

# **Multilingual Semantic Processing**

**Saturday, May 31st 2014**

**Presenters:**

**Roberto Navigli, David Jurgens**

**Sapienza University of Rome**

## Tutorial Programme/Overview

14:00-16:00	Resources for Multilingual Semantic Processing
14:00-14:30	Welcome, Introductory material, and Motivation
14:30-16:00	Building resources for multilingual semantic applications
16:00-16:30	Coffee Break
16:30-18:00	Case Studies in Multilingual Semantic Processing
16:30-16:55	Identifying multilingual concepts and entities in text
16:55-17:20	Adding multilingual semantics to Information Extraction
17:20-17:45	Enriching and Customizing Multilingual Data using Gamification
17:45-18:00	Discussion and Closing Remarks

# Tutorial Description/Outline/Contents

**Overview:** Multilingual data is becoming an increasingly important component of online text.

However, many of the Natural Language Processing resources and algorithms have focused entirely on English. Recently, new multilingual resources have emerged that provide a direct bridge from existing resources such as WordNet to multilingual representations. In this tutorial, we highlight one such resource, BabelNet, to show how to transform existing monolingual semantic processing into the multilingual setting. This half-day tutorial introduces concepts of Multilingual Semantic Processing and provides participants with an in-depth look at several case studies that incorporate multilinguality in their semantic tasks using BabelNet. The tutorial contains five core sessions, designed to help participants familiarize themselves with the necessary concepts and extend their current monolingual semantic processing to the multilingual setting. Each session comes with hands-on examples that are done together, take-home exercises for those interested in working on more involved tasks, and a list of references to related research to further deepen their knowledge on a particular topic.

**Contents:** The tutorial is divided into two parts. Part 1 provides the necessary background in semantic processing and introduces how monolingual semantic processing can be extended to the multilingual setting using new multilingual resources. To demonstrate the challenges and benefits of building and using multilingual resources, Part 1 introduces a case study using BabelNet, which merges multiple common semantic resources, such as Wikipedia and WordNet, to enable many kinds of multilingual semantic processing.

Part 2 consists of case studies in building state-of-the-art multilingual semantic processing applications and customizing resources to meet application-specific needs. The first application demonstrates how to identify multilingual concepts and entities in text, while the second application

shows how Information Extraction can be enriched with multilingual semantics to identify relationships between concepts, rather than text. The last session of Part 2 introduces the challenges of producing new data for multilingual resources and demonstrates ideas on how to design game-like tasks to customize resources with new information or improve the quality of automatically-constructed resources.

**Participation Requirements:** We invite all parties interested in extending their work or research from the monolingual to the multilingual setting, or those seeking to understand the challenges and benefits of working with multilingual data and resources. Participants are not expected to be fluent in multiple languages to benefit from the tutorial.