



The 7th International Conference on Ambient Systems, Networks and Technologies
(ANT 2016)

Semantic Annotation of Arabic Web Resources using Semantic Web Services

Saeed Albukhitan, Ahmed Alnazer, Tarek Helmy*

*Information and Computer Science Department,
College of Computer Science & Engineering,
King Fahd University of Petroleum & Minerals,
Dhahran 31216, Mail Box 413, Saudi Arabia,
[albokhitan,alnazera,helmy]@kfupm.edu.sa*

Abstract

The vision of semantic Web is to have a Web of things instead of Web of documents in a form that can be processed by machines. This vision could be achieved on the existing Web using semantic annotation based on common and public ontologies. Due to exponential growth and huge size of the Web sources, there is a need to have fast and automatic semantic annotation services of Web documents. Since Arabic language received less attention in semantic Web research as compared to Latin languages especially in the field of semantic annotation. This motivates us in this paper to present semantic Web services that support the semantic annotation of Arabic language documents. The services accept documents and ontologies and produce annotations of these documents using different output formats. The proposed services could be used for building semantic Web applications and semantic search engines for Arabic Language. To evaluate the performance of these services, a set of ontologies were used with pre-annotated documents related to those ontologies. The initial results show a promising performance which will support the research in the semantic Web with respect to Arabic language.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Conference Program Chairs

Keywords: Semantic Web Services; Semantic Annotation; Arabic Language; Ontology, OWL-S

* Corresponding author. Tel.: +966-13-860-1967; fax: +966-13-860-2174.

E-mail address: helmy@kfupm.edu.sa

1. Introduction

Web services are application components that can be accessed via the Web programming interface. They provide functionality of a Web information system using standard technologies. However, their common descriptive languages such as WSDL¹, lack semantic richness in order for machines to process them automatically. They require human intervention to interpret their meanings for discovery, composition, and invoking. Manual intervention is an error-prone and time-consuming task. W3C supports the use of software agents for automating the above tasks. An agent is defined as application software acting on behalf of a person, a system or an organization². Accordingly, an agent could discover, compose and invoke Web services. In order for agent to perform those tasks, it requires a reference specification that includes domain informational knowledge, and operational knowledge of how to perform domain tasks. Ontology is an effective way to provide such specification.

Semantic Web service uses ontologies as their data models which bring the benefits of semantics to the executable part of the Web³. It resolves ambiguous description of service functionality and external interface. It reduces human intervention while integrating services in Service-Oriented Architecture (SOA). With semantic Web services, many tasks in the process of using Web services can be automated. SOA dynamism is improved when new services available for use as they appear. There is no need for service consumers and producers to know of each other's existence. Also, stability of the service could be improved because service interfaces are not tightly integrated so even less impact from changes. Services can be easily replaced if they are no longer available.

The amount of research and work done for annotating Arabic content in Web is very limited and non-scalable. One of the biggest challenges facing Arabic research is the availability and accessibility of Arabic resources, such as ontologies, corpora, named list, dictionaries, and NLP tools. This challenge makes collection, analyzes and investigation of such resources laborious especially if the semantic annotation techniques depend on such resources.

In this paper, we present semantic Web services that annotate the Arabic Web resources and produce annotation in different formats. The services support parsing of ontologies stored in different formats including RDF, OWL and N-TRIPLE. The services include document and ontology handling, in addition to entity and relation extraction.

The rest of this paper is organized as follows. Section 2 reviews the existing related work on semantic annotation and Web services for Web sources. Section 3 presents the proposed services. Section 4 presents the performance of the services prototype with the discussion of experiments and results. Finally, we summarize the paper and highlight the future work directions in Section 5.

2. Related Work

A survey of some semantic Web technologies supporting Arabic is presented in Beseiso et al.⁴. Four mostly used semantic Web tools were investigated, namely Protégé, Jena, Sesame, and KOAN. Their investigation focuses in the tool's functionality, type of standards supported and support level of Arabic language. Their conclusion is that those tools do not support Arabic language completely as compared to Latin languages. The most supporting tool for Arabic language was Jena with some limited support for query processing. Arabic language does not get the same support as compared to Latin languages. The common challenges of Arabic language with respect of NLP tasks were highlighted in Abdel Rahman et al. [5]. Arabic language does not have features such as case-sensitivity which is an importance feature used by Latin languages to detect proper names. Arabic words could have more than one affix and can be expressed as combination of affix such as prefixes, lemma and suffixes which make it more difficult for stemming. Arabic words also have diverse types of ambiguities associated to typographic forms and spelling.

The first Arabic semantic annotation tool was presented by Bin Saleh and Alkhalifa⁶. The presented tool was named AraTation to annotate Arabic news in the Web. The tool is capable of extracting named entities using Arabic location ontology built for this purpose. The tool reported an achievement in average precision of 67% and recall of 82% on a set of ten locations over 25 Web documents. We were not able to test the tool since the tool is not publicly available. Another work on Arabic language annotation was presented by Zaidi et al.⁷ using GATE⁸ NLP toolkit. Their system used crescent Quranic Corpus as an input. The system is capable to extract named entities through predefined patterns that use tokenized and morphology analyzed corpus with Part Of Speech (POS) features. Another semantic annotation tool for Arabic sources is presented by Motasem et al.⁹. It was tested on news article

Download English Version:

<https://daneshyari.com/en/article/485395>

Download Persian Version:

<https://daneshyari.com/article/485395>

[Daneshyari.com](https://daneshyari.com)