LanguageWare is a natural language processing (NLP) technology developed by IBM , which allows applications to process natural language text. It comprises a set of Java libraries that provide a range of NLP functions: language identification , text segmentation/tokenization, normalization, entity and relationship extraction , and semantic analysis and disambiguation. The analysis engine uses a finite-state machine approach at multiple levels, which aids its performance characteristics while maintaining a reasonably small footprint.

The behaviour of the system is driven by a set of configurable lexico-semantic resources which describe the characteristics and domain of the processed language. A default set of resources comes as part of LanguageWare and these describe the native language characteristics, such as morphology, and the basic vocabulary for the language. Supplemental resources have been created that capture additional vocabularies, terminologies, rules and grammars, which may be generic to the language or specific to one or more domains.

A set of Eclipse -based customization tooling, LanguageWare Resource Workbench, is available on IBM's alphaWorks [1] site, and allows domain knowledge to be compiled into these resources and thereby incorporated into the analysis process.

LanguageWare can be deployed as a set of UIMA -compliant annotators, Eclipse plug-ins or Web Services .

See also

Data Discovery and Query Builder

Formal language

IBM Omnifind

Linguistics

Semantic Web

Semantics

Service-oriented architecture

Web services

UIMA

References

External links

IBM LanguageWare Resource Workbench on alphaWorks

IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks on alphaWorks

JumpStart Infocenter for IBM LanguageWare on IBM.com

UIMA Homepage at the Apache Software Foundation

UIMA Framework on SourceForge

IBM OmniFind Yahoo! Edition (FREE enterprise search engine) Archived 2007-02-17 at the Wayback Machine

Semantic Information Systems and Language Engineering Group

SemanticDesktop.org

Related Papers

Branimir K. Boguraev Annotation-Based Finite State Processing in a Large-Scale NLP Architecture, IBM Research Report, 2004

Alexander Troussov, Mikhail Sogrin, "IBM LanguageWare Ontological Network Miner"

Sheila Kinsella, Andreas Harth, Alexander Troussov, Mikhail Sogrin, John Judge, Conor Hayes, John G. Breslin, "Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects"

Mikhail Kotelnikov, Alexander Polonsky, Malte Kiesel, Max Völkel, Heiko Haller, Mikhail Sogrin, Pär Lannerö, Brian Davis, "Interactive Semantic Wikis" [permanent dead link]

Sebastian Trüg, Jos van den Oever, Stéphane Laurière, "The Social Semantic Desktop: Nepomuk"

Séamus Lawless, Vincent Wade, "Dynamic Content Discovery, Harvesting and Delivery" [dead link]

R. Mack, S. Mukherjea, A. Soffer, N. Uramoto, E. Brown, A. Coden, J. Cooper, A. Inokuchi, B. Iyer, Y. Mass, H. Matsuzawa, and L. V. Subramaniam, "Text analytics for life science using the Unstructured Information Management Architecture"

Alex Nevidomsky, "UIMA Framework and Knowledge Discovery at IBM", 4th Text Mining Symposium, Fraunhofer SCAI, 2006