

AsIsKnown

**A semantic-based knowledge flow system for
the European home textiles industry**

Work package 3: Common sense ontology engineering

**Deliverable D11 “Reports on the tests and evaluation of common
sense ontology”**

Third version

Lead participant: IPP-BAS
Nature: Report
Dissemination level: PU
Delivery date: 33 PM



This document has been produced in the context of the AsIsKnown Project. The AsIsKnown project is part of the European Community's Sixth Framework Program for research and development and is as such funded by the European Commission. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors view.



Context

WP 3	Common sense ontology engineering
Task 3.7	Testing and evaluation
Dependencies	This deliverable requires user requirement input from 3.2 and 3.4

Contributors: Kiril Simov (IPP-BAS) Petya Osenova (IPP-BAS)	Reviewers: Kiril Simov (IPP-BAS)
--	--

Approved by: Kiril Simov, Bulgaria as WP3 Head



Executive Summary

This report focuses on tests and evaluation of common sense ontology, the creation of the lexicons aligned to the ontology, definition of web services for the provided software and their integration with other components of the AsIsKnown System. The work in the period was done in parallel to the exploration of corresponding modules and resources.

Lexicons for the project languages were created. The languages are: German, French, Italian, Hungarian, and Bulgarian. The English lexicon was created during the previous period of the project.

All the necessary components for the usage and maintenance of the resources were equipped with web services. The integration with other components of the system was finished and tested.



Table of Contents

EXECUTIVE SUMMARY	4
TABLE OF CONTENTS	5
LIST OF ABBREVIATIONS	6
1 INTRODUCTION AND PROBLEM STATEMENT	7
2 LEXICON CREATIONS	8
2.1 Ontology to Lexicon Model	8
2.2 Structure of the Lexical Entry	9
3 WEB SERVICES	11
4 CONCLUSIONS AND OUTLOOK	13
REFERENCES	14



List of Abbreviations

1 Introduction and Problem Statement

In this deliverable we discuss the test and evaluation of the common sense domain ontology in the area of Home textile done during the period from month 22 to month 33 of the project. In this period we have finished the construction of the lexicons for the languages of the project; we have tested and finished the software components and the web services supporting the work with the data (ontologies and lexicons); and we finished the ontology.

The lexicons play two roles in AsIsKnown System – communication with the users and support of language processing. During the above period we have created lexicons for all languages in the project: German, French, Italian, Hungarian, Bulgarian, and English. The model for mapping of ontology to lexicons reflects the idea of language independent ontology.

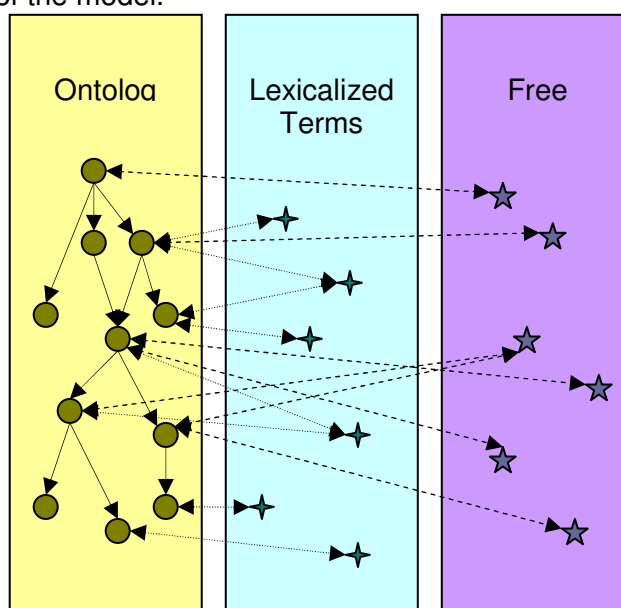
In order the created ontology and lexicons to be used within the AsIsKnown System we have implemented (or reused implementations of other peoples) a number of software modules equipped with web services. Here we list all the services. The documentation is on a public wiki and pointers to it are provided.

2 Lexicon Creations

One of the main tasks for the workpackage for the third year of the project was the creation of lexicons in the languages of the project – German, French, Italian, Hungarian, and Bulgarian. First we have created a lexicon in English which was used as a model for the lexicons in the other languages. The next subsection repeats the elements of the ontology-to-lexicon relation which we are using. Then we present the elements of the lexical entry in the finished lexicon.

2.1 Ontology to Lexicon Model

In this subsection we present briefly the linguistic model of the ontology-to-lexicon relation adopted within the project. The ontology-to-lexicon relation as presented here is a part of the ontology-to-text relations we presented in D11-2. We assume that the ontology is the repository of the lexical meaning of the language. Thus, we have started with a concept in the ontology and we searched for lexical items and non-lexical phrases that convey the content of the concept. There are two possible problems here: (1) there is no lexical item for some of the concepts in the ontology, and (2) there are lexical items in the language without a concept representing the meaning of the lexical item in the ontology. The first problem is overcome by allowing also non-lexical (fully compositional) phrases to be represented in the lexicon. The second problem is solved by the extension of the ontology. The following figure depicts the elements of the model.



We have been using the relations between the different elements for the task of ontology-based search. The connection from ontology via lexicon to grammars is relied on for the concept annotation of the text. In this way we established a connection between the ontology and the texts. The relation between the lexicon and the ontology is used for definition of user queries with respect to the appropriate segments within the documents.

Our approach gains in many respects from such works as WordNet [1], EuroWordNet [2], SIMPLE [3]. In spite of the fact that we employ the experience from these projects (mapping to WordNet and Pustejovsky's ideas in SIMPLE), we also suggest an alternative for the connection between the ontology and the lexicons. Our model is very close to LingInfo model

Third version

(see [4] and [5]) not only with respect to the mapping of the lexical items to concepts, but also with respect to the other language processing tools we connect to the ontology – the concept annotation grammars and concept disambiguation tools. As to WordNet and EuroWordNet, we differ in the direction of the workflow, i.e. we start from ontology to the lexicon, not vice versa. From SIMPLE we differ in using domain ontology instead of a general linguistic ontology. From LingInfo model we differ in the fact that the three components (ontology, lexicon and grammars) are represented by different models.

2.2 Structure of the Lexical Entry

On the basis of the model of the ontology-to-lexicon relation we have created XML DTD for formatting of the lexicons in the different languages. The definition of the DTD is as follows:

```
<!ELEMENT Lexicon (entry+)>
<!ATTLIST Lexicon
    version CDATA #IMPLIED>
```

The lexicon contains entries.

```
<!ELEMENT entry (
    (owl:ObjectProperty | owl:Class | rdf:Description),
    def,
    termg+)>
<!ATTLIST entry
    id ID #REQUIRED>
```

Each entry contains fragment from the ontology (a class, a property or an instance), then a definition in English and a list of term groups (element <termg>). At least a term group for English is presented.

```
<!ELEMENT termg (term+,def?)>
<!ATTLIST termg
    lang (none|bg|de|en|fr|hu|it) "none">
```

Each term group contains one or more appropriate terms in a given language, and a definition in this language. A given term group contains terms from one language which is marked by the attribute *lang*.

```
<!ELEMENT term (#PCDATA*)>
<!ATTLIST term
    shead (1|0) "0"
    type (lex|nonlex) "lex">
```

Each term contains the basic form of a term for the corresponding concept in the language of the term group. A term can be lexicalization of the concept in the language and this fact is marked by the attribute *type* with a value "lex". If the value is "nonlex", then the term is a non lexicalized phrase. Among the terms in one term group one term is selected a head term (or representative term). This term is used when the corresponding concept needs to be presented to a user in the language of the term group. The head term for a term group is marked via the attribute *shead* with value "1".

<!ELEMENT def (#PCDATA)>

The element <def> contains a definition of the concept in English when it is on the level of an entry or in a term group for English terms. Otherwise it contains a definition in the language of the corresponding term group.

On the basis of this structure we generated lexicon template for each language in the project. The template contained an entry for each concept and propriety in the ontology, a definition in English and term group for English. Then the corresponding language expert created term group for his/her language. The language expert also had the possibilities to write comments within each entry and in this way we collected some comments also on the ontology. This was especially useful in the case of Bulgarian and Italian. The language specialists for these languages were knowledgeable with respect to the textile domain. Their suggestions were carefully examined and some of them were incorporated in the ontology. The result of these activities is the lexicons for the languages of the project languages. Here is an entry with term groups for all the languages:

```
<entry id="entry-34">
  <owl:Class rdf:about="http://www.asisknown.org/AIKHT#CarpetOWN">
    <rdfs:comment>a piece of thick heavy fabric (usually with nap or pile) used to cover
      a floor</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class
        rdf:about="http://www.loa-cnr.it/ontologies/OWN/OWN.owl#FURNISHINGS"/>
      </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Class rdf:about="http://www.asisknown.org/AIKHT#FloorCovering"/>
    </rdfs:subClassOf>
  </owl:Class>
  <def>a piece of thick heavy fabric (usually with nap or pile) used to cover a floor</def>
  <termg lang="bg">
    <term shead="1">килим</term>
  </termg>
  <termg lang="en">
    <term shead="1">carpet</term>
    <term>carpeting</term>
    <term>rug</term>
    <def>a piece of thick heavy fabric (usually with nap or pile) used to cover a floor</def>
  </termg>
  <termg lang="it">
    <term shead="1">tappeto</term>
    <def>spesso tessuto di lana, cotone o altra fibra, specialmente con disegni ornamentali,
      destinato a essere collocato sul pavimento per abbellimento o per maggiore
      confortevolezza</def>
  </termg>
  <termg lang="ru">
    <term shead="1">ковѐр</term>
  </termg>
  <termg lang="de">
    <term shead="1">Teppich</term>
    <term>Teppichläufer</term>
  </termg>
  <termg lang="hu">
    <term shead="1">szőnyeg</term>
  </termg>
</entry>
```

```
<def>padlóra való, szőtt vagy csomózott anyagból készült, vastag, súlyos szövet</def>
</termg>
<termg lang="fr">
  <term shead="1">tapis</term>
  <def>panneau plus ou moins épais de fibres d'origine animale, végétale ou synthétique,
    nouées ou tissées selon divers procédés, que l'on pose sur le sol pour le confort
    et la décoration</def>
</termg>
</entry>
```

3 Web Services

In this section we describe the web services defined, implemented, and tested during the period. Here is a list of the services:

1) Ontology Reasoner

This module is published as a Web Service. It is used for inference on an OWL-DL ontology model.

The ERP of that service is: <http://213.191.192.50:8080/aikos/services/Reasoner>.

All the artifacts for that component can be found at:

<http://www.bultreebank.bas.bg/public/aikos/InferenceServices/>.

Users can find there a client implementation and a wrapping Query class in the *Reasoner.jar*. If users need to integrate the web service on their own Tomcat server, they can find the *Reasoner.aar* file at the pointed location. A documentation of the service can be found on our public wiki:

<http://www.bultreebank.bas.bg/wiki/doku.php?id=projects:aikos:ws:inference>

2) Term Resolver (Lexicon Management)

This module is published as a Web Service.

The basic functionality of the module is to translate ontological units (concept, object properties URLs) into terms, using lexicons that contain the URL - term relevance for different languages and to translate terms into ontological units using the lexicons.

The ERP of that service is: <http://213.191.192.50:8080/aikos/services/TermResolver>.

All the artifacts for that component can be found at:

<http://www.bultreebank.bas.bg/public/aikos/TermResolver/>.

Users can find there a client implementation and a wrapping Query class in the *TermResolver.jar*. If users need to integrate the web service on their own Tomcat server, they can find the *TermResolverService.aar* file at the pointed location. A documentation of the service can be found on our public wiki:

<http://www.bultreebank.bas.bg/wiki/doku.php?id=projects:aikos:ws:termresolver>

3) Ontology Translator

This module is published as a Web Service.

The ontology translation and transformation component is used for modification of OWL-DL Ontology model description to an XML document that contains the named concepts, their relations: is-a, property relations, and restrictions on properties. These relations and restrictions on relations are presented as <arc> elements with attributes: *domainConceptURI*, *rangeConceptURI*, *propertyURI*, *modifier*. All the <arc> elements are

children elements of the <Graph> elements (*transformed ontology model*). The translation of all the URIs available as values of the attributes *domainConceptURI*, *rangeConceptURI*, *propertyURI*, *modifier* are translated via lexicon (XML document that contains *URI - terms* correspondence for specific languages). The element that contains the translations is <Vocabulary> that has as children <Term> elements that express the *URI - term* correspondence. If no lexicon is loaded the <Vocabulary> element is empty.

The ERP of that service is: <http://213.191.192.50:8080/aikos/services/OntologyTranslator> .

All the artifacts for that component can be found at:

<http://www.bultreebank.bas.bg/public/aikos/Ontology%20Translation%20&%20Transformation/>

Users can find there a client implementation and a wrapping Query class in the *OntologyTranslator.jar*. If users need to integrate the web service on their own Tomcat server, they can find the *OntologyTranslator.aar* file at the pointed location. A documentation of the service can be found on our public wiki:

<http://www.bultreebank.bas.bg/wiki/doku.php?id=projects:aikos:ws:ontologytranslation>

The translation of the AsIsKnown ontology for the different languages takes between 1 to 2 seconds.

4) Ontology Lexicon Requests

This module is published as a Web Service.

The basic functionality of the module is to allow users post requests for addition of new data to an ontology and a lexicon. The new items that can be requested are: concepts, relations and terms. The data is saved in a *requests.sys* XML document that is available in the default (root) folder of the Tomcat server where the service is deployed.

The ERP of that service is: <http://213.191.192.50:8080/aikos/services/NewOntoLexRequest>

All the artifacts for that component can be found at:

<http://www.bultreebank.bas.bg/public/aikos/Ontology-Lexicon%20Requests/>

Users can find there a client implementation and a wrapping Query class in the *OntoLexRequests.jar*. If users need to integrate the web service on their own Tomcat server, they can find the *OntoLexRequest.aar* file at the pointed location. A documentation of the service can be found on our public wiki:

<http://www.bultreebank.bas.bg/wiki/doku.php?id=projects:aikos:ws:ontolexrequests>

The Ontology Resoner is used by the Smart Profiler for checking consistency. It is also used in the Ontology Translator module in order to construct the simplified representation of the ontology. The Term Resolver is used by the Translator service of RWTH in order to support the transfer of producer data into AsIsKnown system. It is used also in Ontology Translation module for accessing lexical information. Ontology Lexicon Requests module is used by the partner systems for reporting gaps in the ontology and/or lexicons.

4 Conclusions and Outlook

In this report we present the tasks we have done within the period for testing and evaluation of the AsIsKnown Home Textile Ontology. We were concerned with the creation of the lexicons for the languages in the project. We have finished the definition, implementation and testing of the web services. We also fixed all the found bugs.

Also we finished the substitution of the upper part of the ontology with a new simplified version. This is necessary in order to support better usability of the ontology. Also some changes in the domain part of the ontology was done.

References

- [1] Christiane Fellbaum. 1998. Editor. WORDNET: an electronic lexical database. MIT Press.
- [2] Piek Vossen (ed). EuroWordNet General Document. Version 3, Final, July 19, 1999, <http://www.hum.uva.nl/~ewn>
- [3] Alessandro Lenci, Federica Busa, Nilda Ruimy, Elisabetta Gola, Monica Monachini, Nicoletta Calzolari, Antonio Zampolli, Emilie Guimier, Gaëlle Recourcé, Lee Humphreys, Ursula Von Rekovsky, Antoine Ogonowski, Clare McCauley, Wim Peters, Ivonne Peters, Robert Gaizauskas, Marta Villegas. 2000. SIMPLE Work Package 2 - Linguistic Specifications, Deliverable D2.1. ILC-CNR, Pisa, Italy.
- [4] Paul Buitelaar, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Ralf Engel, Massimo Romanelli, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel, Philipp Cimiano LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In: Proc. of OntoLex06, a Workshop at LREC, Genoa, Italy, May 2006.
- [5] Paul Buitelaar, Michael Sintek, Malte Kiesel A Lexicon Model for Multilingual/Multimedia Ontologies In: Proceedings of the 3rd European Semantic Web Conference (ESWC06), Budva, Montenegro, June 2006.