Knowledge base modeling and design procedure

Article · January 2012

DOI: 10.3233/978-1-60750-992-9-331

3 authors, including:

Ladislav Burita

University of Defence

29 PUBLICATIONS **6** CITATIONS

Knowledge Base Modeling and Design Procedure

Ladislav BURITA[1, 2], Petr DO[1], Vojtech ONDRYHAL[1]

[1]CIS Department, Faculty of Military Technology, University of Defence, Brno, Czech Republic
[2]IEaIS Department, Faculty of Management and Economics, Tomas Bata University in Zlin, CR

Abstract. *The article is based on the defense research project "Knowledge Management of the NEC in the Army of the Czech Republic – MENTAL". The theoretical basis of the project is Topic Maps. The key issue for the project solution is designing and creating a suitable ontology. The implementation environment is technology Tovek Tools and ATOM2. The paper describes the procedure from the selection of an Upper Ontology through the Core Ontology design to the processing of the Domain Ontology. Ontology definitions are stated and their meaning is explained. The paper next explains the ways reusing of an existing taxonomy in an ontology construction, presents the possibilities of using the taxonomy built in the selected domain when creating its domain ontology, explains the difference between taxonomies and ontologies in various contexts, and focuses on the description of the specific domain ontology and the use of the existing taxonomy for its building.*

Keywords. Upper-, Core-, and Domain Ontology, Knowledge Management System, MENTAL, Network Enabled Capability, taxonomy, ontology.

1. Introduction

The article describes the experience on the defense research project "Knowledge Management (KM) of the Network Enabled Capability (NEC) in the Army of the Czech Republic (ACR) – MENTAL" [1], which is the first research project dealing with the Knowledge Management (KM) at the Ministry of Defence (MoD) Czech Republic (CR).

The defence research project will result in a Knowledge Management System (KMS) MENTAL. The aim of the MENTAL is [1]: "To carry out the analysis of knowledge approaches, ontologies and ontology languages, and to assess their suitability for using them in the ACR; to evaluate the security state solution; to formalize the ACR NEC strategy and develop an encyclopaedia of NEC terms; to propose a methodology for knowledge systems development in the ACR; to elaborate a knowledge system proposal in the ACR NEC administration and to implement it." The accomplishment of the project is assured by successful cooperation of researchers from the University of Defence (UoD) with the TOVEK and AION CZ companies [4] which technology is used.

The most important activity concerning the knowledge-based system is the design and development of an appropriate ontology, which constitutes a formal framework for storing the knowledge, creating links between knowledge and ontology concepts, and establishing connections to concepts and pieces of knowledge of vital documents, which are connected with the area in focus. Ontology itself, without using the known definitions, can be considered an abstract model of a part of reality - domain for which the knowledge-based system is created. Part of the project is the validation of the methodology for ontology creation. One of the underlying methodological postulates for designing ontology is a logical procedure from an Upper Ontology through a Core Ontology to a Domain Ontology.

The significant project milestone towards creating of the knowledge base is considered domain ontology. A domain ontology building process could consist of the following stages:

- Design the principles of the ontology creation and its connection to other ontologies.
- Preparing a vocabulary of terms - clarification, sorting terminology of the given domain.
- Creating taxonomy in the given domain, i. e. finding relationships between concepts.
- Definition of classes (types, concepts), suggestion appropriate.
- Adding instances to various ontological types.

When building the NEC ontology, it is necessary to keep in mind that it is considered domain ontology, therefore the bottom-up or middle-out approaches are expected. When applying the middle-out approach, the ontology is built from the most frequent to the less frequent concepts.

2. Ontology Modeling and Design

2.1 Upper Ontology Selection

In information science, an Upper Ontology (Top-Level Ontology or Foundation Ontology) is an ontology which describes very general concepts that are the same across all knowledge domains. The most important function of an Upper Ontology is to support very broad semantic interoperability between a large numbers of ontologies accessible "under" this Upper Ontology. The following ontologies are now competing to be used as the foundation for standard [Wikipedia]:

- IFF Foundation Ontology.
- Suggested Upper Merged Ontology.
- OpenCyc.
- Lattice of Theories including the above and the 4D ontology based on ISO 15926.
- The Multi-Source Ontology.

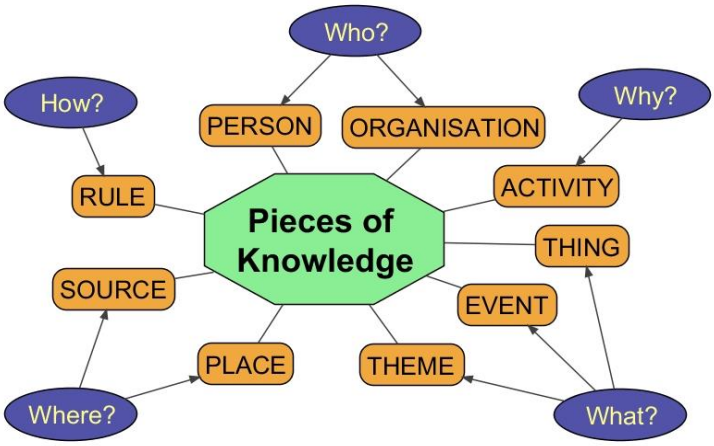

Figure 1. The Upper Ontology for MENTAL

Based on our analysis and recommendations by the cooperating companies, we have selected the Upper Ontology with the theme of Competition Intelligence for the MENTAL project. The general concepts here are PERSON, ORGANIZATION, ACTIVITY, RULE, SOURCE, THING, THEME, EVENT and PLACE; see Figure 1. This ontology corresponds with the NEC theme, which is being solved, and therefore with respect to the Upper Ontology, the ontology can be linked to all projects with a similar approach (Competitive Intelligence).

2.2 Core Ontology meaning and role

In philosophy, a Core Ontology is a basic and minimal ontology consisting only of the minimal concepts required to understand the other concepts. It must be based on a core glossary in a human language, so that humans can comprehend the concepts and distinctions made. Such a Core Ontology is a key pre-requisite to more complete ontology foundation, or a more general philosophical sense of ontology. Core Ontology is a concept that is used in information science as well [Wikipedia].

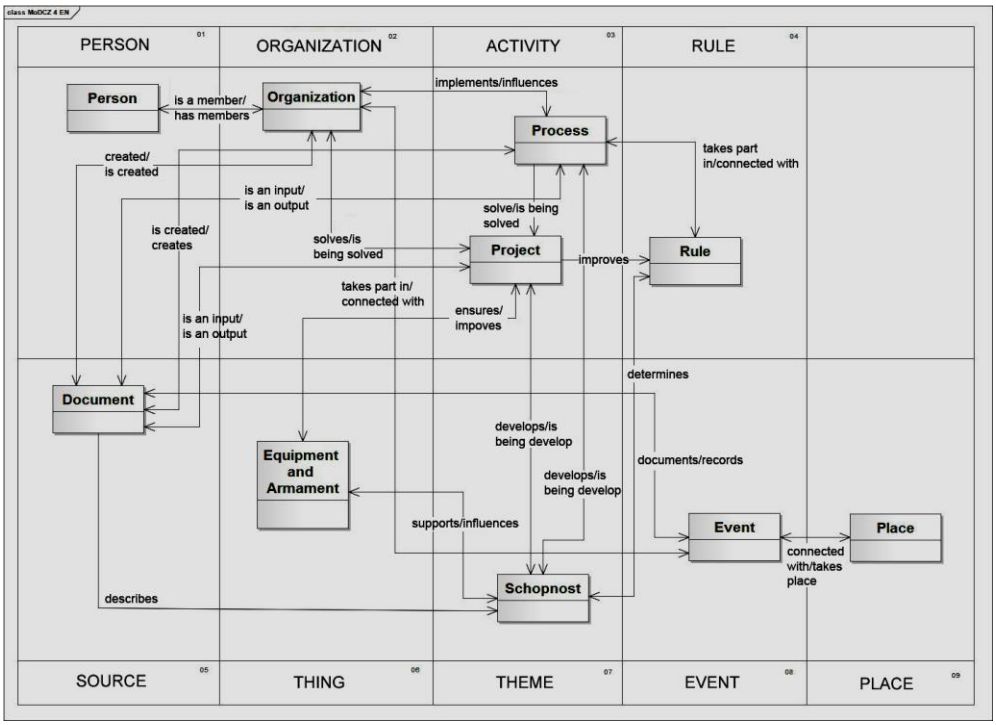

Figure 2. The MoD Core Ontology

Core Ontology has an important position in the interoperability area. It is a central ontology for systems that integrates many ideas from various points of view of the same problem. Other view on the Core Ontology corresponds with the work of representatives from various communities with the goal of harmonizing their knowledge perspectives [2]. The next solution of the Core Ontology is connected with the integration of dictionaries from many fields of the same theme, for example in medicine, in an attempt to find the core part that is the same (or similar) in all fields [3]. In the MENTAL project the Core Ontology is a general model of the military at the Czech MoD, see Figure 2. This ontology should integrate all ideas concerning knowledge management in Czech military area.

2.3 Domain Ontology creation

A Domain (or Domain-Specific) Ontology models a specific domain, or a part of the world. It represents particular meanings of terms as they apply to that domain. Since Domain Ontologies represent concepts in very specific and often eclectic ways, they are often incompatible. As systems that rely on Domain Ontologies expand, they often need to merge Domain Ontologies into a more general representation. This presents a challenge to the ontology designer. Different ontologies in the same domain can also arise due to different perceptions of the domain based on cultural background, education, ideology, or because a different representation language was chosen [Wikipedia].

The NEC Domain Ontology is based on the MoD of the CR Core Ontology. The methodology for creating the ontology should include a preparatory stage, in which a set of documents that sufficiently describes a given domain (document base) will be collected. At this stage, the project team members

were trained in the fundamentals of ontology, and tried to create a working version of their own ontology. Furthermore, it is necessary to clarify the basic concepts of the subject area in focus; for instance, by means of the analysis of the document base, which characterizes the selected domain. Basic concepts of the domain are arranged, e. g. into taxonomy.

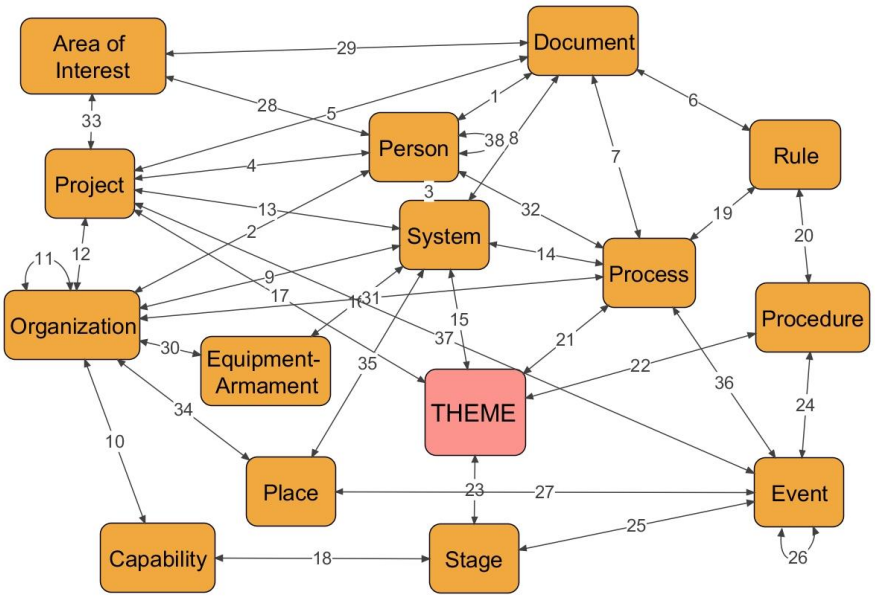

Figure 3. The NEC Domain Ontology – Classes and Associations

Taxonomy is a set of concepts, where concepts of higher levels can be further developed by concepts of lower levels. The depth of the hierarchical structure is set by goals for which the ontology is created. Our research team uses the TOVEK (www.tovek.cz) products for this purpose, especially Tovek Tools Analyst Pack. The document base should be put into a unified form, which assumes the selection of documents by language and format.

A design of an ontology scheme follows above mention analytic process; see Figure 3. A prerequisite to an appropriate ontology design is a good understanding of the subject area (domain) of the future KMS. This is an iterative "top-down and bottom-up" procedure which leads to continuous improvement of the original proposal. The main criterion for the quality of the ontology will be an effective and user-friendly knowledge application.

The ontology design contains a set of ontology classes: area of interest, project, process, document, person, organization, system, equipment and armament, theme, place, capability, stage, rule, procedure and event. Each class has its own definition and attributes that it characterize. The relations between classes (in the diagram in Figure 3 numbered only) and their names and meanings are described in a separate table and set of attributes is related to the classes in which are used, see Figure 4.

3. Taxonomy versus Ontology

Taxonomy is a set of controlled vocabulary terms, which are organized in a hierarchical structure. Each term is included in at least one parent-child relationship. Different types of parent-child relationship can be distinguished, such as whole-part, type-instance, is-a or genus-species. The last one is often used in taxonomy in biology. When building taxonomy, it is useful to set a rule that the parent is of the same type as descendants (this is not applicable to the whole-part type). In taxonomy it is also true that a child has only one parent. If there are more parents, the taxonomies are known as poly-

hierarchical. This approach is applied when a concept is found in several places in the taxonomy. If so, then it is understood as the same concept.

Attribute:			Concept (Class):										
No	Attr Name	Data type	PERSON	ORGANISATION	DOCUMENT	RULE	ARMAMENT	PROCESS	PROJECT	EVENT	PLACE	TOPIC	STAGE
1	name	text	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥
2	name in English	text		♥	♥	♥	♥	♥	♥	♥	♥	♥	♥
3	place	text			♥		♥			♥			
4	date	qq/yyyy			♥		♥	♥	♥	♥			
5	place - address	text	♥	♥						♥	♥		
6	address	text	♥	♥					♥	♥	♥		
7	www - url page	text	♥	♥	♥	♥	♥		♥		♥		
8	E-mail	text	♥	♥									
9	categogies	classifier	♥	♥	♥	♥	♥		♥			♥	♥
10	role in MENTAL	classifier	♥										
11	annotation	text	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥
12	annotation in English	text	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥
13	picture, scheme	picture	♥	♥			♥			♥	♥		
14	TTD	text					♥						
15	start (date)	dd.mm.yy							♥	♥			♥
16	end (date)	dd.mm.yy							♥	♥			♥
17	evaluation	text	♥	♥			♥	♥		♥		♥	♥
18	problems, risk	text	♥	♥			♥	♥		♥		♥	♥
19	taxonomy domain link	link	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥
20	taxonomy term link	link	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥	♥

Figure 4. The NEC Domain Ontology – Classes and Attributes

When building the taxonomy, the number of parents was reduced to one; for building hierarchies, the whole-part or part-instance types were selected. An example of the whole-part relationship is a representation of the organizational structure of departments or institutions; the type-instances relation can be traced in the processing of specific positions and persons in organizations. When building the taxonomy, relations that are not hierarchical were often accessible. These relations are also worth noticing, especially for further processing in the ontology building. A typical example is a simple related-to relationship, the one that is known from building thesauruses.

As already stated, the aim is to create ontologies in the NEC area, a controlled vocabulary which is expressed by means of an ontological language. Ontology thus contains a dictionary of terms that can be adapted from the taxonomy already built; the description of the meaning of the concepts in the context of the given domain, i.e. both among individual concepts and within the context of that domain as a whole [5]. It should be mentioned here that the ontology under construction is supposed to be the basis of the knowledge portal. In this case, the ontology does not necessarily need to be described by a formal apparatus, which ensures the possibility of computer processing and deriving new knowledge. Historically understood, taxonomy is in fact a special type of ontology (a lexical ontology) with a very limited set of rules, and therefore the taxonomy can serve as a knowledge base for the knowledge portal. Within the project, the ontology with a wider set of rules will be built. To ensure the formal correctness, the tools designed for ontologies editing are used, for instance the Protégé using OWL or Ontopia; or ATOM2, the formal apparatus of which is based on Topic Maps technology.

4. Transformation of taxonomy into ontology

This section briefs on the instances dealing with the transformation of taxonomy into ontology. The parts of the NEC taxonomy under development serve as examples. Besides, different types of hierarchical relations are analysed.

4.1 Type-instance relationship

An example of this relationship is a part of taxonomy recording the defence research projects.

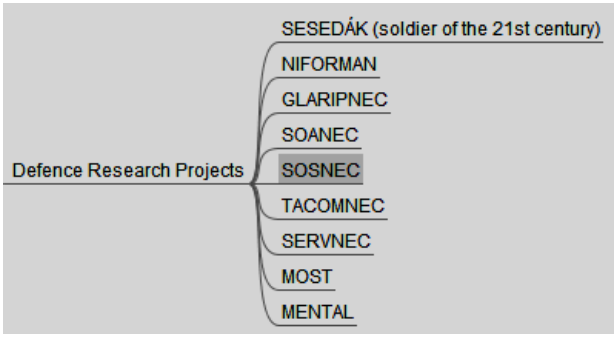

Figure 5. Type-instance relationship example

The term "Defence Research Projects" is broader than the term "TACOMNET". More precisely, "TACOMNET" is an instance of "Defence Research Projects" (Figure 5). In case that this type of hierarchy will be found in the taxonomy, it will be transferred into the ontology in such a way that the broader concept will likely become the basis of the ontology type and the narrower term will constitute one instance of the ontological type.

In the NEC taxonomy other examples of this relationship type can be found, e.g. Authorities of the Czech Army (with instances like NEC Control Board), Institutions in the Czech Army (Communication and Information Systems Division), Documents (NNEC Feasibility study), etc. The examples stated above indicate that suitable ontological types might be, for instance Project, perhaps even Defence Research Project, Authority, Institution and Document.

4.2 Is-a relationship

This type of relationship defines hierarchy as it is known from the object technologies, such as in case when a child has characteristics of an ancestor and new attributes are added. The following terms (Figure 6) in taxonomy serve as an example of such relationships. Each of the specific domains inherits the qualities of the concept at higher level in the hierarchy, i.e. the concept (Area-domain).

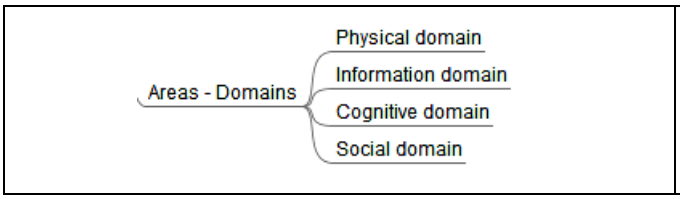

Figure 6. Is-a relationship example

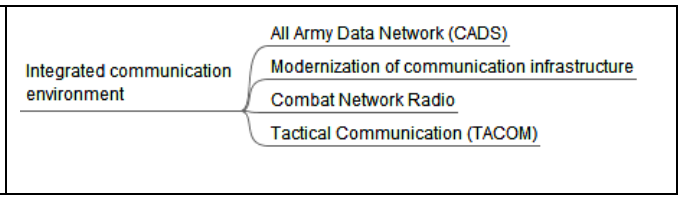

Figure 7. Whole-part relationship example

4.3 Whole-part relationship

Another type of relationship that can be traced in taxonomy is a situation where one concept is inherently included in another. In case the whole-part relationship is found in a hierarchy (Figure 7), it is not possible to transform it smoothly, and it needs to be further analyzed. Likely, the part-of relationship will occur in the ontology; however, it is necessary to assign individual terms in the ontology types correctly. Very likely, the parent and even individual descendants will belong to different ontological types. In developing the taxonomy it is appropriate to avoid this type of relationship; it is suitable to focus on type-instance or is-a relationships.

5. Implementation

Now, the way how taxonomy and ontology is implemented and interconnected in our domain of interest is described. All following examples are based on ATOM2 Topic Maps based system by company AION, the modeling part is done in Enterprise Architect; UML modeling tool.

5.1 Taxonomy

The taxonomy is based on a simple model (Figure 8). The topic *Term* can be accessed in two different roles, the *Broader term* and the *Narrower term*. Between the roles the hierarchy association is defined. Nevertheless, the types of hierarchies as described earlier in the paper are not modelled.

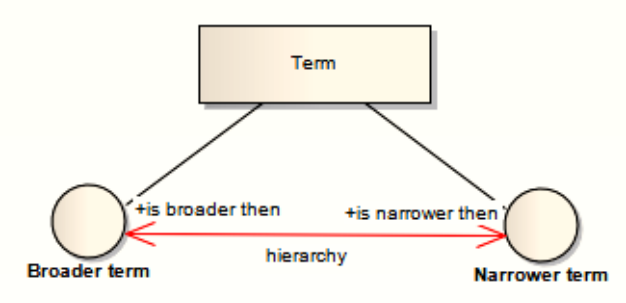

Figure 8. Taxonomy model

The actual implementation in the end user application is as follows (Figure 9). The term *NATO member country* is broader than individual countries like Albania, Belgium, Bulgari, etc.

Figure 9. Taxonomy implementation – NATO member country

The taxonomy implementation is quite simple and straightforward. On the other hand, the ontology implementation is complex; the number of classes and associations is plentiful.

5.2 Ontology

The ontology in our domain concerns all objects connected to NEC. On the Figure 10 is displayed part of the ontology concerning structuring of documents. This part of the model is selected for demonstration on purpose since the idea of document structuring can be reused in any ontology and is understandable across domains. The document structuring can be made using three classes:

- *Document* – is a main class that holds the whole document.
- *Document section* – is a part of document that can hold other sections or document content instances. The Document section class cannot hold document content itself; it is for structuring purposes only.
- *Document content* – instances of this class contains the document text and images. Document content can be associated with a number of other classes in ontology. Such associations hold the knowledge about the document theme.

The structuring of document is done using the two associations. Note that the role details in associations are stripped from the figure to make the figure more readable:

- *Document structuring* association splits the document into parts (sections and content).
- *Section structuring* association splits a single section into other sections or document content instances.

The classes can be seen in different roles based on associations. For example the *Document section* class can act as *Document part*, *Section container* or *Section part* in the ontology.

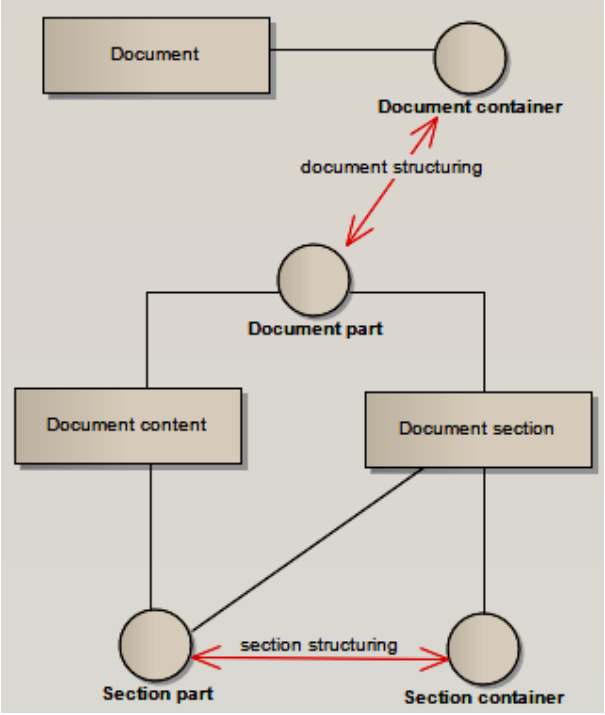

Figure 10. Part of ontology model

For example (Figure 11), the instance of class *Document* the *NNEC FS Executive summary* consists of the two document content instances (text icon) and the two document section instances (clip icon).

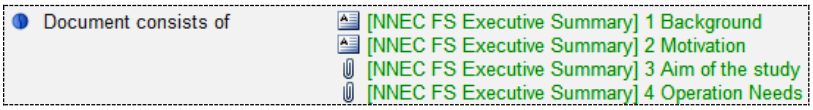

Figure 11. Part of ontology model

5.3 Taxonomy to ontology interconnection

Both the taxonomy and the ontology are implemented as independent structures within the knowledge database. A new *Ontology term* role to all relevant ontology classes is added; the ontology class with the *Ontology term* role can be interconnected with a matching term in taxonomy.

On the next figure (Figure 12) the selection of the three ontology classes (Document, Project and Person) are interconnected in the way described above with terms in taxonomy. Classes *Document section* and *Document Contend* are not selected for the *t-o interconnection* as they don't have the relevant instances matching any term in the taxonomy.

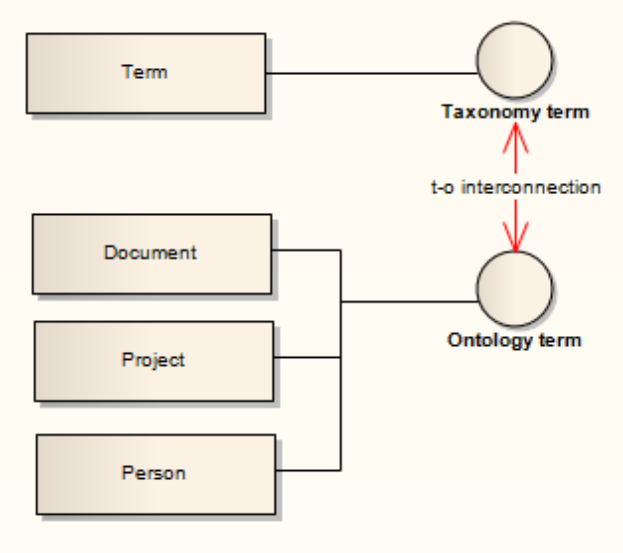

Figure 12. Taxonomy to ontology interconnection

6. Knowledge system

This part documents only the current achievements in the knowledge system (KS) construction. The KS that is being built on top of the ontology consists of two parts. The first part is based on common functionality provided by the supporting technology (e.g. ATOM or Tovek). The second part is created based on custom requests. The custom part includes functionality that is not provided by the supporting tool; it includes:

- Personalization,
- Communities of Interest support,
- Bookmarking and history for person, group, document.

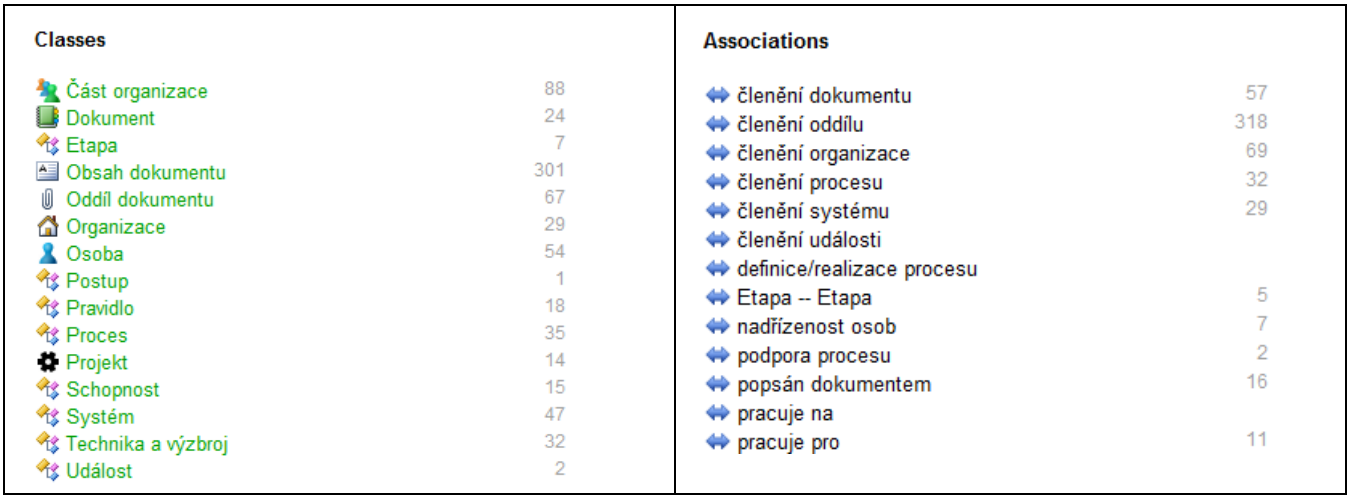

Figure 13. Ontology classes and associations (selection) implemented

The several following images display current stage of KS development based on the taxonomy and ontology. The system is being built for department in MoD of the CR; all views are available only in Czech language; we deeply apologize but we are not able, including the paper writing, transfer the knowledge system content (or part) into English language. The content is too complex and comprehensive result of the several team members.

On the Figure 13, the list of ontology classes and the number of instances is displayed; on the Figure 14, the entry point of the system. The NEC topic on the left (under group label) is the name of the entire ontology. The objects on the right are the main instances, like NEC Domains, NEC Capabilities, etc., of selected classes associated with NEC topic.

On the Figure 15, the example of topic detail is provided. The topic belongs into Document class, the document structuring, the document names and other links based on ontology are available. On the Figure 16 and 17 is an example of faceted search. On the first figure the keyword NATO is entered. The list of results is returned (on the right – 20 results). On the next figure, the result is filtered based on ontology classes and groups (all projects connected to search).

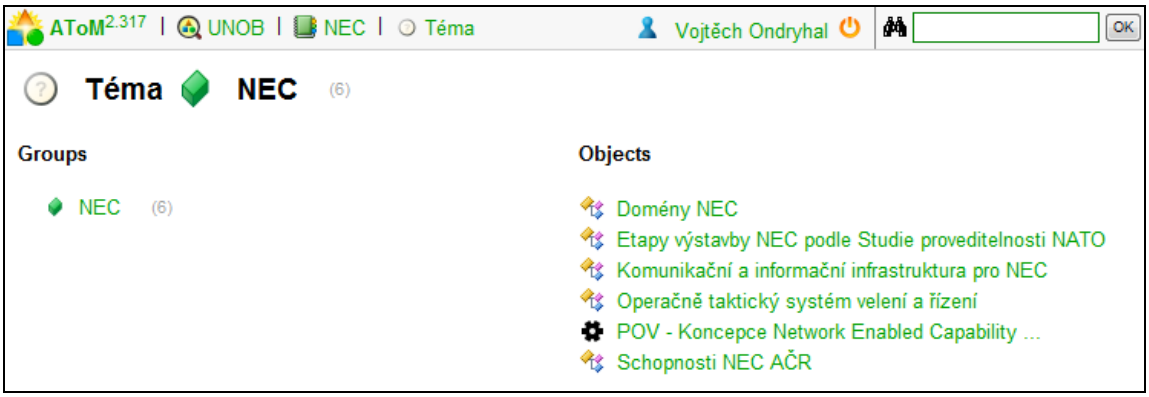

Figure14. Knowledge System Entry Point

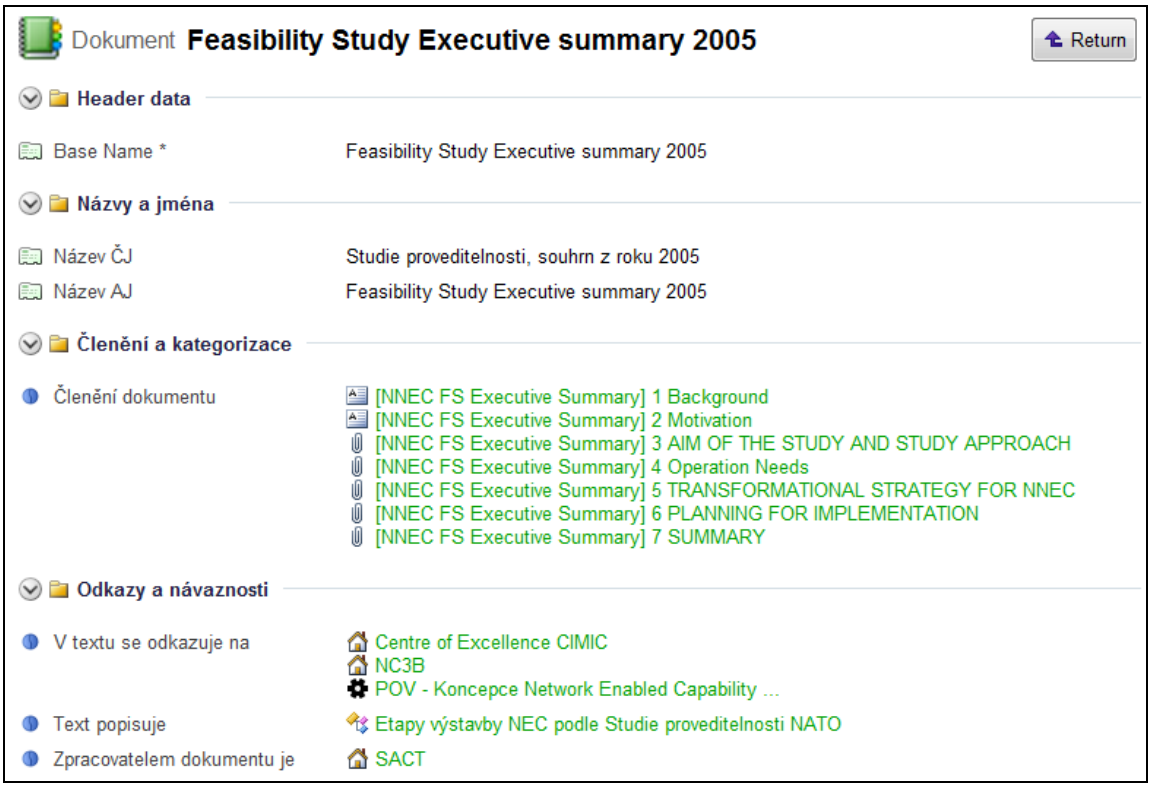

Figure 15. Example of the topic detail (selection)

7.Conclusion

The MENTAL project is still under progress. A significant part of the research task was finished and the same part is still to be developed. During the project development new theory, technology and tools were used. This is the first project dealing with the KM theme at the MoD. The aim of this paper was to point out the pitfalls in the ontology development, in the transformation from taxonomy into ontology using a instance of the knowledge-based system in the NEC domain. In retrospect, the taxonomy built from the perspective of the needs of ontology can be subjected to critical evaluation: the taxonomy often contains relationships that are difficult to be transformed into ontology, such as frequent use of the whole-part relationship. Nevertheless, it is obvious that the creation of taxonomy is a vital stage in the process of building ontology.

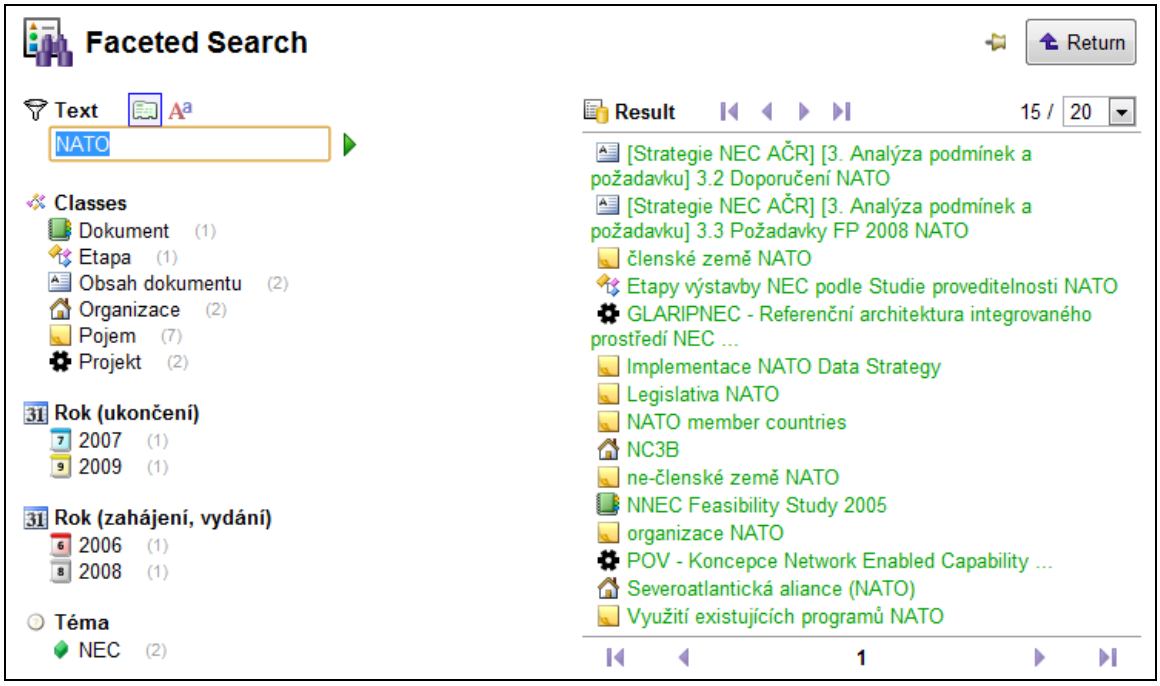

Figure 16. Faceted search – keyword search

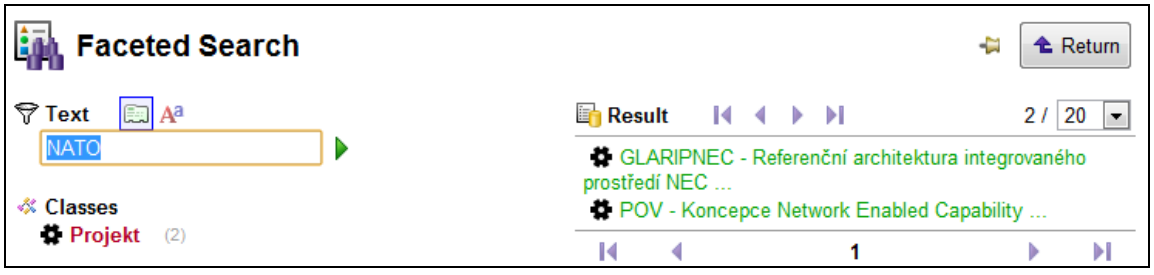

Figure 17. Faceted search – ontology based filtering

Acknowledgement

The article is prepared as a component of the Research Project [1]. It introduces some outcomes of the solutions in C2 IS interoperability field. Our results are part of the education process in the Network Enabled Capability course at the University of Defence in Brno and the Information Management course at Tomas Bata University in Zlin.

References

[1] *Documentation of Research Defence Project "The Knowledge Management (KM) of the Network Enabled Capability (NEC) in the Army of the Czech Republic (ACR) – MENTAL"*. Prague: MoD, 2008-2011, 32 s.

[2] DOERR Martin, HUNTER Jane, LAGOZE Carl. *Towards a Core Ontology for Information Integration.* Journal of Digital Information, Vol 4, No 1, 2003, 22 pp. Available on http://journals.tdl.org/jodi/article/viewArticle/92-Archiv

[3] CHARLET Jean. *The management of medical knowledge between non-structured documents and ontologies.* ANNALES DÉS TÉLÉCOMMUNICATION, 62, n 7-8, 2007, pp. 808-826.

[4] Information sources of the companies TOVEK, www.tovek.cz, and AION, www.aion.cz.

[5] USCHOLD, Mike, JASPER Robert. A Framework for Understanding and Classifying Ontology Applications. Proceedings of the IJCAI-99 workshop on Ontologies and Problem-SolvingMethods (KRR5) Stockholm (V.R. Benjamins, B. Chandrasekaran, A. Gomez-Perez, N. Guarino, M. Uschold, eds.), Sweden, August 2, 1999

[6] ALMEIDA, Jose Joao, SIMOES, Alberto. T2O — Recycling Thesauri into a Multilingual Ontology. In Proceedings of Fifth international conference on Language Resources and Evaluation, LREC 2006, Janova, Italy. May 2006.

[7] Using taxonomies to support ontologies. VocabControl [online]. 4.4.2010, [cit. 2010-09-14]. Available on WWW: <http://www.vocabcontrol.com/?p=210>.

About Authors

Ladislav BURITA is a senior lecturer at the CIS Department, at the Faculty of Military Technology (FMT) University of Defense (UoD) in Brno, Czech Republic. Since 2007 has been a member of Thomas Bata University in Zlín, Faculty of Management and Economics. Finished his academic studies in 1985; became an associate professor in 1991 and a professor in 2003. A member of the UoD and FMT Academic Board, a member of the MoD Board for Defense Research; has been in charge for the Defense Research Project MENTAL; has published several university textbooks and books in the fields of informatics, interoperability and project management; publishes papers and gives presentations at national and international conferences.

Vojtech ONDRYHAL is a senior lecturer at the CIS Department, at the FMT UoD. He received his Ph.D. from the University of Defence in 2010. His research interests include database systems and data mining, information and knowledge systems modeling and web technologies.

Petr DO is a Ph.D. student at the CIS Department FMT UoD. Graduated from the UoD, Robotics and Cybernetics, 2005; Czech Technical University, Electric Power Engineering, 2008; UoD CIS, 2008.