Abstract

In this article, we present MELIS (Meaning Elicitation and Lexical Integration System), a method and a software tool for enabling an incremental process of automatic annotation of local schemas (e.g. relational database schemas, directory trees) with lexical information. The distinguishing and original feature of MELIS is the incremental process: the higher the number of schemas which are processed, the more background/domain knowledge is cumulated in the system (a portion of domain ontology is learned at every step), the better the performance of the systems on annotating new schemas. MELIS has been tested as a component of the MOMIS-Ontology Builder, a framework able to create a domain ontology representing a set of selected data sources, described with a standard W3C language wherein concepts and attributes are annotated according to the lexical reference database. We describe the MELIS component within the MOMIS-Ontology Builder framework and provide some experimental results of MELIS as a standalone tool and as a component integrated in MOMIS.

Keywords: data integration; data semantics; metadata; semantic matching

Introduction

The growth of information available on the Internet has required the development of new methods and tools to automatically recognize process and manage information available in Web sites or Web-based applications. The aim of the Semantic Web is to build a web of data by providing a common framework that enables data sharing and reuse across application, enterprise, and community boundaries. The Semantic Web relies on the use of shared schemas and ontologies, which should provide a well-defined basis of shared meanings for data integration and reuse.

However, practical experience in developing semantic-enabled Web applications and information systems shows that the simple and intriguing vision sketched above is not a
solution to all problems. In particular, we stress the following issues:

- selecting an appropriate ontology for describing an application’s data may be very difficult. Indeed, engineering a new ontology from scratch can be extremely time consuming, and expensive and requires appropriate skills; finding a preexisting ontology that perfectly fits local data is very unlikely, as most available ontologies are either too generic (and therefore semantically poor) or too specific (and therefore not suited for data different from those of the original application). Moreover, there is no standard recommendation or specification for referencing ontologies in information sources and different tools use different languages and techniques to add annotations. Several proposals and tools have been developed for including references to ontologies in HTML pages. However, such operation is typically executed off-line by adding “annotations” to the sources.

- because of the intrinsically distributed nature of knowledge on the Web, different applications may refer to different ontologies to specify the meaning of their data.

The two issues above led the Semantic Web and Database communities to address two very hard problems: ontology learning (inducing ontologies from data/schemas) and ontology matching/integration (bridging different ontologies). There is a vast literature on these topics, and we will review part of it in a later section.

However, we observe that several methods and tools developed to address the two problems rely—in different ways—on the use of lexical information. The reason is simple: beyond the syntactic and semantic heterogeneity of schemas and ontologies, it is a fact that their elements and properties are named using natural language expressions, and that this is done precisely because they bring in useful (but often implicit) information on the intended meaning and use of the schema/ontology under construction. Therefore, it should not come as a surprise that a large number of tools for ontology learning and schema/ontology matching include some lexical resource (mainly WordNet) as a component, and use it in some intermediate step to annotate schema elements and ontology classes/properties with lexical knowledge. To sum up, lexical annotation seems to be a critical task to develop smart methods for ontology learning and matching.

In this context, we developed MELIS (Meaning Elicitation and Lexical Integration System), an incremental method and a software tool for the annotation of data sources. MELIS is based on the integration and the extension of the lexical annotation module of the MOMIS-Ontology Builder and some components from CtxMATCH2.0, a tool for eliciting meaning and matching pairs of nodes in heterogeneous schemas, using an explicit and formal representation of their meaning (Bouquet, Serafini, & Zanobini, 2005; Bouquet, Serafini, Zanobini, & Sceffer, 2006). CtxMATCH2.0 was extended with respect to (Bouquet et al., 2005, 2006) with a set of heuristic rules to generate new annotations on the basis of the knowledge provided by a given set of annotations; WNEditor was modified in order to jointly work with CtxMATCH2.0, by providing a customized lexical database.

The distinguishing feature and the novelty of MELIS is its incremental annotation method: the more sources (including a number of different schemas) are processed, the more background/domain knowledge is cumulated in the system, the better the performance of the systems on new sources.

MELIS supports three important tasks: (1) the source annotation process; that is, the operation of associating an element of a lexical reference database (WordNet in our implementation, but the method is independent from this choice) to all source elements, (2) the customization of the lexical reference with the introduction of new lexical knowledge (glossa, lemma and lexical relationships), and (3) the extraction of
Semantic Verification of Business Process Models: Prospects and Limitations
www.igi-global.com/chapter/semantic-verification-business-process-models/60060?camid=4v1a

Using the Linked Data Approach in European e-Government Systems: Example from Serbia
www.igi-global.com/article/using-the-linked-data-approach-in-european-e-government-systems/203691?camid=4v1a