
The Web of Reference

Harry Halpin

May 20th, 2004



Background

Interests: Computational linguistics with real-world data, philosophy of computation, the Web.

Motivating Methodology: What are people doing everyday and how can computers aid them? Formalize interdisciplinary insights as much as possible to build working computer programs.

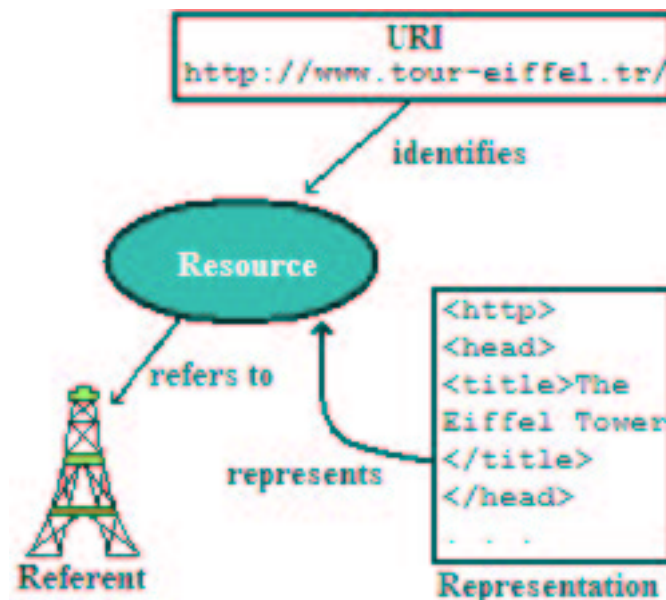
Advisor *Henry Thompson and Johanna Moore*

Homepage *<http://www.ibiblio.org/hhalpin>*

E-mail *hhalpin@ibiblio.org*

Institute for Communicating and Collaborative Systems

People often search for information *about* things, not for particular web-pages
Which web-pages are about which things? Which utterances are about which things in text or dialogue?



Example

Similar to the **use-mention** distinction in natural language.

The problem of reference in RDF:

Gustave Eiffel is the creator of the Eiffel Tower, Gary Feuerstein created a web-page about the Eiffel Tower.

`http://www.paris.org/Monuments/Eiffel
creator`

`http://www.endex.com/gf/`

`http://www.paris.org/Monuments/Eiffel
creator`

`http://www.gustaveeiffel.com/`

Dealing with Reference

Easy: Humans usually can identify reference. Provide them a scheme with common formats that makes the distinction between referent and reference, **Web Proper Names (WPN)**.

Hard: Machines find this difficult. Instead of just inspecting URI, inspect the content of the HTML web-page itself using IR and NLP tools, i.e. use of **GridNLP** tools. In NLP tasks such as dialogue, multimodal information will be required.

Benefits: A level of automated mapping among ontologies and schemas, ontology creation from WPN bookmarks, semantic extraction.

Implementation: Implement and integrate WPN and GridNLP tools into web browser and spider, apply principles to multimodal dialogue data.