

Semants

Towards Converting the Internet into Topic Maps

Lars Heuer <heuer@semagia.com>
Semagia - <http://www.semagia.com>

TMRA'06 • 12.10.2006



Table of Contents



- Introduction
- Architecture
- Wikipedia Semants
- Other Semants
- Further Work
- Summary

Introduction



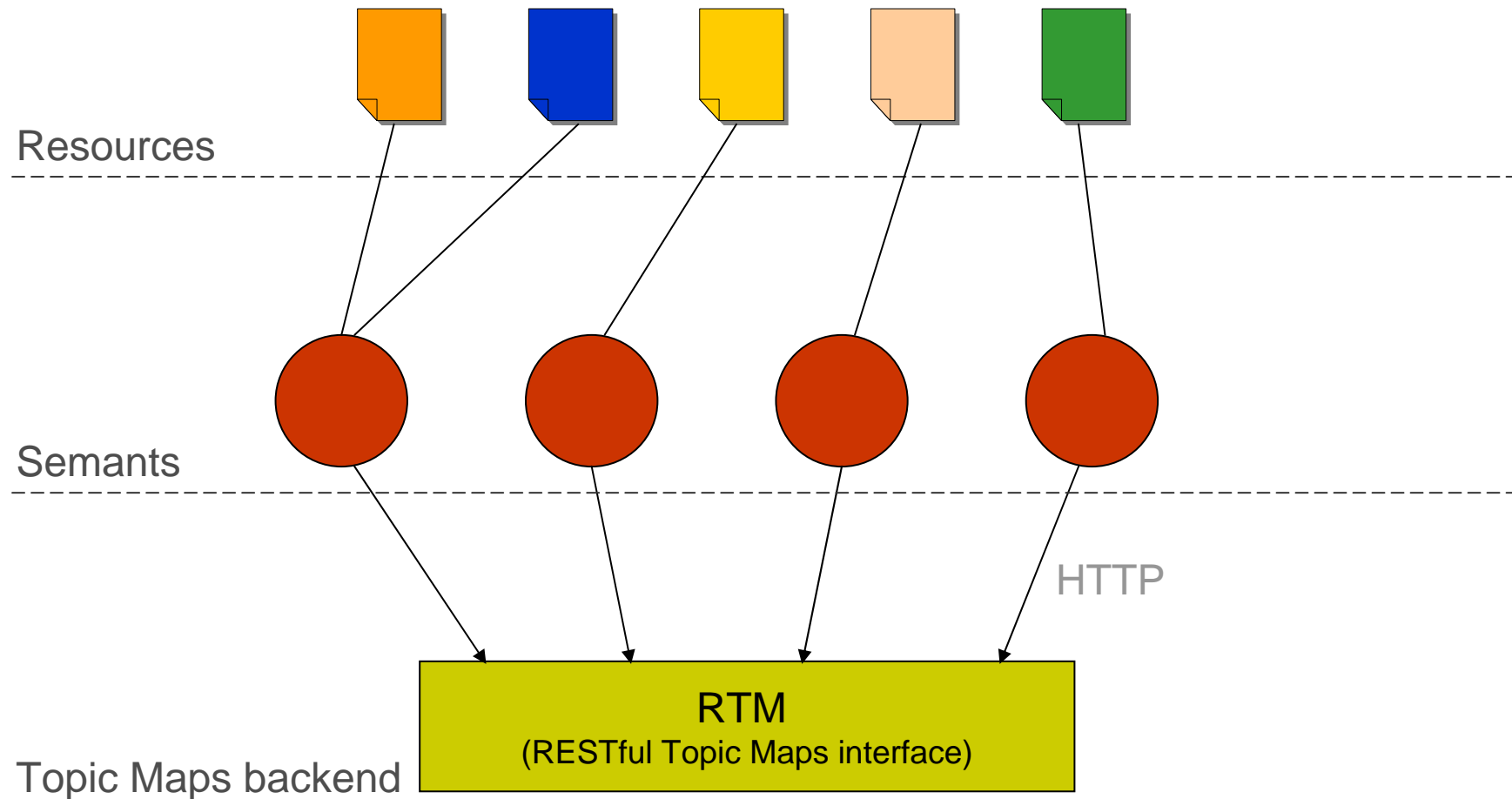
- Wikipedia is full of PSIs
- Conclusion: Let's use Wikipedia as source for subject indicators
- Problems
 - You cannot ask Wikipedia about a particular subject
 - Is http://de.wikipedia.org/wiki/John_Lennon a different subject than http://en.wikipedia.org/wiki/John_Lennon?

Introduction – The Semants



- Semants: Small applications that convert resources into topic maps
- Semants should be programming language independent
- RESTful

Architecture



Wikipedia Semants



- Semants to convert persons and locations
- Links to other Wikipedia sites become subject identifiers and occurrences
- Wikipedia categories are connected via associations

Wikipedia Semants



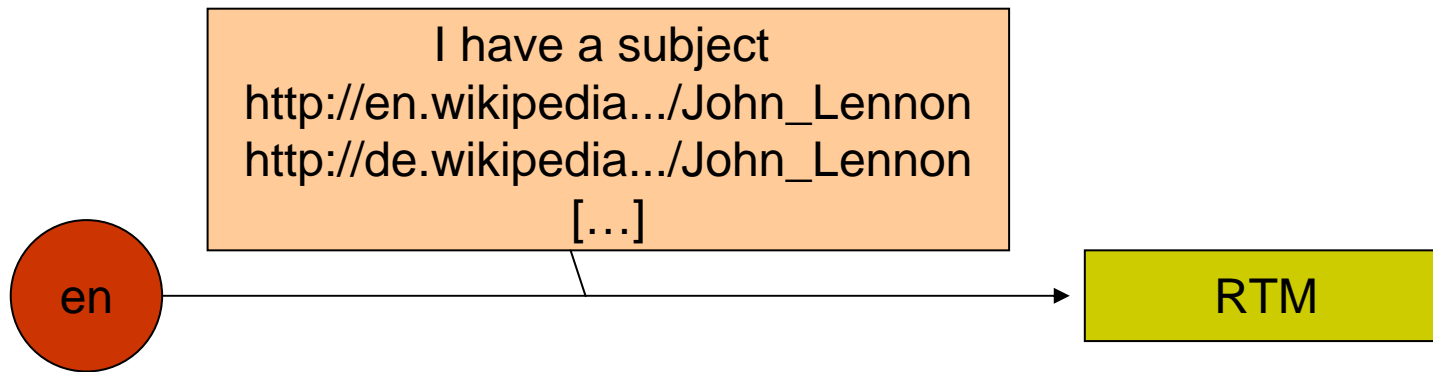
Extract from http://en.wikipedia.org/wiki/John_Lennon

```
{ { Infobox_Biography
|subject_name='''John Lennon'''
[... ]
|date_of_birth=[[October 9]], [[1940]]
|place_of_birth=[[Liverpool]], [[England]]
|date_of_death=[[December 8]], [[1980]]
|place_of_death=[[New York City]], [[USA]] }
```

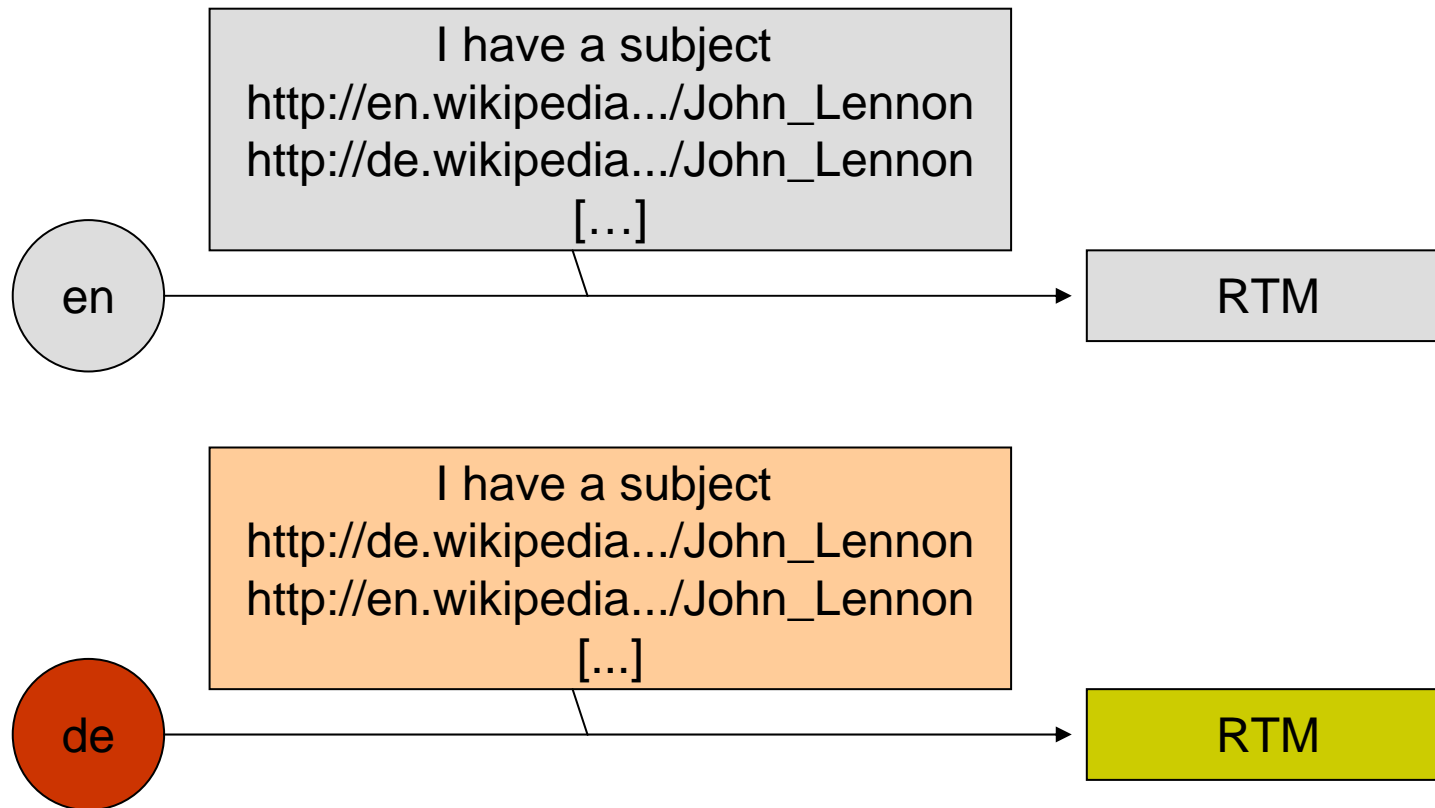
AsTMa=

```
http://en.wikipedia.org/wiki/John\_Lennon isa person
born 1940-10-09 in http://en.wikip.../Liverpool
died 1980-12-08 in http://en.wikip.../New\_York\_City
name: John Lennon
```

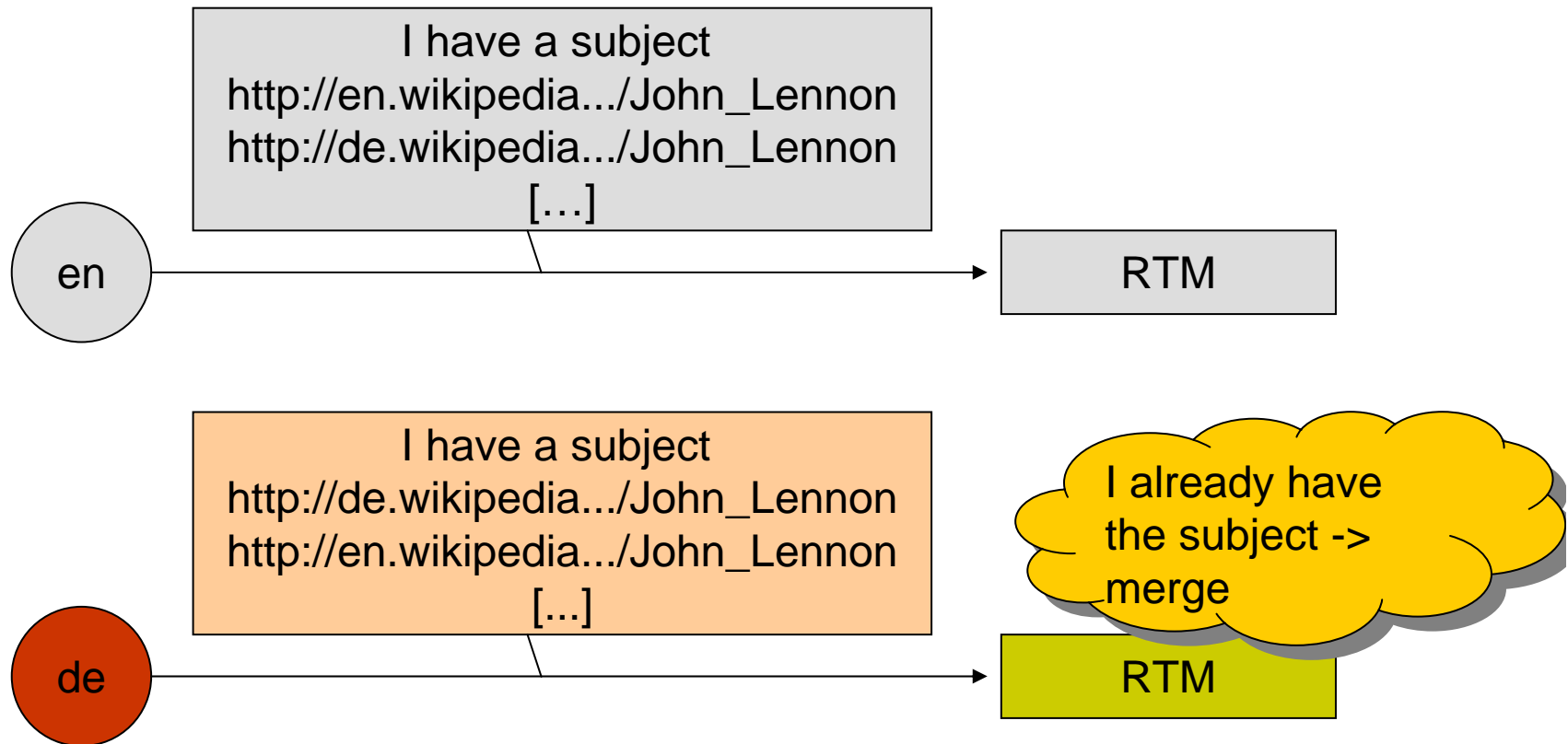
Merging



Merging



Merging



Wikipedia Semants - Issues



- Semants do not know that John Lennon is a musician and was a member of The Beatles
 - Paul McCartney is not connected to Lennon, they are just persons
- Persons are isolated, they are only connected through dates and locations and Wikipedia categories

Wikipedia Semants – Issues II



- Each Wikipedia site uses its own conventions
 - A Semant that works in en.wikipedia does not work in de.wikipedia
- The en.wikipedia uses a lot of different templates (i.e. Infobox_Biography vs. Infobox_Philosopher)

Other Semants



- Internet Movie Database (IMDB)
- E-Mails
- ...
- ... the Internet 😊

Further Work



- Provide a Semants framework
- Decouple Semants from RTM
 - Semants should provide a Topic Maps view on a resource
- User defined ontologies
- Connect other databases to the Wikipedia subjects (Amazon, ...)
- ...

Outlook



- Open Source Topic Maps engine November / December 2006
- Persons and locations topic map December 2006
- Semants framework Q1/2007

- Watch this (currently ugly) space:
<http://www.semagia.com> 😊

Summary



- Wikipedia is a good source for PSIs but it is complicated to extract more information from it
- Converting the whole Internet may take some time 😊

References



- Wikipedia
<http://www.wikipedia.org/>
- R. Fielding *"Architectural Styles and the Design of Network-based Software Architectures"*, PhD., University of California, Irvine, 2002
<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>
- R. Barta, L. Heuer
"AsTMa= 2.0 Language Definition", 2005
<http://astma.it.bond.edu.au/astma=-spec-2.0r1.0.dbk>

Discussion



Questions?

Answers! 😊