

CONTESSA: A CONTENT Semantic Service Agent

Schahram Dustdar, Harald Gall, Gerald Reif
Technical University of Vienna
A-1040 Wien, Argentinierstrasse 8
{dustdar,gall,reif}@infosys.tuwien.ac.at

Klaus Niederacher, Alexander Wahler
NIWA Web Solutions
A-1090 Wien, Porzellangasse 13/16
{niederacher,wahler}@niwa.at

1 Motivation

An emerging direction of today's content management is leading towards document tagging empowered by the knowledge about the semantic domain of document idioms and the linguistic knowledge. The CONTESSA project addresses the area of semantic web and domain ontologies from such a distinctive angle: documents are structurally and semantically analyzed by a service agent that learns from and modifies the underlying ontology through this semantic tagging process. Therefore, ontologies are progressively evolved and transformed based on the processed documents and tags. The goal of the CONTESSA project is to build such a content-semantic service agent to support content managers or content management system users in semantic analysis, attribution, tagging, and interweavement of XML documents. Heuristics-based tag system transformation will significantly contribute to evolving ontologies.

2 Goals and proposed Solution

The CONTESSA service agent will be developed to support several key requirements for effective content-semantic document management:

- *linguistic analysis of document structures*
- *semantic analysis of document domains*
- *assessment of quality attributes of ontologies*
- *learning from the past by evolvement of ontologies based on enrichment of domain and linguistic knowledge*
- *powerful attribution, tagging and interweavement of documents*

For that, the CONTESSA service agent will be based on a novel approach integrating several technologies: structural tagging, analysis of topic/comment structures, semantic tagging and discourse analysis, text mining, and heuristics-based tag-system transformation. Combined in a uniform architecture and prototype tool, such techniques will allow operating a service agent for content management that fully supports the user in managing large quantities of documents in a semantically progressive and advanced instrumentation.

Recently, a few techniques have been proposed to design and build ontologies, resolve redundancies across

documents, or to evaluate existing ontologies. The CONTESSA project combines technologies ranging from text mining to heuristics-based ontology transformation. This technology merging will define a significant step in the upcoming area of semantic web. Figure 1 depicts the main aspects of the project.

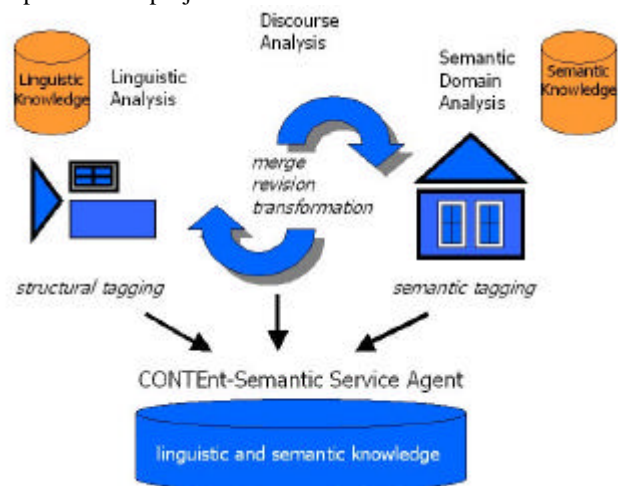


Figure 1 A Content-semantic service agent

3 Conclusions

Advanced content management will be enabled by a service agent that supports a user in a way such that textural contents will be analyzed and semantically related based on ontologies and their semantical progression. The solution will combine text mining with heuristics based ontology evolution and transformation.

References

- [Berners-Lee, 2001] Berners-Lee T., Hendler J. and Lassila O. The semantic Web. Scientific American (May2001).
- [Byron 1999] Byron, Donna & James Allen. 1999. "Applying Genetic Algorithms to Pronoun Resolution". Proceedings of the 16th National Conference on Artificial Intelligence (AAAI-99).
- [Ceri 2002] Ceri, S. Fraternali, P. and Matera, M. Conceptual modeling of data-intensive Web Applications, IEEE Internet Computing, to appear (2002).
- [Stephens 2001] Larry M. Stephens, Michael N. Huhns Consensus Ontologies Reconciling the Semantics of Web Agents, IEEE Internet Computing 9 (2001).