

## SEMANTIC WEB RELATED TO ITS SYSTEMS

*Abstract: This paper describes a way to better navigate in an ever increasing number of standards in the field of ITS. Short overview about fundamental ITS models is presented, the principles of related knowledge system are described. Ontology approach is used, especially as domain-oriented ITS ontology, where a new hybrid method is explained.*

**Keywords:** ITS, standardization, ontology, knowledge system, knowledge unit.

### 1. Problem declaration

The elaboration of ITS (Intelligent Transport Systems) standards strongly supported by common European policy is relatively successful. It is possible to identify about 180 active standards within European standardization committee CEN/TC278 at present. Together with ISO/TC204, there are more than 310 standards. Some of them are being elaborated, some are under revision and a significant part of standards has been finished. They are focused on the different market applications, starting from CEN/TC278/WG1 “Electronic Fee Collection” (EFC) up to the ISO/TC204/WG16 “Wide Area Communications/Protocols and Interfaces” (CALM) communication standards as examples.

There are two basic problems limiting successful utilization of ITS standards in real praxis. They are very complex (in general, a standard often has almost 100 pages). There are more or less 20 000 pages of ITS standards, which is a huge quantity for comprehensive reading. The standards are frequently written in the Universal Modelling Language (UML) or eXtensible Markup Language (XML) conventions and they need an advanced reader. Even for experts in ITS standardization it is often difficult to follow such a complex set of standards. Generally said, standards are not readable at all, it is a problem especially for investors and decision makers.

This article gives short overview about fundamental models describing complex ITS systems. Briefly describes principles of knowledge system developed in the frame of ZNALSYS project supported by Ministry of Transport of the Czech Republic. Main focus is on ontology; especially domain-oriented ITS ontology. New hybrid method is explained. It uses not only terms creation based on different ITS dictionaries but ontology is completed by terms coming from object-oriented models of standards describing processes within them.

### 2. User-oriented web applications

The fundamental goal of presented research is linked to end user who needs to find out appropriate information within set of standards. The ontology enables to create semantic meaning of the term which is under his/her consideration. Web searching mechanism looks for these qualified questions. Their quality is done by the quality of ontology.

The aim of the user-oriented web application is to satisfy users and to ensure their positive experience with the application. Reference [1] shows that the users never spend more than 50 % of their on-line time with any web application. Any user is more satisfied if he is carried out through given subject by fixed rules. Domain-oriented ontology provides a fixed taxonomy with well-defined semantic meaning of terms as shared framework of standards. Web-oriented search system was tested in practice in the frame of ZNALSYS research project supported by Czech Ministry of Transport. Brief look at the semantically oriented web-searching system is presented in chap. VII.

### 3. Model of ITS systems

ITS systems (IT systems generally) are modelled by data models. These models identify and describe a part of real word formally (with help of mathematics, formal text, graphs ...) using terms like entities, relationships etc. Data models are created on some abstraction levels:

- *Conceptual model* (which is also called the semantic one) is situated at the highest level. The model is not dependent on the specific programming language or software/hardware platform. Crucial entities and relationships are identified in this model. The typical representative is just E-R model (Entity-Relationship model).

\* Pavel Pribyl<sup>1</sup>, Vit Fabera<sup>2</sup>, Vladimir Faltus<sup>2</sup>, Lukas Tyfa<sup>1</sup>

<sup>1</sup> Department of Transporting Systems, Faculty of Transportation Sciences, Czech Technical University, Prague, Czech Republic, E-mail: pribyl@fd.cvut.cz

<sup>2</sup> Department of Transport Telematics, Faculty of Transportation Sciences, Czech Technical University, Prague, Czech Republic

- *Logical model* is expanded conceptual model and it is created by adding specifications of entities, attributes and relationships. Ranges of values and parameters of attributes are specified. The model is still independent on the technological platform.
- *Physical model* is adapted to the realization. Structures suitable for implementation are created and technical details are added to realize IT.

Three well-known data models are:

- *E-R model (Entity-Relationship model)* is a conceptual model used in the database area. They are described by E-R diagrams. E-R diagrams contain entities, relationships and attributes. Top-down method is usually applied when E-R model is created (modelled subject is decomposed into smaller parts until indivisible entities are identified).
- *Relation model* is derived from E-R model. All entities and relations are implemented by (database) relations. Relational algebra defines operations over relations.
- *Object-oriented data models* are used to describe real world systems when objects are searched to be similar to objects in real world. Object is a real or abstract entity containing data (attributes) and descriptions how to manipulate with them (methods). The object is defined by its state, behaviour and its identity.

Object-oriented models are notated with UML. UML uses structure diagrams (Class diagrams, Component diagrams ...) and behavioural diagrams (Activity diagrams, Use case diagrams ...).

In connection with other considerations, we will stream to create a semantic domain-oriented model of ITS. To do this exactly we will use object-oriented models to be more precise.

#### 4. Role of ontologies

The ability to search and fuse information from heterogeneous standards significantly contributes to the discovery of added value knowledge that is unreachable using classical searching methods. In order to provide an efficient information mining from standards a knowledge-based model is an optimal solution. Using an ontology approach, a coherent, consistent and non redundant knowledge model could be designed.

In information science, ontology is defined as a formal representation of knowledge as a set of individuals (instances, terms), classes (concepts), attributes, and relations between those concepts. Instance depicts the basic object and it is indivisible. In the sense of the paper an instance unit of text in an extract could be represented as a semantically uniform term. It is also possible to speak about hierarchical categorization which describes the same model as the ontology is.

In theory, an ontology is a "formal, explicit specification of a shared conceptualization", [2], expressed as a shared vocabulary, which can be used to model a certain area. There are universal ontologies, as for example SUMO (Suggested Upper Merged Ontology). SUMO ontology creates high-level ontologies for different

domains – finance, computers, geography, transport etc. and it is too generic to help us mining appropriate information from standards.

It is possible to divide ontologies into following groups:

- *Terminological ontologies* – they are used especially in librarianship and in next regions oriented to text sources. Terms play crucial role, relations carry taxonomies, i.e. specify relationships of general or special terms.
- *Information ontologies* – they are superstructures over relational-database sources; they ensure conceptual abstraction level necessary to conceptual querying.
- *Knowledge representation (KR) ontologies* – classes and instances are systematically defined using formal language.

According to the subject of formalization it is possible to use next classes:

- *Domain ontologies* – are the most used ones, they are focused on the specified area;
- *Generic ontologies* – they describe general concepts and relations across several areas;
- *Task ontologies* – ontologies that (unlike the other ontologies covering the state) are focused on deduction processes, diagnosis etc.;
- *Application ontologies* – ontologies that are focused on the specific application, they involve both domain and task ontologies.

In the following consideration we will present methodology how to create domain-oriented ontology focused on ITS systems.

##### A. ITS standards and ontologies

The design of methodology should start under assumption that it is possible to define the ITS systems, including their boundaries over which it is realistic to create ontology. Selected assumptions that influenced the development of methodology are summarized and commented in the following text:

- Analyses show that no complex domain ontological model has been created for the present to conceptualize terms over all ITS area. It is not realistic to finish it in next years due to complexity and dynamics of ITS field.
- Next, generic ontology is too general to be used practically.
- Domain and application ontologies can be used in practice. Ontologies are known to be used in the area of location-based services, [3] or to analyze research documents, [4].
- Analyses created within ZNALSYS project show that it is better to focus on domains which are described by generally accepted form. ITS standards have this form.
- The advantage of ITS standards is their similar syntactic form (structure) – the documents are more or less formalized and have therefore formally similar structure.

Created ontology only describes a fraction of hierarchy of concept of ITS terms in ZNALSYS project. Tree structure of this hierarchy captures subordination of concepts but no additional relations and no additional assertion about concepts and terms. Just relations and terms in the first-order logic are basics of knowledge.

### 5. Domain-oriented ontology – new approach

Authors describe several methodologies how to construct ontologies, [5]. Presented methodologies both depend on areas to cover by ontologies and on the way how the reality is informally described. The METHONTOLOGY, [6], is a suitable technique for our purposes. This methodology describes the ontology creation in eight tasks (steps). The first four tasks represent coarse steps of the ontology creation process, the tasks number 5 – 8 determine how to define relations, attributes etc. in details.

The first possibility is to build a glossary of terms to be included in the future ontology. This task has been already done for ITS because the Czech and English explanatory dictionary of ITS terminology had been published, [7]. It will also be able to extract Concept taxonomies (task 2) and ad-hoc binary relations (task 3) from this dictionary.

ITS Standards are specific and well-structured documents. Our idea is to find sets of typical (and analogous) statements in ITS standard and then to collect a set of rules how to transform these typical statements into first order logic sentences.

There is a wide discussion within literature whether is better to create ontology based only on “terminological” procedure (as described above) or is it better to use object-oriented models. The newly proposed procedure is not based on the approach “either/or”, but it is complementary approach based on hybrid model. It uses both components but in the three stages:

- (A) *The basic ontology* is created by using terms which are usually defined at the beginning of the standard. They can be corrected and supplemented according to technical dictionaries. The output from this stage is hierarchically organized tree of terms.
- (B) The standards are described as *object-oriented models*, i.e. for a specific process described in a standard the model is created using object-oriented diagrams. Expert or knowledge engineer decides whether to use object-oriented models and which kind of model will be used (structure diagrams and/or behavioural diagrams, e.g. class diagrams, activity diagrams, sequence diagrams, etc.). These models reveal number of new terms.
- (C) Basic ontology (A) is completed by new terms generating by object models (B) and the semantic meaning of all terms is unified.

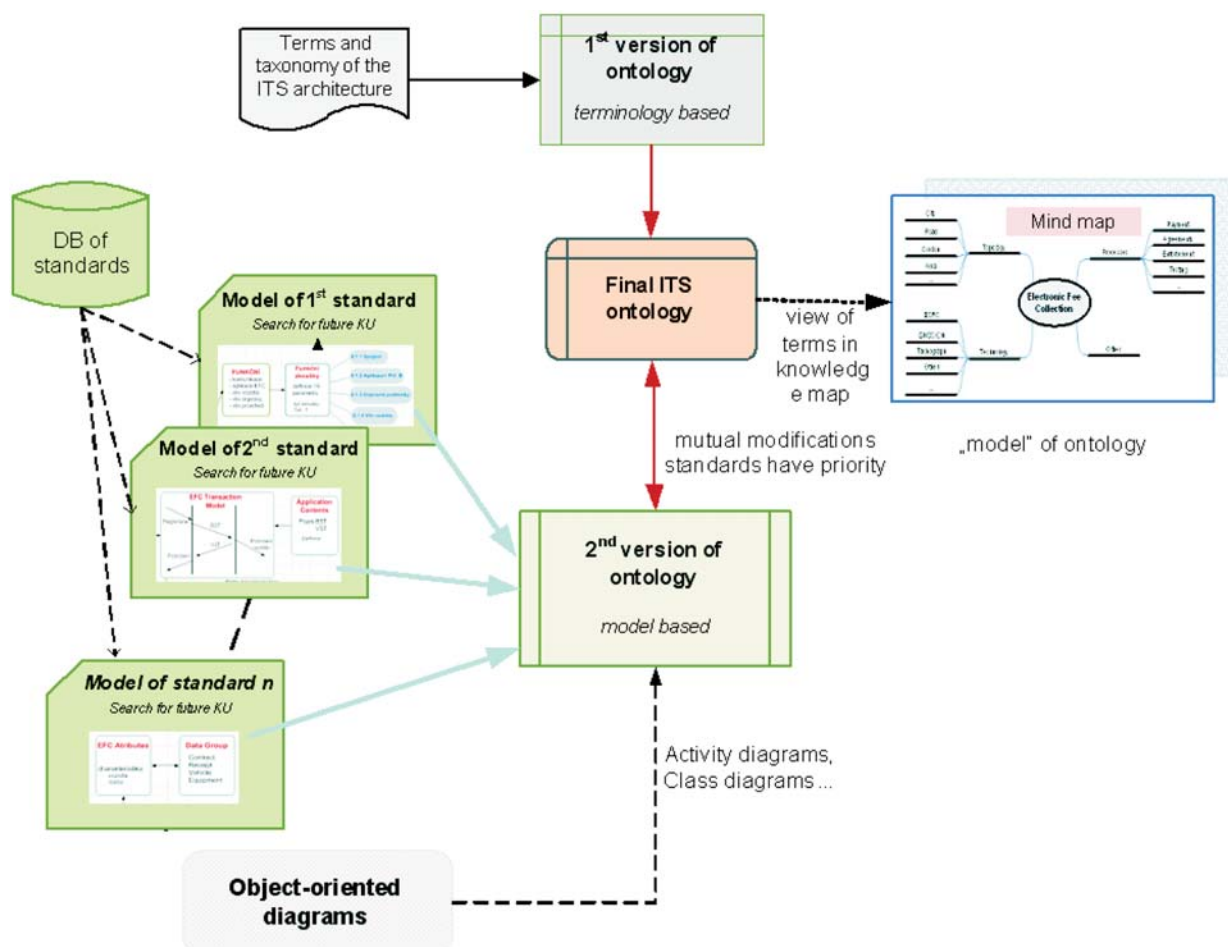


Fig. 1 Hybrid model of creation of domain ontology for ITS

The model created according to this algorithm is valid generally, is well-comprehensible and verifiable. It covers ITS domain more complex by unmasking terms which are not present in primary model (created ad-hoc). Newly found terms are subsequently included in knowledge map. Knowledge map creation is not necessary but its graphical form makes possible back verification of hierarchy and the completeness of the semantic model.

The creation of the basic ontology is depicted in Fig. 1. First stage uses terms defined in standards or in dictionaries. There are minimally two disadvantages: terms coming from standards or dictionaries did not cover the whole area and in addition they have not any hierarchical structure. The structure has to be created more or less intuitively. The advantage is to acquire many new terms. Using Mind Map graphical tools helps to prepare best possible hierarchy and better understand a structure of a system.

Second version of ontology is elaborated on the basis of models of standards. Models are usually created with help of CASE tools, for example Enterprise Architect. It is possible to use static description like Class diagrams or dynamic description (Activity diagrams, State diagrams). The deployment of knowledge expert is necessary to build up suitable diagrams.

On the other hand, these diagrams can detect many disadvantages in created ontology so we consider adding, verify or correct terms on the basis of object-oriented model (OOM) in these hybrid models. If it is very difficult to create OOM due to complexity of concrete standard it is recommended to use different

kinds of simplification. This approach is possible to see in Fig. 1. As a result of this stage is that the missing terms are added into ontology.

### 6. Test of proposed method

In the verification stage of proposed method, which should take in account both the above approaches (see A and B) a few standards will be analyzed, i.e. the standards of group CEN/WG1 “Electronic Fee Collection”: 14906 “Application interface for DSRC”, 14907-1, 2 “Test procedures”; 17574 “Safety frames” and 17575 “GNSS interface”. Only first two will be explained more in detail.

#### A. Domain ontology - terminology based approach

As primary source of terms, design of basic ontology uses standards which cover domain of our interest. Keywords are listed at the beginning of standards. Very good source is also technical report ISO/DTR 14812 “ITS glossary of standard technologies for the transport information control sector”. This rather older document (2002) could give good overview on 350 pages.

The basis for Czech ITS domain-oriented ontology is “Terminological dictionary” prepared by SDT (Traffic Telematics Association) which include several hundred (450 pages) of terms organized in the clusters according to WG in the CEN committee. Fig. 2 depicts small part of the chapter “Electronic fee collection”.

English terms and definitions	Termíny a definice
<p><b>B.12 apportionment</b> allocation of money to transport service operators according to the consumption of the services provided, e.g. a bus operator being paid an amount based on the number of a particular type of customer carried ENV ISO 14904</p>	<p><b>B.12 rozdělení/přidělení/dávkování</b> přidělení peněz dopravním operátorům v závislosti na poskytovaných službách, např. provozovatel autobusu je placen dle množství přepravených cestujících</p>
<p><b>B.13 architecture boundary</b> divides the interface classes from those classes, which form the actual architecture, namely the control classes and the information classes; in the sequence diagrams developed in later clauses there is often an implicit interaction across the system boundary involving an actor; this may be implied whenever a message is initiated or terminated at an interface class ISO 14813</p>	<p><b>B.13 hranice architektury</b> hranice rozděluje třídy rozhraní tvořící aktuální architekturu, a to na řídicí a informační; v sekvencním diagramu často předpokládá vzájemné působení přes systémová rozhraní zahrnující nějaký aktor; to nastává vždy, když je zpráva inicializována nebo ukončena na rozhraní (hranice systému) viz K.3</p>
<p><b>B.14 architecture element</b> a definable element of a system, which forms part of a component or system, but does not necessarily have independent operational functionality ISO 14813</p>	<p><b>B.14 prvek architektury</b> definovaný prvek systému, který tvoří část systému a nemusí mít nutně nezávislou provozní funkčnost viz K.4</p>

Fig. 2 Terms and their definition (semantic meaning) also in Czech language

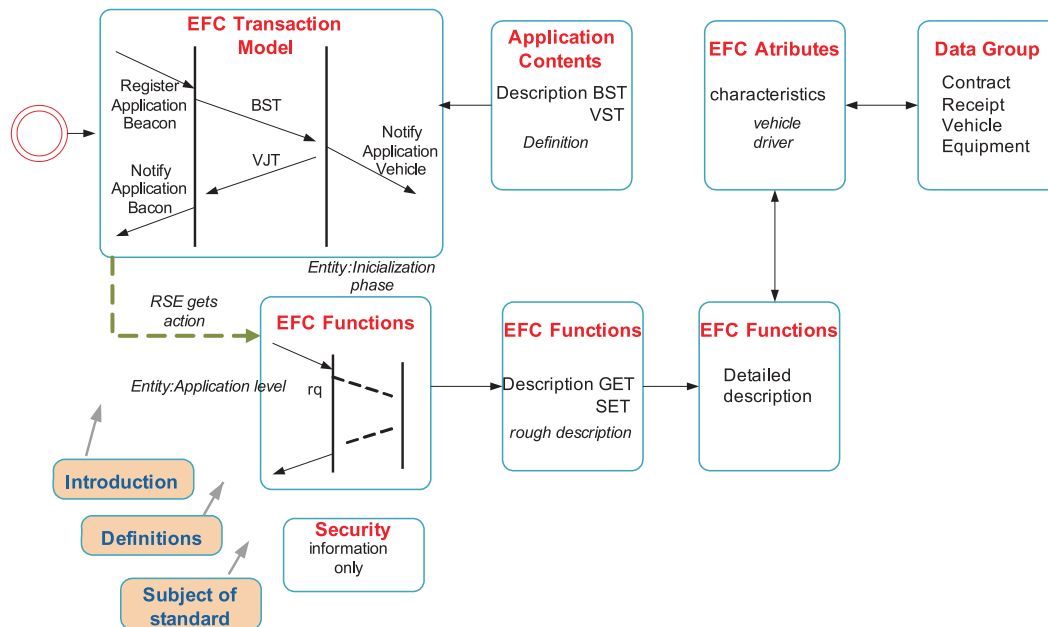


Fig. 3 Structure and logical relation as model of 14906 (product-oriented standard)

Practical experience shows that terms are usually (significantly) different from the following aspects:

- level of detail (while one term describes insignificant detail, other may describe system properties);
- frequency in the documents (importance);
- semantic meaning (that is often different according to the context).

*B. Domain ontology - object-oriented model of standards*

This chapter clarify topic B - “The standards are described as object-oriented models”. ISO DIS 14 906 and 14705-1 will demonstrate the process of simple object-oriented model. The aim is to process simplified model of standard, in the first stage as model of classes as it is known in UML.

In order to distinguish the differences of subjects among multiple standards, the standard 14906 is first analyzed, which almost exclusively addresses the manner and form of communication between on-board unit (OBU) and the infrastructure facilities (RSE). Standard 14 705-1 is focused only on testing of OBU and RSE.

A closer look at Fig. 3, it is evident that the connection is first initiated by “EFC Transaction Model” block whose parameters are transmitted in the “Application Content” block. After establishing and confirming the connection, the data exchange itself follows, while a rough diagram of the transaction is in the first “EFC Functions” block, followed by two blocks of the same name with more and more detailed description of the transmitted data.

“EFC Attributes” block describes parameters that are transmitted in data files and “Data Group” block clusters it into blocks.

Relatively small and unspecific mention is about security (“Security” block).

Discussion to “model” of standard 14906:

- This simplified model (more activity diagram than class diagram) gives a good idea about the possible creation of knowledge-based segments linked to appropriate terms.
- There is a problem with a link to ontology - in terms of creating entities in a real standard there is necessary to logically identify potential knowledge unit (KU) and a KU with unambiguous interpretation of the meaning ⇒ it can be inferred that the “blind” creating of model of standard without accordance with pre-built ontology or real needs would not lead to the goal of developing the most competent ontology.
- In this configuration and at this level of resolution it is necessary to still think how competency questions may be formulated. Respective terms need to be targeted to help to prepare competitive question.

Further examined standard, see Fig. 4, describes the test procedures of dedicated short-range communication. The structure is clear, there are three parallel testing procedures consisting of practically the same blocks.

Discussion to “model” of standard 14705-1:

- Model is similar to Classes diagram, but it is not the same. Terms linked to testing can be formulated very well at the highest level.
- If it is necessary to expand a list of terms and to go more in detail, more detailed descriptions of each test will be needed and detailed model shall be formed as Fig. 4 shows.
- Next group of terms can be directed to the requirements for certification and documentation requirements for DSRC testing.

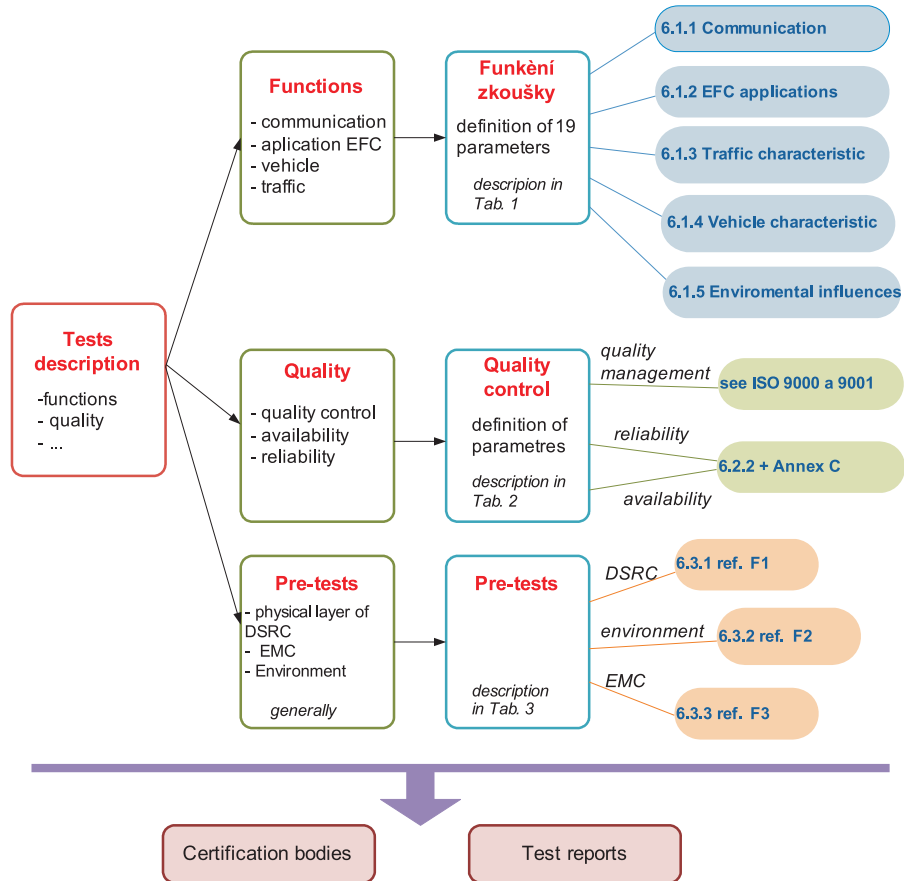


Fig. 4 Structure model of 14 705-1 (standard for testing)

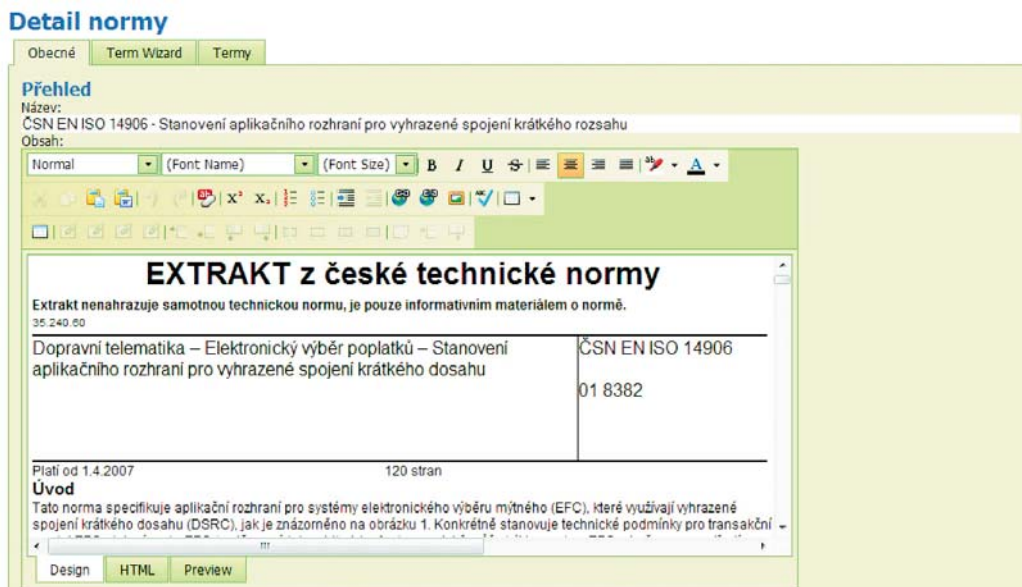


Fig. 5 Example of knowledge unit provided by Czech web application

- The disadvantage of this simple model is that the created ontology is rather general. Nevertheless to go more in a detail within different tests described in standard will significantly extend levels of ontology.

## 7. Pilot application of semantic web

The basic aim of the pilot application was to present knowledge units (well defined short texts of standard linked to ontology) for selected categories of users. The web application promoted the use of knowledge units (KUs) in ITS standards, see Fig. 5, so the ontology model and its units (terms) have offered not only the definitions, but also the links to the knowledge units through competitive questions. The end user found out a set of KUs according to his questions and KUs gave him overview about a problem which is under solution.

The web-oriented search engine was worked out to mind out much focused knowledge contained in the standards. The result was in a form as an on-line application with the possibility to start with a term or uncertain need and to end with a set of knowledge units coming from the source standards.

The web application promoted the use of ITS standards on a simple way available for non-experts too.

## 8. Conclusion and next works

The article presents the hybrid model of creation of domain-oriented ontology of ITS systems. The basic idea is built on ITS standards. These standards cover significant part of ITS applications. Standards are also very proper due to their context form which is very similar also for very different applications.

In theory, an ontology is a “formal, explicit specification of a shared conceptualization”, expressed as a shared vocabulary, which can be used to model a certain area. ITS systems are usually connected through data interfaces and it works quite well. On the other hand there is not commonly used ontology covering this area. It means that there is no interface on semantic level.

Presented hybrid model uses two basic steps. The first one, more traditional, creates fundamental ontology coming up from different technical dictionaries containing appropriate terms of discussed domain. This basis is completed and verified by new terms derived from object-oriented models within step two. In terms of finding out terms which could complete basic ontology, the simplified models of standards are developed by experts.

The idea of this approach is under discussion with Japan. The ministry MLIT provides some financial support to bi-lateral project oriented to formulate ontology of ITS systems. Czech experts will discuss next development of proposed methodology in Tokyo in April this year.

## References

- [1] KREMENOVA, I., GALOVIC, M.: Proposal of Key Factors for User-oriented Web Application, *Communications - Scientific Letter of the University of Zilina*, ISSN 1335-4205, 2, 2010, pp. 72-75.
- [2] GRUBER, T.: *A Translation Approach to Portable Ontology Specification*, Knowledge System Laboratory - Stanford University, CA, Technical Report KSL 92-71, 1992, pp. 192.
- [3] PFOSE, D., TRYFONA, N.: *The Use of Ontologies in Location-based Services: The Space and Time Ontology in Protege*, Research Academic Computer Technology Institute, Research Unit 3, Athens, Greece, 2002, pp. 15.
- [4] WENG, S., CHANG, H.: Using Ontology Network Analysis for Research Document Recommendation, Science Direct, *Expert System with Applications* 34, 2008, pp. 1857-1869.
- [5] PEREZ, A., LOPEZ, M., CORCHO, O.: *Ontological Engineering*, Springer, London, ISBN 1-85-233-551-3, 2004, pp. 403.
- [6] LOPEZ, M., PEREZ, A., JURISTO, N.: *Methodology: From Ontological Art Towards Ontological Engineering*, Spring Symposium on Ontological Engineering of AAAI, Stanford University, California, 1997, pp. 33-40.
- [7] PRIBYL, P. et al.: *Terminological Dictionary of Traffic Telematics*, SDT, Praha, 2006, ISBN 80-239-7780-6, 2006, pp. 480 (Czech-English).