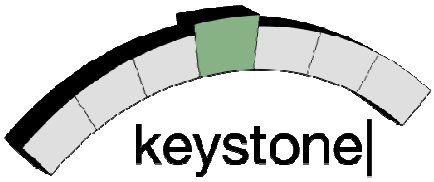

COST Action IC1302
Semantic keyword-based search on structured data sources (KEYSTONE)

Minutes of the PROFILES KEYSTONE Meeting (WG1/2)

Aldemar Royal Mare Resort, Hersonissos, Crete, Greece

25 May 2014

Organizers
Elena Demidova, L3S Research Center, Hannover, Germany
Julian Szymanski, Gdańsk University of Technology Poland
Stefan Dietze, L3S Research Center, Hannover, Germany
Raquel Trillo-Lado, University of Zaragoza, Spain

Meeting participants
Konstantin Todorov, LIRMM, France
Thanassis Tiropanis, University of Southampton, United Kingdom
Nattiya Kanhabua, L3S Research Center, Hannover, Germany
Khali Belhajiame, Université Paris-Dauphine, France
Karol Draszawka, Gdańsk University of Technology, Poland
Bernardo Pereira Nunes, PUC-Rio, Brazil
Dragan Ivanović, University of Novi Sad, Serbia
Maria Del Carmen Rodriguez Hernandez, University of Zaragoza, Spain
María Teresa Lozano Albalate, Centro Universitario de la Defensa Universidad de Zaragoza, Spain
Dragana Dudić, University of Belgrade, Serbia
Sergej Zer, L3S Research Center, Hannover, Germany
Besnik Fetahu, L3S Research Center, Hannover, Germany

1. Reception of the participants and coffee

The meeting took place in the Aldemar Royal Mare Resort, Hersonissos (Crete). The meeting started at 11:00, on 25 May, 2014.

Dr. Elena Demidova, Chair of the Working Group 2 of the Action, welcome the participants and presented the agenda of the meeting. The agenda of the meeting was adopted (http://www.keystone-cost.eu/keystone/outreach/meetings/profiles-meeting/).

2. Keynote talk " Web Observatory Architectures" by Dr. Thanassis Tiropanis

Session chair Dr. Elena Demidova presented Dr. Thanassis Tiropanis from University of Southampton (UK) who talked about Web Observatories: their goals, teams and architecture. The talk started with the presentation the evolution of the web and the web science. After that Dr. Thanassis Tiropanis presented some examples of how the Web have changed the way people socialize, as currently people want web data not only to get information (to download the data) but also to interact with that data and build applications. Then, web observatories and their design principles were presented. Finally the main concepts of the infrastructure of web observatories were explained: architecture, technologies, etc.

The abstract of the talk: Web Observatories are conceptualised as global distributed resources that can engage communities with analytics and make the most of existing and emergent datasets including those of the linked data cloud, social media, online archives and media archives. There is a plethora of Web Observatories engaging communities in longitudinal studies over different datasets. However, as the number and size of datasets and analytic tools are increasing, scalable Web Observatory architectures and standardisation will be critical. This talk will discuss those requirements and architectural responses.

The slides of this presentation are available from the KEYSTONE webpage at: http://www.keystone-cost.eu/keystone/wp-content/uploads/2014/04/keystone-tiropanis.pdf

3. Introduction of the participants

Dr. Elena Demidova and Dr. Julian Szymanski proposed to the participants the following challenge: presented themselves by using 2 or 3 keywords.

- **Elena Demidova**: Semantic Search, LOD.
- **Konstantin Todorov**: Multilingual, Ontology Matching, Imprecision definition, Data Quality.
- **Nattiya Kanhabua**: Temporal Information Retrieval.
- **Khali Belhajiame**: DataSpaces.
- **Julian Szymanski**: Wikipedia, Visualization and Representation of Data and Information, Information Retrieval.
- **Karol Draszawka**: Computer Science, Matching Learning, How people interact with the data and Web.
- **Raquel Trillo-Lado**: Semantic Search, generation of formal queries from keyword queries, Ontology Matching.
- **María del Carmen Rodríguez**: Mobile Context, Recommendation Systems
- **María Teresa Lozano Albalate**: Pattern recognition, Image Retrieval.

- **Dragana Dudić** Searching and Reasoning.
- **Sergej Zerr** Linked Data, security.
- **Stefan Dietze** Linked Data, Data Profiling, Knowledge Representation, Entity Linking.
- **Thanassis Tiropanis** Web science.
- **Bernardo Pereira Nunes:** Ontology Matching, Entity Interlinking, The quality of the interlinking
- **Dragan Ivanović.**

4. Lunch break

Participants went to have lunch at the Snack Restaurant Allegro.

5. Keynote talk " From data to knowledge – profiling and interlinking datasets on the Web" by Dr. Stefan Dietze

Dr. Elena Demidova presented the second keynote speaker Dr. Stefan Dietze from L3S Research Center (Germany) who talked about Semantic Web and Linked Data, in particular he talked about how to move from data to knowledge and how to link web datasets (interlinking) and how to create dataset profiles. Dr. Stefan Dietze presented several research results related to those topics and indicated some references to research papers to delve into them.

The abstract of the talk: While the Web of Data, and in particular Linked Data, has seen tremendous growth over the past years, take-up, usage and reuse of data is still limited and is often focused on well-known reference datasets such as DBpedia. Datasets vary heavily with respect to their characteristics, such as the amount, quality, domain coverage or currency of exposed data. While this heterogeneity is natural, given the distributed nature of Linked Data, it also poses challenges for finding, adopting and reusing data across the Web. This problem is further elevated by the lack of reliable information (or metadata) about such dataset characteristics. Hence, judging the suitability and trustworthiness of arbitrary datasets for a given task is challenging. In addition, the evolving nature of Linked Data calls for scalable and efficient techniques which are able to automatically (re-)compute dataset metadata (profiles) and links. This talk will give an overview of ongoing research on dataset profiling and interlinking. Profiling involves the automated extraction of metadata, in particular, about the type or topic coverage of datasets. Interlinking, on the other hand, investigates methods for linking not only entities across the Web of data but also to identify candidate datasets for interlinking tasks. The talk will introduce methods, techniques and resulting datasets, aimed towards the overall goal of improving take-up and reuse of Web datasets.

The slides of this presentation are available from the KEYSTONE webpage at: http://www.keystone-cost.eu/keystone/wp-content/uploads/2014/04/keystone-datasetprofiling-dietze.pdf.

6. Work in groups

Participants were split in two discussion groups: group 1 and group 2. The **group 1** was chaired by Dr. **Elena Demidova** and Dr. J**ulian Szymanski** and include the following participants**:** Konstantin Todorov, Thanassis Tiropanis, Nattiya Kanhabua, Khali Belhajiame, Karol Draszawka, Bernardo Pereira Nunes and Dragan Ivanović. The **group 2** was chaired by Dr. **Stefan Dietze** and Dr. **Raquel Trillo-Lado** and include the following participants: María del Carmen Rodríguez, María Teresa Lozano Albalate, Dragana Dudić , Besnik Fetahu and Sergej Zerr.

The two groups discussed several topics during two sessions. In particular they focused on these topics:

- Semantic keyword search on structured data (group 1 and group 2)
- Linked Data: federated keyword search, mining and preservation (group 1)
- Representation and analytics of structured data sources (group 1 and group 2)
- Aspects of Semantic Keyword Search in Big Data (group 1 and group 2)
- Scenarios of interactions for improvement of keyword based retrieval (group 2)

And on answering the following questions:

- Own research of participants related to the topics.
- Advantages with respect to their research on the topics they envision from collaborations.
- Scenarios they would like to address in their research on the topics.
- Open challenges on the topics.
- Datasets available to them and/or suitable for researching on the topics.
- Evaluation methods that are used in different contexts related to the topics.

7. Closing

Dr. Elena Demidova and Dr. Raquel Trillo-Lado presented the results of the groups During the closing session, conclusions and follow-up actions were discussed and agreed upon, including the intention to start group work on topics such as:

- Surveying methods for search, data discovery and dataset profiling;
- Collecting datasets and results for evaluation and experimentation (as a general resource to KEYSTONE members and the general public)
- Surveying schemas and vocabularies for dataset profiling or link representation

The meeting closed at 18:00 25th May 2014.