NeOn: Lifecycle Support for Networked Ontologies

Integrated Project (IST-2005-027595)

Priority: IST-2004-2.4.7 – "Semantic-based knowledge and content systems"

D3.7.2 Advanced context-sensitive visualization

Deliverable Co-ordinator: Dunja Mladenić

Deliverable Co-ordinating Institution: J. Stefan Institute (JSI)

Other Authors: Boštjan Pajntar (JSI), Blaž Fortuna (JSI), Luka Bradeško (JSI), Dunja Mladenić (JSI), Martin Dzbor (OU)

This deliverable provides a software prototype that enables context sensitive visualization of ontologies. The idea is to enable the user to gain a broad overview and abstract insight into the mesh of networked ontologies by projecting an ontology onto different landscapes and projecting different ontologies onto the same landscape. The software prototype also enables the generation of different landscapes based on a document collection and personalized styling of landscapes. Furthermore the generated landscape can be stored for later usage by other users. The system enables loading of arbitrary ontologies as well as predefined landscapes and the visualization of the ontology onto the landscape.

This is one of the three core technologies for contextualization. It focuses on providing means for visualizing any ontology in the context of any general background knowledge encoded in textual form.

Document Identifier:	NEON/2009/D3.7.2/v1.1	Date due:	February 29, 2009
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	September 15, 2009
Project start date:	March 1, 2006	Version:	V1.1
Project duration:	4 years	State:	Final
		Distribution:	Public

NeOn Consortium

This document is a part of the NeOn research project funded by the IST Programme of the Commission of the European Communities by the grant number IST-2005-027595. The following partners are involved in the project:

Open University (OU) – Coordinator Knowledge Media Institute – KMi Berrill Building, Walton Hall Milton Keynes, MK7 6AA United Kingdom Contact person: Martin Dzbor, Enrico Motta E-mail address: {m.dzbor, e.motta} @open.ac.uk	**Universität Karlsruhe – TH (UKARL)** Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Englerstrasse 11 D-76128 Karlsruhe, Germany Contact person: Peter Haase E-mail address: pha@aifb.uni-karlsruhe.de
Universidad Politécnica de Madrid (UPM) Campus de Montegancedo 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez E-mail address: asun@fi.upm.es	**Software AG (SAG)** Uhlandstrasse 12 64297 Darmstadt Germany Contact person: Walter Waterfeld E-mail address: walter.waterfeld@softwareag.com
Intelligent Software Components S.A. (ISOCO) Calle de Pedro de Valdivia 10 28006 Madrid Spain Contact person: Jesús Contreras E-mail address: jcontreras@isoco.com	**Institut 'Jožef Stefan' (JSI)** Jamova 39 SI-1000 Ljubljana Slovenia Contact person: Marko Grobelnik E-mail address: marko.grobelnik@ijs.si
Institut National de Recherche en Informatique et en Automatique (INRIA) ZIRST – 655 avenue de l'Europe Montbonnot Saint Martin 38334 Saint-Ismier France Contact person: Jérôme Euzenat E-mail address: jerome.euzenat@inrialpes.fr	**University of Sheffield (USFD)** Dept. of Computer Science Regent Court 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk
Universität Koblenz-Landau (UKO-LD) Universitätsstrasse 1 56070 Koblenz Germany Contact person: Steffen Staab E-mail address: staab@uni-koblenz.de	**Consiglio Nazionale delle Ricerche (CNR)** Institute of cognitive sciences and technologies Via S. Martino della Battaglia, 44 - 00185 Roma-Lazio, Italy Contact person: Aldo Gangemi E-mail address: aldo.gangemi@istc.cnr.it
Ontoprise GmbH. (ONTO) Amalienbadstr. 36 (Raumfabrik 29) 76227 Karlsruhe Germany Contact person: Jürgen Angele E-mail address: angele@ontoprise.de	**Food and Agriculture Organization of the United Nations (FAO)** Viale delle Terme di Caracalla 1 00100 Rome Italy Contact person: Marta Iglesias E-mail address: marta.iglesias@fao.org
Atos Origin S.A. (ATOS) Calle de Albarracín, 25 28037 Madrid Spain Contact person: Tomás Pariente Lobo E-mail address: tomas.parientelobo@atosorigin.com	**Laboratorios KIN, S.A. (KIN)** C/Ciudad de Granada, 123 08018 Barcelona Spain Contact person: Antonio López E-mail address: alopez@kin.es

NeOn

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts: JSI, OU.

Change Log

Version	Date	Amended by	Changes
0.1	20-01-2009	Dunja Mladenic	Overall structure of the report
0.2	10-02-2009	Dunja Mladenic	Approach description, summary
0.3	25-02-2009	Boštjan Pajntar	User documentation
0.3	05-03-2009	Blaz Fortuna, Luka Bradeško	Example usage
0.4	17-03-2008	Boštjan Pajntar	Read-me file
0.6	28-03-2009	Dunja Mladenić	Finalization
1.0	01-04-2009	Boštjan Pajntar	Overall refinements (internal review)
1.1	26-08-2009	Boštjan Pajntar	Revision due to requests of the review

Executive Summary

This report is describing a software deliverable developed as NeOn toolkit plug-in based on NeOn deliverable D3.7.1. The main idea implemented in the presented plug-in is context sensitive visualization of ontologies is the existence of a landscape (context) defined by the domain of interest onto which we can project ontologies. In that process we abstract from the structure of the ontology and focus on its content and coverage of different domain content as provided in the landscape. One can imagine that having a landscape that covers a rather broad topic area enables comparison of general as well as specific ontologies; the former covering a whole landscape and the latter covering just a part of it.

The idea of using landscapes as context for ontology visualization is implemented in two stages. First, the user selects the desired landscape or generates a new one, thus representing the domain of interest. The landscape is generated based on a collection of documents describing the landscape topic and stored in a library of predefined landscapes. The developed plug-in for context sensitive visualization enables loading of a predefined landscape and an arbitrary ontology. The system then automatically projects the ontology onto the landscape and enables browsing through the visual representation of the projection.

New landscapes can be generated from any collection of textual documents. The choices for such corpus can stem from many sources: it can be documents bound to an ontology, that will have to be annotated with it, or they can even be a low quality documents, for example results returned by a search engine for a suitable query.

This is one of the three core contextualizing technologies stemming from WP3. It provides means of using various non-structured, general background knowledge, to serve as a context for visualizing any ontology. The other two core contextualizing technologies are complementary: the context can be provided by one networked ontology for the other. This was implemented in OntoConto (D3.2.2, D4.5.2), consuming Alignment Server (D3.3.1, D3.3.2) and second, an ontology can provide means for the contextualization of general knowledge. This will be implemented in SearchPoint (D3.2.4, D4.5.3)

The software is implemented in Java as a Neon Toolkit plug-in. It covers the main functionalities proposed in D3.7.1.

The report contains basic documentation on the approach description and user support documentation. The software can be downloaded from NeOn website.

Table of Contents

List of Figures

1. Introduction

Many existing tools for handling ontologies provide reasonable editing facilities and some visualization/navigation support. However, the existing user interfacing support for visualization and navigation of large ontologies is far from ideal. Several drawbacks are pointed out in D4.5.4 including the following issues:

- facets like complex network graphs and the lengthy trees starting at the abstract levels as the primary visual elements, were generally found to be a poor metaphor for user needs, and only tended to add to the already steep learning curve for a non-expert user

- ontology visualization tools are often too complex and do not reflect users' models of what they would normally want to see in unfamiliar ontologies

- non-experts found themselves less efficient due to the lack of simple visualization and navigation support compared with experts, especially in terms of obtaining overviews of provided ontologies.

- the existing tree-based visualizations of ontology hierarchies in particular were criticized

The main idea implemented in the presented plug-in is context sensitive visualization of ontologies is the existence of a landscape (context) defined by the domain of interest onto which we can project ontologies. Visualization of ontology as proposed in D3.7.1 and implemented in here as NeOn toolkit plug-in is by visualizing the content covered by ontology while abstracting the ontology structure. Related deliverable in WP4 (D4.5.4) proposes visualization of ontologies by taking the ontology structure into account. Content of ontologies is represented by names of concepts and relations as well as instances. Here we focus on using the textual part of ontology representation and providing a visualization of that text. Moreover, not only the ontology is abstracted to textual data but also the context that is used to provide the context sensitive visualization. We have introduced term landscape in D3.7.1 for context in which the ontology is visualized. Landscapes are either pre-calculated, based on a document collection that captures content of the landscape that we want to use, or they can be generated by the user from any textual corpus from within the plugin.

The rest of this deliverable briefly describes the approach, example usage of the plug-in and provides user support documentation.

2. Approach Description

As proposed in D3.7.1, the main idea behind the proposed approach to context sensitive visualization of ontologies is the existence of a landscape (context) defined by the domain of interest onto which we can project ontologies. In that process we abstract from the structure of the ontology and focus on its content. Visualization is based on showing how the ontology content covers a domain of interest as captured in the landscape.

Landscapes are generated from a document collection describing the domain of interest. All the landscapes are stored in a library of predefined landscapes that can be shared by different users. Both landscape and ontology are loaded in NeOn Toolkit and the plug-in presented here automatically projects the ontology onto the landscape and enables browsing through the visual representation of the projection.

2.1 Landscape generation

The landscapes are generated by mapping all the documents into a two dimensional vector space so we can plot them on a computer screen. Ideally they would be positioned in such a way that the distance between two documents would correspond to the content similarity between them. The approach is based on document collection visualization approach proposed in (Fortuna et al. 2006) where combining the two methods (linear subspace and multidimensional scaling) is applied for visualization of document collection.

First textual corpus that will serve as a context landscape must be selected by the user. This can be done in several ways. Often the unstructured corpus exists prior to the ontology. The ontology is created in order to classify or annotate the corpus. Such a corpus is of course the most obvious choice for use a context landscape, as it will clearly show which topics are well modeled by the ontology and even by which concepts. Sometimes however, ontology needs to be studied in a different context than the one that it was created in. If there is no preexisting corpus for the new context, any textual documents will do. Even snippets returned by any internet search engine for an appropriate query give good results for a new landscape. After textual documents are manually collected, they are then automatically first represented as vectors in a high-dimensional vector space, where each dimension corresponds to one word from the vocabulary (Salton 1991). Then Latent Semantic Indexing (Deerwester et al. 1990) is used to transform the vectors into lower dimensional space (in our case 100 dimensions are used). Vectors that form the basis can be also viewed as concepts and the space spanned by these vectors is called the *Semantic Space*. Then Multidimensional scaling (Carroll and Arabie 1980) is applied to enables dimensionality reduction by mapping original multidimensional vectors onto two dimensions. What follows is a description of that algorithm:

2.2 Projecting ontology onto semantic landscape

The procedure defined in the previous sections can be used to make a map of topics from a given collection of documents. The map is calculated for each document collection separately with no connection to the other document collections. In our case we would like to visualize a given ontology on a predefined set of topics, e.g. for comparing two ontologies on a predefined set of topics. To handle such scenarios we developed a notation of *semantic landscape* which defines the set of topics over which the ontology will be visualized. Essentially it is a pre-given set of documents, which talk about the desired topics (e.g. biology, fishing industry, pharmaceutical products) projected onto a two dimensional map. The documents which define the semantic landscape are called *landmarks*.

Notice that the landmarks do not appear in the visualization, they are just used for positioning the given collection of documents on the map as follows. In order to position a document on a map, the top N most similar landmarks are calculated and the position of the document is then a convex combination of these landmarks' positions. More specifically, let μ_i be the cosine similarity of the document to the i-th most similar landmark and A_i be the location of the landmark on the 2D map. Than the location of the document is calculated as

$$\left(\frac{1}{\sum_{i=1}^{N} \mu_i} \right) \sum_{i=1}^{N} \mu_i A_i. \tag{1}$$

The positioning is computationally very inexpensive which in turn means that once the semantic landscape is calculated the visualization of new documents can be done in near real time.

The semantic landscape can be also be used for a more efficient visualization of larger corpora, when the multidimensional scaling (MDS) algorithm would take too long to finish. In such cases a semantic landscape is generated from landmarks, which are extracted from the corpora. In our experiments we used a k-means (Jain and Flynn 1999) clustering algorithm for clustering the corpora in a larger number of clusters, usually several hundreds. Since k-means is more efficient than MDS, this significantly improves the time performance.

2.3 Visualization enhancements

Several additions, besides the positions of ontology concepts, can be added to the visualization to increase the amount of information the user can get from it at a fast glance or while exploring the map. First of all, the background can be used to depict, for example, the density of landscape documents or ontology documents in a particular part of the map. Second, the content of the landscape documents can be used for generating and displaying the keywords which are relevant for a particular part of the map.

The background depicting the density can be either generated based on the documents which were provided for the background landscape or the ontology's concepts. The density is estimated as follows. Each point from the landscape is assigned height using the formula

$$h(x, y) = \sum_i \exp(-\sigma \, ||(x,y) - (x_i, y_i)||^2), \tag{2}$$

where σ defines how wide is the influence of one point.

Similar to density, a set of most important keywords can be assigned to each point of the map. The keywords are selected by averaging the TFIDF vectors of documents which appear within a predefined distance to the point.

Apart from this, several additions have been made to the user interface to enhance the user experience. It is possible to zoom in to a certain region, by either double clicking or drawing a square. Apart from more distinct visibility of the zoomed in region, all of the currently visible documents are listed in a separate view. This helps the user with the annotation of the documents, if this is the task at hand.

To boost the navigation, the user can click on any concept to navigate to NTK view where it is possible to edit that concept. This enables the user to visually explore an ontology in different contexts, with the possibility to easily shift into an editor mode when irregularities are found.

3. Usage of the system

The usage of the Document Atlas NeOn Toolkit plugin is straightforward and simple. In this section we first show how to install the plugin and describe plug-in's functionalities. Next we present some possible scenarios of use and finally we showcase the scenario on the screenshots as would be performed by the user.

3.1 Installation and basic description of the functionalities

Both binary and source version are available at http://kameleon.ijs.si/documentatlas/. To install the binary version it is first necessary to download the archive with a prefix »BIN_« and extract it to the plugin folder of a working NeOn Toolkit installation. NeOn Toolkit must be restarted in order to recognize the plugin.

The utility for automatic landscape generation is part of the plugin. The input of this utility is a set of textual documents. The documents have to be given as a text files, stored inside a single folder. Output of the command line utility is an .xml file automatically storing the landscape generated for visualization and future reuse. This generated file is placed to the "data" directory inside the plugin which enables easy distribution and sharing of custom landscapes.

There are three ways how to start the plug-in. The simplest way is to invoke the action button under the main menu. This button can be found by its alternate text »Visualize ontology in context« and is showed on Figure 1. The second way is to select the view from the menu »Visualization« and the third is by opening DocumentAtlas View from the default menu location. This can be done by selecting Window -> Show View -> Other...-> Visualization -> DocumentAtlas View. In all the three cases the Document Atlas View will open in the main area of the screen.

The graphical user interface that takes care of the navigation is positioned at the top of the view (Figure 2). On the left it consist of two select boxes and a button »Refresh ontologies«. First select box selects between the ontologies stored in the NeOn Toolkit and second is for selecting a semantic landscape. On the right there is a button for creating new landscapes. We give a short description for each of the UI components.

The button »Refresh ontologies« is used to reload all the ontology data that are stored in the NeOn toolkit. If the NeOn Toolkit starts with DocumentAtlas view open, no ontology data is loaded, since the toolkit needs more time to load all the ontologies than to load the view. When the ontologies get loaded, it is easy to forward them to the DocumentAtlas plugin by clicking on this button. This button can also be used when there has been any change with the ontologies (editing, importing, deleting).

The left select box is for selecting an ontology that will be visualized. If this box is empty (on NeOn Toolkit start), press refresh ontologies. After that, selecting ontology from this box sends a request to visualize it. After ontology is visualized, the user can visualize another ontology by another selection in this selectbox. What actually happens between selecting an ontology and visualizing it on the screen is hidden from the user. We explain it here in order to give a better understanding of the functionality. First, each ontology concept is transformed into a high dimensional space (created from textual annotations and the underlying landscape) and this representation of the concept is consequentially used for calculating the most similar position on the semantic landscape, where the concept will finally be visualized. Also SubConceptOf relations are visualized between the corresponding concepts in order to retain the sense of the core ontology structure.

The right select box is used in order to select the pre-calculated semantic landscapes. These landscapes are stored in the "data" folder of the plugin in an xml format. To add a new landscape it is enough to add the file to the folder and restart the DocumentAtlas plugin. Upon selecting a new landscape the process of calculating the position of each ontology concept described in the

previous paragraph is restarted. To choose another landscape for the visualization it needs to be again selected from the box.

The button for creating new landscapes brings up a folder selection dialog (Figure 3). After the folder with textual documents gets selected, the new landscape is calculated and will be available in the landscape select box from now on.

3.2 Discussion on use of the Document Atlas plug-in

Here we would like to discuss on the three basic ways the Document atlas can be used and the rationale for each of these possibilities. The big picture as explained before is always the same - an ontology at hand is put into the context of the landscape. However, three different scenarios of use can benefit from this.

The first possibility is that of an ontology creator/editor (ontology engineer) working on an ontology. Usually the creator/editor is modelling a concrete task which usually reflects on the ontology engineering. Since the creation of the ontology is bound to the concrete task, this can be viewed as a specific context in which the ontology is growing and can be very unbalanced in another context in which the ontology will be later used. For example we can imagine a marine biologist engineering an ontology of fish populations and marine life. It is highly probable that this ontology will later be used in the fishing context, therefore it is highly beneficial for the engineer to view the evolving ontology visualized on the correct (fishing) landscape, which immediately reveals, which parts of the semantic space is already covered by the evolving ontology and which semantic parts from the context (not necessarily known to the marine biologist) could or should also be modelled with the ontology.

The second possibility of use is that of a user unfamiliar with an ontology trying to understand it. By visualizing the ontology inside a context or several contexts that are familiar, it is easy to quickly position the domain the ontology models (by visualizing it on a landscape generated from general high level domains) and further to profile the ontology parts (by using a more specific landscape). Even more important than this in a scenario where a certain unknown ontology is designated to be used for annotation of a domain, Document Atlas is a great tool by first providing means to generate a semantic landscape out of domain documents (instances) that need to be annotated, and even more so, for giving a good overview of what instances (documents) should be annotated by what concepts in the ontology.

The third possibility is that of choosing the right ontology for a given domain. First the landscape is generated out of the domain. On top of this landscape ontology candidates can be visualized one by one and furthermore, studied in a visual way to see which candidates have a good architecture for the given domain and also how good is the coverage of the domain (i.e. how much the ontology will have to be edited).

3.3 Showcase

In this section we present DocumentAtlas plug-in with step by step actions the user must perform to harness its full potential. The workflow is accompanied with screenshots of the key moments in the process.

The scenario we will mimic is that of an ontology engineer adapting an existent ontology that needs to be applied on a new domain not familiar to the engineer.

After the NeOn toolkit is started, the engineer opens DocumentAtlas view by clicking on the action button (Figure 1). Once the view is opened, the user first selects the landscape from the landscape selection box.

In our case the user would like to study his ontology in the context of Money, Banking and Finances for which a landscape is already prepared and distributed together with the DocumentAtlas plug-in. This landscape is generated from an extensive set of financial news freely published by Yahoo Finance.

Upon selecting the desired landscape, the landscape gets immediately visualized on the screen (Figure 4) and offers the user an opportunity to study the domain, its topics and relations (similarity) between them.

The user is now ready to visualize the ontology he is working on in the context of this landscape. He selects the ontology from the select box and after a short while, this ontology gets visualized on top of the semantic landscape. Several observations can be made from studying the ontology in the context of the selected landscape (Figure 5).
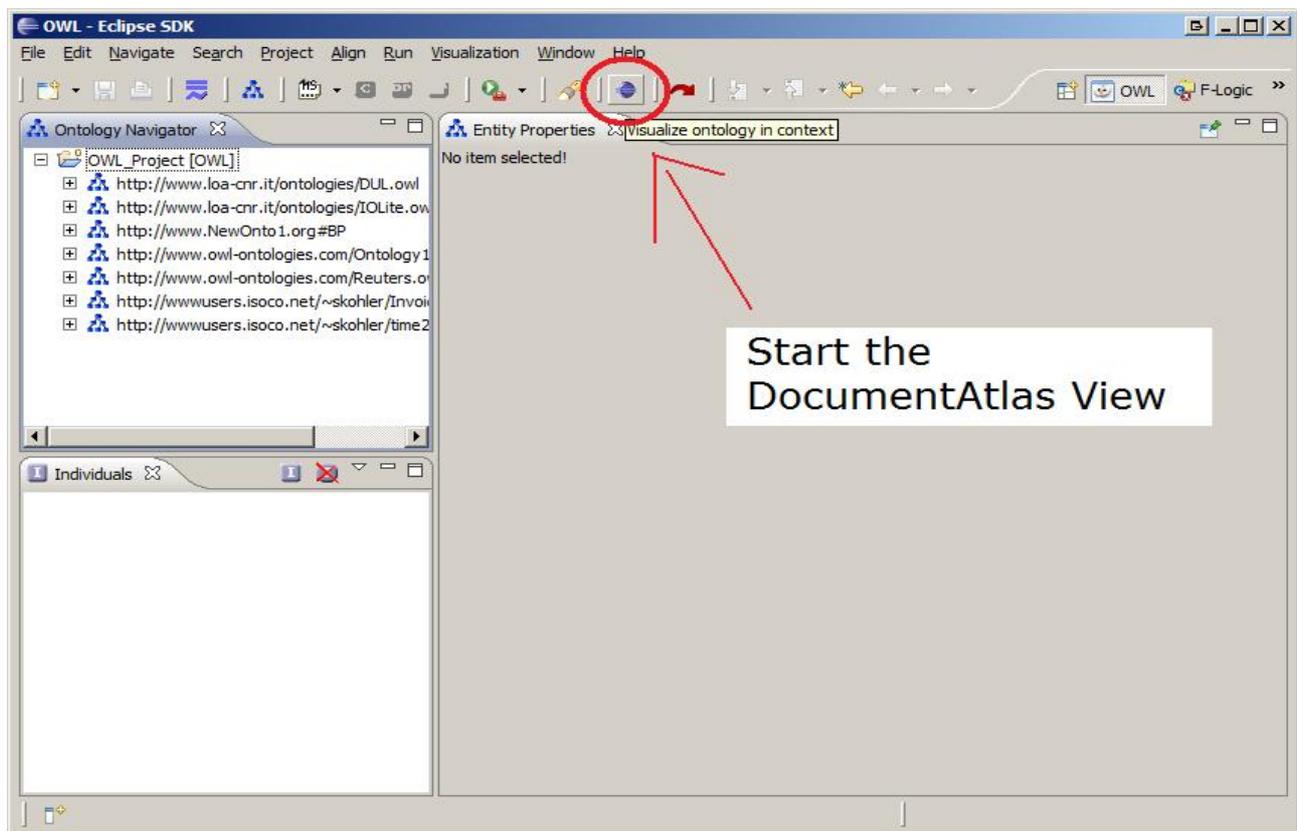
Figure 1: Starting the DocumentAtlas View in NeOn Toolkit

The DocumentAtlas is seamlessly integrated into the NeOn Toolkit platform as a plug-in. It can be started in several ways, most simply by clicking on the action button marked on the figure.

Figure 2: Close up of the Navigation GUI

The user interaction with the plug-in is a simple two step process. In the left select box any available ontology can be selected. The list is automatically obtained from all the NeOn Toolkit ontologies. In the right select box the semantic landscape is chosen. This box is automatically filled with all the landscapes that are stored in the "data" folder of the plugin. The button "Refresh

ontologies" is used when ontologies were added, removed and modified or when the plug-in was loaded before the ontologies into the Toolkit.

The generate landscape button opens a new dialog box, where the folder containing the textual corpus is selected.
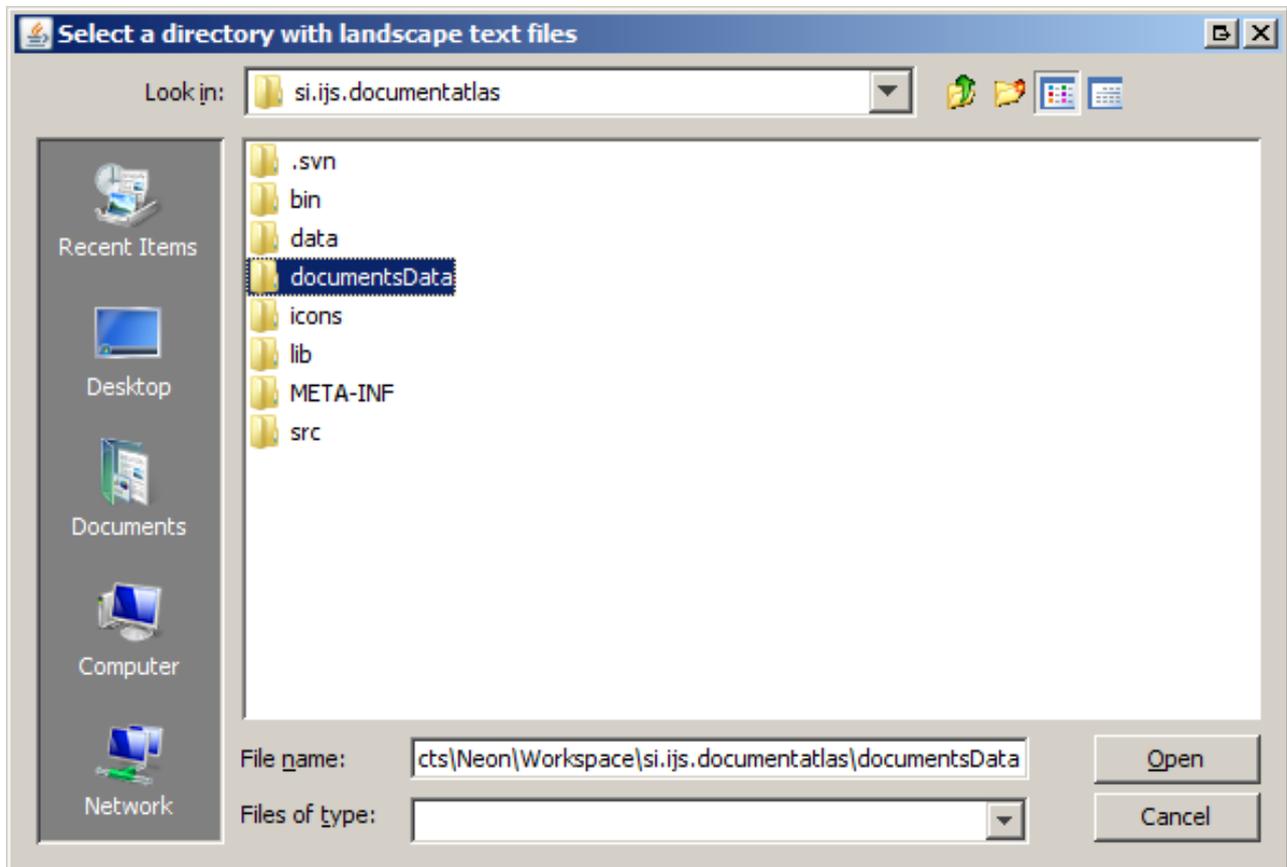
Figure 3: The file open dialog for selecting the context folder

After clicking on the "Generate landscape" button, the file open dialog is opened. The user has to navigate to the folder where the textual documents for generating the landscape are stored. When the folder is selected, the automatic generation of the landscape takes place. This may take some time, however it has to be done only once. After the landscape is created, it becomes available in the right select box among other landscapes and can also be shared with other users as an xml file.
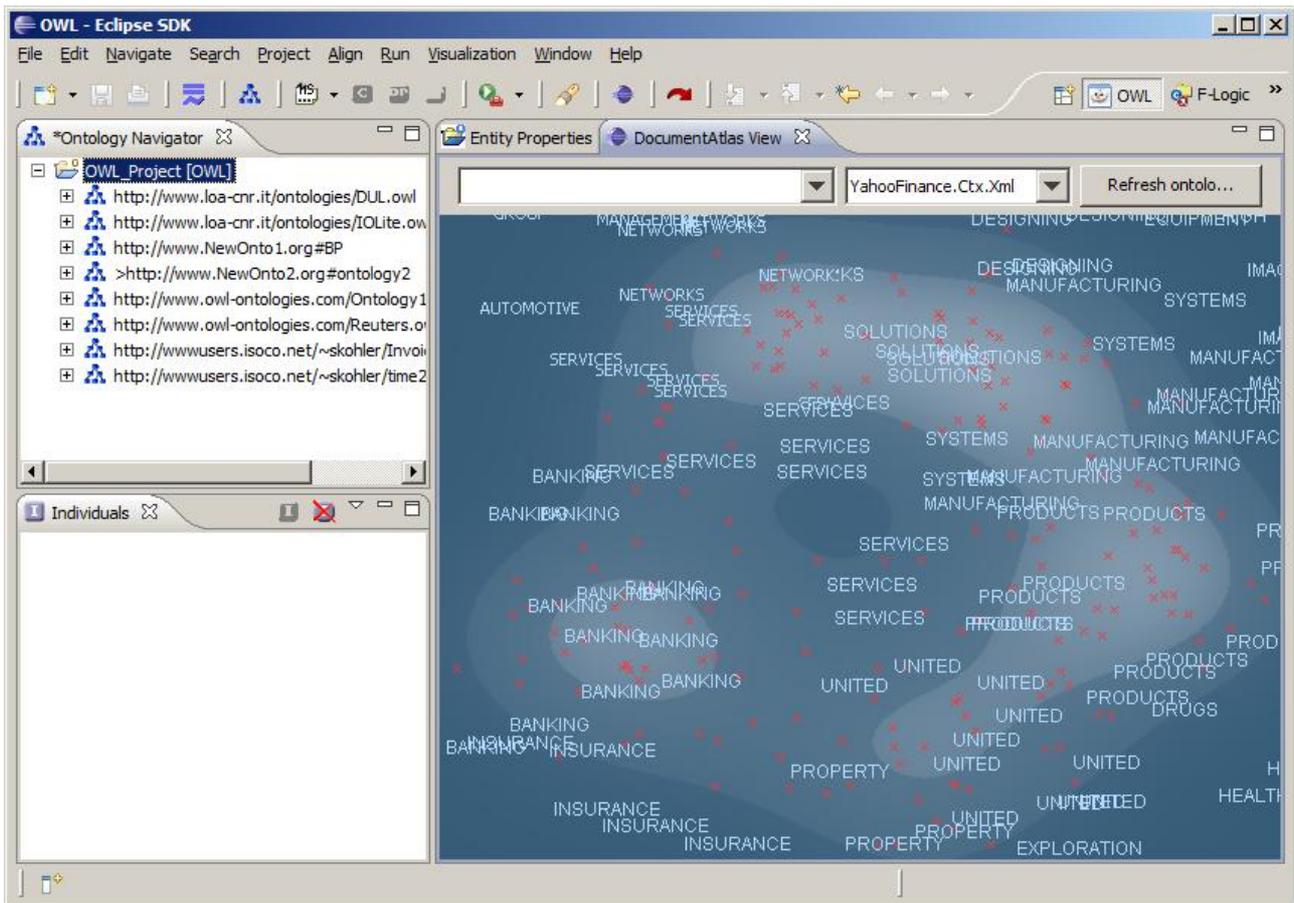
Figure 4: Visualized Semantic Landscape

The user has selected the YahooFinance.Ctx.Xml for the semantic Landscape. This context landscape is generated out of large corpus of financial news obtained from Yahoo Finance.

The study of this landscape shows a Banking and Insurance sector (bottom left). This part seems more separated from other topics. On the rest of the landscape, most prominent topics are Network Services (top middle), Manufacturing and Products (centre right) and Property and Exploration (bottom centre). These are the most prominent categories in the context of Yahoo Finance news.
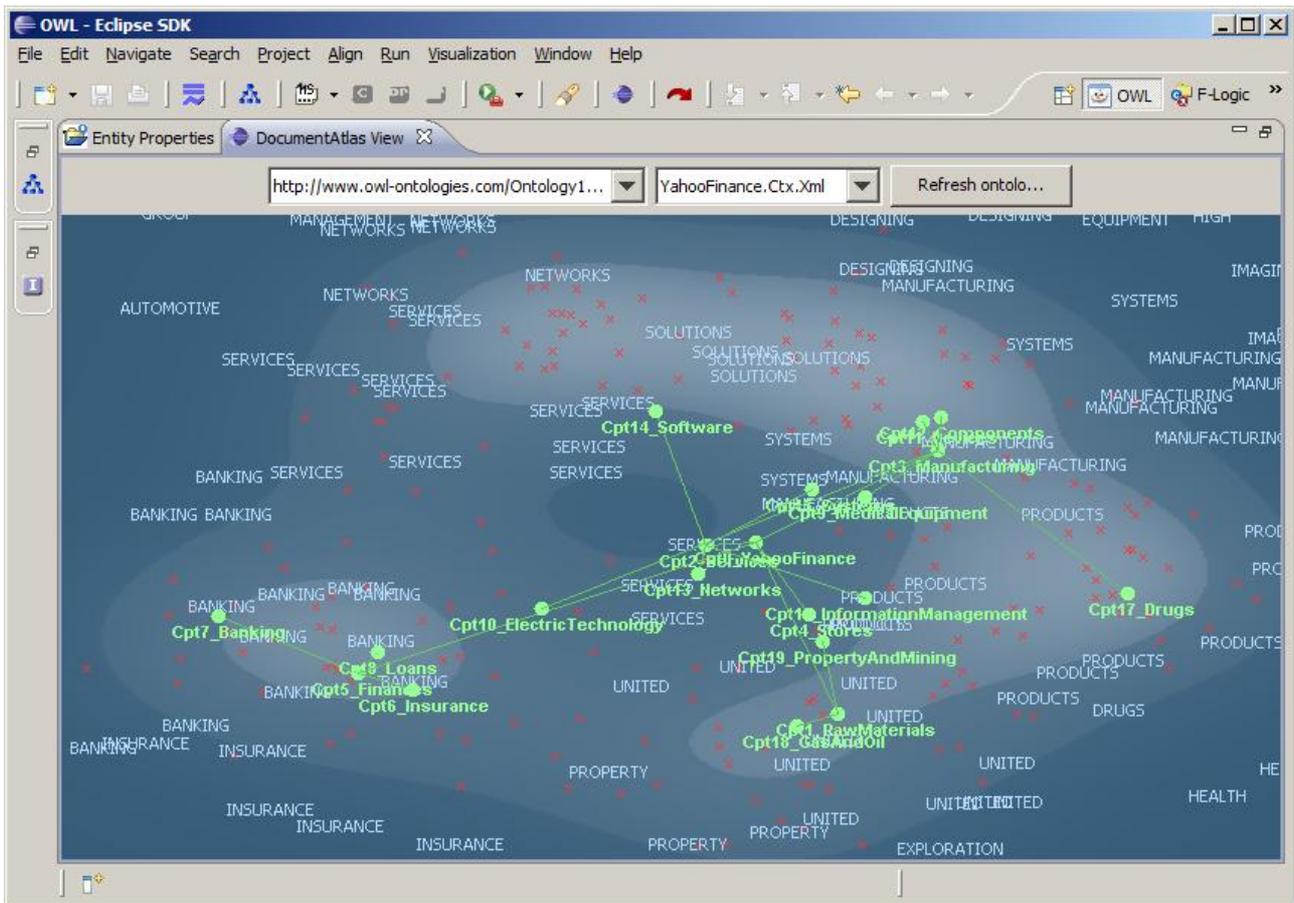
Figure 5: Ontology in Context

The user has visualized the current state of the ontology onto the semantic landscape. He can now study this ontology to see how well it covers the financial domain. The first impression is that most of the domain is already nicely covered. Closer inspection however, reveals good and weak points of the current ontology state.

The Banking and Insurance sector (bottom left) is covered by the concepts "Cpt5_Finances", "Cpt6_Insurance", "Cpt7_Banking" and "Cpt8_Loans". The architecture of this part of the ontology seems to be in accordance with the context provided by the landscape.

The Property and Exploration part (bottom center) is similarly well covered by "Cpt19_PropertyAndMining", "Cpt21_RawMaterials" and "Cpt18_GasAndOil".

In the Manufacturing and Products part (centre right) the user observes nice coverage of Manufacturing by "Cpt11_Devices", "Cpt12_Components", "Cpt13_Manufacturing" and less than optimal coverage of Products by "Cpt17_Drugs". Perhaps it would make sense to extend the ontology with products, or even repair the architecture, since "Cpt17_Drugs" is far away from the rest of the cluster.

The last observation is that the part of Networks, Services and Solutions is not covered at all by the current ontology. If this part needs to be modelled, the "Cpt14_Software" concept would probably be a good starting point where to achieve this.

4. Conclusion and future work

We have implemented a plugin for the NeOn Toolkit that enables the user to study ontologies in different contexts. This is achieved by visualizing ontology on top of the precalculated semantic landscape. Concepts in the ontology that are similar in the context of the landscape are drawn close. This way it is easy to study how well ontology covers the domain (presented with the landscape) and how well the architecture of the ontology is in accordance to topic profilation of the landscape.

This plugin serves as a graphical user interface for the Toolkit. It is very intuitive and easy to use even for the general public. This is in accordance to a requirement of both case study partners calling for tools that enable experts of the domain; however non-experts in ontologies to work with networked ontologies.

The preparation of the document corpus could be simplified. One idea is to gather documents from web search – i.e. for a given query we designate hits returned by a search engine as the documents of the corpus.

Appendix A: User support documentation

In this section we state some formal notifications for the user (licensing, copyrights and liability information). Apart from this we also provide some technical information about the software. All of this information is also present in the *readme.txt* file that is located in the repository together with the software.

Formal notifications:

- DocumentAtlas Plugin was developed and implemented by Jozef Stefan Institute (JSI) as part of the effort under the NeOn project.

- The source code of the plug-in is released under the LGPL license.

- Neither the NeOn consortium nor JSI in particular accept any liability for using this software.

Technical information:

- The DocumentAtlas plug-in was created to be used with the NeOn Toolkit platform.

- Software can be obtained at: http://kameleon.ijs.si/documentatlas/ in both the binary and source versions.

- To install the plugin into an instance of the Toolkit, the binary version needs to be extracted into the plugin directory. The Toolkit needs to be restarted.

- For questions or bug reports send an e-mail to: bostjan.pajntar@ijs.si

References

Mladenić D, Fortuna B, Grobelnik M, Dzbor M   Context sensitive visualization of multiple ontologies. Deliverable 3.7.1, NeOn, 2008.

Salton G (1991) Developments in Automatic Text Retrieval. Science, Vol 253: 974-979

Deerwester S, Dumais S, Furnas G, Landuer T, Harshman R (1990) Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science

Carroll JD, Arabie P (1980) Multidimensional scaling. In M.R. Rosenzweig and L.W. Porter (Eds.), Annual Review of Psychology, 31: 607-649

Fortuna B, Mladenić D, Grobelnik M (2006) Visualization of text document corpus. Informatica 29: 497-502

Fortuna B, Grobelnik M, Mladenic D. Semi-automatic Data-driven Ontology Construction System. Proceedings of the 9th International multi-conference Information Society IS-2006, Ljubljana, Slovenia.

Grobelnik, M., Mladenic, D., Text Garden, Springer 2008

Grobelnik M, Brank J, Fortuna B, Mozetic I. Contextualizing ontologies with ontolight: a pragmatic approach. Proceedings of the 10th International multi-conference Information Society IS-2007, Ljubljana, Slovenia.

Jain AK, Murty MN, Flynn PJ (1999) Data Clustering: A Review, ACM Comp. Surv.

Wang X, McCallum A (2006) Topics over time: a non-Markov continuous-time model of topical trends. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 424 – 433, Philadelphia