

Manuscript ID : 00000-65932

International Journal of Computer Engineering and Technology

Volume 7, Issue 5, September – October 2016, Pages 65-76, Page Count - 12



Source ID : 00000005

## FIOBODA - SEMANTIC ANNOTATION FRAMEWORK FOR WEB EXTRACTED DATA

C. Gnana Chithra <sup>(1)</sup> E. Ramaraj <sup>(2)</sup>

<sup>(1)</sup> Equity Research Consultant, Angeeras Securities, Chennai, India.

<sup>(2)</sup> Professor, Department of Computer Science and Engineering, Alagappa University, Karaikudi, India.

### Abstract

*Semantic annotation of web pages is the state of art technology for achieving the unified objective of attaining Semantic web Universe, which enables sharing, and reusing the document content beyond the boundaries and applications. Web is a treasury of knowledge and efficient tools should be designed to explore the structured and unstructured data. Annotating million of web pages manually is an impossible task. For high information retrieval rates, automatic annotation of documents is mandatory. Metadata is added to the web pages to make it intelligent for processing in content based intelligent applications. This paper analyses the problems with the current Semantic annotation systems and proposes a new Ontology based Automatic annotation system Framework. Ontology based semantic annotation is one of the best methods for extracting data from the Knowledge Base.*

*The integration of Modified Manning's Sentence boundary detection algorithm and Noun Phrase Collocation algorithm and classification using machine learning techniques in the Information Extraction module, and developing a new data model and ontology for Structured Ontology engineering model is contributed in this paper. Annotation module annotates the output of the information extraction module with the aid of ontologies and dictionaries and stores the resultant annotated data as RDF triples in the Annotation database. Reasoning is made on the Annotated data by the RDF repository interface. FIOBODA is abbreviated as the Financial Instruments ontology based open document annotation. Web pages extracted from the Financial securities domain are mapped with the Finance ontology to extract the subject, predicate and object. SVM classifier is used to classify the correct and incorrect annotations. The correct output annotation data is stored in Annotation data base and RDF repository for later use. The proposed framework to an extent solves the problem of knowledge bottleneck due to its reusability and interoperability features.*

### Author Keywords

Dublin Core, FIOBODA, Financial Securities Ontology, Metadata, Semantic Annotation Framework

**ISSN Print:** 0976-6367

**Source Type:** Journals

**Publication Language:** English

**Abbreviated Journal Title:** IJCTET

**Publisher Name:** IAEME Publication

**Major Subject:** Physical Sciences

**Subject area:** Software Development

**ISSN Online:** 0976-6375

**Document Type:** Journal Article

**DOI:**

**Access Type:** Open Access

**Resource Licence:** CC BY-NC

**Subject Area classification:** Computer Science

**Source:** SCOPEDATABASE