



UNIVERSITÀ  
DEGLI STUDI  
DEL MOLISE

**Remo Pareschi**  
**University of Molise**

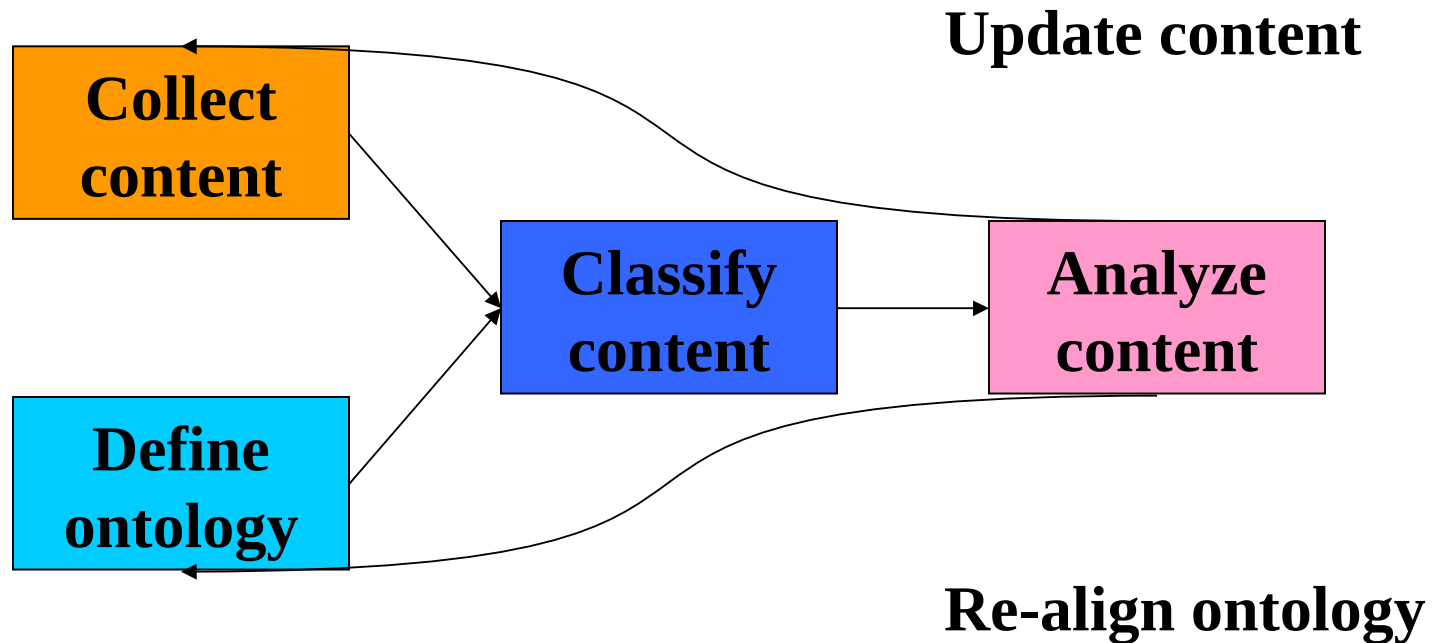
# **An inside-out perspective on lightweight ontologies**

**Joint work with F. Arcelli Fontana, M. Rossetti, F.  
Stella**

# Lightweight Ontologies

Simple conceptual trees organized according to the subsumption relationship  
By and far the type of ontologies of most widespread use for practical purposes  
Current focus of use: content analysis and classification

# The collection-classification-analysis cycle



# The collection-classification- analysis cycle

**PLUS:** well-established technique, effective returns from its application, expertise derived through many years of practice

**MINUS:** heavy maintenance requirements (content upgrading, ontology conceptual re-alignment)

# An orthogonal view of ontologies: approximate-navigate-specialize

View ontologies as approximations of content

Use concepts in ontologies as content attractors

Return content regions of interest as navigation spaces

Specialize ontologies

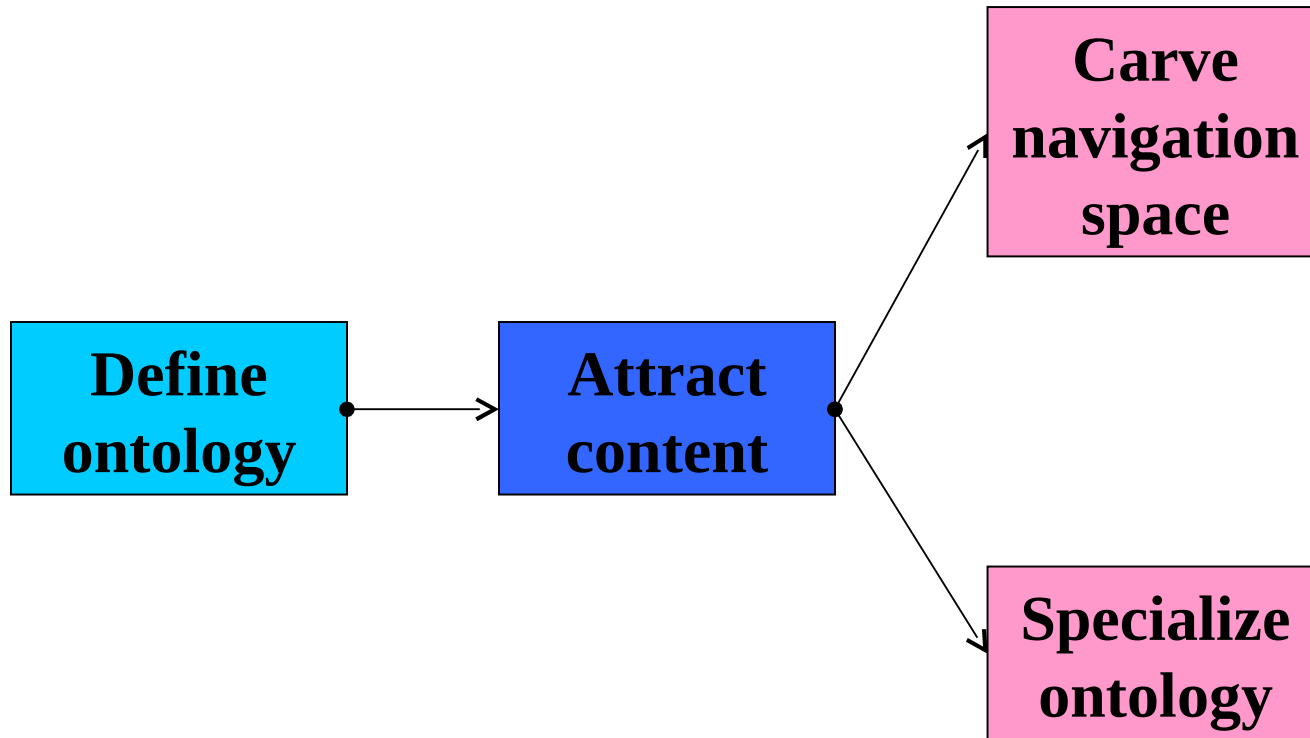
# Underlying ‘philosophy’

**Radical Conceptualism:** everything is a concept

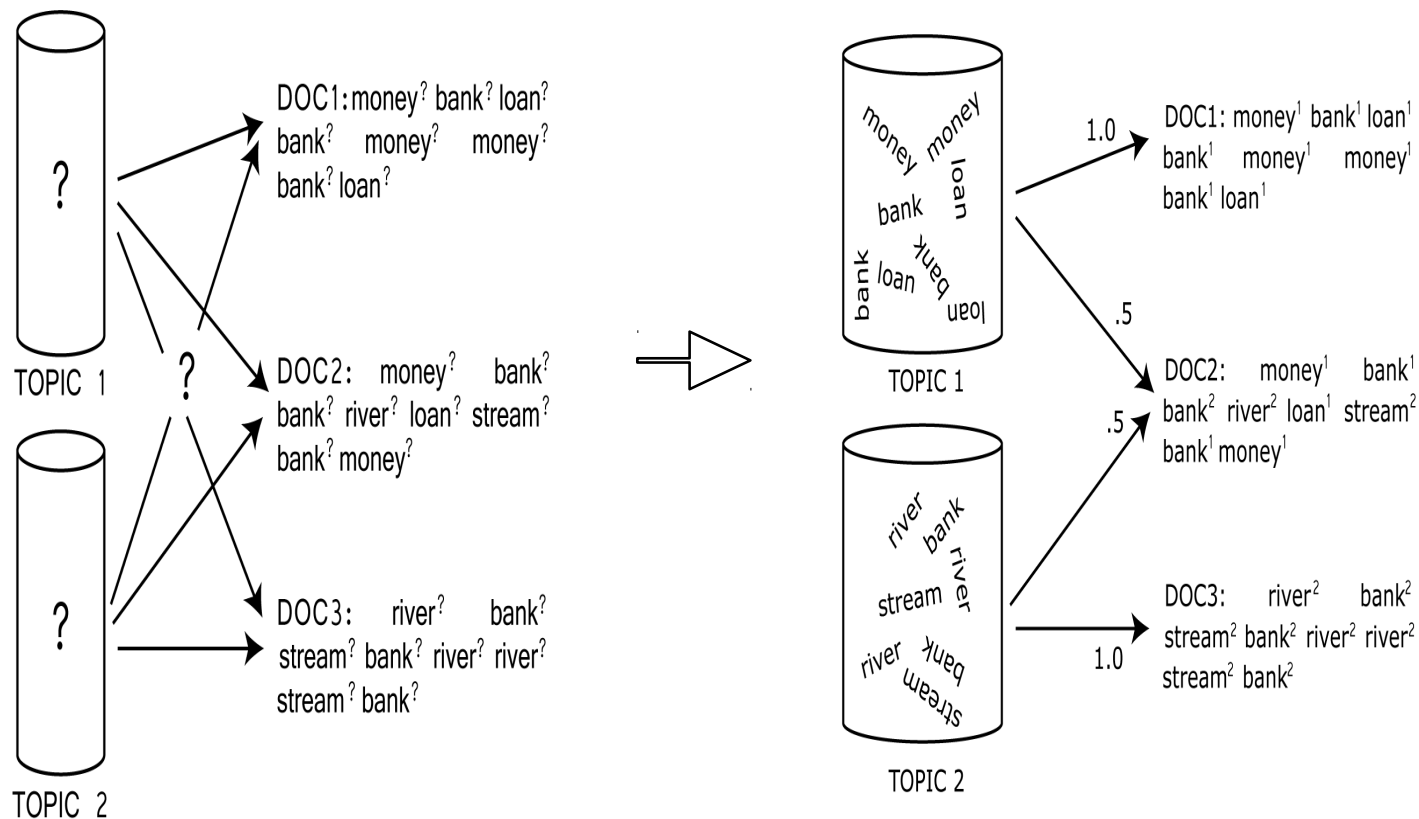
**Equivalent to Radical Lexicalism:** every lexical item or combination of items is potentially a concept

In principle, no distinction between **Language** (content) and **Meta-language** (Tags, Labels, Concepts)

# The approximation-navigation-specialization process



# Basic Methodology – Topic extraction





# Approach of the experiment (described in Master thesis by Marco Rossetti, University of Milano Bicocca)

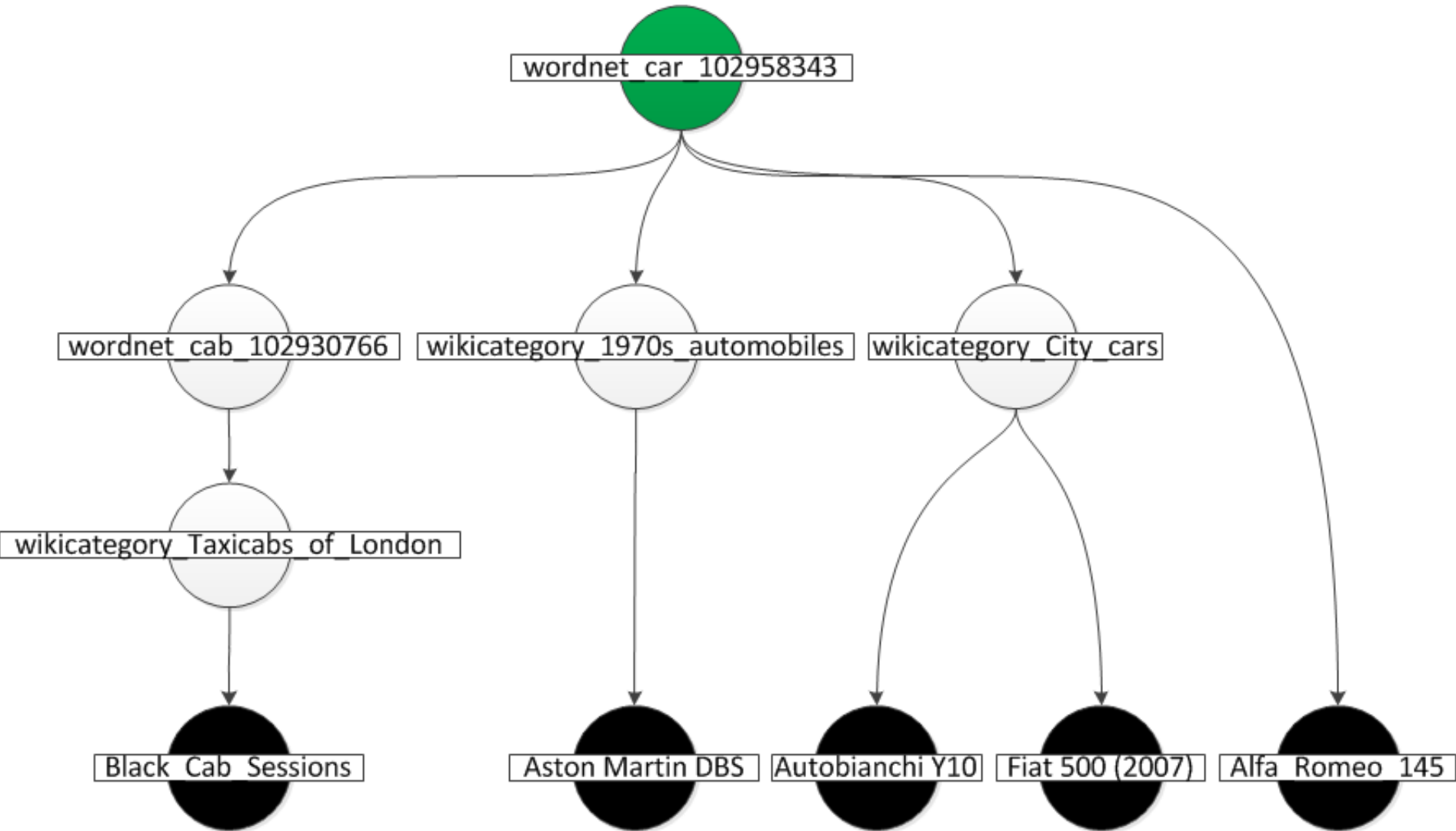
- *Basis for ontology bootstrap: YAGO2*

*"YAGO2 is a huge semantic knowledge base, derived from Wikipedia, WordNet and GeoNames. Currently, YAGO2 has knowledge of more than 10 million entities (like persons, organizations, cities, etc.) and contains more than 80 million facts about these entities."*

*(YAGO2 web site, 2011)*

- By navigating the links of the Wikipedia pages related to the start-up ontology, pages are clustered according to degrees of semantic and structural similarity.

# Bootstrapping of core ontology through YAGO2



# Acquisition of pages

## External links

- [Official Aston Martin Site](#)
- [Aston Martin Flash Site](#)
- [Official Aston Martin Used Car Locator](#)
- [Newport Pagnell 6 Cylinder Cars](#)
- [Top Gear.com article: Now pay attention 007](#)
- [Original Aston Martin brochures](#) - Original sales literature for Aston Martin viewable online, including the DBS

## Aston Martin DBS

From Wikipedia, the free encyclopedia  
(Redirected from [Aston martin dbs](#))

*This article is about a 1967-72 car.*

The **Aston Martin DBS** is a *GT car* produced by the British manufacturer [Aston Martin Lagonda Limited](#). Originally produced from 1967–72, it featured in the 1969 *James Bond* film *On Her Majesty's Secret Service*. A new version, based heavily on the [Aston Martin DB9](#), is featured in the 2006 film *Casino Royale* and the 2008 film *Quantum of Solace* (see [Aston Martin DBS \(2007\)](#)).



Aston Martin DBS

GT car

Aston Martin DB9

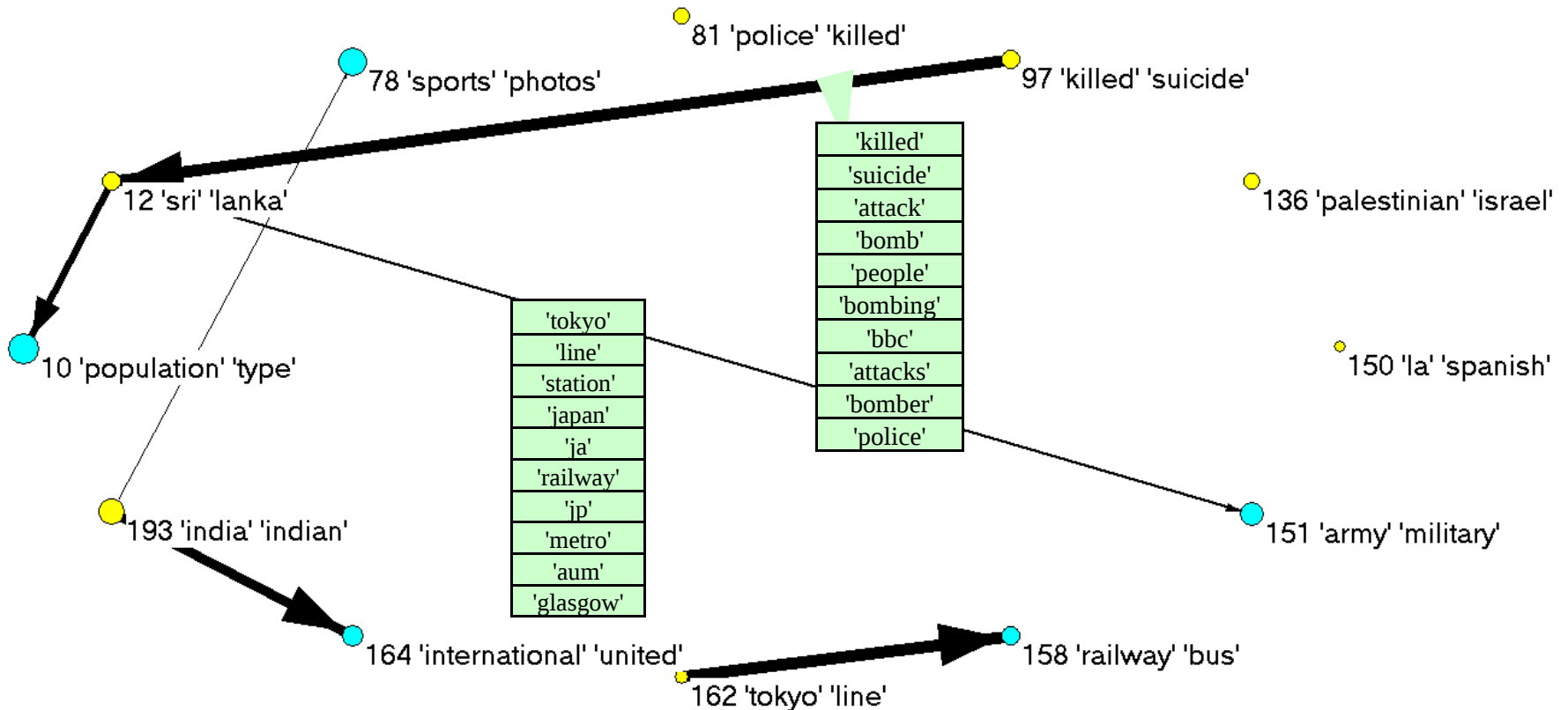
Casino Royale

<http://www.astonmartin.com/>

[http://www.topgear.com/...](http://www.topgear.com/)

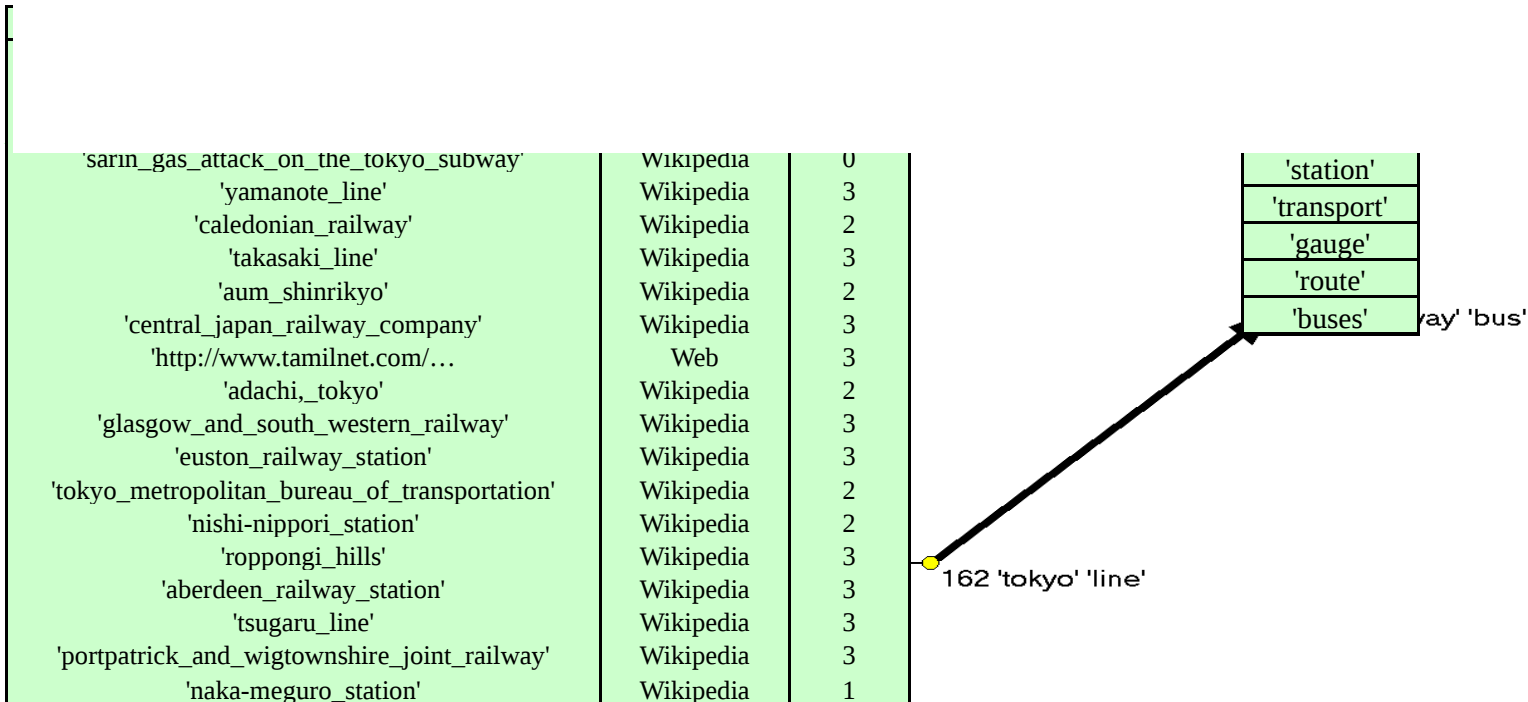
# Example of navigation space [1]

Carve navigation space starting from core ontology  
Terrorism



# Topic navigation [2]

$\alpha = 0.3$



# Topic navigation [3]

'india'
'mumbai'
'delhi'
'indian'
'bengal'
'chennai'
'web'
'kolkata'
'bombay'
'state'

80 'india' 'mumbai'

162 'tokyo' 'line'

41 'australia' 'australian'

158 'railway' 'bus'

165 'company' 'vehicle'

'australia'
'australian'
'au'
'abc'
'sydney'
'victoria'
'melbourne'
'queensland'
'gold'
'coast'

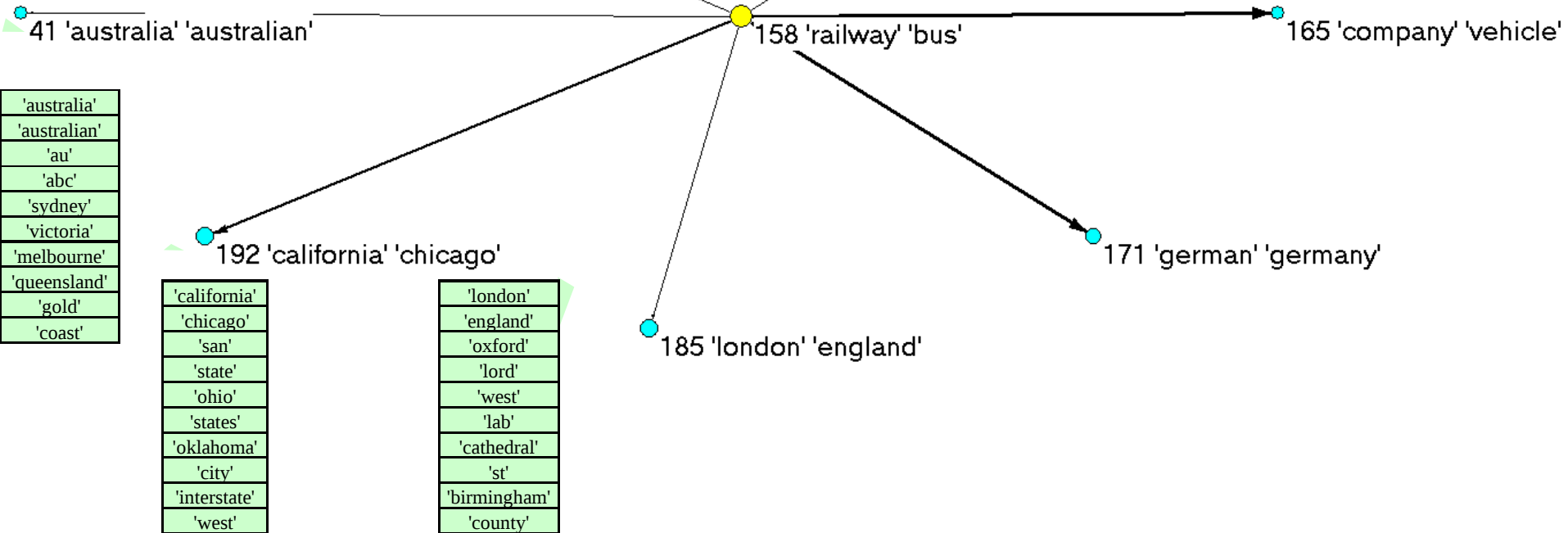
192 'california' 'chicago'

171 'german' 'germany'

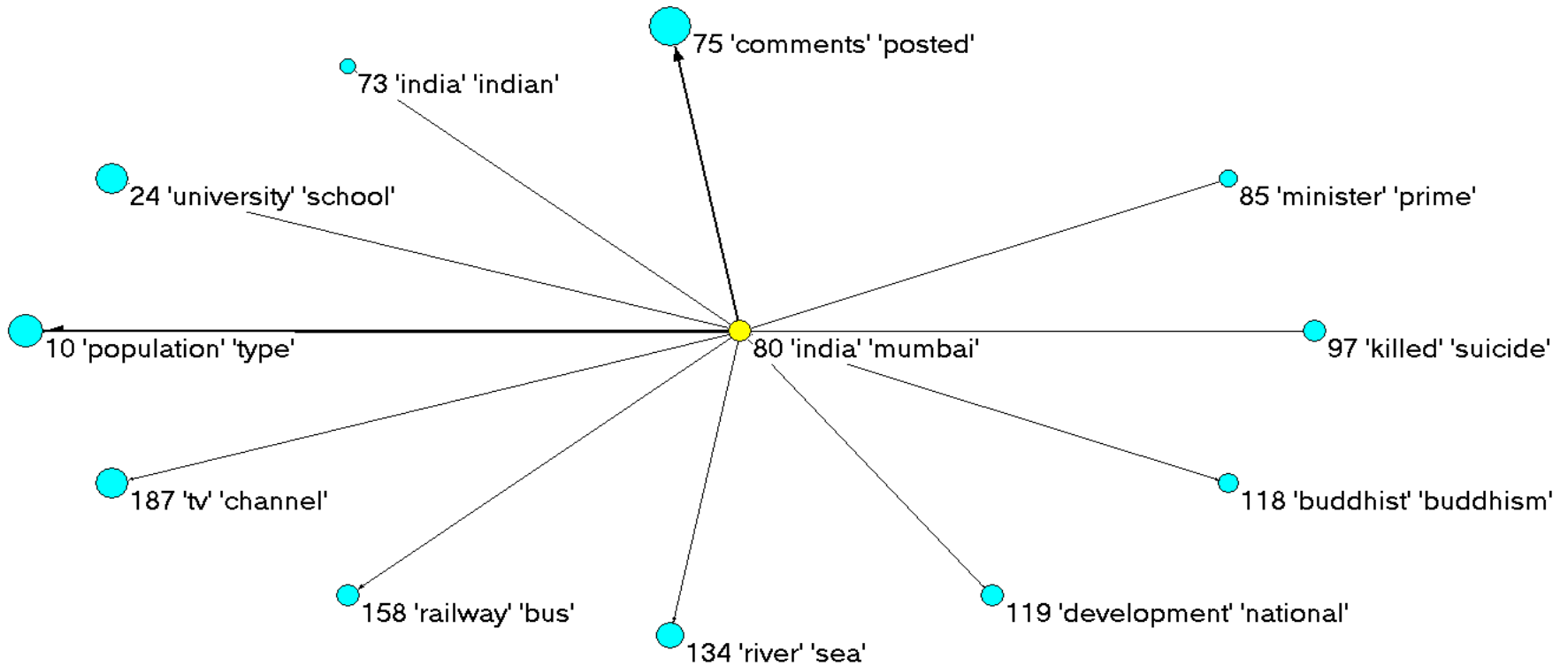
'california'
'chicago'
'san'
'state'
'ohio'
'states'
'oklahoma'
'city'
'interstate'
'west'

'london'
'england'
'oxford'
'lord'
'west'
'lab'
'cathedral'
'st'
'birmingham'
'county'

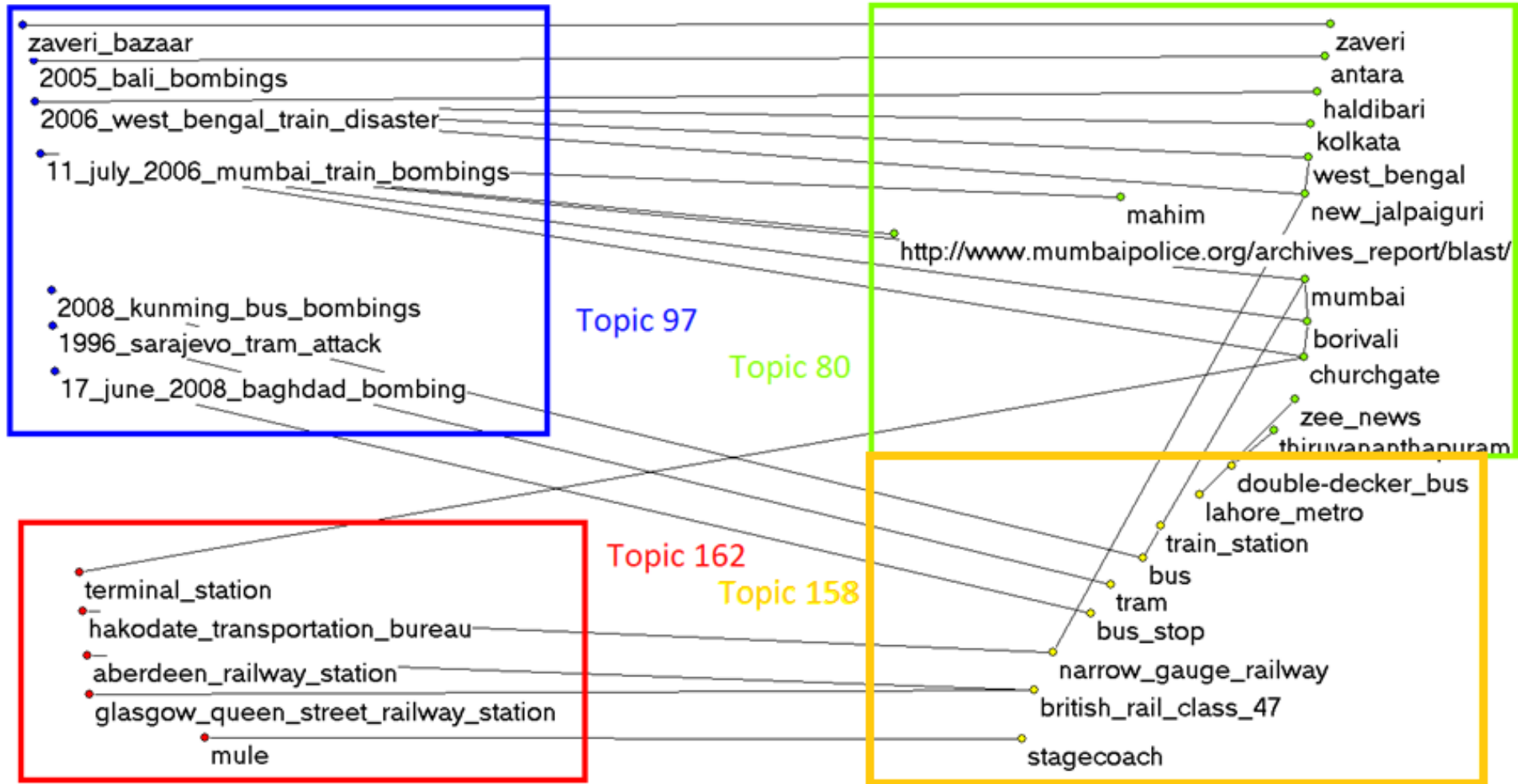
185 'london' 'england'



# Topic navigation [4]



# Topic navigation [5]

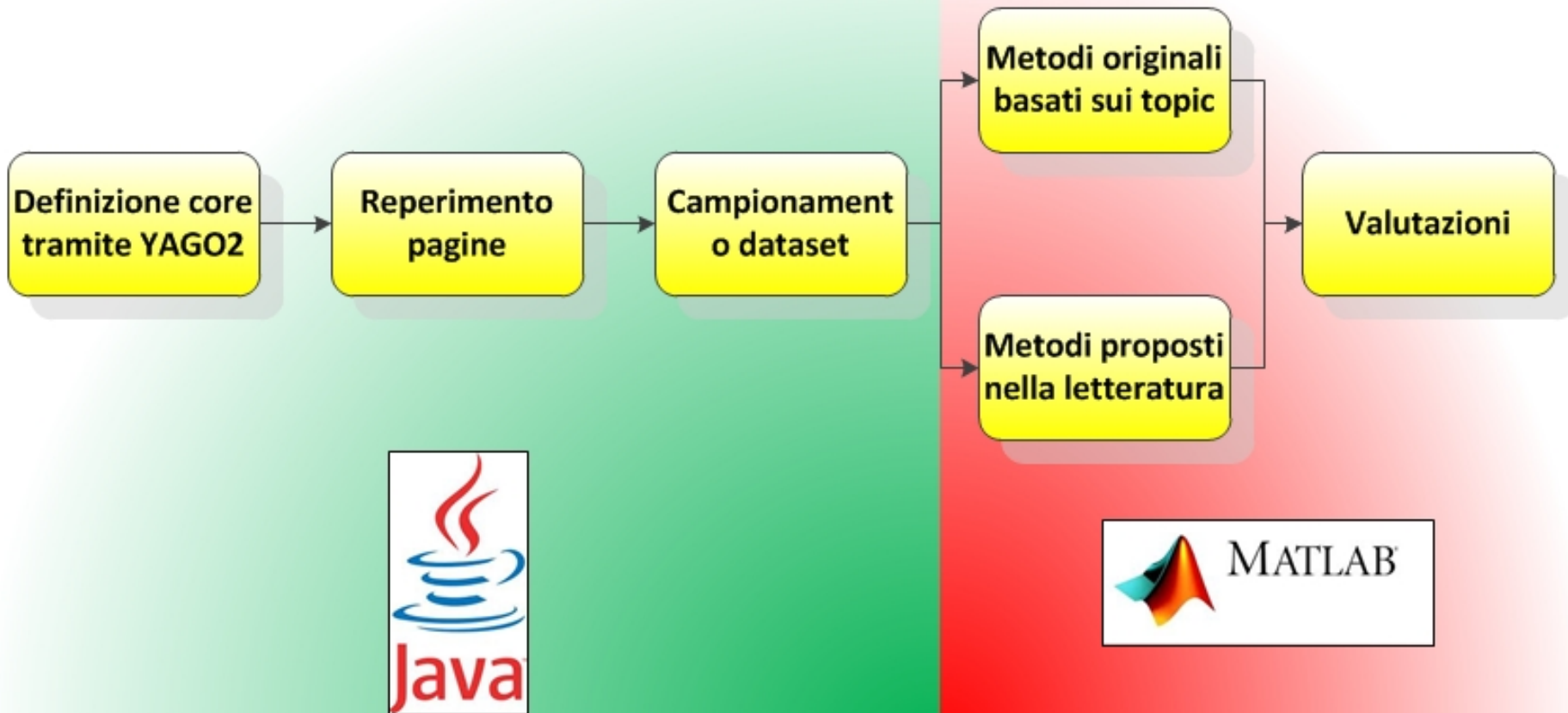




# Softyware environm,ent

Java and Matlab have provided the basic tools

MALLET, Crawler4j, MWDumper, OWLIM, Sesame, SAX, H2 Database Engine, Language-detection, ...



# Future developments

- Design and implementation of an advanced user environment letting users navigate the graph and extract useful information
- Experiments in ontology cross-breeding, eg cross fashion and consumer electronic ontologies to carve common navigations spaces (connection with LL multiplicative fragment?)
- Develop an environment for ontology specialization and verticalization (from general ontologies to vertical ontologies)