

Towards ONTO6 Framework for Concept Elicitation

Uldis Straujums

University of Latvia, Faculty of Computing, Raina bulvaris 19, Riga, LV-1586, Latvia

Abstract

The article proposes an approach to simplify a process of identifying significant concepts for a given domain. The author describes his ONTO6 methodology based on a semi-informal meta-ontology. The stages of applying the ONTO6 methodology are: the development of a meta-ontology instance appropriate for the domain to be informatized; the development of an initial ontology from the meta-ontology instance; and the gradual detailing of domain concepts that appear in the initial ontology – the development of an enriched initial ontology. The transition of ONTO6 methodology to ONTO6 framework by usage of tools – LVTagger, Cellfie – is demonstrated.

Keywords

Ontology learning, domain-specific modeling, text analysis tool

1. Introduction

The article proposes an approach to simplify a process of identifying significant concepts for a given domain. The author expands on his previous research in the specific field of informatization [1, 2]. Informatization is understood as an analysis of the business processes, the specification of requirements and the development of software.

In his previous research the author had proposed a unified description of methods and suggestions to identify the essential concepts of the domain to be informatized to introduce notations for the various levels of detail, and to specify details for the informatization aspects. The author's approach helped overcome the difficulties observed during implementation of several informatization projects and to implement several improvements:

- The development of a unified understanding about the domain to be informatized, particularly about the essential concepts and their interpretation
- The introduction of a suitable notation for various aspects of informatization which are necessary for users involved in the project and appropriate for different levels of competence
- The proposal of a general methodology for performing informatization.

The proposed ONTO6 methodology appears to be expandable to other domains with several enhancements. Namely, several tools have to be added to make the process of supplying source information more convenient for the user.

Firstly, a tool for entering information into the instance of the metaontology for the particular domain has to be developed.

Secondly, an API has to be specified and implemented to allow a fine-tuning of the essential concepts elicitation according to the particular domain.

In the article, the author gives the specifications of needed enhancements of ONTO6 methodology and describes the current state of their implementation.

Thus, the original informatization-specific ONTO methodology is transformed into a ONTO6 framework suitable for the essential concept elicitation for a given domain.

Baltic DB&IS 2022 Doctoral Consortium and Forum, July 03-06, 2022, Riga, LATVIA

EMAIL: uldis.straujums@lu.lv (U. Straujums)

ORCID: 0000-0002-2212-5435 (U. Straujums)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

2. Roots of the ONTO6 methodology

The need to identify and record significant concepts for a given domain is well accepted. Scientists have developed several knowledge representation approaches – controlled vocabularies, thesauri, classification schemes, taxonomies, topic maps, frame languages, logical theories and meta-models. All of these approaches, as well as many others, form the basis of so-called ontologies.

2.1. Concept of ontology

The concept of ontology was formally defined by Thomas Gruber in a Stanford University publication [3], in which he redefines the concept of ontology as generally applied in philosophy.

The definition of ontology as formulated by Thomas Gruber is:

“An ontology is the explicit specification of a conceptualisation for a domain”.

Over time the concept of ontology has been defined more precisely [4]:

“An ontology defines a set of representational primitives with which to model a domain of knowledge or discourse”.

2.2. Formal ontologies

As pointed out by Ruth Wilson [5], the differences lie in the possibilities of describing terms and to define relationships between them. These differences at the formalization level represent an ontology spectrum. A meta-model is a clearly defined model of the domain of interest, comprising concepts and rules that are essential for the construction of specific models. Each meta-model is an ontology, but it is a far richer notion – it can be used as a set of building blocks and rules, as a model for the domain of interest and as an instance of another model.

2.3. Semiformal ontologies

At one end of the ontology spectrum is a controlled vocabulary – a list of enumerated terms. Ideally, each term should have only one meaning. In practice, however, terms are accorded different meanings in different domains. If several terms have one and the same meaning, then one term is selected as the preferred one, while the others are classified as synonyms or aliases. With controlled vocabularies, more advanced ontologies can be constructed. For example, a thesaurus is built up by adding associative relationships to the controlled vocabulary. Frame languages have the ability to express the properties, logical constraints and detailed relationships of terms. Ontologies can be used to express and analyze taxonomical relationships as suggested by Christopher Welty and Nicola Guarino [6].

2.4. Ontology as an understanding

Ontologies can be depicted in several ways, but the presentation must be suitable for the target user. Furthermore, they must be capable of adapting to users from different backgrounds and abilities (desire) to get a grasp of formal constructions. These requirements mean that the developer of the ontology and its users must come to a common understanding regarding the level that the user can comprehend. Thomas Gruber, the inventor of the concept of ontology, has advocated this approach with exceptional clarity in the 2004 publication “Semantic Web & Informations Systems” [7].

2.5. Ontology clusters

A domain-specific ontology is usually built by a team of several people who have diverse skills within the framework of the particular domain. This approach has been described by researchers Pepijn R. S. Visser and Valentina A. M. Tamma [8], who recommend taking advantage of individual team

members, who have mutually complementary knowledge about the concepts of the particular domain. A hierarchical ontology is created with an application-specific ontology at the root. The definitions of the terms in this application-ontology are derived from an existing top level ontology, which the abovementioned authors have chosen to be the English language lexicographical data base Wordnet [9]. The new ontology cluster is a derived ontology that defines new concepts using those concepts already defined in the upper ontology.

3. Methodologies of ontologies development

Methodologies in the development of ontologies reflect the formal background of the ontology developer.

3.1. Logical theories

The specification can be developed in the form of a logical theory that describes the intended meaning. Nicola Guarino [10] implements this principle, "An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models."

3.2. Linguistic relativism

A specification can be developed using the concept of linguistic relativism, an approach suggested by Boris Wyssusek [11]. It is understood that the concept of linguistic expression is not uniquely definable since separate elements can have different interpretations. The criterion for the adoption of an interpretation is its correspondence to the real world. A common understanding of the language needs to be attained, such common understanding is a prerequisite for a stable interpretation of the language.

Abel Browarnik and Oded Maimon [12] propose models for ontology learning based on linguistic knowledge and existing, wide coverage syntactical, lexical and semantic resources – ASIUM, Text-To-Onto, TextStorm/Clouds, Syndicate, OntoLearn, CRCTOL and OntoGain.

3.3. Analysis of taxonomical relations

Concepts unified through taxonomy are analyzed according to their meta-characteristics: identity, rigidity, unity, dependence, thereby revealing more readily the intended meaning of the taxonomical relations. This is the course followed by Christopher Welty and Nicola Guarino [6].

3.4. Methodologies specific to information systems

Typical concepts of information systems are formalized: the system, the subsystem, unification. Yair Wand and Ron Weber [13] use the formal model to confirm whether the system is properly divided into components.

If an information system can be regarded as a branch of science, then it can be analysed with a methodology that looks at several processes important to development: inclination, learning, influence of culture, consolidation, as shown in the work of Brian O'Donovan and Dewald Roode [14].

For the analysis of an operating system with an existing descriptive ontology, Peter Green and Michael Roseman suggest changing the ontology into an ER-based meta-model. The meta-model then permits the form of the central concept of the ontology to be determined – function, activity or thing [15].

Researcher Mauri Leppänen [16] proposes the following methodology for the analysis of the output of information systems – "A system of perspectives is composed of five perspectives. These are the

systelological perspective, the infological perspective, the conceptual perspective, the datalogical perspective, and the physical perspective.” Māris Treimanis [17] recommends an aspect-oriented approach when structuring the output of an information system.

For the building of a taxonomy for modeling method requirements researchers Dimitris Karagiannis, Patrik Burzynski, Wilfrid Utz, Robert Andrei Buchmann [18] propose a metamodel CoChaCo (Concept-Characteristic-Connector) for representation and management of modeling methods, including an evaluation protocol.

3.5. Methodological applications

Ontologies are developed using various means and differ in the way they depict the world. Standards in ontology are necessary for the regulation of the following:

- What should be included in an ontology
- What are the basic categories and entities
- How are the entities depicted taking into account the knowledge level of the prospective user.

A great variety of backgrounds is needed for the development of ontologies. The author’s aim is to develop a methodology for the user who is an expert in the problem-domain albeit without any special knowledge in formalized engineering knowledge systems.

The term “methodology” here is rather ambiguous. There are several definitions of methodology. The author has chosen the definition:

methodology – an assorted coordinated succession of techniques or methods that constitute a general system theory or prescribes how thought-intensive activity is to be achieved [19].

This definition of methodology has been chosen for the author’s approach, i.e., a succession of techniques or methods has been developed that defines the thought-intensive activities for the concept elicitation.

Methods comprising the methodology consist of procedures, which have been proposed for the creation of a knowledge model, for the acquisition of a conceptual scheme from a knowledge model, and for the detailed description of the aspects of the given domain.

Techniques comprising the methodology include processes that have been developed for the application of methodology methods, as well as recommendations for completing the stages of the methodology – the elicitation of knowledge in the development of the knowledge model, the derivation of a conceptual schema, and the choice of aspect level of detail.

The author’s ONTO6 methodology is developed for the user who is an expert in the problem-domain without any special knowledge in formalized engineering knowledge systems.

4. ONTO6 methodology

The development of the ONTO6 methodology was influenced by a “6W” approach based on six questions, which, it seems, was first mentioned by the Greek rhetorician Hermagoras already in the year 1 B. C. [20].

The 6W approach can be considered as a means of obtaining essential information by asking the questions - What, Where, When, How, Why, Who.

The author has named his methodology ONTO6, a name that was chosen not only because ontology is used to define a knowledge model, but also because the development of the ontology was influenced by the 6W approach.

The 6W approach has been adapted to the organization of business knowledge [21], the depiction of business structures [22], [23], journalism, police work [24], the organization of brain-storming sessions [25], the sphere of architectonic design [26], user modeling [27], the planning of information systems [17], [28], but it is not known to be used in the area of informatization.

The ONTO6 methodology makes use of the 6W framework:

What, Where, When, How, Why, Who.

It is aimed at identifying concepts, determining the interaction between objects corresponding to those concepts and determining the functionality of the objects.

The ONTO6 methodology is based on a semi-informal meta-ontology.

The stages of applying the ONTO6 methodology are:

- the development of a meta-ontology instance appropriate for the domain to be informatized
- the development of an initial ontology from the meta-ontology instance; and
- the gradual detailing of domain concepts that appear in the initial ontology – the development of an enriched initial ontology.

The end result is an ontology cluster, comprising a meta-ontology, a meta-ontology instance, an initial ontology, and an enriched initial ontology.

The ontology cluster is examined for its comprehensibility and its suitability for the domain, thus obtaining answers to several questions of competence.

To achieve a sufficiently general methodology, one that can be applied to the conceptualization of diverse domains to be informatized, a base structure has been incorporated into the methodology as well as a process for obtaining a useful model of the conceptualization of a particular domain from the base structure. This base structure in the ONTO6 methodology is a knowledge model that contains the meta-concept – aspect space. Aspect space describes all possible aspects of the domain to be informatized by grouping them into subsets.

For a given aspect set $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$, where $i = 1$ to n , where n is a natural number and a_i is an aspect of the domain to be informatized, the aspect space (A) is the set of all the subsets of the aspect set A (power-set). Therefore any element of the aspect space is a subset of an aspect set.

The aspect space remains constant for any domain to be informatized, however a suitable aspect space element must be allocated to the domain. From the knowledge model a usable model for the particular domain to be informatized can be derived. ONTO6 knowledge model is built (see **Figure 1**). Figure 1 shows a set of knowledge of various types of the domain to be informatized, namely, aspect subset, sub-aspects, concept instances. The relations among them are determined by procedures eliciting sub-aspects and concept instances from the textual information on the domain.

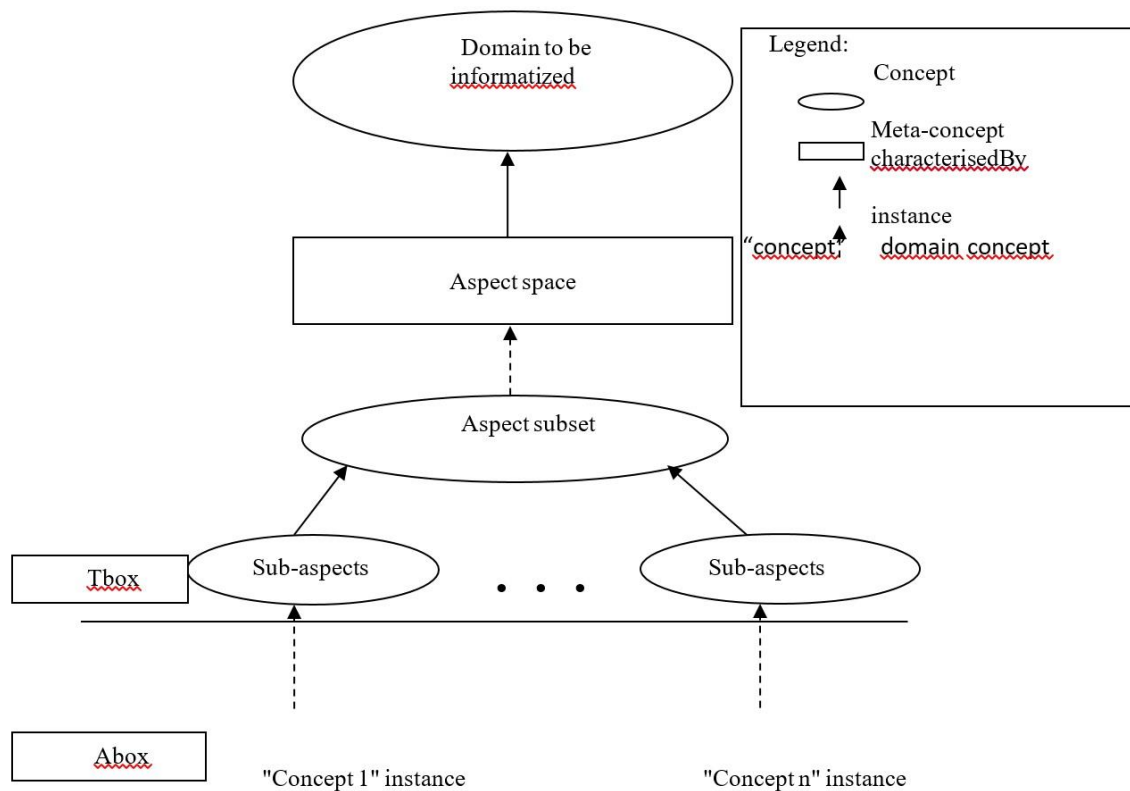


Figure 1: ONTO6 knowledge model

In order to obtain the model for the domain to be informatized, a procedure is applied to the knowledge model for determining the aspect subset (an instance of the aspect space) – the frequency of terms corresponding to a particular aspect is calculated, least frequent aspects are not included in the aspect space. A subjectively chosen threshold value is used for determining the essential aspects; a procedure for adding sub-aspect class instances is developed using text morphological analysis.

In line with the six question approach [20], the aspect set, A, is chosen to be

A = { What, Where, When, How, Why, Who }.

It is proposed to depict the concepts of the knowledge model in the language OWL with classes. For example, the term “Who” is shown as follows in the syntax of OWL RDF/XML:

```
<owl:Class rdf:about="#Who">
  <rdfs:subClassOf rdf:resource="#Aspect set"/>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Who</rdfs:label>
</owl:Class>
```

The meta-concept "Aspect space" is depicted as a class of classes with restrictions on the class elements. In OWL RDF/XML syntax this appears as:

```
<owl:Class rdf:ID="Aspect space">
  <rdfs:comment>
    This is the power-set.
  </rdfs:comment>
  <owl:sameClassAs>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#rdfs:subClassOf"/>
      <owl:hasValue rdf:resource="#Aspect set"/>
    </owl:Restriction>
  </owl:sameClassAs>
</owl:Class>
```

The relationship is shown as a property of the object or data type.

For example, the relationship "characterisedBy" is shown in the syntax of the language OWL RDF/XML as a property of the object "Infodomain" as follows:

```
<owl:ObjectProperty rdf:ID=" characterisedBy ">
  <rdfs:range rdf:resource="#Aspect space"/>
  <rdfs:domain rdf:resource="#Infodomain"/>
</owl:ObjectProperty>
```

The relationship between the concepts of the knowledge model is depicted in an ontology, which is referred to as a meta-ontology because it contains the meta-concept “Aspect space”, whose instance is the concept “Aspect subset”. Meta-ontology is an essential tool of the ONTO6 methodology. The ONTO6 methodology prescribes the development of a meta-ontology instance in conformance with the domain to be informatized, the development of an initial ontology from the meta-ontology instance and the enrichment of the initial ontology in subsequent informatization. The initial ontology does not change during informatization process. A visualization of the ONTO6 meta-ontology can be built (see Figure 2). The circles denote the possible aspects, while the arrows show possible relationships between the aspects. In meta-ontology aspect space instances, some of the arrows between the aspects as well as some aspects themselves may be absent along with the arrows.

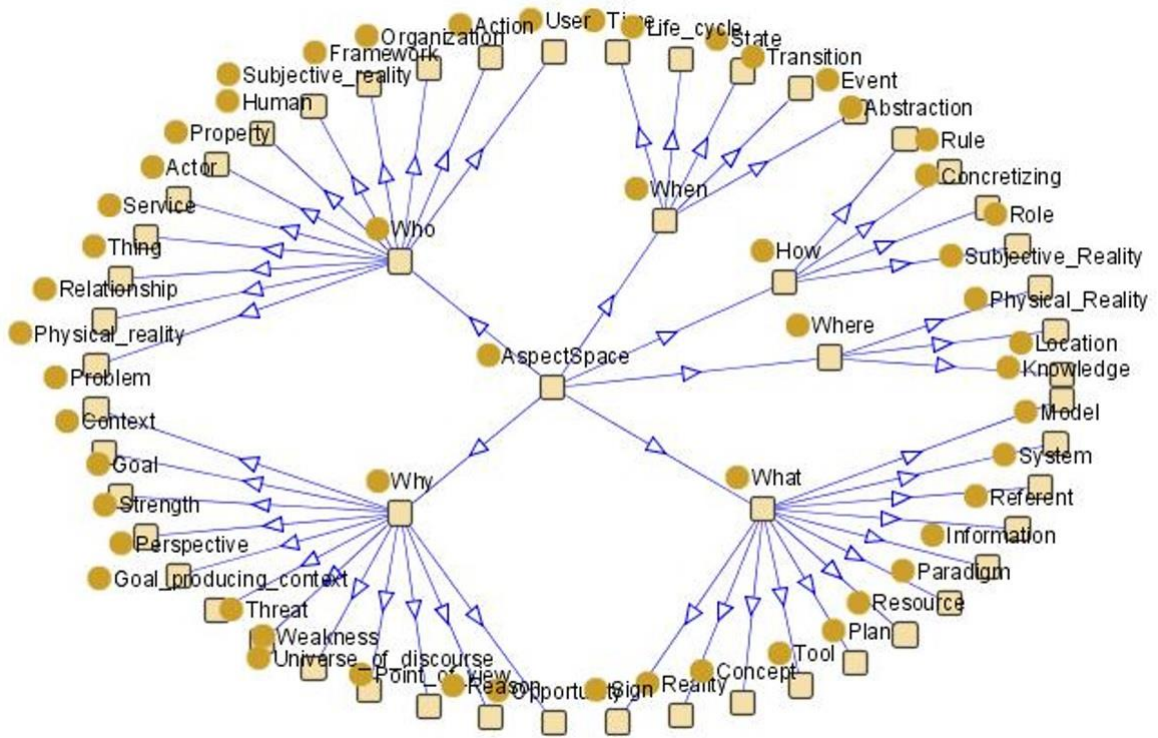


Figure 2: ONTO6 Meta-Ontology Highest level Simplified Visualization

Author’s ONTO6 methodology was successfully applied to several domains to be informatized including the Latvian Education Informatization System (LIIS). It became clear as a result applying the ONTO6 methodology that the LIIS domain has only two essential aspects – Where and What. Therefore LIIS can be added to that class of domain to be informatized, which has as its subspace instance the aspect subset {Who, What}. In **Figure 3** the essential LIIS domain concepts can be shown: education content, teaching, management, schools, ministry ,society, etc. .

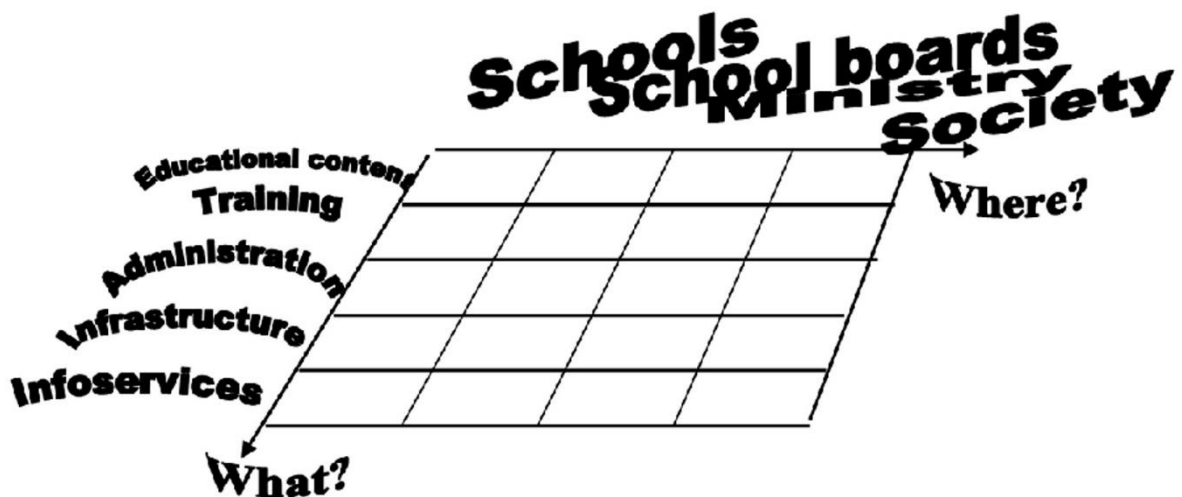


Figure 3: The essential LIIS domain concepts

The ontologies gained as a result of the ONTO6 methodology stages provide answers to questions of competency formulated by necessity in the development of the methodology:

- what are the essential concepts in the given problem domain? (the meta-ontology instance contains only the essential aspects – Who, Where)
- what are the relevant sub-concepts of the essential concepts? (the meta-ontology instance contains some sub-aspects of the essential aspects)
- which aspects of informatization must be examined in more detail? (the initial ontology includes the sub-aspect instances – Abox elements – schools, school boards, ministries, society, educational content, training, administration, infrastructure, information services)
- what kind of functionality is inherent (desired) in the specific aspect? (refined ontologies and visualizations agreed with the user describe in detail the desired functionality)
- which problem domains are similar to the given domain? (it is natural to consider as similar those domains which have the same essential aspects as the LIIS domain).

With ONTO6 methodology it is possible to find essential concepts for different domains.

5. From methodology to framework

Some constraints of ONTO6 methodology usage are: a fixed algorithm for concept elicitation, tiresome manual work to add ontology class instances (individuals) into an ontology, manual comparison of results with expert results.

The author has looked at several tools which could help at concept elicitation, namely, tools for finding word patterns – AntCone, WordSmith Tools, #LanesBox, SCP, corpkit, TextStat and LVTagger. Author has decided to use the Latvian language text analysis tool LVTagger developed by Peteris Paikens [29] because the fine-tuning of the essential concepts elicitation according to the particular domain can be easily accomplished using LVTagger.

The author has decided to use the Cellfie Plugin for Protégé 5 [30] for automatically entering class instances into an ontology for a particular domain.

5.1. Concept elicitation with LVTagger

As an input for LVTagger a text relevant to a particular domain could be given. As an output an information of text morphological analysis is produced. Several output formats are supported: CONLL-X, tab-delimited columns. The author has applied the LVTagger for an annotated word list creation in the CONLL-X format from a large text document describing a state level project for informatization (see **Figure 4**).

```
E:\Uldis\LU\LU2022\LU_konference_2022\Straujums_04-02-2022\ONTO6_framework\Onto6_stuff\LVTagger
java -jar tagger-2.2.1-jar-with-dependencies.jar <informatika98_all.txt >infovardi.txt
Input type : PARAGRAPH
Output type : CONLL_X
Loading classifier from models/lv-morpho-model.ser.gz ... Loading the lexicon from an input stream...
Leksmai 'pagaidas' #233437 galotne neatbilst paradigmai
-
      irka ID = 225524
      irka cilvklasmais ID = pagaidas:1
      Pamatforma = pagaidas
Leksmai 'pustrs' #275722 galotne neatbilst paradigmai
-
      FreeText = pamata skait.
      irka ID = 265618
      irka cilvklasmais ID = pustrs:1
      Pamatforma = pustrs
Lexicon 2.2.1 loaded
done [10,3 sec].

E:\Uldis\LU\LU2022\LU_konference_2022\Straujums_04-02-2022\ONTO6_framework\Onto6_stuff\LVTagger
>
```

Figure 4: LVTagger usage for CONLL-X format word list creation

The output of LVTagger serves as an input for process adding ontology class instances to an ontology. The adding of instances is done with the Cellfie Plugin for Protégé ontology editor.

5.2. Adding individuals to a class with Cellfie plugin

The list of individuals created by LVTagger can be easily converted to an Excel spreadsheet. An Excel spreadsheet can be given to the Protégé plugin Cellfie. The author has added individuals to the LIIS ontology with Cellfie (see **Figure 5**).

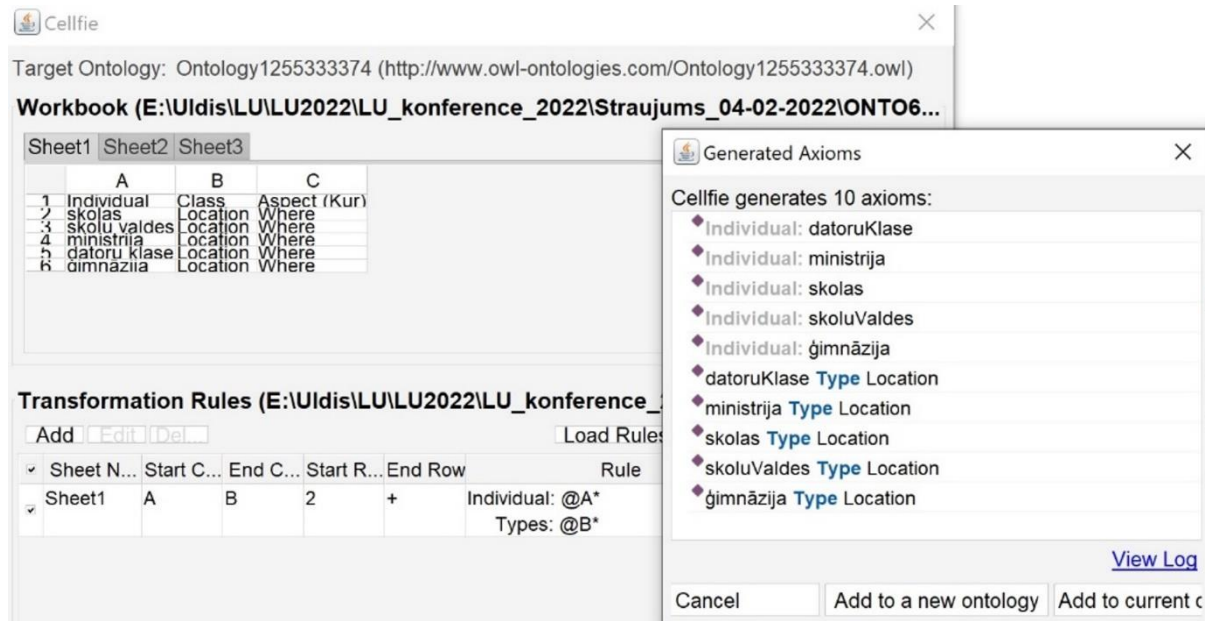


Figure 5: Adding individuals to an ontology with Cellfie plugin

6. Conclusion

The ONTO6 methodology has proven to be useful in situations where a compact view is desired of a complicated domain. It has shown itself to be well-suited to the development of a unified user understanding of the domain and for the creation of a description of the essential domain characteristics.

The ONTO6 framework will serve as a convenient way to apply the ONTO6 methodology.

7. References

- [1] U. Straujums, Conceptualising Informatization with the ONTO6 Methodology, in: volume 733 of Acta Universitatis Latviensis. Computer Science and Information Technologies, University of Latvia, Riga, 2008, pp.241-260.
- [2] U. Straujums, ONTO6 Methodology, Ph.D. Thesis, University of Latvia, Riga, 2010.
- [3] T. R. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing. KSL-93-04, Knowledge Systems Laboratory, Stanford University, 1993.
- [4] T. Gruber, Ontology. Encyclopedia of Database Systems, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2009. URL: <http://tomgruber.org/writing/ontology-definition-2007.htm>.
- [5] R. Wilson, The Role of Ontologies in Teaching and Learning. JISC Technology and Standards Watch Report TSW0402, 2004.
- [6] Ch. Welty, N. Guarino, "Supporting ontological analysis of taxonomic relationships." Data and Knowledge Engineering 39(1), (2001): 51-74.

- [7] T. Gruber, "Interview Tom Gruber." AIS SIGSEMIS Bulletin 1(3), (2004): 4-8.
- [8] P. R. S. Visser, V. A. M. Tamma, An experience with ontology clustering for information integration, in: Proceedings of the IJCAI-99 Workshop on Intelligent Information, Stockholm, Sweden, 1999.
- [9] Wordnet: An Electronic Lexical Database. Ed. by Christiane Fellbaum. Bradford Books, 1998.
- [10] N. Guarino, Formal Ontology and Information Systems, in: N.Guarino (ed). Formal Ontology and Information Systems. Proceedings of FOIS'98, IOS Press, Amsterdam, Trento, Italy, 1998. pp. 3-15.
- [11] B. Wyssusek, Ontology and Ontologies in Information Systems Analysis and Design: A Critique, in: Proceedings of the Tenth Americas Conference on Information Systems, 2004, pp. 4303-4308.
- [12] A. Browarnik, O. Maimon, Ontology Learning from Text, in: ALLDATA 2015: The First International Conference on Big Data, Small Data, Linked Data and Open Data, Barcelona, Spain, 2015.
- [13] Y. Wand, R. Weber. "An Ontological Model of an Information System." IEEE Transactions on Software Engineering 16(11), (1990): 1282-1292.
- [14] B. O'Donovan, D. Roode. "A framework for understanding the emerging discipline of information systems." Information Technology & People 15(1), 2002: 26-41.
- [15] P. Green, M. Rosemann, Ontological Analysis of Business Systems Analysis Techniques, Business Systems Analysis with Ontologies. UQ Business School, Australia; Queensland University of Technology, Idea Group Publishing, Australia, 2005.
- [16] M. Leppänen, An Ontological Framework and a Methodical Skeleton for Method Engineering. Helsinki, 2005.
- [17] M. Treimanis, ISTechnology – Technology Based Approach to Information system Development, in: Proceedings of the Third International Baltic Workshop "Databases and Information Systems", vol. 2, Riga, 1998, pp. 76-90.
- [18] D. Karagiannis, P. Burzynski, W. Utz, R. A. Buchmann, A Metamodeling Approach to Support the Engineering of Modeling Method Requirements, in: 2019 IEEE 27th International Requirements Engineering Conference (RE), Jeju Island, Korea (South), 2019, pp. 199-210.
- [19] IEEE Standard Glossary of Software Engineering Terminology. IEEE Computer Society. IEEE Std 610.121990, New York, 2002.
- [20] D. W. Robertson, Jr., "A Note on the Classical Origin of 'Circumstances' in the Medieval Confessional." Studies in Philology 43(1), (1946): 6-14.
- [21] Organizing Business Knowledge: The MIT Process Handbook. /Ed. Malone, Thomas W., Crowston, Kevin, Herman, George A. MIT Press, 2003.
- [22] J. F. Sowa, J. A. Zachman, "Extending and formalizing the framework for information systems architecture." IBM System Journal 31(3), (1992): 590-616.
- [23] John Zachman, A. Framework2. The Concise Definition, 2008.
- [24] SixWs. Online Encyclopedia Wikipedia, 2022. URL: http://en.wikipedia.org/wiki/Six_Ws.
- [25] Mindtools. Starbusting template, 2022. URL: http://www.mindtools.com/pages/article/newCT_91.htm.
- [26] Ju-Hung Lan, A Preliminary Study of Knowledge Management in Collaborative Architectural Design, in: CAADRIA2004, Seoul, Korea, 2004, pp. 35-47.
- [27] M. Yudelson, T. Gavrilova, P. Brusilovsky, Towards User Modeling Meta-Ontology, in: UM2005, LNAI 3538, Edinburgh, UK, 2005, pp.448-452.
- [28] J. Iljins, M. Treimanis, From Organization Business Model to Information System: One approach and Lessons Learned, in: 19th International Conference on Information Systems. Prague, Czech Republic, 2010.
- [29] P. Paikens, Latvian morphological tagger, 2022. URL: <https://github.com/PeterisP/LVTagger>.
- [30] Cellfie Plugin, 2022. URL: <https://github.com/protegeproject/cellfie-plugin/wiki>.