LREC2012 Tutorial: "Bootstrapping ontology evolution: a generic approach relying on ontology-based information extraction"

Conference venue, Lütfi Kirdar Istanbul Exhibition and Congress Centre

Monday, 21 May 2012 – Morning Session

Summary

This tutorial provides a detailed introduction to the research area of ontology evolution. After a short introduction to the problem of ontology evolution and the presentation of the current state of the art (Part I), the tutorial will present in detail the ontology learning approach that has been developed in the context of the BOEMIE EU-funded research project. The tutorial will present an ontology-based information extraction system and how this system is exploited to learn an ontology in a synergetic, semi-automated approach, employing bootstrapping (Part II). The third part of the tutorial (Part III) will focus on how internal information (encoded in instances) and external knowledge sources (i.e. other ontologies and hierarchies) can be exploited in order to enhance proposals for new concepts, through instance matching. Finally, the tutorial will conclude with the state of the art in ontology evaluation, and evaluation results of the described approach on the thematic domain of athletics (Part IV).

Motivation

In recent years, ontologies have become extremely popular as a means for representing machine-readable knowledge. The difficulty of extracting information from the Web, that was created mainly for visualising information, has driven the birth of the Semantic Web, which will contain much more resources than the Web and will attach machine-readable semantic information to these resources. Realizing the difficulty of designing the grant ontology for the world, research on the Semantic Web has focused on the development of domain or task-specific ontologies, which have made their appearance in fairly large numbers.

Having provided an ontology for a specific domain, the next step is to annotate semantically related Web resources. If done manually, this process is very time-consuming and error prone. At the same time, acquiring domain knowledge for ontologies is also a resource demanding and time consuming task. Thus, the automatic or semi-automatic construction, enrichment and adaptation of ontologies, is highly desired. To this end, the evolutionary aspects of ontologies have received significant research attention during the last years, as ontology engineering has reached a certain level of maturity, considering the vast amount of contemporary methods and tools for formalizing and applying knowledge representation models.

Overview

This tutorial aims to provide a comprehensive and up-to-date overview to the developments and research issues of ontology evolution, presenting the state of the art in the field, describing in detail a semi-automated ontology evolution approach relying on ontology-based information extraction, ontology learning and matching. Evaluation methods focusing on the evaluation of the proposed approach will also be presented.

In the first part of the tutorial, we will start by giving a general overview of the field and discussing the main tasks addressed therein. We will then focus on the state-of-the-art of the field. We will review some of the currently available ontology learning frameworks and systems, discussing their main features as well as drawbacks.

In the second part, we will present a semi-automated approach for ontology learning and we will show how this ontology evolution approach overcomes both the problem of manually annotating large corpora of multimedia documents and the problem of providing a semantically rich description of resources. This is done by supporting domain experts in visualizing resource interpretations and in enriching the domain knowledge of the system. This triggers a new cycle of document interpretation and improves the capability of the system to provide a correct description of new bunches of documents.

The third part of the tutorial will be devoted to the role of ontology and instance matching in the ontology evolution process. Actually, the capability of comparing different ontology versions and their instances, by formally describing and measuring their differences and similarities through automatic matching techniques, plays a crucial role both in the evolution process and in the evaluation of ontology learning systems. In order to address these issues, we will briefly define the ontology and instance matching problems by presenting the main state of the art contributions in the field, and then we will focus on the application of these methods to ontology evolution.

In the last part of the tutorial, we will present the current approaches in the evaluation of ontology learning systems and we will then focus on the evaluation of the proposed ontology learning approach. The evaluation scenario to be presented concerns the complete bootstrapping approach. This way, we can evaluate both the capacity of the system to propose concepts/relations/rules, as well as the capability of the new concepts to explain the multimedia documents analysis results.

Outline of the tutorial

Part I. Introduction to Ontology Learning

- Definition of Ontology Learning
- Existing ontology learning frameworks and systems

Part II. Semi-automated approach for ontology learning

- Ontology-based Information Extraction
- Ontology Population
- Ontology Enrichment
- Bootstrapping approach

Part III. Ontology & Instance Matching

- Ontology matching
- Instance matching
- Application on Ontology Evolution

Part IV. Evaluating Ontology Learning Methods

- State-of-the-art evaluation approaches
- Evaluation of the bootstrapping approach

Part IV. Wrap-Up & Discussion

The Presenters

- Dr. Georgios Petasis (NCSR "Demokritos", Greece)
- Dr. Anastasia Krithara (NCSR "Demokritos", Greece)
- Dr. Alfio Ferrara (University of Milano, Italy)
- Dr. Vangelis Karkaletsis (NCSR "Demokritos", Greece)

Dr. George Petasis holds a PhD in Computer Science from University of Athens (2011) and he is a research associate at the Software and Knowledge Engineering Laboratory of the Institute of Informatics and Telecommunications of N.C.S.R. "Demokritos". His research interests include the areas of named-entity recognition, information extraction, ontology learning, linguistic resources, speech synthesis, natural language infrastructures, grammatical inference and machine learning.

Dr. Anastasia Krithara holds a PhD in Machine Learning from Pierre and Marie Curie University, Paris 6 (2008) and she is a research associate at the Software and Knowledge Engineering Laboratory of the Institute of Informatics and Telecommunications of N.C.S.R. "Demokritos". Her research interests include machine learning, information retrieval, ontology learning and knowledge discovery and engineering.

Dr. Alfio Ferrara is assistant professor of Computer Science at the Computer Science Department of the University of Milano, where he received his Ph.D. in Computer Science in 2005. His research interests include database and semi-structured data integration, Web-based information systems, ontology engineering, and knowledge representation and evolution.

Dr. Vangelis Karkaletsis is a Research Director at NCSR "Demokritos", and head of the Software and Knowledge Engineering Laboratory of the Institute of Informatics and Telecommunications. He holds a PhD in computer Science from University of Athens. His research interests are in the areas of language and knowledge engineering, as applied to content analysis, natural language interfaces, ontology engineering, personalisation.