



## Research Networking Programmes

Short Visit Grant  or Exchange Visit Grant

(please tick the relevant box)

### Scientific Report

The scientific report (WORD or PDF file – maximum of eight A4 pages) should be submitted online within one month of the event. It will be published on the ESF website.

***Proposal Title:*** 3rd MUMIA Training School on Information Retrieval and Interactive Information Access

***Application Reference N°:*** 6734

#### 1) Purpose of the visit

The purpose of this visit to FORTH Institute of Computer Science, Heraklion, was to attend to the 3rd MUMIA Training School on Information Retrieval and Interactive Information Access, which was held in that institution on July 21-25. This event was organized by the Foundation for Research and Technology-Hellas (FORTH) and the IC1002 COST Action MUMIA (Multilingual and Multifaceted Interactive Information Access), and was intended to give its participants grounding in the core topics of Multilingual and Multifaceted Interactive IAc. It was mainly focused on PhD/MSc students or post-doctoral researchers interested in the fields of information management, information retrieval and related subjects. The MUMIA Training School was the most relevant summer school in the field of Information Access (IAc) in Europe in 2014.

Regarding to my attendance to this training school, it was due to:

1. My research work being focused on IAc within the context of Microblogging and Online Knowledge Base systems and their evaluation. In this respect, assisting to the courses announced in

- the program of this training school was very useful for completing my training on IAc-related topics, with special interest for those courses focusing on evaluation of this type of systems.
2. My interest in interacting with other researchers and students from other countries and establish new contacts to share with our experiences, something essential for a novel PhD student such as me.
  3. My participation in a poster session held at this event which allowed me to present some of my research group's work to the assistants of the training school.

## 2) Description of the work carried out during the visit

The main goal of this short visit was to attend and actively participate in the courses announced in the training school program, with special interest for those focusing on evaluation of IAc systems:

- Task-based Information Interaction Evaluation by Prof. Kalervo Jarvelin, member of the Steering Committee of ELIAS Network. Course totally focused on evaluation.
- IR Evaluation++ by Dr. Mihai Lupu. The same applies.
- Cross-language and Multilingual IR by Prof. Martin Braschler, a well-known specialist in IR application evaluation, member of the Steering Committee of the CLEF (Cross-Language Evaluation Forum) Initiative, organization sponsored by ELIAS Network.
- Bridging the Web of Documents with the Web of Data by Prof. Yannis Tzitzikas. This course focuses on search systems integrating semantic information (as in the case of the Knowledge Bases I work with) and their evaluation.

On the other hand, I also had the chance to present my research group's current work on the topic of microtext normalization during a poster session held at this training school. My presentation focused around the proposed system's architecture which allows us to test a wide range of techniques and thus study the best approach to the problem of normalizing microtexts written in Spanish language. This project's aim is to provide as output a normalized version of a given microtext input (such as tweets) for further processing by other systems developed within my research group, Grupo LyS, such as a sentiment analysis system, available online at <http://miopia.grupolys.org/> .

### 3) Description of the main results obtained

From the courses mentioned before, these are the topics I found more interesting for my research work:

- Theoretical overview of a user centered test framework and techniques for IAc systems evaluation.
- Statistical methods and metrics for measuring the usual IR evaluation parameters, focusing on effectiveness.
- Common problems in IR systems development, distinguishing between monolingual and multilingual systems.
- Usage of semantic data in IR systems to improve results through techniques such as entity mining, semantic analysis and further interaction with the user, and their evaluation.

With respect to my poster presentation, several of the participants showed much interest in the way we made use of a good architecture design to provide an efficient framework to easily test multiple tools and techniques, allowing us to find a better way to solve the problem of Spanish microtext normalization. The poster also showed an overview of the Natural Language Processing (NLP) techniques in use by the system, which led to most of the participants wondering about how these techniques could be applied, if possible, to their corresponding native languages. We could then discuss about these and other NLP techniques that were being applied to their native languages, and how they could be translated, once again, if possible, to Spanish.

### 4) Future collaboration with host institution (if applicable)

In this regard, I maintained several informal meetings with other participants in order to discuss about our work, our group's and the possibility, or not, of future collaborations.

### 5) Projected publications / articles resulting or to result from the grant *(ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant)*

There are no planned publications to be written as a direct result of the assistance to this training school, mostly due to the type of activity. However, the knowledge acquired during this event will prove very useful in my future work within my research group.

### 6) Other comments (if any)

Finally, I would like to note that the visit lasted a day longer than expected. This was to benefit from a lower flight ticket prize.