PoolParty Thesaurus Management based on Open Semantic Web Standards

Andreas Blumauer
a.blumauer@semantic-web.at

May, 2011 **http://www.poolparty.biz/**

Introduction

PoolParty is a web-based thesaurus management system which is completely built on top of W3C´s Semantic Web standards.

Collaborative thesaurus management

Thesauri in the age of the web most often should be engineered and maintained in a collaborative manner.

Thesaurus Management and Linked (Open & Closed) Data

The rise of Linked Data indicated by the enormous growth of the Linked Open Data cloud is an important argument for many organisations to publish their own data at least partly as Linked Open Data.

PoolParty Enterprise Vocabulary and Metadata Management

PoolParty is an enterprise ready system, which offers high reliability, usability, performance and mechanisms like failover which guarantees smooth workflows and protection from loss of data.

Text Mining and Semantic Search

PoolParty offers a variety of options to ease thesaurus management by means of text mining as well as semantic search capabilities.

Vertical Search Solutions: PoolParty Product Family

PoolParty product family consists of three components: PoolParty Thesaurus Management (PPTM), PoolParty Extractor (PPX) and PoolParty Semantic Search (PPSS).

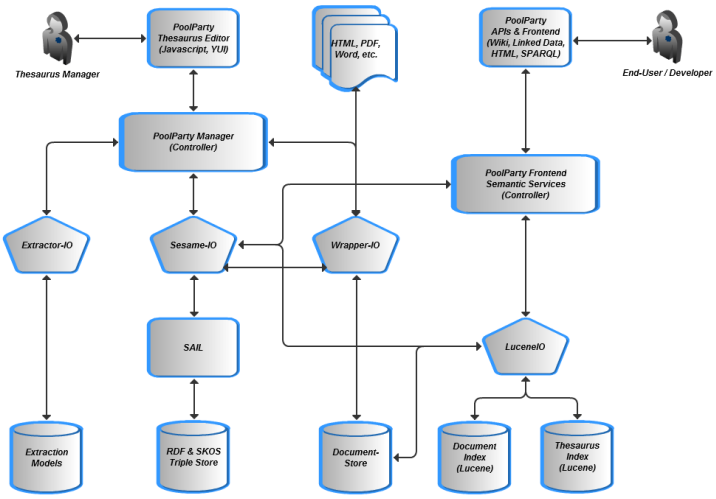
Summary

PoolParty Product Family offers a wide variety of options to deploy semantic technologies. The big three topics are: Semantic Search, Thesaurus Management and Linked Data.

References

Introduction

PoolParty Thesaurus Manager (PPTM) [1] is a web-based thesaurus management system which is completely built on top of W3C´s semantic web standards [2]. In its core PoolParty uses RDF [3] to represent SKOS [4] and other vocabularies like Dublin Core [5] or FOAF [6], therefore an RDF triple store is used as its technological basis. Compared to other systems which still rely on relational databases PoolParty is ready to consume and to publish Linked Data out-of-the-box. Besides the possibility to publish any PoolParty based thesaurus via a Linked Data front-end, the system offers a SPARQL endpoint [7, 8] to execute queries over each thesaurus project. This technology can be used to integrate a thesaurus with other platforms (wikis, CMS, etc.) or search engines.



PoolParty Thesaurus Management System - Overview

Collaborative thesaurus management

Thesauri in the age of the web most often should be engineered and maintained in a collaborative manner. PoolParty is fully web-based; administrators need only a web browser to do all typical CRUD operations like adding new concepts or relations. PoolParty also offers a wiki frontend for each thesaurus project so that more people can be involved in the thesaurus development process. Linking concepts is another flexible way to build thesauri in decentralised structures. Based on the linked data principles [9] thesauri can be maintained at different places but still can be connected to each other indicating that several concepts are similar or even identical to each other.

Thesaurus Management and Linked (Open & Closed) Data

The rise of Linked Data indicated by the enormous growth of the Linked Open Data cloud [10] is an important argument for many organisations to publish their own data at least partly as Linked Open Data. PoolParty´s Linked Data frontend [11] is an easy-to-manage way to do exactly this by as it also offers options to customise the own publishing process. Since PoolParty is not only a system serving government organisations but also enterprises with metadata management solutions, PoolParty´s Linked Data mechanisms can be used as a data integration technology also behind the corporate firewalls.

PoolParty also makes use of existing Linked Data sources, e.g. concepts can be aligned and enriched with additional information from sources like Dbpedia [12], Sindice [13] or others. To generate seed-thesauri for a certain domain the PoolParty team has developed a method to extract such structures from DBpedia automatically.

PoolParty Enterprise Vocabulary and Metadata Management

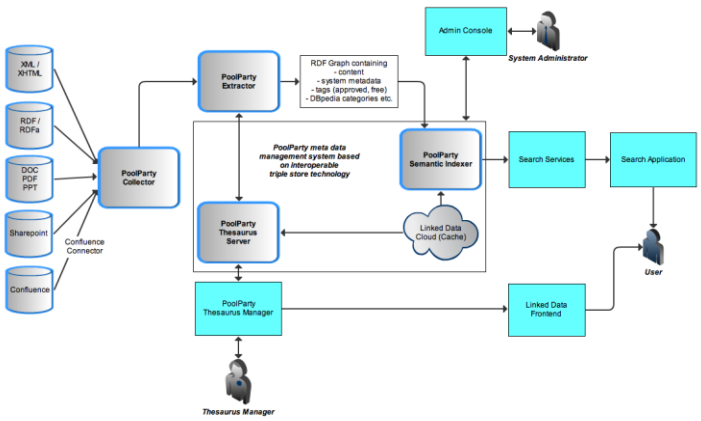
PoolParty is an enterprise ready system, which offers high reliability, usability, performance and mechanisms like failover which guarantees smooth workflows and protection from loss of data. Typical enterprise systems like Linux or Windows servers are supported. A constantly ongoing quality assurance process around the product including high quality documentation

accompanies the overall development of PoolParty. Enterprise Vocabulary and Metadata Management is fully supported and open standards guarantee a high investment security. The integration of PoolParty thesauri with enterprise systems can be realised on top of standard APIs.

Text Mining and Semantic Search

PoolParty offers a variety of options to ease thesaurus management by means of text mining as well as solutions to make semantic search solutions possible. PoolParty can analyse different text formats like HTML, PDF or Word and can detect significant terms within a document either based on existing thesauri or to serve as a new candidate term to further expand a thesaurus. With PoolParty Thesaurus Management document repositories can be indexed and searched in a semantic way out-of-the-box.
The PoolParty product family consists of two other components which together with thesaurus management are the basis for enterprise semantic search solutions.

Vertical Search Solutions: PoolParty Product Family



PoolParty Product Family - Overview

PoolParty product family consists of three components: PoolParty Thesaurus Management (PPTM), PoolParty Extractor (PPX) and PoolParty Semantic Search (PPSS). Combined these elements form the basis for true semantic search and vertical search solutions. PoolParty can index unstructured, semi-structured and structured information and can integrate different sources on top of a semantic thesaurus.

PoolParty Semantic Search is shipped with a full blown Search API which can be used for integration with into existing enterprise platforms. The API supports catgeorised auto-complete, faceted search, full-text search and search assistants which are based on thesauri representing the background knowledge of the domain expert. PPSS can handle millions of documents and is very fast and ready for vertical search applications also in large companies. PPSS can also be used for the development of search assistants typically used within web shops or help desk and call center applications.

Summary

PoolParty Product Family offers a wide variety of options to deploy semantic technologies. The big three topics are: Semantic Search, Thesaurus Management and Linked Data. PoolParty uses in its core Semantic Web technologies which are built around open standards and state-of-the art technologies. Professional metadata management is the key for efficient information management in large organisations and on the web. PoolParty combines Semantic Web, text mining and collaborative knowledge engineering to make applications smarter.

References

[1] http://www.poolparty.biz
[2] http://www.w3.org/standards/semanticweb/
[3] http://www.w3.org/RDF/
[4] http://www.w3.org/2004/02/skos/
[5] http://dublincore.org/
[6] http://xmlns.com/foaf/spec/
[7] https://grips.punkt.at/display/public/POOLDOKU/SPARQL+Endpoint
[8] http://www.w3.org/TR/rdf-sparql-query/
[9] http://www.w3.org/DesignIssues/LinkedData.html
[10] http://richard.cyganiak.de/2007/10/lod/
[11] https://grips.punkt.at/display/public/POOLDOKU/PoolParty%27s+Linked+Data+frontend
[12] http://dbpedia.org/
[13] http://sindice.com/