The BOEMIE Semantic Browser: An Application Exploiting Rich Semantic Metadata *

Sofia Espinosa Peraldi, Atila Kaya, Ralf Möller

The Semantic Web vision promises to pave the way for the development of intelligent applications and services. To this end the semantics of information in the Semantic Web should be available. Nowadays the majority of information in the Web reside in web pages. In the last decade, information retrieval from the Web has become the underlaying basis of daily information access. At the same time, natural language processing techniques have reached a high level of maturity such that text analysis tools with good performance become widely-used in practical applications. Nowadays, tools such as Ellogon\(^1\) provide a comprehensive infrastructure for natural language processing. Facilitated by the improvements in natural language processing technologies, commercial interest arose for applications that automate the annotation of documents with metadata. Thus, companies such as ClearForest\(^2\) and Reuters\(^3\) provide annotation services about various entities such as persons, organizations, locations as well as facts and events from textual parts of documents. But current annotation services should be further developed such that they generate richer metadata if one is to develop ‘intelligent’ applications that offer more valuable services than possible today.

Rich metadata means symbolic descriptions that represent a deeper level of information i.e. more abstract entities and relations among those. Rich metadata is also multimodal thus describes the content of multimedia documents. This paper has two main contributions: First we describe use cases for ‘intelligent’ applications that exploit rich metadata automatically extracted from multimedia documents. Second, the presentation of the BOEMIE Semantic Browser (BSB), a semantic application that implements these use cases in order to demonstrate possible exploitation scenarios of rich semantic metadata. Furthermore, to better understand what is meant with rich semantic metadata, we introduce a framework for the extraction of such rich metadata. In this framework analysis tools are responsible for the extraction of metadata from modality specific content, e.g., text, image and video. This metadata is represented through Description Logics (DLs)-based symbols which are a subset of first-order logic. The resulting modality specific metadata is later used as input for a process of interpretation and fusion that finally results in rich metadata. In conclusion we describe what rich metadata is and how they are usable for the creation of ‘intelligent’ applications.

---

\(^*\)This paper has been partially supported by the BOEMIE Project, contract FP6-027538, under the 6th EU Framework Programme.

\(^1\)http://www.ellogon.org/

\(^2\)www.clearforest.com

\(^3\)www.reuters.com