

Statement of Interest: Antonio Pareja-Lora
Facultad de Informática, UCM
Ciudad Universitaria, s/n, Madrid, Spain, 28040
apareja@sip.ucm.es

I am a member of LIA (Artificial Intelligence Laboratory) at UPM, currently working in a Ministry of Science and Technology (MCyT) funded project called "ContentWeb: Semantic Web Technologic Platform: Ontologies, Natural Language Analysis and E-Business", focusing on ontologies and linguistic annotations for the Semantic Web.

The purpose of *ContentWeb* is the creation of an ontology-based Semantic Web platform to enable users to query e-commerce applications by using natural language, performing the automatic retrieval of information from web documents annotated with ontological and linguistic information. Besides, a prototype in the entertainment domain will be developed. *ContentWeb* objectives can be enunciated as follows:

1. Semi-automatic building of ontologies in the domains of e-commerce and of entertainment, reusing existent ontologies and international e-commerce standards and joint initiatives.
2. Elaboration of *OntoTag*, a model and environment for the hybrid –linguistic and ontological– annotation of Semantic Web documents.
3. Development of *OntoConsult*, a natural language interface based on ontologies.
4. Creation of *OntoAdvice*, an ontology-based system for querying and information retrieving from annotated Semantic Web documents in the entertainment domain.

Following the guidelines of the Semantic Web initiative [Berners-Lee et al., 99], AI researchers have found in *ontologies* [Gruber, 93], [Guarino et al., 95], [Studer et al., 98] the ideal knowledge model to formally describe web resources and its vocabulary and, hence, to make it explicit in some way the underlying meaning of web page concepts. Many systems and projects have been developed towards this aim hitherto: SHOE [Luke et al., 00] proposes HTML page semantic annotation with a Horn clause-based language also called SHOE; the (KA)² initiative [Benjamins et al., 99] seeks to annotate HTML documents with ontological

information, taking Knowledge Acquisition Community ontologies as a basis; PlanetOnto [Motta et al., 99] aims at automatically annotating an organisation's HTML news pages through the information obtained from an event-ontology-based knowledge base; finally, within the Semantic Community Web Portals project [Staab et al., 00] an ontology-based architecture for an easier editing and maintenance of web portals is being developed.

However, these researchers have been neglecting, somehow, the decades of work and the results obtained in the field of Corpus Linguistics on corpus annotation, not only in the semantic level, but also in other linguistic levels. These other linguistic levels, not being intrinsically semantic, can add some semantic information to help a computer understand a text or, in our case, web pages. The purpose of our project, ContentWeb, is to bring together and join the results from both fields (AI and Corpus Linguistics) in a hybrid ontological and linguistic annotation model that will require the use of up-to-date ontologies for the Semantic Web, on which this workshop will focus.

Related Publications

Benjamins, V.R., Fensel, D., Decker, S., Gómez-Pérez, A. (KA)²: *Building Ontologies for the Internet: a Mid Term Report* in IJHCS, International Journal of Human Computer Studies, 51, pp. 687-712. 1999.

Berners-Lee, T., Fischetti, M. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. Harper. San Francisco. 1999.

Gruber, R. *A translation approach to portable ontology specification*. Knowledge Acquisition. #5, pp. 199-220. 1993.

Guarino, N., Giaretta, P. *Ontologies and Knowledge Bases: Towards a Terminological Clarification* in N. Mars (ed.) *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995*. IOS Press, Amsterdam, pp. 25-32. 1995.

Luke S., Heflin J. *SHOE 1.01. Proposed Specification*. SHOE Project. February, 2000. <http://www.cs.umd.edu/projects/plus/SHOE/spec1.01.htm>

[Motta et al., 99] Motta, E., Buckingham Shum, S. Domingue, J. *Case Studies in Ontology-Driven Document Enrichment*. Proceedings of the 12th Banff Knowledge Acquisition Workshop, Banff, Alberta, Canada, October 16-22, 1999.

Staab, S., Angele, J., Decker, S., Erdmann, M., Hotho, A., Mädche, A., Schnurr, H.-P., Studer, R. *Semantic Community Web Portals*. WWW'9. Amsterdam. 2000.

Studer, R., Benjamins, R., Fensel, D. *Knowledge Engineering: Principles and Methods*. *DKE* 25(1-2), pp. 161-197. 1998.