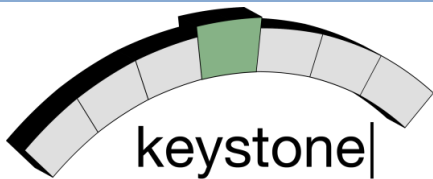
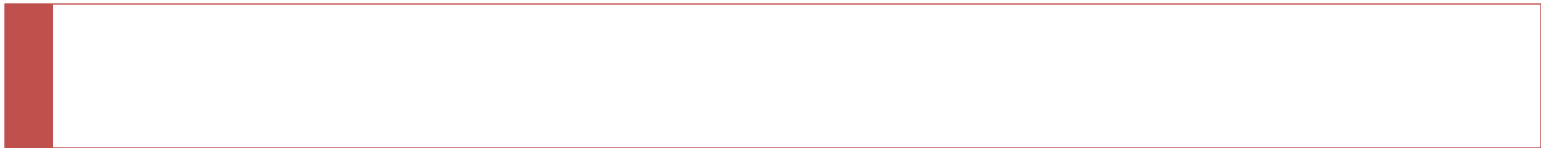


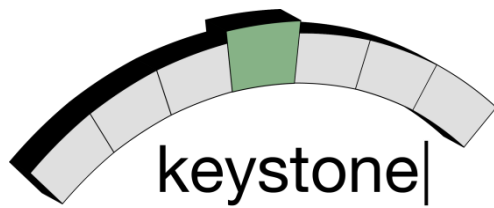
22-23 February 2016
Winter WG Meeting
Marseille (FR)



keystone|
IC1302

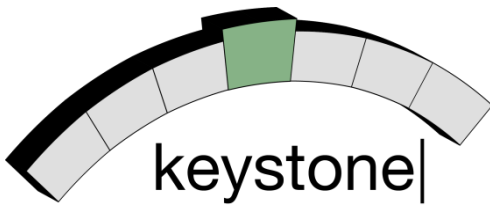
Theoretic Hackathon





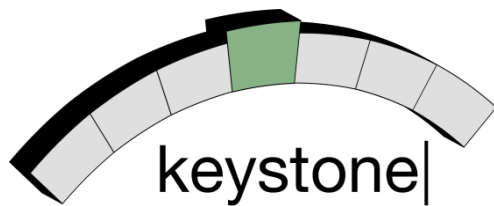
Why a hackathon

- ▶ In previous meetings, we experimented brainstorming sessions aiming at discussing challenges, solutions, ideas about keyword search in structured data.
- ▶ Why a hackathon now?
 - ▶ We conceive it as an advanced brainstorming more oriented to a practical result
 - ▶ Papers (for IKC 2016?)
 - ▶ Projects (STSMs?, a H2020 proposals?)
 - ▶ The proposals will be analyzed in a plenary session
- ▶ Agenda
 - ▶ The problem (Mauro)
 - ▶ The rules of the hackathon (Francesco)



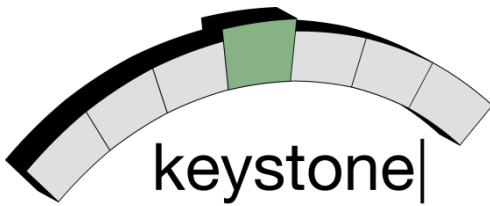
The problem

- ▶ Topic: Keyword search over a multi-dimensional representation of documents
- ▶ Documents and queries structure:
 - ▶ Textual layer: natural language text
 - ▶ Metadata layers
 - ▶ Entity Linking
 - ▶ Predicates
 - ▶ Roles
 - ▶ Timing Information
 - ▶ ...
- ▶ Problems:
 - ▶ How to compute the score for each layer?
 - ▶ How to aggregate such scores?
 - ▶ How to weight each layer?
 - ▶ Which further layers could be used?



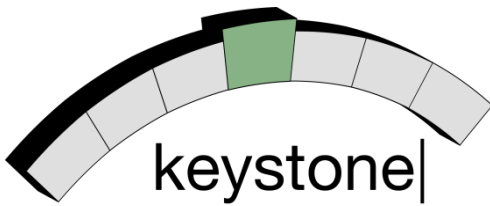
The dataset and the queryset

- ▶ Small dataset containing 331 documents and 35 queries.
 - ▶ Documents come from blog posts
- ▶ Natural language content is enriched with four metadata/semantic layers
 - ▶ URI Layer: links with entities detected into the text and mapped to DBpedia entities
 - ▶ TYPE Layer: conceptual classification of the named entities detected into the text and mapped with both DBpedia and Yago knowledge bases
 - ▶ TIME Layer: metadata related to the temporal mentions found into the text by using a temporal expression recognizer (ex. “the eighteenth century”, “2015-18-12”, etc.)
 - ▶ FRAME Layer: output of the application of semantic role labeling techniques. Generally, this output includes predicates and their arguments describing a specific role in the context of the predicate.
Example:
“He has been influenced by Carl Gauss” →
[framebase:Subjective_influence; dbpedia:Carl_Friedrich_Gauss]



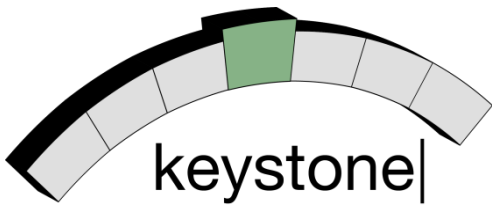
An example

- ▶ Text: “astronomers influenced by Gauss”
- ▶ Layers
 - ▶ URI Layer: “dbpedia:Carl_Friedrich_Gauss”
 - ▶ TYPE Layer: “yago:GermanMathematicians”, “yago:NumberTheorists”, “yago:FellowsOfTheRoyalSociety”
 - ▶ TIME Layer: “day:1777-04-30”, “day:1855-02-23”, “century:1700”
 - ▶ FRAME Layer: “Subjective_influence.v_Carl_Friedrich_Gauss”



The Goal

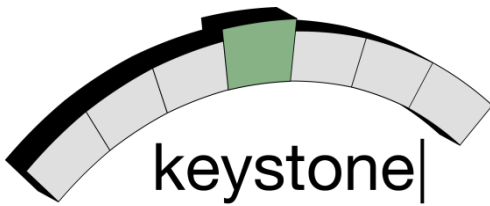
- ▶ The final goal of the theoretic hackathon is to develop a **functional architecture** for a system addressing the problem in hand.
 - ▶ Each component of the system has to be described in terms of
 - ▶ Input
 - ▶ Output
 - ▶ Process / Algorithm implemented
 - ▶ The system has to be described in terms of
 - ▶ Input
 - ▶ Output
 - ▶ External resources needed (if any)
 - ▶ Experimental evaluation (datasets, measures, ...)
- ▶ Format
 - ▶ Presentation (powerpoint or similar)



The application

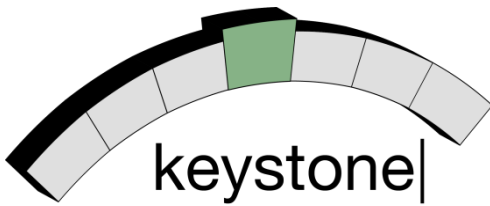
- ▶ Functional Architecture (picture) and description of the main components
- ▶ A table mapping team members' expertise in the aspects related to the solution to the query

Name	Step 1: access to the data sources	Step 2: solution to the query	Step 3: evaluation



The application (2)

- ▶ Additional questions (to be answered) enabling the discussion
 - ▶ Question 1: Datasets
 - ▶ Are there **Open Data sets** (e.g., EU open-data.europa.eu/) that can queried by means of the developed keyword-based search methods?
 - ▶ What are the problems associated with **accessing data sets** (e.g., copyrights, formats)?
 - ▶ Question 2: Query Methods
 - ▶ How can language **understanding technologies** and **keyword-based query** methods help **non-expert users**?
 - ▶ **Which** Semantic Web technology can support this task?
 - ▶ Question 3: Evaluation
 - ▶ How can technologies and keyword-based query methods working on the Semantic Web be **evaluated** using open data sets?
 - ▶ How can their **scalability, accuracy, and feasibility** be established?
 - ▶ What kind of **benchmarking environments** can be created to compare the proposed with different keyword-based query methods?
 - ▶ Which types of **convincing competitions** can be organized to foster and evaluate keyword-based query methods?



Organization

- ▶ The participants will be divided in 4-5 groups that will work on the development of a project.
 - ▶ Each group will be coordinated by a member who will create the presentation of the application
 - ▶ We have to appoint the coordinators
 - ▶ We have to decide how to create groups
 - ▶ The works will be presented in the final plenary session
- ▶ Timing
 - ▶ 23/2 from 9.45 to 11.00 1h
 - ▶ 23/2 from 11.20-13.00 1.40 h
 - ▶ 23/2 from 14.45 – 16.00 1.15 h
 - ▶ 23/2 from 16.45 – 17.30 Plenary Session