lieu of those characteristics promoting the reproduction or imitation of the thing.

• **protocol**: A convention for transmitting information between two or mobile agents.

• **proxy**: a cache that is not stored on the user-agent itself, but are shared among multiple user-agents by a server or group of servers.

• **public language position**: The Web is a form of language, and language exists as a public mechanism among multiple agents, then the meaning of a URI is the use of the URI, which must be a public mechanism that easily fits in the form of life of agents on the Web, which lets them in turn establish, find, and re-use URIs.

• **purpose**: the intended use of a thing as given by its creator, regardless of its history of natural selection or imitation.

• **Open World Assumption**: statements that cannot be proven to be true cannot be assumed to be false.

• **Open World Principle**: See *Principle of the Open World*.

• **owner**: the agent that have the ability to create and alter the Web representation accessible from the URI.
- **outbound links**: Links that are inserted into Web representations directly and go beyond the local Web representation to an distal ending resources.

- **realization**: Synonym for exemplification.

- **receiver**: See information.

- **relative URIs**: a URI in a scheme where the path component itself is enough to identify a resource within certain contexts.

- **reference**: the relationship of an thing to another thing to which one is immediately causally disconnected.

- **referent**: the distal thing referred to by a representation.

- **representations**: information-bearing messages having distal content. Note that this word “representation” has a distinct meaning in terms of its usage in Web standards, which we disambiguate by using the term Web representation. See Web representation for details.

- **representational cycle**: Simply put, process $S$ is connected with process $O$. The process $S$ is connected with $R$. An input procedure of $S$ puts $R$ in some causal relationship with process $O$. Processes $O$ and $S$ change in such a way that the processes are disconnected. Due to some local change in process $S$, $S$ uses its connection with $R$ to initiate local behavior that is in part caused by $R$. See Section 3.5 for details.

- **resource**: any thing capable of having identity. A resource is typically not a particular encoding of the information but the content of the information that can be given by many encodings.

- **request**: In HTTP, the method used by the agent and the headers, along with a blank line and an optional message body.

- **response**: In HTTP, the combination of the status code and the entities.

- **REST (Representational State Transfer)**: An architectural style in which all state where the information state of the interaction between the between the server and client is stored on the client.
• **role**: A URI that can be attached to a link to provide information about the ending resource.

• **satisfaction**: An interpretation that defines whether or not every statement in the language can be given a content in the world

• **scheme**: A name of a protocol or other naming convention. Unlike protocols, a scheme does not have to be capable of transmitting information.

• **sentence**: Any combination of terms that is valid according to the language’s syntax.

• **semantics**: A system in which the content of information-bearing messages are related to each other.

• **Semantic Web**: The use of the Web as a language to represent analog things.

• **Semantic Web resource**: Some resource that is analog.

• **Semantic Web URI**: A URI for a Semantic Web resource.

• **server**: The agent that is responding to the request.

• **specific resource**: A Web resource equivalent to a particular realization to a Web representation that should not change.

• **standard**: A convention for the encoding and possibly interpretation of information.

• **statement**: Any combination of terms that has an interpretation according to the language’s semantic

• **status code**: One of a finite number of codes gives the user-agent information about the server response itself.

• **sound**: The inference procedure of a language if every inferred sentence can be satisfied.

• **source**: See information.

• **starting resource**: The resource that the link is directed from.
- **state**: information about a resource that is not given as part of its identity, so it is information that may change over time.

- **syntax**: a system in which the content of information-bearing messages are related to each other.

- **system**: A synonym for *thing* and *process*, used to emphasize the fact that on a level of abstraction, one thing (the system) can be given as multiple things.

- **term**: Regularities in marks.

- **thing**: Synonym for *process*.

- **types**: In a particular interpretation or for a given purpose, an equivalence class of marks, encodings, or content.

- **spooky**: anything that appears to violate these common-sense intuitions about physics and causation.

- **user-agent**: A client in the context of the Web.

- **URI** (Uniform Resource Identifier): A unique identifier whose syntax is given in ?, that may be used to either or both refer to or access a resource.

- **URI Collision**: When the same resource has multiple URIs.

- **URI Opacity**: a URI should never itself have an interpretation, only the information referred to or accessed by that URI should have an interpretation.

- **URL** (Uniform Resource Locations): A scheme for locations that allows user-agents to via an Internet protocol access a realization of information.

- **URN** (Uniform Resource Name): A scheme whose names that could refer to things outside of the causal reach of the Internet.

- **Web representation**: the encoding of the content given by a resource given by a resource given in response to a request.

- **world**: All things and processes.

- **WWW** (World Wide Web): An information space in which resources are identified by URIs.