

# Release Notes - PoolParty 5.2

Here you can find the release notes for PoolParty release 5.2. In the release notes you can find major developments, improvements and changes made in the respective release. The release notes are divided in three chapters:

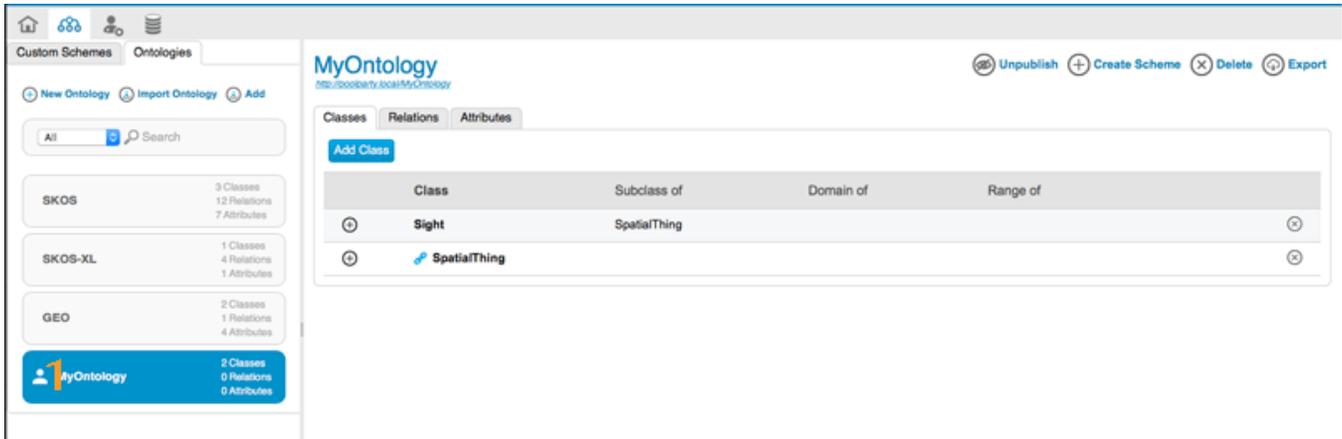
- [Highlights](#)  
Newly developed functions or major improvements of existing functions.
- [Improvements](#)  
Minor improvements of existing functions.
- [Fixes](#)  
Fixes for problems found in previous releases of PoolParty.

## Highlights

### PoolParty Thesaurus Server (PPT)

#### Custom Ontology/Scheme Management

PoolParty's Custom Scheme management capabilities have been extended providing now a clear distinction between ontologies and custom schemes. Ontologies can still be added from a list of predefined ontologies. In addition, custom ontologies can be created allowing to define classes, relations and attributes that are not covered by the predefined set of ontologies. Custom ontologies are visually distinguished from predefined ontologies by an icon in the (1) ontology selection on the left.



#### Custom Ontologies

Custom schemes allow to create 'subsets' or 'views' on the available ontologies by mashing and reusing classes, relations, and attributes from available ontologies.



#### Custom schemes

In addition, an import export functionality for ontologies and schemes has been added, allowing to export to the SWC RDF format or to OWL in all relevant serialization formats and to import ontologies and schemes that have been exported from another PoolParty server.

#### Improved Corpus Management Workflow & Cooccurrences

The **Corpus Management** workflow has been redesigned to make integration of new terms based on corpus analysis as easy as possible. The tree view in **Corpus Management** now also provides a view on the thesaurus so no switching between the two views is necessary anymore.

The screenshot shows the 'Social Semantic Web' interface. On the left is a 'Thesaurus' tree with categories like 'Data resources', 'Information Sources', 'Organisations', etc. The 'Corpora' section is expanded to show 'Social Semantic Web'. The main panel displays 'Corpus Analysis Summary' with a status message: 'Corpus Analysis has been performed on "04.11.2015 - 20:47"'. Below this is a table of statistics:

Concepts	1,009
Extracted Terms	511,469
Concept occurrences	84,887
Suggested Term occurrences	1,135,773

To the right is a 'Corpus Summary' table:

Stored Documents	1160
Overall Filesize	6.98 MB
Language	en
Created by	nagyhei
Created	04.11.2015 - 20:34
Last Modified	04.11.2015 - 20:42

### Thesaurus tree in corpus management

Candidate Terms have become **Candidate Concepts** and are stored per project and not per corpus anymore. **Candidate Concepts** are still part of the tree and each candidate concept has become its own resource with a node below of the **Candidate Concepts** node. **Extracted Concepts** and **Extracted Terms** list have become tabs in the corpus details view. Candidate concepts are created by double-click on an extracted term or by selecting several terms and adding them via the respective button. Of course you can also add candidate concepts from terms provided viewing a document in corpus management.

The screenshot shows the 'Social Semantic Web' interface with the 'Extracted Terms' view active. The left 'Thesaurus' tree shows 'Candidate Concepts' expanded to include 'natural language', 'working group', 'ontology', 'subclass', 'classification', and 'berkeley db'. The main panel displays a search bar and a table of extracted terms:

Term	Relevance	MIS	CTS	Frequency	Similar Terms	Blacklist
search	416.8	0	13.786	3047	Show	<input checked="" type="checkbox"/>
document	347.99	0	13.32	2275	Show	<input checked="" type="checkbox"/>
xml document	324	94.96	8.877	136	Show	<input checked="" type="checkbox"/>

### Candidate Concepts

Co-occurrences are calculated for candidate concepts and for thesaurus concepts. When you select a candidate concept, the calculated co-occurring terms are displayed on the top. They can be turned into alternative or hidden labels by drag and drop or turned into candidate concepts by double-click. Other candidate concepts can be related as narrower or related concepts or merged as alternative or hidden labels. Thesaurus concepts can be selected as broader concepts to connect the candidate concepts to your thesaurus project. A green light indicates that a candidate concept is ready for integration, clicking the respective tab on top will turn all those candidate concepts into concepts of your thesaurus project.

The screenshot shows the 'classification' candidate concept details. The left 'Thesaurus' tree shows 'Candidate Concepts' expanded to include 'natural language', 'working group', 'ontology', 'subclass', 'classification', 'berkeley db', 'internet', and 'universal decimal classification'. The main panel displays 'classification' details with suggestions and integration options:

**Suggestions:** decimal, soil, system, scheme, facet, library classification, decimal classification, dewey decimal, climate, chinese, film, subject, climate classification, congress, classify, dewey decimal classification, library of congress classification, soil classification, universal, board.

**Broader Concepts:** 08.01.01.General.concepts (Thesaurus Concept)

**Narrower Concepts:** universal.decimal.classification

**Related Concepts:** [checked]

**Preferred Label:** classification

**Alternative Labels:** classification system

**Hidden Labels:** [empty]

### Integrate candidate concepts

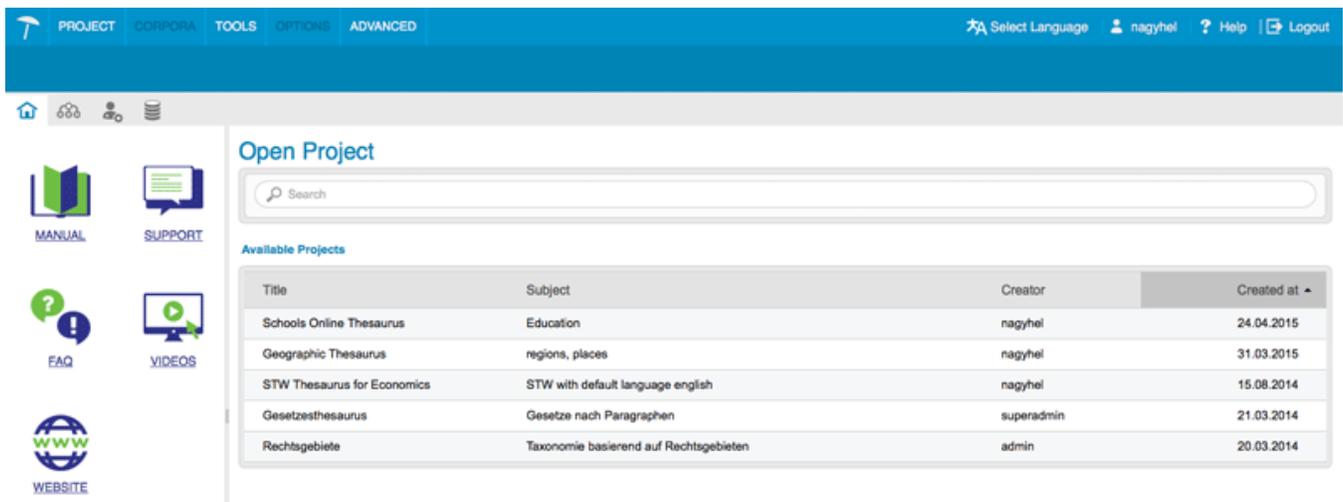
## New Login Page and Home Cockpit

The PoolParty Login Screen has been redesigned and displays now in addition the PoolParty news feed.



### New login page

When you login the new **Home** cockpit is displayed offering all projects available for the respective user on the server sorted by creation date. You can double click a project to open it.



### Home dashboard

## PoolParty Extractor (PPX)

### Configuration of stop word lists and lemmatisation files

For each language stop words and lemmatisation lists need to be specified. Those affect the quality of term extraction and the recall rate in concept extraction. PoolParty extractor ships with defaults for major European languages. It is now possible to change these defaults (i.e. use a different set of stop words) and also to configure this information for languages where it is not present.

### Lemmatization of extracted terms

Free terms (in the PPX extract call of the API) and extracted terms (in the corpus analysis) are now lemmatised to unite different forms of the same term. Now, if terms are extracted that are just variants (e.g. car vs. cars) then they are treated as the same term and also the counting and scoring takes this into account. This improves the output of the extract service by reducing redundancy and reduces work in the corpus analysis because the same term has not to be treated multiple times.

## Synchronise word forms with corpus

To detect variations of concepts in text, PoolParty extends the names of concepts as they are written in the thesaurus with word forms for each of the words. This allows us to detect for example "school children" when there is a concept name "school child". For languages like German or Spanish this increases detection performance quite considerably. Now PoolParty has been improved to also handle highly variable languages like Russian well. The generated word forms are checked against a reference corpus and only the relevant forms are added to the extraction model keeping it performant and light-weight.

## Improvements

### PPT

#### New PPT API services

The PPT API has been extended to cover all functionalities and methods we found missing in projects and that have been requested by customers over the last releases.

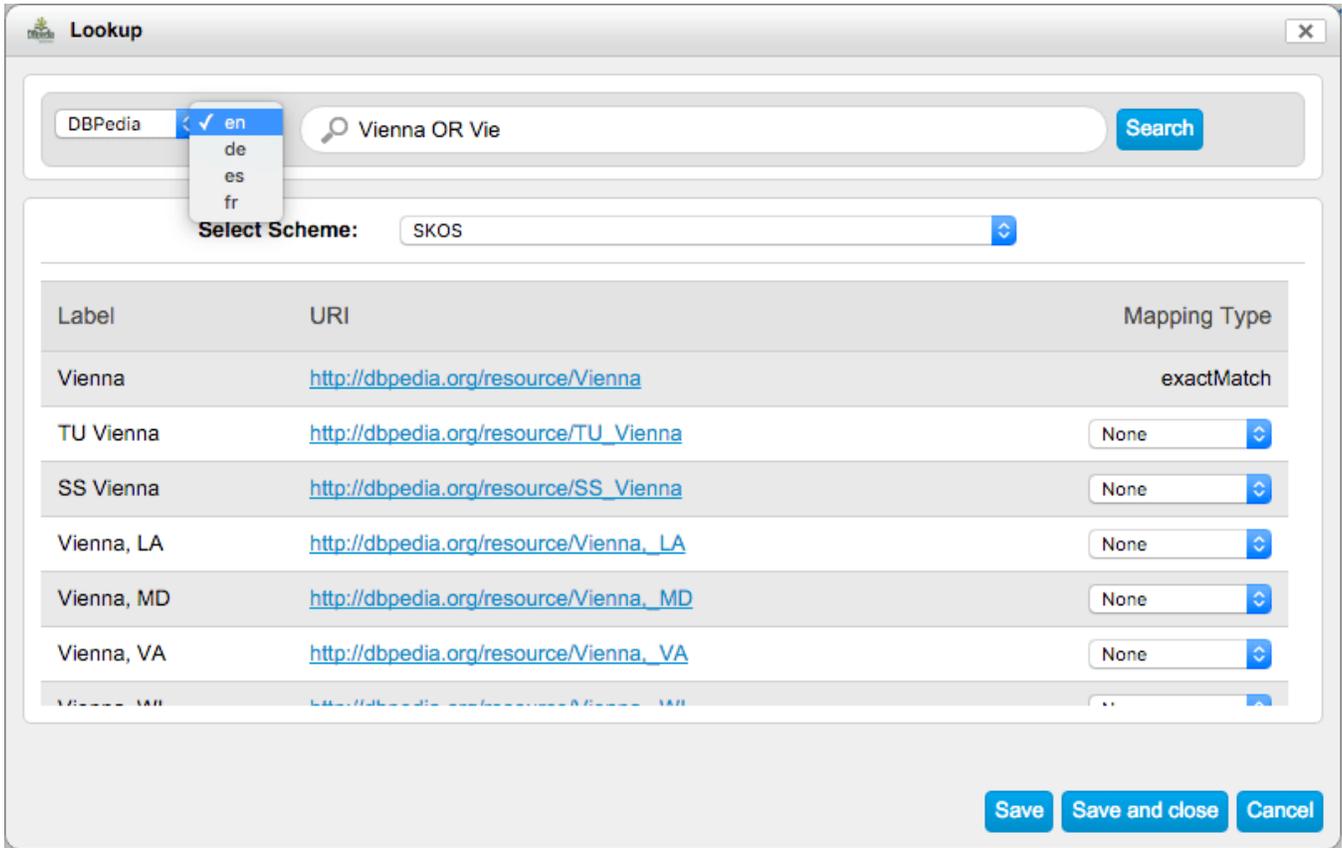
- **Thesaurus Service**
  - **subtree:** Returns the subtree of all narrower concepts with the provided concept as root.
- **Corpus Service**
  - The endpoint changed: /PoolParty/api/corpus/ -> /PoolParty/api/corpusmanagement/
  - **blacklist:** concepts and free terms can be blacklisted via API
  - **Corpus management:**
    - Get corpus list: get the list of corpora of a specified project.
    - Get document list for corpus: list of corpus documents, with the possibility to filter by time range.
    - Delete document in corpus: allows you to delete documents from the corpus.
    - Get metadata of corpus: retrieve metadata information for a given corpus like number of documents, last modification date, language, number of concepts, quality, ...
    - Get detailed results of corpus analysis: collect the list of matching concepts and free terms and details like URI, frequency, hierarchical information for concepts and relevancy values for free terms.
  - **Approval workflow:**
    - The approval workflow can be accessed by assigning concepts to users and for approval/rejection of changes. In every case you can leave notes to make the decision clear for other users.
  - **Concept suggestions:**
    - suggestConcept is the replacement for the outdated method suggestFreeConcept.
    - Collect details of all suggested concepts of a project or of a specific suggested concept.
    - Find out the current state of a concept: SUGGESTED, REGULAR, MERGED, DELETED
  - **History:**
    - Retrieval of history events for a specified time frame for a whole project or for specific concepts.
  - **Extraction model:**
    - A new up-to-date check: allows to check if extraction model of a project is up-to-date.
  - **Notes:**
    - Create and retrieve notes assigned to a concept
  - **Collections:**
    - Retrieve all collections of a project
    - or retrieve all members of a collection
    - Create, delete and update collections by adding and deleting members
  - **SKOS-XL support:**
    - All necessary services that allow the management of SKOS-XL labels in a PoolParty project
  - **Generic RDF services:**
    - Collect all properties of a given resource (concept, collection, ..)
    - Retrieve values for a given property of a given resource.
- **Improvements**
  - All Thesaurus basic services that respond concept details, like getConcept, getChildren, getTopConcepts, etc. now provide more details when requested:
    - Possible properties are extended by: 'skosxl:prefLabel', 'skosxl:altLabel', 'skosxl:hiddenLabel', 'skos:exactMatch', 'skos:closeMatch', 'skos:broaderMatch', 'skos:narrowerMatch', 'skos:relatedMatch' or 'all' to fetch all properties.
    - The workflow status can be retrieved for the specific concept.
- **Deprecated API service**
  - suggestFreeConcept is replaced by suggestConcept (see above "Concept suggestions")



The API will probably never be complete but we are happy to get your suggestions for missing features and will always point you to the SPARQL endpoint that can do whatever the API provides and also whatever the API does not provide.

## Improved LOD linking

The LOD linking has been improved and we reviewed all LOD sources by removing those, which we found that they are not frequently used (Umbel, Yago, DMOZ). We've added new ones (Wikidata) and the DBpedia lookups have been made multilingual: German, French and Spanish lookups have been added. DBpedia resource and DBpedia categories lookup have been combined and the language can be selected when linking is done.

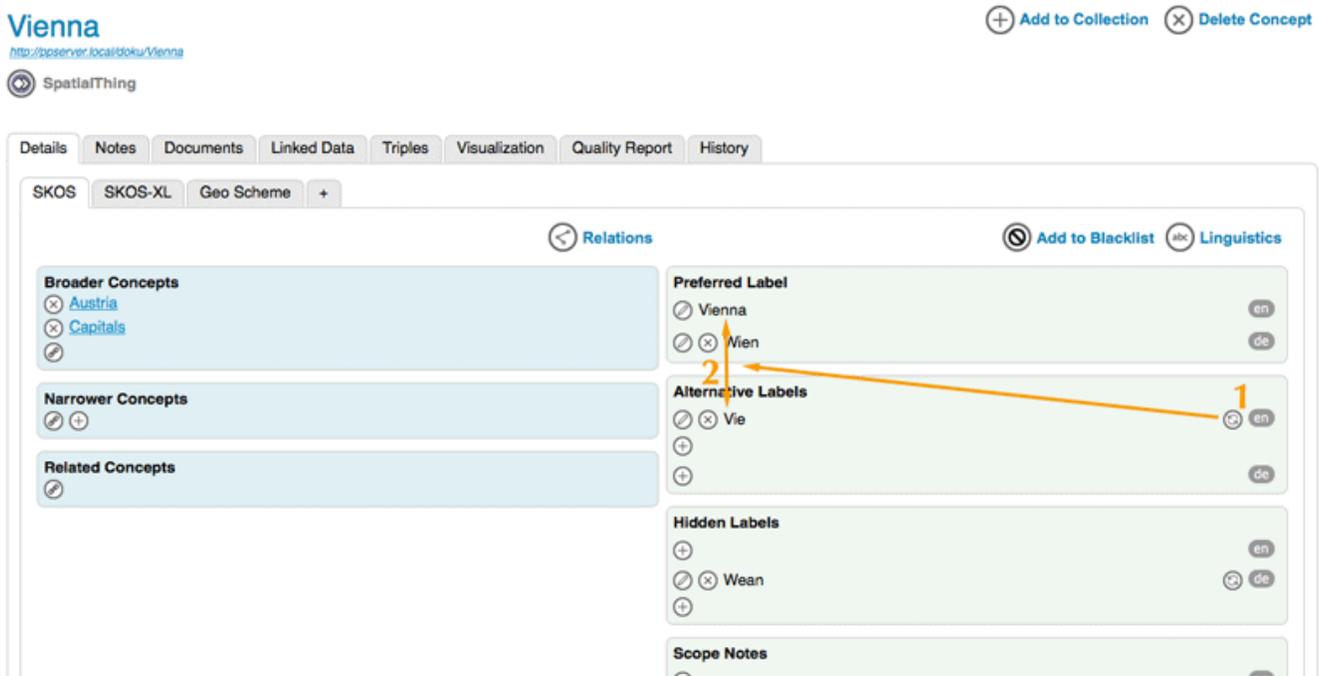


DBpedia lookup with language selection

In addition the handling of redirect links has been improved

### Change Preferred, Alternative and Hidden Label

You can now change the preferred with an alternative or hidden label simply clicking the respective icon (1) in the details tab of the concept.



## Change alternative/preferred label

### Additional Improvements

- The **Quick Open** menu entry has been renamed to **Recently Opened Projects**, now showing the ten projects that have been last modified.
- The **Menu** structure has been streamlined.
- You can now define if custom schemes or custom ontologies are published.
- The reasoning implementation has been changed to use SPARQL based reasoning. This improvement provides significant performance improvements.
- The base URLs for custom schemes and ontologies and for projects can now be set via properties.
- Creating relations in the EXCEL import can now also be done providing the URL of the respective concept.
- Projects are now sorted by creation date in the **Open Project** dialogue.
- Corpus analysis has become a long running tasks and now can be canceled in addition the document name is displayed in the log if a document cannot be uploaded.
- The blacklist is now available per default without having to create a corpus first.
- Categorization of concepts for auto population allows now to multi-select categories.
- Improved snapshot behavior validating project integrity.

### Fixes

- Displaying dates in the Snapshot Dashboard works now correctly in IE and Safari
- Fix problem in Excel import when importing date information