Introduction

PoolParty Semantic Suite (http://www.poolparty.biz/) is a world-class semantic technology suite that offers sharply focused solutions to transform knowledge organization and content business. As a semantic middleware, PoolParty enriches all kinds of information with valuable metadata and links business and content assets automatically.

PoolParty technology platform consists of several components and can be configured and extended to meet individual requirements. PoolParty Thesaurus Server supports web-based taxonomy and ontology management. It is completely built on top of W3C’s Semantic Web
standards (http://www.w3.org/standards/semanticweb/). In its core, PoolParty uses RDF to represent SKOS and other vocabularies like Dublin Core or FOAF, 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 (http://www.w3.org/TR/rdf-sparql-query/) to execute queries over each thesaurus project. This technology can be used to integrate knowledge graphs with content platforms (Wikis, CMS, etc.) or search engines.

Additionally, PoolParty supports highly scalable and precise entity extraction and other text mining services. Its ability to transform structured and unstructured information into RDF offers new options for data analytics. PoolParty Semantic Integrator, which is the most advanced configuration level of PoolParty Suite, is a solution for data analytics built on top of Linked Data Warehouses with the power of SPARQL engines in its core.

In addition to full support of SPARQL, PoolParty APIs offer ‘traditional means’ to integrate semantics into enterprise information systems and web platforms. Based on JSON REST, developers can make use of all CRUD methods necessary to maintain a taxonomy from within a third-party application like a CMS.

PoolParty integrations have been implemented with content platforms like Drupal, SharePoint, Confluence, Alfresco, or WordPress. As an additional result, guidelines have been developed, which can be reused for other integration projects. PoolParty was also successfully integrated into search engines like Solr, Elasticsearch, Mindbreeze or Intrafind.

The subsequent figure illustrates the most important technical components of PoolParty technology platform:

- As a basic layer, various data formats and sources can be consumed and transformed, either to generate knowledge graphs from it, or as an incoming information stream to get automatically linked to the knowledge graphs in the RDF store.
- PoolParty Semantic Middleware, which is deployed in Apache Tomcat, offers various semantic services, GUIs and APIs. Core services are built on top of the Spring Framework (http://spring.io/). Taxonomists and thesaurus managers take care of the knowledge graphs. Information architects make sure that relevant content sources are properly linked to those, which is a precondition to establish applications and services like semantic indexing on top.
- PoolParty APIs are used by developers to generate smart applications and semantic services like recommender systems, automatic tagging facilities, or semantic search applications. Linked Data Warehouses (Remote RDF Graph Databases) can be attached to the platform to store all resulting RDF graphs to make them available via powerful SPARQL queries and reports.
PoolParty Semantic Suite - Technical Overview
PoolParty Semantic Platform

The technical architecture above provides an overview of the technical components that provide the basis for the PoolParty Semantic platform. From an application perspective the platform can be divided in different functional components that are combined to specific product bundles offering various integration options.
PoolParty Thesaurus Server

Taxonomies and 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 creating, deleting or editing concepts or relations in a knowledge graph. Workflows and approval processes can be activated if desired.

PoolParty Thesaurus Server - Graphical User Interface

The PoolParty graph modelling approach intertwines taxonomy and ontology management in a new way. SKOS taxonomies can be extended by ontologies like schema.org or FOAF. PoolParty users can create their own custom ontologies and schemas by reusing existing ontologies.

By means of text corpus analysis and automatic quality checks, PoolParty supports taxonomists to make sure that resulting thesauri comprise the content base in a meaningful way. Automatic suggestions for the supervised extension of taxonomies are generated through elaborated text mining algorithms.

The rise of Linked Data indicated by the enormous growth of the Linked Open Data cloud is an important argument for many organisations to maintain their own data at least partly based on Linked Data standards. PoolParty makes use of existing Linked Data sources, e.g. concepts can be aligned and enriched with additional information from sources like Dbpedia, Geonames, Wikidata or others. To generate seed-thesauri for a certain domain, DBpedia can be used a source to extract a taxonomy automatically.

© Semantic Web Company 2015

http://www.poolparty.biz/
A wiki frontend for each thesaurus project helps to involve other people than taxonomists in the thesaurus development process. Linking concepts is another flexible way to build thesauri in decentralised structures. Based on the linked data principles, 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.

PoolParty is an enterprise-ready system, which offers high reliability, security, performance and mechanisms like failover, which guarantees smooth workflows and protection from loss of data. The use of open standards guarantee a high investment security. The integration of PoolParty thesauri with enterprise systems can be realised on top of standard APIs.

PoolParty Thesaurus Server is natively built on top of RDF triple stores. Its graph management facilities can be integrated with all graph stores providing a SAIL API (http://rdf4j.org/sesame/2.7/docs/users.docbook?view).

PoolParty Extractor

The PoolParty Extractor analyses documents and text and extracts meaningful phrases, named entities, categories or other metadata automatically with high throughput and accuracy. Different data or metadata schemas can be mapped to a SKOS thesaurus that is used as a unified semantic knowledge model. During this process the extracted entities are linked to the knowledge model (the thesaurus in the PoolParty Thesaurus Server) via URIs that provide a direct way to integration following Semantic Web principles. The PoolParty Extractor is implemented as a pipeline of annotation units where each specific unit adds to the final result. This keeps the system flexible and allows it to be adapted quickly to new requirements.

Advanced linguistic features include classification, corpus statistics and disambiguation. Documents are classified along the structure of a thesaurus what allows the user to flexibly change the classification criteria. Corpora (sets of domain specific documents) are a great way to add background knowledge to text mining process. They provide term frequencies and distributions that improve the scoring of entities and drive the detection of new relevant entities from text. Ambiguity can greatly reduce the precision of entity extraction when identical terms are used to refer to different entities. Such ambiguities can be modeled in PoolParty and improve extraction quality and in the end the experience of the users that interact with the annotation results.
The text mining functionality of the PoolParty extractor is integrated with other systems via a web service API that follows the RESTful principle and produces results in JSON. The API is designed for high throughput. In situations with special requirements in terms of high availability or scalability the system can be operated in clustered mode, too. Out of the box, the system comes with connectors to RDF graph databases that enable easy integration of the results of text mining processes with other RDF data.
PoolParty PowerTagging

PoolParty PowerTagging is a unified integration approach to semantically enrich 3rd party CMS and DMS solutions and provide advanced search features. From the PoolParty Server side it is based on the Thesaurus Server API and the Entity Extractor API and requires the respective integration on the CMS side. Existing integrations are in place for Atlassian Confluence, Drupal, Alfresco and MS SharePoint. The advantages of this approach are manifold:

- **Automatic concept tagging**: Annotate your content and attachments with concepts from your thesaurus and add additional tags if you like.
- **Consistent metadata**: Benefit from consistent tagging by the provision of auto-completion based on controlled vocabularies.
- **Enhanced search**: Extend your CMS’s search capabilities by search facets, precise similarity search, automatic query expansion, sentiment analysis, and trend diagrams
- **Bulk-tagging**: Existing CMS and its whole content base can be tagged automatically at once via bulk-tagging
- **Multilinguality**: Multilingual thesauri and therefore multilingual tagging and search is supported.

PoolParty UnifiedViews

PoolParty UnifiedViews as part of the PoolParty Semantic Integrator provides a framework to develop, execute, monitor, debug, schedule, and share RDF data processing tasks. It can be seen as a Extract-Transform-Load (ETL) framework for RDF, although it doesn’t strictly follow ETL processes as, for example one process can trigger the next one.

Data processing tasks are modelled as pipelines via a graphical interface and can consist of several Data Processing Units (DPUs). PoolParty UnifiedViews comes with a predefined set of DPUs and offers a well documented API to develop Custom DPUs (plugins) on demand.
Pipelines can be scheduled to run on a timely basis or can be triggered by other pipelines. The Execution Monitor gives detailed information on the execution of a pipeline. All data processed is stored in separate graphs and can be reviewed for debugging. The underlying triple store is configurable. Per default, a built-in memory store is used but basically any triple store supporting the SAIL API can be integrated. The Scheduler also includes a notification system that allows to send information about the outcome of scheduled data processing tasks.

PoolParty UnifiedViews is based on the open source UnifiedViews project as SWC is a member of the core development team (see http://unifiedviews.eu). PoolParty UnifiedViews is a SWC-maintained and supported build of UnifiedViews including DPUs that allow easy integration with the other PoolParty Semantic Integrator server components.
PoolParty Graph Search Server

The usual workflow of a Graph Search project starts with the gathering of structured and unstructured data from various sources by using Unified Views and/or by transforming documents into RDF by means of PoolParty Extractor.

The processed RDF data is stored in triple stores, optimally organized in named graphs to provide a separation of data. By this, you can aggregate and manage large volumes of information like DBpedia, WordNet, Geonames etc. to integrate those into your search and analytics applications.

A typical semantic search application based on PoolParty Graph Search Server uses special data structures in the data that has to be searched in combination with optimized SPARQL queries. By this you can generate a combination of ‘traditional’ document search applications with additional beneficial features like synonym search and hierarchical drill down based on the knowledge graph that is managed with PoolParty Thesaurus Server.

Since all relevant data can be acquired from the triple store, interactive visualizations or other forms of analytics, like reports can be built using SPARQL queries. Customizing integrated linked knowledge graphs and adapting SPARQL queries allows you to adapt and modify your analysis applications in a very dynamic and agile way. Like you do mashups, you can combine different data sets and formulate queries to retrieve data accordingly to your needs.
Integrating PoolParty

Its APIs make PoolParty ready to be integrated into any Enterprise Information System:

- **JSON RESTful**: Semantic technologies ready to be integrated based on popular web technologies
- **CRUD**: Create, read, update, and delete – full-blown API for all kinds of interactions with taxonomies and knowledge graphs
- **Secure**: PoolParty API is fully integrated into PoolParty’s security layer based on Spring
- **SPARQL endpoint**: In addition to the standard API, PoolParty’s SPARQL endpoint is used to execute more complex queries and integrate data in a highly flexible manner

PoolParty Semantic Integrator - High-level architecture

Summary

PoolParty Product Family offers a wide variety of options to benefit from semantic technologies. The major topics covered are: Semantic search, taxonomy and ontology management, data integration 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 methodologies from the Semantic Web with text mining algorithms and approaches for collaborative knowledge engineering to make applications smarter and to improve user experience.

For more information on PoolParty Semantic Suite, please visit our homepage at [www.poolparty.biz](http://www.poolparty.biz) and contact our Consultant Team:

PoolParty Consultant Team