

Snoussi, H., Magnin, L., Nie, J.-Y. « Heterogeneous Web Data Extraction using Ontologies » In *Third International Bi-Conference Workshop on Agent-oriented information systems (AOIS-2001)* In *Proceedings of the 5th International Conference on Autonomous Agents 2001*. Montréal, Canada, May 28, 2001

Abstract

Multi-agent systems can be fully developed only when they have access to a large number of information sources. These latter are becoming more and more available on the Internet in form of web pages. This paper does not deal with the problem of information retrieval, but rather the extraction of data from HTML web pages in order to make them usable by autonomous agents. This problem is not trivial because of the heterogeneity of web pages. We describe our approach to facilitate the formalization, extraction and grouping of data from different sources. To do this, we developed a utility tool that assists us in generating a uniform description for each information source, using descriptive domain ontology. Users and agents can query the extracted data using a standard querying interface. The ultimate goal of this tool is to provide useful information to autonomous agents.

Keywords: data extraction, WEB, ontology, agent, XML.