

Motivations behind software integration for NLP

- For high level projects involving...
 - ...many existing linguistic resources and linguistic software (*lingware*), such as lexical databases, tokenisers, syntactic parsers, MT systems, etc.
 - ...backend software (databases, web servers, etc.)
 - ...end-user software (offline or online GUI)
- Many of them conceived by different teams, following heterogeneous approaches, for various purposes...
- What & how do we integrate this at LIG-GETALP team ?
 - 2 examples : OMNIA project & AXiMAG project

Outline

- The KISS approach and REST implementation
- OMNIA project (terminated)
- AXiMAG project (reimplementation in progress)

The KISS principle

- Stands for : *Keep It Simple, Stupid*
 - Acronym created by war plane manufacturers
 - A general design principle, applied to IT

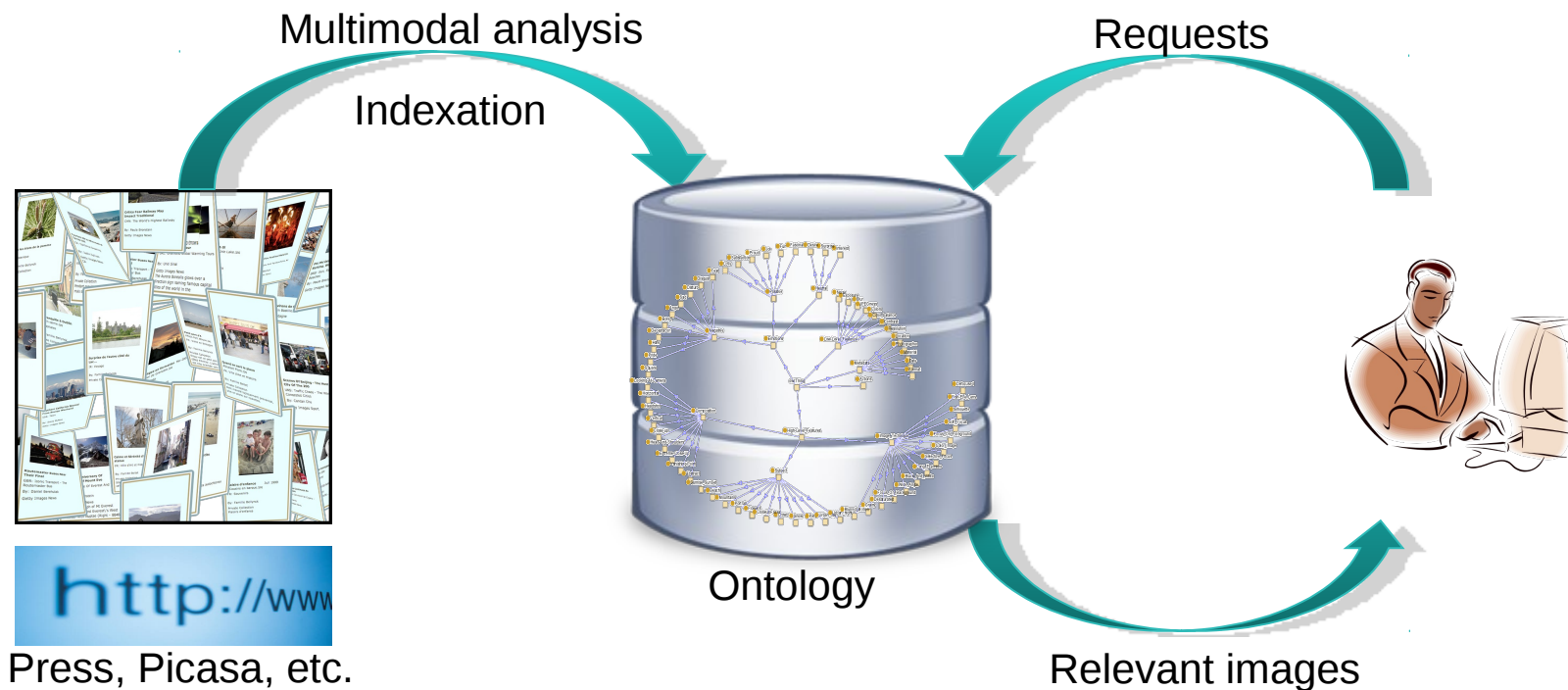
KISS for integration : REST architecture

- Stands for : *REpresentational State Transfert*
- Integration of services, through HTTP protocol
 - Client-server (independent of any GUI)
 - Stateless (services have nothing to remember)
 - Cachable (for better scalability)
 - Layered (redirections permitted without restrictions)
- Service call via :
 - Any web browser
 - Any programming language supporting HTTP requests
 - Command line with popular utilities like cURL

Lingware integration in the OMNIA project

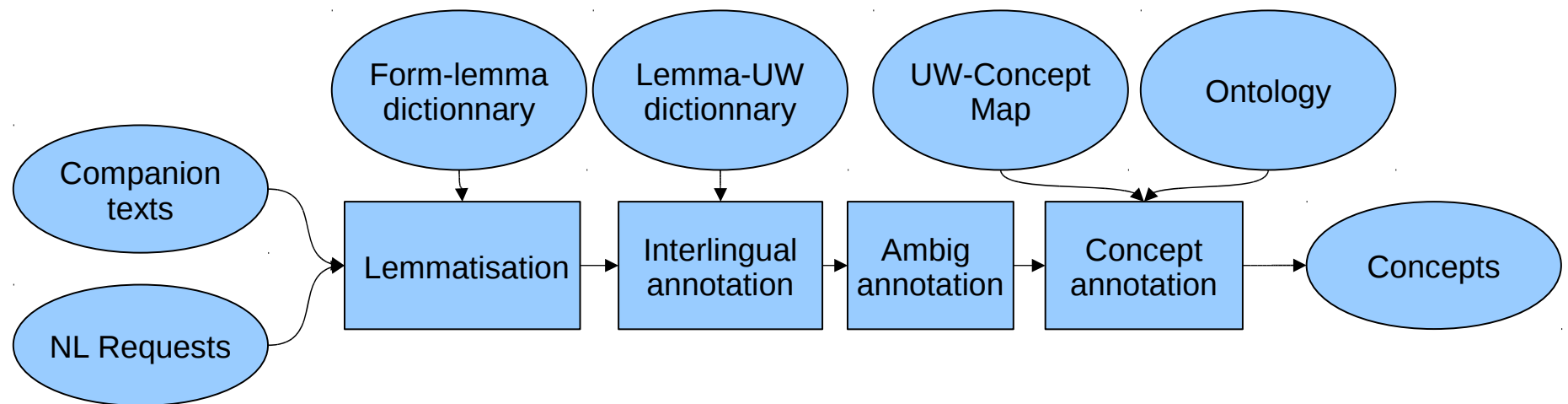
OMNIA project (2008-2010) : multilingual information retrieval

- From *companion texts* (eg picture captions) in several languages, to interlingual concepts
- Ontology-based indexing



OMNIA modular architecture : from text to concepts

4 modules & 4 resources



Example companion text

AWA05 - 20020924 - BAGHDAD, IRAQ : Iraqi women sit under a portrait of Iraqi President Saddam Hussein in a waiting room in Baghdad's al-Mansur hospital 24 September 2002. Saddam Hussein is doggedly pursuing the development of weapons of mass destruction and will do his best to hide them from UN inspectors, the British government claimed in a 55-page dossier made public just hours before a special House of Commons debate on Iraq. Iraqi Culture Minister Hamad Yussef Hammadi called the British allegations "baseless."
EPA PHOTO AFPI AWAD AWAD

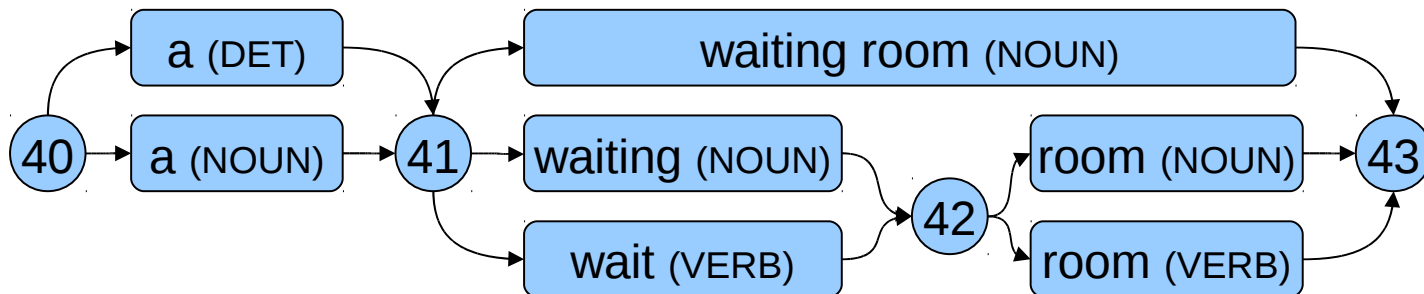


Lemmatisation process

Iraqi women sitting in **a waiting room** of Bagdad hospital

(Iraqi, NOUN) (woman, NOUN) (sit, VERB) (in, PREP) (**a, DET**) (**waiting room, NOUN**)
(of, PREP) (Bagdad, NOUN) (hospital, NOUN)

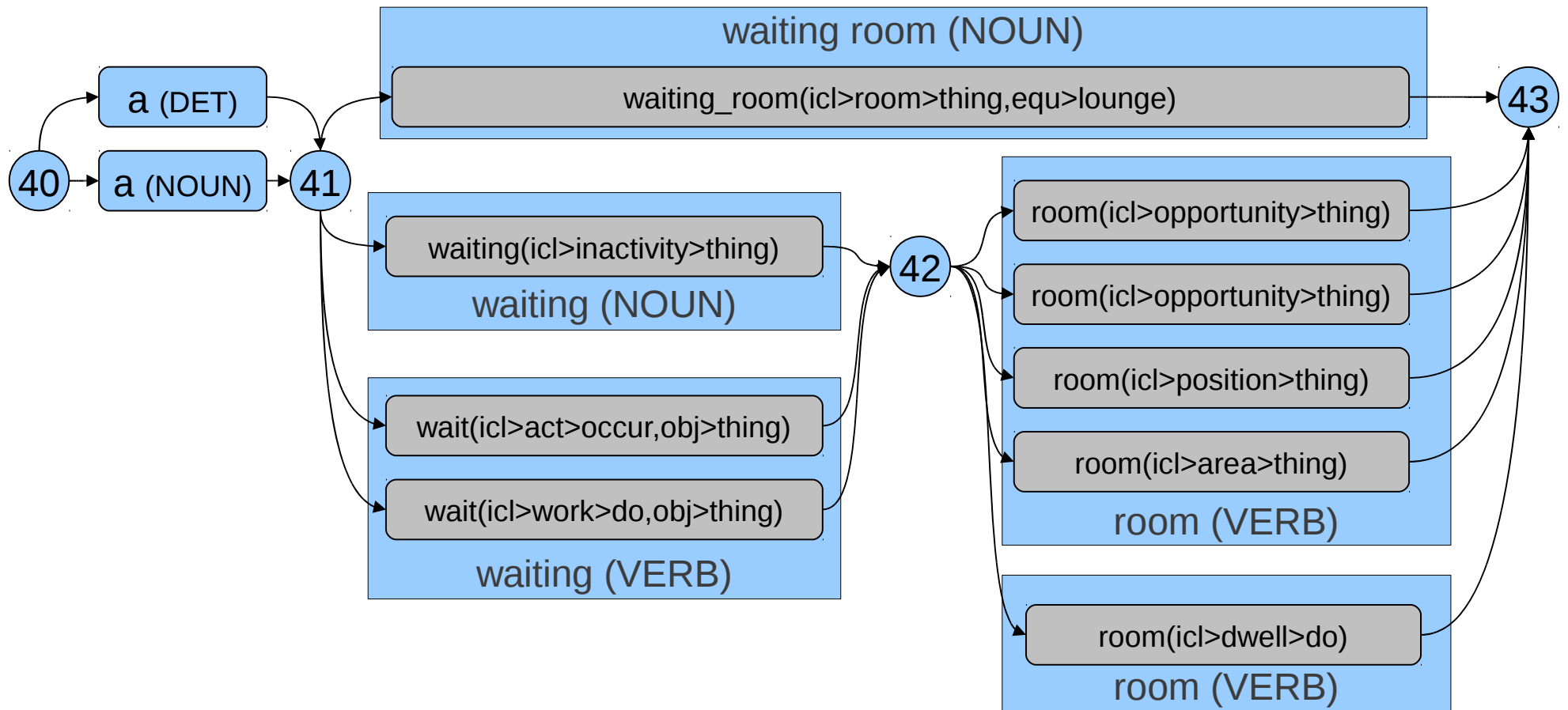
Graph structure with ambiguities :



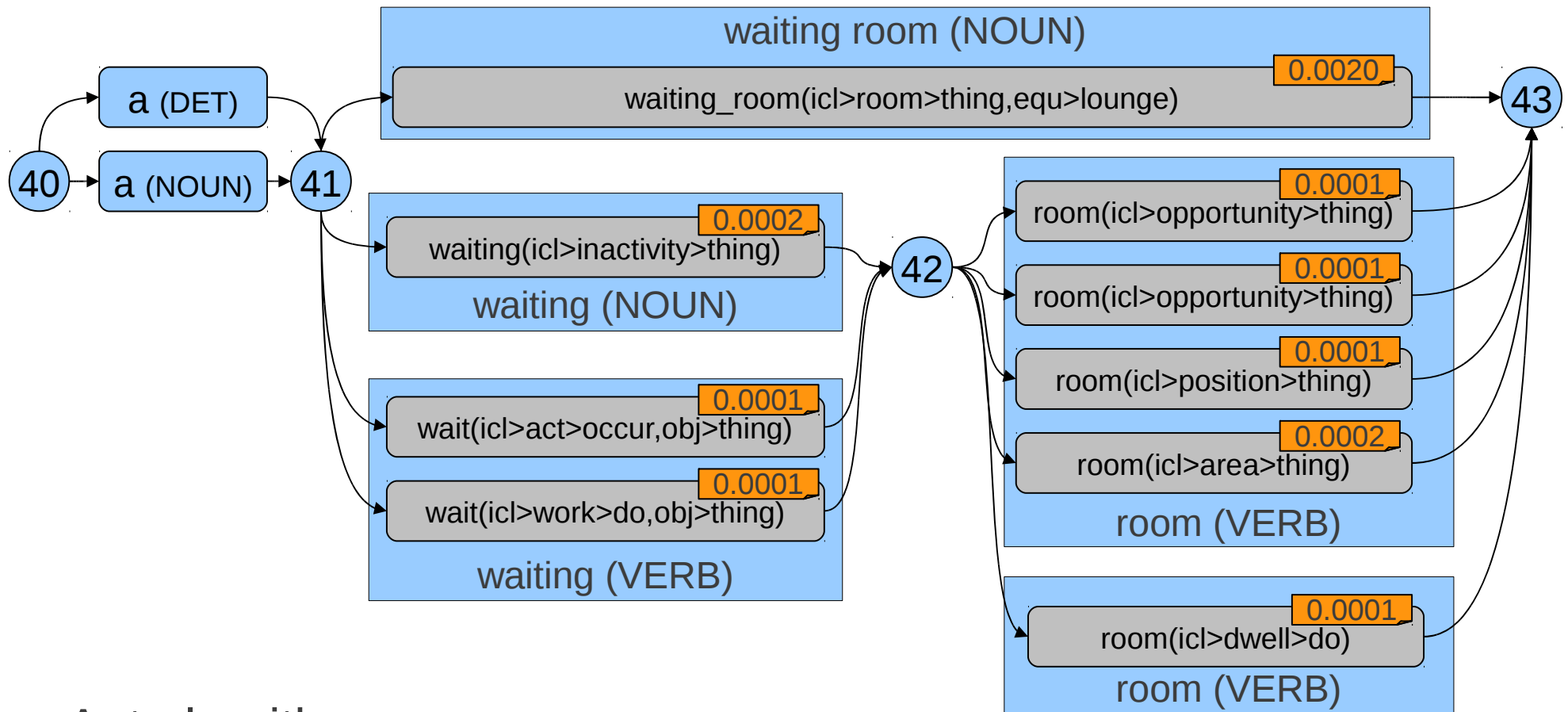
Interlingual lexicon : Universal Words (UW)

- Universal Words (UW)
 - represent acceptions without ambiguities
 - headword with meaning restrictions
 - Examples :
 - book(icl>thing)
 - book(icl>do, agt>human, obj>thing)
 - ikebana(icl>flower arrangement)
 - 200.000 UW++ built from WordNet Synsets

Interlingual annotation process



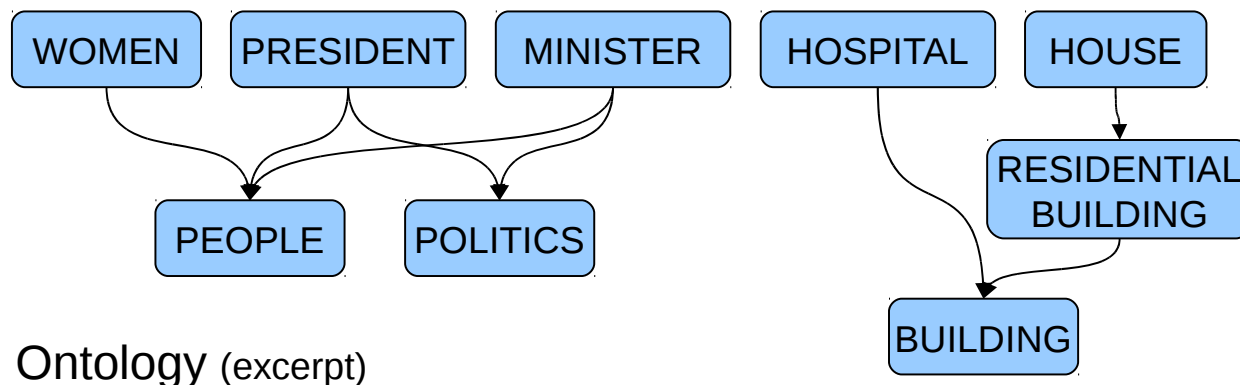
Automatic disambiguation process



Ant algorithm
(Schwab & Lafourcade 2007)

Ontology for conceptual annotation

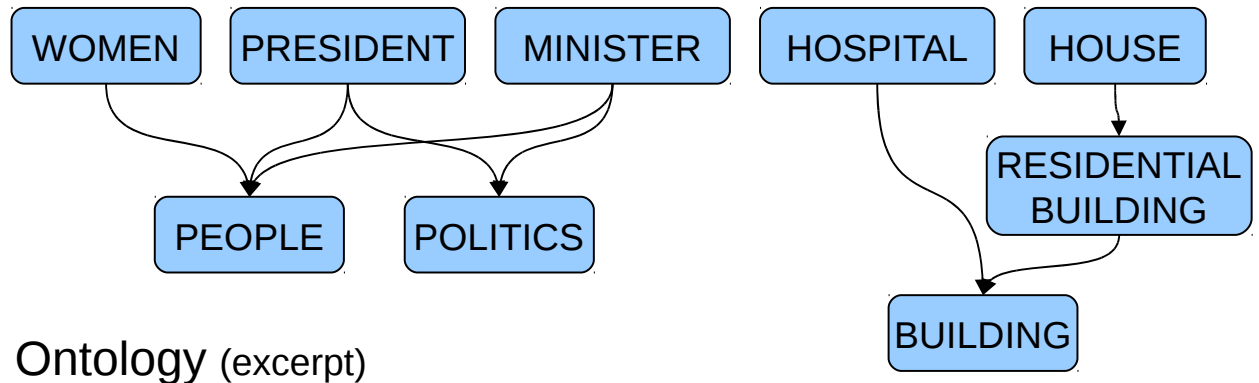
- Ontology
 - Domain dependant
 - Concepts hierarchy
 - Interlingual
- Alignment with interlingual lexicon
 - Manual or automatic (Rouquet & Nguyen, 2009)



Extracted concepts

- Output for whole text

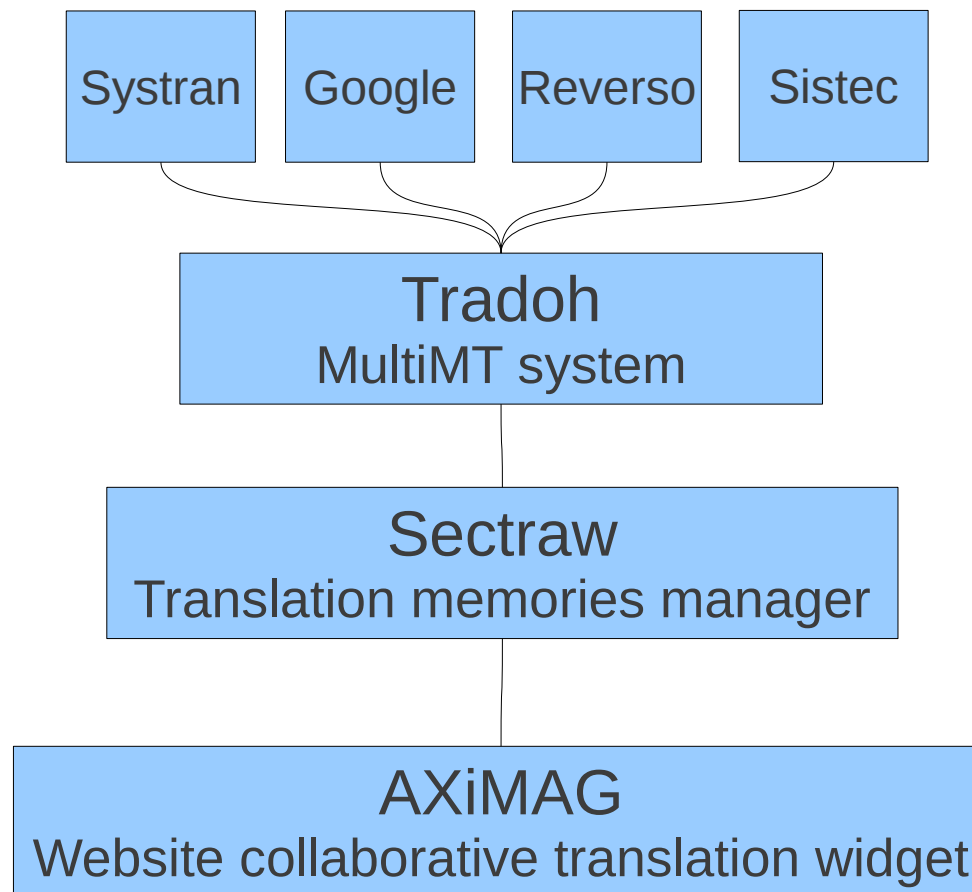
Concept	Score
RESIDENTIAL BUILDING	0.0002
PRESIDENT	0.0004
BUILDING	0.0004
HOUSE	0.0002
HOSPITAL	0.0002
POLITICS	0.0044
MINISTER	0.0040
WOMAN	0.0002
PEOPLE	0.0146



OMNIA demo

Lingware & software integration in the AXiMAG project

AXiMAG version 2 project (2011) : website collaborative translation



AXiMAG demo

Design issue in service-oriented architecture

- Request building & request processing
 - May lead to DoS
- Means to avoid this issue
 - Scope of the service
 - Internal cache