International Journal of Innovative Computing, Information and Control
Volume **7**, Number **11**, November **2011**

ICIC International ©2011 ISSN 1349-4198
pp. 6115–6133

ONTOLOGY POPULATION: AN APPLICATION FOR THE E-TOURISM DOMAIN

Juana María Ruiz-Martínez[1], Jose Antonio Miñarro-Giménez[1]
Dagoberto Castellanos-Nieves[2], Francisco García-Sánchez[3]
and Rafael Valencia-García[1]

[1]Facultad de Informática
Universidad de Murcia
ES-30100, Murcia, Spain
{ jmruymar; jose.minyarro; valencia }@um.es

[2]Depto. de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada
E-18071, Granada, Spain
dcastellanos@ugr.es

[3]Escola Tècnica Superior d'Enginyeria
Universitat de València
Avda. Dr Moliner 50, Burjassot 46100, Valencia, Spain
francis6@uv.es

Received May 2010; revised December 2010

Abstract. *The Semantic Web aims to extend the current Web standards and technologies so that the semantics of Web contents is machine processable. For the Semantic Web vision to become real, methods and mechanisms that assist in the creation of an initial pool of semantically described Web resources have been developed. However, these methods suffer from some problems such as lack of scalability and automation. Besides, most of them focus on English language resources and the cost of initial requirements is too high. In order to solve these drawbacks, this paper proposes a methodology for extracting semantic content from textual web documents to automatically instantiate a domain ontology (i.e., ontology population). In a first stage, the system obtains, through the GATE framework, a set of semantic annotations which are considered as ontology instance candidates. In a second stage, the semantic ambiguities are solved, and the annotations are related with their corresponding ontological entities. The methodology has been tested in a tourism domain corpus and the results of the validation process seem promising in terms of precision and recall.*

Keywords: Ontology population, Natural language processing, Named entity recognition

1. **Introduction.** The information contained on Web pages was originally designed to be human-readable, and so, most of the knowledge currently available on the Web is kept in large collections of textual documents. As the Web grows in both size and complexity, there is an increasing need for automating some of the time consuming tasks related to Web content processing and management. In 2001, T. Berners-Lee and his colleagues defined the Semantic Web as an extension of the current Web, in which information is given well-defined meaning, enabling computers and people to work better in cooperation [1]. The Semantic Web vision is based on the idea of explicitly providing the knowledge behind each Web resource in a manner that is machine processable. Ontologies [2] constitute the standard knowledge representation mechanism for the Semantic Web. The formal semantics underlying ontology languages enables the automatic processing of the