Article

Knowledge System Supporting ITS Deployment

Pavel Přibyl¹, Aleš Janota^{2,*} , Juraj Spalek² and Vladimír Faltus¹

¹ Faculty of Transportation Sciences, Czech Technical University, 110 00 Prague, Czech Republic; pribypav@fd.cvut.cz (P.P.); faltus@fd.cvut.cz (V.F.)

² Faculty of Electrical Engineering and Information Technology, University of Žilina, 010 26 Žilina, Slovakia; juraj.spalek@uniza.sk

* Correspondence: ales.janota@uniza.sk; Tel.: +421-415133356

Abstract: Intelligent transportation systems are one of the most rapidly evolving areas, requiring an appropriate response from standardization bodies and adequate support from EU regulations. This results in a high and ever-expanding volume of technical standards, which makes their practical use difficult and the harmonization process unsustainable. Therefore, standardization bodies are officially required to supply public information, based on which potential users can decide whether they need to buy or use a particular standard or not. The authors demonstrate how to solve this problem and achieve sustainability in terms of operating in a more intelligent and efficient manner. The proposed solution relies on the creation of standard extracts using a hybrid method that combines syntactic and semantic analysis and assumes human expert involvement. The paper presents the practical experience and results obtained from a long-term national project. A practical example is included so that the reader can comprehend a basic idea of the achieved results. The authors believe the proposed method can be adopted across other professional domains and other European countries.

Keywords: standardization; intelligent transport system; extract; language; hybrid analysis; project

Citation: Přibyl, P.; Janota, A.; Spalek, J.; Faltus, V. Knowledge System Supporting ITS Deployment. *Sustainability* **2021**, *13*, 6407. <https://doi.org/10.3390/su13116407>

Academic Editor: Armando Carteni

Received: 30 April 2021

Accepted: 23 May 2021

Published: 4 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

We have all probably heard, through various media, about autonomous and automated vehicles, cooperative technologies, cyber-physical systems, smart cities, artificial intelligence, and other emerging technologies. These systems create opportunities for integrated mobility in our cities and urban environments. According to [1], intelligent transportation systems (ITS) describe “transportation systems where vehicles interact with the environment, and with each other, to provide an enhanced driving experience, and where intelligent infrastructure improves the safety and capacity of road systems”. The more recent CEN/TC 278 definition refers to efforts to collect, store, and provide real-time traffic information to maximize the utilization efficiency; provide convenient, safe transport; and reduce energy by applying advanced electronics, information, and telecommunication technologies in roads, automobiles, and goods [2]. We must not leave these systems to chance and innovation alone, without regulation and standardization. The potential of ITS can only be realized if its deployment is transformed from the current limited and fragmented implementation into EU-wide or worldwide implementation [3]. Within ITS standardization, there are technical committees, some of which are of special interest for the European domain. These are CEN TC 278, ETSI TC ITS, and ISO TC 204. Of these, the European CEN TC 278 and ETSI TC ITS are of particular interest because they have a special focus on European legislation, although major work is also developed in ISO TC 204, which works jointly with CEN TC 278. The ITS Coordination Group (ITS-CG) between CEN and ETSI was established to ensure ongoing coordination of the standardization activities within these two standardization developing organizations (SDOs). ISO, IEC, and ITU are global SDOs who standardize ITS on a global level. Many of the working groups (WGs) in CEN TC 278 overlap with WGs in ISO TC 204. Joint meetings between

CEN 278 and ISO TC 204 are held twice a year in order to harmonize and achieve good and fruitful cooperation.

Standardization dramatically affects life as it leads to progress, shapes technology, defines the terms of commerce, sustains our environment, and makes us safe. Standardized technological solutions make life easier because they can ensure simplification, interchangeability, communication, the use of proper symbols and coding, safety, consumer and community interest, and a reduction in trade barriers [4], as well as many other important aspects ensuring efficiency, sustainability, safety, and security. Standardization of the technological solutions for road transport is an important aspect of increasing the uptake of ITSs to reap the estimated benefits. In addition, ITS standardization can produce the following benefits:

- Enables interoperability of systems/services and between different implementations that will give use seamless plug-and-play functionality;
- Encourages innovation, fosters enterprise, and opens up new markets for suppliers;
- Creates trust and confidence in products and services. This includes a test and quality that will assure that products/solutions are safe, healthy, secure, flexible, and of appropriate quality;
- Expands the market, brings down costs, and increases competition.

1.1. ITS Standardisation

ITSs have become the focus of several policies and legislative initiatives in Europe. The motivation is to accelerate the deployment of new innovative transportation technologies in Europe by laying down the legal framework. ITS standardization in the EU is based on several key documents. The Action Plan for the Deployment of ITSs in Europe [4] aims to accelerate and coordinate the deployment of ITS in road transport, including interfaces with other transport modes. The Action Plan outlines six priority areas for action and identifies a set of specific actions and a clear timetable for each area. The six priority areas build on input from public and private stakeholders and assume that ITS applications to be deployed in the short-to-medium term should be mature, sufficiently interoperable, and able to create a catalytic effect across Europe. Regulation (EU) No. 1025/2012 of the European Parliament and of the Council [5], the first document that deals with the issue of European standardization coherently, supports all types of creation and development of new European standards and encourages the use of standards, especially by small and medium-sized enterprises (SMEs). We need a uniform platform to make the mass implementation of ITSs possible. Thus, standardization is considered as a method for the integration of ITSs. The regulation pays close attention to the strategic significance of standardization, supports all sorts of creation and development of new European standards, promotes the use of standards by small and medium-sized enterprises, etc. In several parts of the Regulation, it is implicitly stated that national standardization bodies should seek to promote informative texts on standards publicly available on websites:

“1. National standardization bodies shall encourage and facilitate the access of SMEs to standards and standards development processes in order to reach a higher level of participation in the standardization system, for instance, by:

(e) making available free-of charge abstracts of standards on their website.”

The Communication from the Commission to the European Parliament, the Council, and the European Economic and Social Committee of the Regions [6] emphasized the need to take full advantage of smart digital solutions and ITS and the enormous potential of connected, cooperative, and automated systems for the functioning of the whole transport system, its sustainability, and safety goals. The guideline [7] provides up-to-date information on global standardization and deployment in the domain of ITS, with a focus on cooperative ITS (C-ITS). C-ITS is a new paradigm in ITS. Table 1 illustrates the recent (2020) situation in Europe based on mandate applications.

Table 1. Sharing common resources for various ITS applications [8] (p. 131).

Application	Technology
Infotainment / Sustainability system	Radio box (3G/LTE, WiFi, Bluetooth, GPS, ...)
EETS system ¹	Radio box (2G/3G, GPS, 5.8 GHz, ...)
eCall system	Radio box (2G/3G, GPS, ...)
Active safety system	Radio box (GPS, 5.9 GHz, ...)

¹ European Electronic Tolling Service.

The most common understanding is that C-ITSs are needed to move away from the multitude of proprietary stand-alone boxes invading the driver environment. It is not sustainable to put a new box with antennas, display, keyboard, etc., for each new application in a car. This is too costly, is too unsafe, does not provide interoperability, and is just not sustainable from a windshield real estate point of view [8].

ITS standardization ultimately aims to address the lack of harmonization between ITS solutions deployed across the EU Member States, accelerate ITS deployment, and open the internal market in a consistent way. Improving public and private sector awareness of European and international ITS standards can create opportunities both for innovative SMEs in the ICT sector and for SMEs that can make better and compatible use of ICT to improve their products, services, and business processes.

This paper is focused on the process of creating several-page extracts that are representative of full-text standards, i.e., solving the problem of information extraction, which is a difficult task due to the ambiguity of written natural language [9]. Generally, there are two applicable approaches—knowledge-based techniques and machine learning and statistical techniques [10]. The former requires large volumes of training and validation data that are not always available or must be prepared manually or semiautomatically. The latter utilizes predefined conceptual knowledge, which is usually not obtainable without domain experts. To obtain a complete description of the domain (ITS in our case), the authors can utilize different types of domain-modeling knowledge that can be represented using an ontology [11]. Ontology-based models of the world can help derive machine-processed knowledge and evaluate it [12]. The ontology defines a common set of terms to represent the basic semantically related concepts built on a limited number of predefined relations and terms of a domain [13]. These terms and concepts can be visualized to represent both syntactic and semantic data [14]. The source [15] reviews how definitions of ontologies changed and evolved over time. In the context of standardization considerations, the standardization of ontologies is another current trend, supported by the adoption of semantic technologies in industry. A general review of standardized ontologies can be found in [16]. In the field of ITS domain, ontologies are the subject of increasing research interest, but the reason for this interest is outside the scope of the topic under our discussion. Ontologies have been acknowledged as the most common tools to ensure the interoperability of different cooperating systems to achieve a joint objective function. The scientific literature offers various classifications of ontologies. For example, [15] discusses six groups based on the type of information to be captured: top-level ontologies, the practical use of which is rather limited; domain ontologies that are reusable only in the given domain; task ontologies proposed for generic tasks and activities; domain-task ontologies defining domain-level ontologies for domain-specific tasks and activities; method ontologies helping with the reasoning process; and application ontologies primarily designed to provide knowledge in a specific application. Another term that is used is linguistic ontology. Instead of modelling a specific domain, this kind of ontology describes semantic constructs and originates from natural languages [17]. Using the richness of the internal structure as another dimension for classification, ontologies can be classified along a linear spectrum [18], ranging from the most simple ontologies (e.g., catalogue/ID, terms/glossary, thesauri “narrower term” relation, and informal is-a) to the most complex and powerful (formal is-a, formal instance, frames, values restrictions, general logic constraints, disjoints, inverse, part of, etc.). Similarly, [19] discusses lightweight ontologies (concepts, concept taxonomies, rela-

tionships between concepts, properties describing concepts) and heavyweight ontologies (lightweight ontologies with axioms and constraints added). Thus, ontology-based information extraction has recently emerged as a subfield of information extraction [20], which is applicable when extracting from both unstructured and semistructured domains [10]. The process of exploring and maintaining domain-specific documents can be supported by proper software tools; however, there are certain limitations (e.g., existence of rich and specialized semantic annotations [i.e., semantically enhanced information extraction]) [21]. These semantic annotations couple language entities with their semantic descriptions and connections from a knowledge graph. Software tools for targeted search and ontology management, driven by an ontology-based information extraction system, have been discussed (e.g., in [22]) and compared in [23]. One of the recent approaches worth mentioning is the syntactic-semantic treebank for domain ontology creation. The idea is that the domain knowledge is represented in the domain data, but via treebanking, more linguistic patterns can be extracted, which are then mapped to concepts and relations in a domain ontology [24].

The latest research on information extraction does not cover full-text standards mining and is mainly focused on obtaining information from various media, natural scenes, and images as sources of primarily textual information [25–28].

1.2. Problem Definition

Standards, similar to encyclopedias or directories, have a massive amount of extensive and important data. To understand what they are about, entire documents usually need to be read. The process by which these standards are set and the forces that are re-shaping that process have been discussed (e.g., in [29,30]). The paragraph above explains why ITS standardization is important for reaching pan-European interoperability. This paper addresses the problem of the volume and variability of ITS technical standards, which are permanently growing and becoming more difficult for practical use. The situation will worsen in the coming years, especially in relation to the mentioned C-ITSs. It will no longer be in the power of individuals to follow all relevant ITS standards when considering their volume, scope, detailed contents, relevancy for ITS tenders, creation, deployment, operation, etc. The problem is how to inform potential users of standards in an efficient way and how to inform them about the focus of those standards. When the standard language is different from the national language, the resulting language barrier can negatively affect the orientation in and understanding of the system of standards. As referenced by [31], at least half of Generation Z (preteens and teens born roughly between 1996 and 2010), which will represent 75% of the workforce by 2030, will work in positions where they will need a certain understanding and knowledge of standards and standardization in their specific context.

The authors' intention is to articulate a solution to this problem. This solution relies on the creation of standard extracts and the use of a hybrid method that combines syntactic and semantic analysis. This paper presents the practical experiences and results obtained from the solution of a long-term national project. Such an approach can contribute to sustainability achievement (considering all its coequal parts—economy, equity, and environment [32]).

2. Applied Methodology

This section briefly introduces a particular long-term, practice-proven project and clarifies the approach and methodology applied within it.

2.1. The STANDARD Project

Text summarization has become the subject of intensive research because of the explosion of the amount of textual information available [33,34]. Generally, two categories of summaries are subject to research: extracts and abstracts. Extracts are summaries created by reusing important portions of the input verbatim, while abstracts are created by

regenerating the extracted content [35]. However, research in the field has shown that most of the sentences (80%) used in an abstract are those which have been extracted from the text or which contain only minor modifications [33,35]. This and the difficulty of abstracting are the reasons why the majority of research is still focused on text extraction [36]. The creation of both abstracts and extracts is realized in several stages: for example, according to [35], the stages of summarization are topic identification, topic fusion, and summary generation. Unlike the extraction methods, abstraction requires using bulky tools for natural language processing, including grammars and lexicons for parsing and generation [37].

The project in question, named STANDARD, was an original idea of the Czech National Technical Standardization Committee TNK 136. It has received permanent support from the Ministry of Transport of the Czech Republic. Since 2008, the STANDARD project has aimed to introduce Czech users to the content of standards in its explicit form. English is not a native language in the Czech Republic, so the resulting project output (extracts) is published in Czech, which facilitates its interpretation in the national environment. Thus, the project represents a sophisticated knowledge mining system covering ITS standards. The project complies with the requirements of Regulation No 1025/2012 [5].

Figure 1 depicts the development of a number of standards in the first 5 years of the project's life. The number of elaborated extracts is shown in yellow. The number of active standards worldwide is depicted by a solid line, and the volume of standards transformed in the Czech Republic environment is displayed in aqua (for total) and pink (for the standards translated into Czech).

Figure 1. Development of the number of extracts in the first 5 years of the STANDARD project.

The aim of the long-term project has been to link ITS standardization with the public domain and industrial sectors with a view to increasing the competitiveness of SMEs. The project started in 2008, when there were 73 new standards available in about 5000 pages of text in principle; this meant studying each standard and producing approximately five pages of extract out of the whole text (see principal sketch in Figure 2). The process of creating the extract involves analyzing the whole standard, selecting knowledge, and synthesizing it.

Figure 2. The principal output of the STANDARD project.

The need to deal with a large range of ITS standards is evident today. In November 2020, there were approximately 310 ITS standards, which together with their proposals added up to a total of approximately 20,000 pages (quantity), frequently written in UML or XML conventions (quality) that are not readable, especially for investors and decision makers. Deployment in non-English speaking countries is very limited. In 2020, 23 extracts were in development.

Figure 3 schematically shows the authors' view of how the knowledge of a standard is represented in three hierarchical layers of standardization—from the highest level, made up of technical committees for standardization, to the basic layer, made up of users of standards. The extracts alone effectively facilitate public access to the knowledge contained in the standards. The extract has become a new type of document created exclusively to raise awareness of existing ITS standards. To ensure consistency and unification of the form, the context and content of the extracts were considered, and a specific methodology based on syntactic/semantic text analysis was used.

Figure 3. Awareness of the content and determination of the standard in terms of standardization hierarchy.

2.2. Basic Knowledge Unit—Extract

The extract is the essence of the knowledge contained in the standard and expresses the content of the document. It is based on a contextual decomposition of the original standard according to definitions, terms of figures, chapters of paragraphs, and appendices. In this context, it is appropriate to define the extent of the extract. While the abstract of a standard is only a brief summary in a few paragraphs, the extract is, in terms of extent, a

kind of intermediate between the abstract and the standard itself; it contains structured content that expresses the essential ideas of the standard.

An extract is a document created by reducing the standard to approximately five pages, prepared according to a methodology that deals with the content and formal elements of the standard in such a way that it provides experts with adequate information about the scope of the standard.

A website [38] has been created which describes the methodological procedure in more detail and allows searching for the extracts created and made available to the public. The site is written in the Czech language, and the essential parts are translated into English. An example of a specific extract, including a direct web link, is available in Appendix A and explained in the Results Section. The extract (originally in Czech) has been translated into English to be accessible to international readers.

2.3. Initial Methodology: Knowledge Mining by Computer Analysis of the Text

The first idea was to save the time required for manual text analysis and to search for knowledge hidden in the standards automatically using algorithms for text analysis. This is a very time-consuming process. First, the authors focused on nouns and adjectives. Finding them in English texts is not a difficult task because they are always associated with definite or indefinite sentence elements. As a second possibility, individual ITS terms defined by the group of experts were searched in the text. The third option was to search for coherent pairs or triplets of words found in the sentences of the source.

The example given in Table 2 shows an extract relying on the above three categories, which was created in 2013 by an expert team in Prague.

Table 2. Automated knowledge search result example, ISO/TS 17575-1:2010.

Rank	Noun/Adjective or Abbreviation	Count	ITS Term	Count	Pairs or Triples	Count
#1	Data	1394	data element	455	data element	455
#2	Element	589	front end	202	data type	322
#3	ISO	509	Road	182	ISO TS	263
#4	Optional	436	charge report	111	front end	202
#5	Type	383	back end	91	data elements	187
#6	Toll	369	Cordon	54	charge report	111
#7	EFC ¹	344	charge object	46	back end	91
#8	End	336	Layout	45	context data	79
#9	Defined	276	road usage	38	optional data	76
#10	TS ²	265	System	37	EFC context	74
#11	Charge	249	Contract	31	EFC attribute	70
#12	Elements	240	area pricing	29	optional data element	60
#13	Context	226	Proxy	27	toll charger	60
#14	Front	224	data set	26	service provider	57
#15	Time	219	road usage data	17	FE ³ application	52

¹ Electronic Fee Collection. ² Technical Specification. ³ Front End.

The method of rough parsing of the text was used, followed by the selection of relevant words and sorting according to the frequency of occurrences. The standard ISO/TS 17575-1 first edition 2010-06-15 was elaborated, which was named “Electronic fee collection—Application interface definition for autonomous systems—Part 1: Charging”. The total length of the standard is 23 pages. Table 2 gives examples of the 15 most common results in each of the three categories.

These methods proved not to be applicable because the results were scattered throughout the whole document and did not give an exact idea of the message conveyed by the document. The other two methods for text analysis were considered.

3. Results

The results are in the form of conceptual models of the standard. The conceptual model is a knowledge model that allows the transfer of knowledge from the standard to the extract.

Creating this model requires an in-depth analysis of the standard's text. Unlike the approach described in the previous chapter, the authors tested a combination of syntactic and semantic analyses—two of the primary techniques that lead to an understanding of natural language. Syntax, as a grammatical structure of the text, refers to the ways in which specific words are ordered to create logical, meaningful sentences. Semantics is the analysis of the meaning of words.

3.1. Semantic Analysis

The first method, semantic analysis, finds all terms carrying information in the text and arranges such concepts taxonomically, which leads to an ontology. Creating a representative ontology requires a skilled domain expert. Thus, the experts making up the standard team are preferably ITS experts. Therefore, it was not certain that the ontologies of heterogeneous systems would be uniform so that the extracts would have the same basis.

Creating the right ontology structure, if we have 100 to 200 pages of dense text to be processed, requires an ontology expert who has experience in creating ontologies of complex systems. Therefore, it was decided to acquire the knowledge encoded in the text using semantic analysis. Such analysis is usually used for sentence analysis; within the STANDARD project, it was modified from sentence analysis for the whole document to simplify the analysis and speed up the synthesis of the extract.

3.2. Syntactic Analysis

In computer science and linguistics, syntax analysis is the process of analyzing the sequence of formal elements to determine their grammatical structure in relation to a predetermined (although not necessarily explicitly expressed) formal grammar. This analysis is mostly sentence oriented. In Czech, this means finding the subject, the suffix, and other sentence components.

Standards, like the sentence, always have a similar construction. There are always terms and definitions at the beginning that reveal much about the information contained in the standard. The content structure carries hierarchically oriented concepts. The text associated with any picture also carries important information. Semantic analysis that creates a relatively accurate conceptual model of knowledge requires the work of a specialist.

Therefore, a hybrid procedure was chosen wherein the first step is for formal parsing to create the framework of the conceptual model.

3.3. Workflow for the Process of Creating a Standard's Extract

After taking into account the experience gained, including unsuccessful approaches, the authors finally proposed a hybrid method for creating extracts of the standards. The basic idea characterizing our approach is summarized as a workflow (see Figure 4) consisting of three basic steps:

- Syntactic analysis of the entire text of the given standard results in a basic knowledge matrix (whose image is a basic ontological tree);
- Optional search for additional terms to create an extended knowledge matrix;
- Creation of an extended ontological tree as a cognitive model of the standard.

In general, it is difficult to study an extensive document and keep its contents in mind for later analysis. Therefore, efforts were made to acquire a visual picture of the standard for subsequent analysis without having to start with an ontology. The following steps were used to create a visual of the standard. For graphical elaboration, a freely available SW for mind mapping, x-Mind, was used.

Figure 4. Proposed workflow for the creation of standard extracts.

3.3.1. First Step: Syntactic Analysis

The hybrid method uses a formal procedure in the first step to generate a basic skeleton of the extract. Parsing usually analyzes a sentence: the subject, predicate, and other components. Our method is to perform parsing on the whole document, assuming that there are multilevel headings included. It is probable that the headings contain the

most important terms and the knowledge that we can then extract, at this point without a hierarchy. The authors then organize this knowledge into a basic knowledge matrix (BKM), consisting of the following: order (position) in the first column, a short name (nickname) in the second column, and a link to the particular place in the original document (page) in the third column.

Thus, the following clearly distinguishable entities carry information about the document:

- Chapters, subchapters, sub-subchapters, etc.;
- Text associated with figures.

Both of these entities are inserted into the mind map. In Figure 5, they are marked in yellow. The output of the first step is a skeleton visualizing the structure of the standard. Since the titles of the chapters carry important information, it is possible to say that the output (image) of the first step is a kind of semiontological tree, which should be supplemented by important terms in the second step.

Figure 5. Structure of expressions in a language and the rules governing the structure of a language.

3.3.2. Second Step: Searching for Additional Concepts

This step is optional, but if it is done, it will significantly help with the synthesis of the extract. When studying the standard, key terms appear in individual paragraphs (for example, we can analyze captions). These new terms (concepts) should be inserted directly into the branches of the ontology. In Figure 5 they are depicted in green.

3.3.3. Third Step: Synthesis of the Final Extract

Since the output of the second step takes the form of a visualized knowledge system reflecting the analyzed standard content, the expert can consider it as a cognitive model of the standard (i.e., the form which is well understood by the human brain). Therefore, at this stage, it is no longer a problem for the expert to compress the original content of the analyzed standard into the expected five pages of extract. To perform a synthesis offline, without being able to see the knowledge structure, would certainly lead to inaccuracies. The approach resulting from a known knowledge structure has proved to be feasible, leading

to good results. It is worth noting that formal approaches to linguistic text analysis, such as formal concept analysis, focus more on formal structures than on cognitive-linguistic phenomena, and linguists argue that formal concepts are quite different from cognitive processes related to natural language [39].

For the sake of reflection, Table 3 contains a selection of standards on a specific topic. It is difficult to imagine that a manager of a small traffic company will spend money to buy all standards given (e.g., in Table 3) and that he or she will find somebody able to analyze roughly 800 pages of dense text.

Table 3. Standards on cooperative vehicles with extracts finalized in 2019.

Standard		
ISO 24097-1:2009	Intelligent transport systems—Using web services (machine-machine delivery) for ITS service delivery—Part 1: Realization of interoperable web services	15 July 2014
ISO 14814:2006	Road transport and traffic telematics—Automatic vehicle and equipment identification—Reference architecture and terminology	15 July 2014
ISO 14813-6:2009	Intelligent transport systems—Reference model architecture(s) for the ITS sector—Part 6: Data presentation in ASN.1	15 July 2014
ISO/TR 25104:2008	Intelligent transport systems—System architecture, taxonomy, terminology, and data modeling—Training requirements for ITS architecture	28 January 2008
ISO/TR 25102:2008	Intelligent transport systems—Systems architecture—‘Use Case’ pro forma template	7 February 2008
ISO/TR 25100:2012 (Ed. 2)	Intelligent transport systems—System architecture—Harmonization of ITS data concepts	13 September 2012
ISO/TR 24532:2006	Intelligent transport systems—Systems architecture. Taxonomy and terminology—Using CORBA (Common Object Request Broker Architecture) in ITS standards, data registries, and data dictionaries	9 June 2006
ISO/TR 24529:2008	Intelligent transport systems—Systems architecture—Use of unified modeling language (UML) in ITS International Standards and deliverables	8 April 2008
ISO/TR 17465-1:2014	Intelligent transport system—Cooperative ITS—Part 1: Terms and definitions	16 October 2014
ISO/TR 17452:2007	Intelligent transport systems—Using UML for defining and documenting ITS/TICS interfaces	17 April 2007
ISO/TR 12859:2009	Intelligent transport systems—System architecture—Privacy aspects in standards and systems	28 May 2009
ISO 24531:2013 (Ed. 2)	Intelligent transport systems—System architecture, taxonomy and terminology—Using XML in ITS standards, data registries, and data dictionaries	14 May 2013

To demonstrate the results, Appendix A provides a particular example of a part of EN ISO 14825 (590 pages—extensive standard) created in 2009. Those interested in the standard will find the complete extract at the following link <https://silmos.cz/standard> (accessed on 27 May 2021). The extract for the standard used as an example can be found directly at the following link: <https://silmos.cz/standard/#f=2&norma=%C4%8CSN+EN+ISO+14825> (accessed on 27 May 2021). As mentioned above, the primary language is Czech. The extracts are in the Czech language, as the purpose of the project is to make the content of the standards accessible to Czech citizens and companies, which is why the Czech Ministry of Transport also supports the project. Due to the extensive volume of the original standard, the whole extract is six pages (instead of five pages, as recommended above).

Since this paper is intended for an international reader, the entire abstract has been translated into English, except for figures, which remain intentionally unchanged. The numbering of the chapters corresponds to the numbering in the original standard; the numbering of the tables and figures is preserved as in the extract, but supplemented by the “A” letter in front of the number to be separated from the rest of this paper.

The deployment of ITS standards into real practice is hampered by their large number and lack of information, notwithstanding the EU’s efforts to make Europe competitive in this field. Society is shifting from being information-based to being knowledge-based. Based on the above facts and practical experience, the authors have tried to answer the following question: “Which presentation of a standard is decisive for working with it?” The authors conducted a survey of members of the Czech ITS Association, asking them how high the information content is (in terms of percentage) in different presentation forms to contribute to the decision to purchase the standard (see Table 4). The following options were available, and the results are shown on the right:

Table 4. The information content in different presentation forms of a standard.

#	Presentation Form	Information Content
1	Only name of a standard	less than 5%
2	Abstract: paragraph describing the idea of a standard	about 10%
3	Extract	more than 70%

The result of the survey is depicted in Figure 6.

Figure 6. Influence of the form of presentation on the purchase and use of the standard.

4. Discussion

In this section, the authors interpret and describe the significance of our findings in light of our understanding of the research problem as stated in Section 1.2.

National standardization bodies are officially required by regulation [5] to supply public information for users deciding whether to buy and use a particular standard. This duty implicitly results from [5]. The cardinal problem is in which form this information should be offered and how efficiently the permanently growing numbers of ITS standards, representing thousands of pages, including UML and XML notations, should be processed and made accessible. Therefore, research has been done based on the original idea of the Czech National Technical Standardization Committee TNK 136 (ITS field) and with permanent government support.

Initially, the original methodology, based on making extracts by domain experts via predefined rules, failed. Given the increasing number of standards and the workload of a limited number of domain experts, this method proved unsustainable in the long term. Therefore, further development focused on the study of knowledge mining via computer analysis of the text. The methods of information retrieval examined indicated that there are no methods specially designed for text extraction from full-text standards, nor for creating short extracts or similar text structures. Therefore, our strategy depends on the application of common syntactic methods: automatic text parsing, search for knowledge hidden in the text, and searching for the frequency of words/pairs/triplets/ITS terms, operating on the premise that deeper information is hidden in nouns and adjectives. However, this approach failed since the results were scattered throughout the whole document and no clear idea of the text's message was generated. The authors concluded that the use of a syntactic approach to the extraction of full-text standards does not lead to the desired results.

Our new hybrid method for extract creation as presented above represents a balanced combination of syntactical and semantic approaches, using a semiontological structure (tree) with the active involvement of a domain expert. Machine-based formal parsing creates only partial results, enabling one to get a basic conceptual model. It is followed by sentence structure analysis and the capture of the hierarchical structure of the standard, allowing the authors to create a basic skeleton of the extract. The semantic approach makes possible the arranging of found suitable words taxonomically, which helps a skilled domain expert to make a final synthesis—his or her work here is irreplaceable. Thus,

entities (chapters, subchapters, figure captions) can be inserted into the mind map. In terms of previously discussed ontology types, our ontological structure can be considered a linguistic application ontology since most of the linguistic ontologies use words as grammatical units and can be used for natural language processing and generation. The meanings of grammatical units in natural languages tend to be ambiguous, so our ontology can be considered relatively heterogeneous in terms of heterogeneous standards. From the viewpoint of ontology richness, our model belongs to the category of lightweight ontologies. The visualization knowledge system seems to be extremely helpful for experts in the process of final extract synthesis.

5. Conclusions

Our hybrid method is a novel and valuable approach for mining full-text standards and fills a hole in the market of special purpose methods used for information retrieval. Despite its narrow and specific focus (here in relation to the ITS domain), its principle is potentially applicable across other professional domains because it conforms to regulation [5] requirements. The success of the approach was proved by the survey of SMEs, the main (but not the only) target group. Publicly available extracts significantly reduce the time needed to study standards and make decisions on their potential use.

On the other hand, the number of extracts processed in the proposed manner still lags behind the growing number of standards, despite permanent support from central entities. The number of experts needed throughout the process is limited and represents the main obstacle to large-scale deployment.

Our recommendations for the future are as follows:

1. Knowledge is a cornerstone for ensuring sustainable mobility.
2. The development of the European economy is connected with the acceptable development of mobility through the development of ITS.
3. The significance of the STANDARD project is verified by its 12-year operation.
4. The project can be taken over in a 1:1 ratio in any country or at the level of the whole of Europe.
5. To a certain extent, the hybrid methodology will make it possible to create extracts even for nonknowledge systems specialists.

Author Contributions: Conceptualization, P.P.; methodology, P.P.; software, V.F.; validation, P.P., A.J., and J.S.; formal analysis, P.P.; investigation, P.P., A.J., J.S., and V.F.; resources, P.P. and A.J.; data curation, J.S. and V.F.; writing—original draft preparation, P.P., A.J., J.S., and V.F.; writing—review and editing, P.P., A.J., and V.F.; visualization, P.P. and A.J.; supervision, P.P. and J.S.; project administration, P.P.; funding acquisition, P.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Slovak Research and Development Agency, grant number APVV-17-0014.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are partially contained within this study and partially are openly available at [<https://silmos.cz/standard>] (accessed on 27 May 2021).

Acknowledgments: To the Ministry of Transport of the Czech Republic for long-term support of the STANDARD project. The project has been chaired by the Centre of Standardization Silmos Brno (Igor Večerka since the very beginning).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

EXTRACT from the technical standard

The extract does not replace the technical standard itself; it is only informative material about the standard.

03.220.01, 35.240.70

Intelligent transport systems —

EN ISO 14825

Geographic Data Files (GDF)

01 8281

Effective from 11.2004

590 pages

Introduction

This standard is a part of the standards focused on navigation and location systems and related applications.

The standard was initiated by the manufacturers and users of digital road maps in the 1980s that were looking for a format for the routine exchange of data and the interoperability of the systems that use this data. The activities culminated in the draft of a European GDF standard, which was based on well-known regional standards, such as in Japan and the US.

This standard specifies a conceptual and logical data model and an interchange format for geographic data for ITS applications. The conceptual data model is application independent. This is a prerequisite for the future harmonization of this standard with other geographic database directives. As can be seen from the content of the standard, GDF represents a part of the real world around it, which in itself is dynamic. In this context, the process of monitoring real changes and their reflection in the GDF is implemented through regular updates.

Work on the standard continues, so an updated version of the standard with the working designation XGDF is currently being created. The final draft of the updated standard is to be available in early 2009. The standard includes the incorporation of amendments in the areas of: 3D urban model, 3D terrain model, traffic flow direction model, extension of attributes for commercial vehicles, attribute catalogue, additions to UML diagrams, etc. For example, a city map data model should contain the attributes: building (floor plan), building diagram, building details (outlines), block name, block group, subway representation, railway type, etc.

Applications

The content of the standard belongs to the field of navigation and location systems and related applications. Its applications can be found mainly in the field of navigation and location systems, provision of transport services, traffic news, traffic management systems, and active vehicle systems, or ADAS (advanced driver assistance systems) applications.

For government bodies, this standard specifies the format and scope of data provided by road administrators for the needs of a variety of applications and services.

For equipment manufacturers and telematics system suppliers, this standard is used as a common data format in map products supplied by TeleAtlas Company (TomTom), Navteq (Garmin), CEDA, and in the myriad applications that use these map materials. Recently, cartographic companies have also been using this format and switching to it. This format creates preconditions for unifying the data structure of navigation maps from

various manufacturers. Finally, it is already implemented in a number of commercial GIS products, which allows for its further expansion.

Table A1. Examples of the use of the extended GDF format in the field of active vehicle systems.

Application		
Longitudinal Control	Adaptive cruise control (ACC)	Automatic control with regard to speed and distance from vehicles ahead
	Adaptive headlight adjustment control	Dynamic headlight control
	Speed warning systems	They warn against cornering at excessive speeds
	Visibility improving	Systems increasing visibility in adverse lighting and climatic conditions
	Visual and acoustic signalling to the driver	The systems provide visual and acoustic information to warn of potentially dangerous locations
	Collision warning systems	Warn drivers of the possibility of a collision
Lateral Control	Lane departure warning	Warn driver against unintentional change of direction
	“Track holding” assistant	Helps the driver maintain the vehicle’s driving path
	Lane change assistant	Informs driver about vehicles in adjacent lanes
	Autonomous driving	Supports fully automatic control

Related standards

ISO/IEC 8859-1 Information technology—8-bit single-byte coded graphic character sets—Part 1: Latin alphabet No. 1

1 Scope

The standard is divided into eight basic parts: Overall Conceptual Data Model, Feature Catalogue, Attribute Catalogue, Relationship Catalogue, Feature Representation Rules, Metadata Catalogue, Logical Data Structures, and Media Record Specifications.

3 Terms and definitions

3.2.3 Data file: collection of related data records having a homogeneous structure.

3.4.2 Attribute: characteristics of a Feature which is independent of other Features (nongeometric property of the element—street name, descriptive number, postal code...).

3.4.11 Feature: database representation of a real-world object. (A Feature is a basic spatial entity that is further indivisible into elements of the same type and that is described by spatial data. The Features make up the environment in which the human moves.)

7.3.52 Location referencing: a method of marking positions to easily exchange position information between different systems.

12.1.3 Media record: the name of a group of sequential data fields which belong together. (A GDF media record has a length of 81 or 82 characters, depending on whether one or two control characters are used to terminate the record.)

UML (Unified modelling language): the tool for the description and design of information systems. In this standard, UML is used as a tool to express structural relationships and specific relationships using graphical elements. The full definition of UML is contained in ISO 19501.

ESN (Data description language): the language that enables data types to be constructed of any complexity from a set of elementary types.

Metadata: data describing the content, representation, scope (spatial or temporal), spatial reference system, quality, and administrative aspects, possibly also business aspects of the use of digital data.

5 Overall conceptual data model

This chapter presents the concept of a data model, which explains the basic building blocks of GDF and their relationships. The individual types of topological structures supported by this standard are described here, as well as the approach by which the elements of the real world are represented in the database.

6 Feature Catalogue

In the following chapters Feature Catalogue, Attribute Catalogue, Relationship Catalogue, the standard describes simple and complex Features; their characteristics, which we refer to as Attributes; and topological and nontopological relationships between them. Attributes can also be simple, represented by a single value, or composed of a number of so-called subattributes. Nontopological relationships between Features are represented by relational relationships. The standard approaches the representation of a Feature as a di-mensionless, one-, or two-dimensional entity. The three-dimensional presentation is also supported by the document, but not geometrically.

Figure A1. Relationships of driving priorities when passing through an intersection controlled by vertical traffic signs. For a light-controlled intersection, these cannot be determined.

This part also describes the organization of Features in GDF. Conceptually, Features are divided according to themes. Physically, Features are grouped into groups or layers according to the area they represent.

Figure A2. Data model of the Feature Catalogue (captions in Czech).

The standard contains a summary of Features that are supported by this document. In addition, the inclusion of undefined Features, so-called user-defined Features, is also specified here. However, this Feature specification is not strictly prescribed. The actual GDF content, unlike the minimum set of records, is considered a defined item between the customer and the supplier.

8 Relationship Catalogue

The list of relational relationships defines various nontopological relationships that can occur in Features. Naturally, even here there is the possibility to define relational relationships by the user.

As it is clear from the foregoing, the standard does not specify the actual content of the GDF, but only helps the user to more easily define Features. In some cases, different ways of modelling and presentation are offered.

10 Metadata Catalogue

The standard is conceived as a comprehensive document that does not need external interpretation to understand. To ensure this, the standard defines ways to describe GDF using metadata, see definitions.

11 Logical Data Structures

In this chapter, to facilitate the mechanism of data exchange and definition, a logical view of the data is used, which is then transferred to a logical data structure. Data structures are described using the ESN data descriptive language.

12 Media Record Specifications

This chapter describes the basic concepts of data record interchange format specification.

Example of EN ISO 14825 – Data model for a selected field of attributes

Figure A3. Data model for attributes of road equipment (captions in Czech).

Annex A (normative) Semantic codes

Annex A describes the procedure for managing and assigning identifiers to structures CS1, CS2, and CS8. Article A.1 sets out a hierarchy of authorities (administrators) for assigning identifiers. The main emphasis is on maintaining data consistency. The Annex A also contains codes for the representation of Features, Attributes, and relationships used in real GDFs.

Annex B (informative) Metadata codes

Annex B contains a list of codes used in the formulation of metadata.

Annex C (informative) Services

Annex C contains specifications for the Features of the Services topic.

Annex D (normative) Syntax for Time Domains

Annex D describes the rules for specifying variable aspects of geodetic information.

Annex E (normative) Sectioning GDF Datasets

Annex E describes the process of organizing the segmentation of data records in GDF to facilitate the processing of large amounts of data. The subject of segmentation is only the organization of data for better processing, not the definition of database coverage of the area. The area covered by a particular GDF is defined independently of its logical division into dataset units, layers, and sections.

Annex F (informative) Rules for the formation of Layer 2 Features of Roads and Ferries

Annex F describes Features of road, junction, and level crossing as composite Features, which are made up of different combinations of road elements and connections. The functional character of the intersection is related to the decision on the direction of travel (e.g., “turn left”). The functional nature of the level crossing is generally related to a set of navigation decisions (e.g., at the Frankfurter crossing, turn towards Würzburg).

Practical examples of representation in GDF:

Figure A4. Representation of multilane road on Layer 1 and 2 (left); Three-valence crossing with three connections and one intersection (right).

Annex G (informative) Types of administrative area names and country numbers

Annex G contains the types of administrative area names and country numbers.

Annex H (informative) Definition of measuring methods

Annex H pays attention to the qualitative aspect of geographical information, as a number of erroneous representations of Features can never be ruled out. This is addressed by the attached methodology, with which the concept of quality can be statistically quantified. The methodology does not specify what the quality should actually be, but only sets out how to describe and express qualitative characteristics.

Annex I (informative) Specification of the possible use of attribute types for Features and relationships

Annex I specifies the possible use of attribute types on Features and relationships. This annex contains a table of possible attributes for Features and relationships. The table is for information only. It is intended as a guide to the use of attribute types with different Features and relationships for the predominant use of attributes in this context.

References

1. Williams, B. *Intelligent Transport Systems Standards*, 1st ed.; Artech House, Inc.: Boston, MA, USA, 2008; p. 11.
2. CEN/TC 278. ITS standardization. Available online: <https://www.itsstandards.eu> (accessed on 18 December 2020).
3. Action Plan for the Deployment of Intelligent Transport Systems in Europe. Brussels, 16. 12. 2008 COM(2008) 866 Final, 14 p. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52008DC0886&from=EN> (accessed on 3 June 2021).
4. Spivak, S.M.; Brenner, F.C. *Standardization Essentials. Principles and Practice*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2001; pp. 29–30.
5. Regulation (EU) No 1025 of the European Parliament and of the Council of 25 October 2012. Available online: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:316:0012:0033:EN:PDF> (accessed on 3 June 2021).
6. Sustainable and Smart Mobility Strategy—Putting European Transport on Track for the Future. COM/2020/789 Final. 12 December 2020. Available online: https://eur-lex.europa.eu/resource.html?uri=cellar:5e601657-3b06-11eb-b27b-01aa75ed71a1.0001.02/DOC_1&format=PDF (accessed on 22 December 2020).
7. Cooperative Intelligent Transport Systems (C-ITS). Guidelines on the Usage of Standards. Edition 26 June 2020. Available online: <https://www.itsstandards.eu/app/uploads/sites/14/2020/10/C-ITS-Brochure-2020-FINAL.pdf> (accessed on 18 December 2020).
8. Evenson, K.; Schmitting, P. *D3.5b—Standardization Handbook. Report of the FP7-ICT-2011-8_317547 Project*; ERTICO: Brussels, Belgium, 2014; p. 145.
9. Gutierrez, G.; Dou, D.; Fickas, S.; Wimalasuriya, D.; Zong, H. A hybrid ontology-based information extraction system. *J. Inf. Sci.* **2015**, *42*, 798–820. [CrossRef]
10. Deokar, A.V.; Sen, S. Ontology-Based Information Extraction for Analyzing it Services. In Proceedings of the ICIS 2010 Proceedings, St. Louis, MI, USA, 12–15 December 2010; p. 216.

11. Gayathri, R.; Uma, V. Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey. *ICT Express* **2018**, *4*, 69–74.
12. Lourdasamy, R.; John, A. A Review on Metrics for Ontology Evaluation. In Proceedings of the 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 19–20 January 2018; pp. 1415–1421.
13. Koutsomitropoulos, D.A.; Kalou, A.K. A standards-based ontology and support for Big Data Analytics in the insurance industry. *ICT Express* **2017**, *3*, 57–61. [\[CrossRef\]](#)
14. Hussan, B.K. Comparative Study of Semantic and Keyword Based Search Engines. *ASTESJ* **2020**, *5*, 106–111. [\[CrossRef\]](#)
15. Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. In *Advanced Information and Knowledge Processing*, 1st ed.; Jain, L., Wu, X., Eds.; Springer: London, UK, 2004; pp. 293–362.
16. Li, W.; Tropea, G.; Abid, A.; Detti, A.; Le Gall, F. Review of Standard Ontologies for the Web of Things. In Proceedings of the 2019 Global IoT Summit (GloTS), Aarhus, Denmark, 17–21 June 2019; IEEE: New York, NY, USA, 2019; pp. 1–6.
17. Styltsvig, H.B. Ontology-Based Information Retrieval. Ph.D. Thesis, Computer Science Section, Roskilde University, Roskilde, Denmark, 2006.
18. Lassila, O.; Guinness, D. The role of frame-based representation on the semantic web. Technical report kls-01-02. *Linköping Electron. Artic. Comput. Inf. Sci.* **2001**, *6*, 2001.
19. Corcho, O.; Fernández-López, M.; Gómez-Pérez, A. Methodologies, tools and languages for building ontologies: Where is their meeting point? *Data Knowl. Eng.* **2003**, *46*, 41–64. [\[CrossRef\]](#)
20. Wimalasuriya, D.C.; Dou, D. Ontology-based information extraction: An introduction and a survey of current approaches. *J. Inf. Sci.* **2010**, *36*, 306–323. [\[CrossRef\]](#)
21. Lévy, F.; Tomeh, N.; Ma, Y. Ontology-based Technical Text Annotation. In Proceedings of the SADAATL 2014, Dublin, Ireland, 24 August 2014; Association for Computational Linguistics and Dublin City University: Stroudsburg, PA, USA, 2014; pp. 21–30.
22. Maynard, D.; Yankova, M.; Kourakis, A.; Kokossis, A. Ontology-Based Information Extraction for Market Monitoring and Technology Watch. In Proceedings of the ESWC Workshop ‘End User Aspects of the Semantic Web’, Heraklion, Greece, 29 May–1 June 2005.
23. Alatrish, E. Comparison Some of Ontology Editors. *Manag. Inf. Syst.* **2013**, *8*, 18–24.
24. Osenova, P.; Simov, K. Syntactic-semantic treebank for domain ontology creation. *Cogn. Stud.* **2011**, *11*, 213–226. [\[CrossRef\]](#)
25. Sharma, N.; Nidhi. Text Extraction from Images: A Review. In *Information and Communication Technology for Sustainable Development. Lecture Notes in Networks and Systems*, 1st ed.; Mishra, D., Nayak, M., Joshi, A., Eds.; Springer: Singapore, 2018; Volume 10, pp. 153–165.
26. Pooja; Dhir, R. Video Text Extraction and Recognition: A survey. In Proceedings of the IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 23–25 March 2016; IEEE: New York, NY, USA, 2016; pp. 1366–1373.
27. Deepa, R.; Lalwani, K.N. Image Classification and Text Extraction using Machine Learning. In Proceedings of the 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 12–14 June 2019; IEEE: New York, NY, USA, 2019; pp. 680–684.
28. Gera, D.; Jain, N. Comparison of Text Extraction Techniques- A Review. *IJIRCCE* **2015**, *3*, 621–626.
29. Schoechle, T. *Standardization and Digital Enclosure: The Privatization of Standards, Knowledge, and Policy in the Age of Global Information Technology*, 1st ed.; IGI Global: Hershey, PA, USA; University of Colorado: Boulder, CO, USA, 2009; 304p.
30. BSI Standards Publication. *A Standard for Standards—Principles of Standardization*; BSI Standards Publication: London, UK, 2011; 44p.
31. Mijatovic, I. Teaching Standardization to Generation Z-Learning Outcomes Define Teaching Methods. In *Sustainable Development. Knowledge and Education About Standardisation*; Idowu, S.O., Henk de Vries, J., Mijatovic, I., Choi, D., Eds.; Springer Nature: Cham, Switzerland, 2020; pp. 191–208.
32. Portney, K.E. *Sustainability (The MIT Press Essential Knowledge Series)*, 1st ed.; MIT Press: Cambridge, MA, USA, 2015; pp. 6–7.
33. Orasan, C. Comparative Evaluation of Modular Automatic Summarization Systems Using CAST. Ph.D. Thesis, University of Wolverhampton, Wolverhampton, UK, 2006.
34. Radev, D.E.; Hovy, E.; McKeown, K. Introduction to the Special Issues on Summarization. *Comput. Linguist.* **2002**, *28*, 399–408. [\[CrossRef\]](#)
35. Hovy, E. Text Summarization. In *The Oxford Handbook of Computational Linguistics*, 1st ed.; Mitkov, R., Ed.; Oxford University Press: Oxford, UK, 2003; pp. 583–598.
36. Kapetanios, E.; Tatar, D.; Sacarea, C. *Natural Language Processing. Semantic Aspects*; CRC Press: Boca Raton, FL, USA, 2014.
37. Hahn, U.; Mani, I. The challenges of automatic summarization. *IEEE Comput.* **2000**, *33*, 29–36. [\[CrossRef\]](#)
38. STANDARD. Standards of Intelligent Transport System. ITS STANDARDS SEARCH (Czech & English Versions). Available online: <https://silmos.cz/standard> (accessed on 19 December 2020).
39. Ganter, B.; Wille, R. *Formal Concept Analysis. Mathematical Foundations*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 1999.