Conceptualising Informatization with the Onto6 Methodology

Uldis Straujums
University of Latvia, 19 Raina Blvd., Riga, LV-1586, Latvia
uldis.straujums@lu.lv

Abstract This paper presents a way of conceptualising informatization through a new methodology that is called Onto6. Informatization is defined as the maintained process of creating the technical, economic, and social conditions which are necessary for the fulfillment of information needs. The Onto6 methodology identifies objects and determines their interaction and functionality. The methodology is based on meta-ontology, and it involves the creation of an instance in accordance with the domain. The initial ontology is created from the ontology instance, and it is extended during the informatization. In the initial ontology, root metaphors and the relationship among same must be defined by the planner of informatization. More work is being done by the implementers of informatization to refine the initial ontology. This creates an ontology cluster which consists of meta-ontology, a meta-ontology instance, the initial ontology, and the refinements of that ontology. This reflects a conceptualisation of informatization in a particular domain. The author has taken part in the conceptualisation and maintenance of informatization in several domains. The Onto6 methodology that is proposed in this paper has been applied to several domains at the national level.

1 Defining Informatization

Use of the term “informatization” on the Internet has increased continuously over the course of time. The Google search engine found 40,400 links to the word in November 2004, 258,000 in January 2006, and 1.24 million in September 2007. The Alta Vista engine found only 1,868 references in April 1998. Despite this fact, however, it is difficult to come up with a precise definition of the term. Different sources provide slightly different definitions. The South Korean National Computing Agency [Lim01] defines “informatization” as “converting the main goods and energy of a social economy to information through the revolution of high data communication technology and utilising information produced by gathering, processing and distributing data within the vast fields of the society.” A report prepared for the United Nations Economic Commission for Europe [Hal03], the definition says that “informatization” is the process whereby information technologies transforme economies and societies.

In this paper, we shall use the definition produced by Bicevskis, Andzans, Ikaunieks, Medvedis and Straujums in Education Media International [BAIMS04]:
"Informatization is the maintained process of creating the technical, economic and social conditions for the fulfilment of information needs."

All of these definitions make it clear that while informatization covers the area of computerisation, it is, in fact, a broader term than just that. This author has participated in the conceptualisation of informatization and the maintenance of informatization for several domains, ranging from domains at the national level to the domain of an educational institution. These are the case studies which are examined in this paper:

- The Latvian Education Informatization System (LIIS) project, which deals with the whole range of information issues – educational content, management, information services, infrastructure, as well as user training at several levels – schools, school boards, and the Ministry of Education and Science. The LIIS project is an essential component of the Latvian National Informatics Programme;
- The Unified State Library Information System project (VVBIS), which was developed on the basis of the requirement of the National Informatics Programme that Universal Information Services be created;
- The informatics curriculum standard for Latvia’s general education schools. The analysis of the needs of the people of Latvia in terms of an ICT-educated workforce has been conducted, and solutions have been proposed as to the content of informatics courses. Detailed plans on how to provide the necessary training for existing and future informatics teachers have been drawn up.

All of the projects which are examined in this paper have actually been implemented over the course of many years and with considerable success.

2 An Ontological Approach to the Conceptualization of Informatization

Conceptualisation of informatization is a process with several phases – analysis of the existing situation, setting of goals, planning of activities, and evaluation of costs. This tends to be an iterative process, and several alternatives are compared. The consequence is that a framework must be established so as to correlate the knowledge that has been acquired during the developmental process. The requirements for such a framework include:

- The ability to organise the essential quality in a hierarchical form;
- The mapping of the qualities in the context of recognised standards;
- The interaction of organisational, domain-specific and technological aspects of the process.

All of these requirements led to the introduction of several different approaches in accumulating and then reusing knowledge – controlled vocabularies, thesauri, classification schemes, taxonomies, topic maps, frame languages, logical theories, and meta-models. There are many other approaches called ontologies. In formal terms, an ontology is the explicit specification of a conceptualisation for a domain [Gru95]. This specification can take the form of a logical theory, accounting for the intended
meaning [Gua98], or it can strive to make use of the notion of linguistic relativism [Wys04]. According to Crubézy and Musen [CM04, p. 321], “ontologies support the creation of repositories of domain-specific reference knowledge – domain knowledge bases – for communication and sharing of this knowledge among people and computer applications.” The differences lie in the ability to describe terms and to define relations among them. The differences at the level of formality create an ontology spectrum [Wil04].

The controlled vocabulary, i.e., the list of enumerated terms, is at one end of the spectrum. Ideally, each term should have just one meaning. In practice, however, terms are qualified in accordance with different meanings in different domains. If several terms have the same meaning, one is preferred, while the others are classified as synonyms or aliases. The controlled vocabulary is used to build up more advanced ontologies. For example, a thesaurus is built up by adding associative relationships to vocabulary. Frame languages have the ability to express the properties, logical constraints, and detailed relationships of terms. A meta-model is an explicit model within a domain of interest, containing terms and rules that are needed to build specific models. A meta-model is an ontology, but it is a richer notion – it can be used as a set of building blocks and rules which apply to the construction of models, as a model of a domain of interest, or as an instance of another model.

3 Onto6 – a Methodology for the Conceptualization of Informatization

This author has developed a methodology that can be used in the early conceptualisation of informatization in different domains. The target audience for this methodology is made up of the users of the domain which is undergoing informatization. These users are usually unaware of formal means for describing systems – UML, OWL, GRAPES, OMT, etc.

If we take into account the skills and the knowledge of these clients, we understand that the methodology which is proposed to them must be simple, understandable, and extendable. The Onto6 methodology was inspired by several sources, particularly the GRAPES-86 modelling language, the Object Modelling Technique (OMT), and the 6W approach.

GRAPES (Graphical Engineering System) [GRA02] is a method for system development which supports the entire software development process, from problem analysis to implementation. The GRAPES-86 modelling language is the central element, making it possible to specify the structure, behaviour, and data of information processing systems, particularly distributed systems such as company organisations and network architectures. GRAPES-86 has a formal, defined text and graphic syntax, which means that it can be statically assessed. Furthermore, GRAPES-86 includes a dynamic processing model. The dynamic behaviour of GRAPES models can, therefore, be simulated and analysed. A modelling and development environment GRADE implementing GRAPES is developed [PKB93].
The level of formalism is too high, however, and the learning curve is too steep for typical informatization clients to make effective use of GRAPES-86. More concise means for the conceptualisation of informatization had to be defined. The attractive approach developed by Rumbaugh [Rum97] is called Object Modelling Technology, or OMT. It is simpler than GRAPES, using object-oriented modelling to think clearly about problems and to draw three kinds of diagrams. The need to study the precise notion of diagrams and the preoccupying attention to software development, however, make this approach difficult for the typical informatization user to understand.

The 6W approach is extensively used in various areas – knowledge management [KMO07], context analysis [Mot00], architectural design [Lan04], etc. The approach involves six questions about the topic – what, where, when, how, why, and who. The approach was created by Rudyard Kipling back in 1902 [Kipl02]. This may seem to oversimplify the complexity of the topic that is being considered, but in reality there is a vast amount of sub-questions which arise from the points of view and expectations of those who are doing the asking. The author proposes a methodology based on the 6W approach as a top-level structure which can be understood by every client, one which supplies the most typical sub-questions (terms) for further investigation of the relevant topic.

3.1 Meta-ontology

The developed Onto6 methodology identifies the object and determines interaction and functionality. The methodology is based on a meta-ontology which creates an instance in accordance with the domain. This instance remains stable and unchanged during the informatization process. The initial ontology is created from the ontology of the instance, and then it is extended during the process of informatization. The proposed meta-ontology is shown in Figure 1.

![Fig 1. The Meta-ontology of the Onto6 Methodology](image-url)
The meta-ontology contains the most abstract terms and the relations among those terms, presenting them in a simple form. The level of simplicity is based on the level of competence in the sense of the formal notations of the system’s users, i.e., the clients of the process of informatization.

The top-level terms that have been chosen by the author are inspired by the 6W approach. Each term has its attributes and sub-terms. There can be vast numbers of sub-terms which characterise each top-level term. Lower-level terms themselves can have sub-terms. This means the emergence of a hierarchy, possible one that is recursive.

The author has proposed a set of terms for top-level term attributes. These reflect the most abstract aspects of each top-level term and should be mapped in relation to the real entities when creating the initial ontology. Table 1 shows the attributes of the top-level terms. The attributes are introduced into the Onto6 meta-ontology with a selective subset of attributes from a number of component ontologies [Goed99, Lep05, Hoss06, Sowa06]. The root metaphor and related concepts described in [Bar94, Mot00, VB01, Fon02, Gaz02, Pul03, Lan04, VZSS04 and Sowa07] are also taken into account. The goal is to create a relatively small set of attributes that can be understood by unsophisticated users and that is sufficient for the conceptualisation of informatization.

Table 1. Top-level Terms in the Onto6 Meta-ontology and their Attributes

<table>
<thead>
<tr>
<th>Top level term</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>Concept, Sign, Referent, Reality, Resource, Plan, Tool, Knowledge, Information, System, Model, Paradigm</td>
</tr>
<tr>
<td>Where</td>
<td>Subjective reality, Physical reality, Location</td>
</tr>
<tr>
<td>When</td>
<td>Time, State, Transition, Event, Life cycle</td>
</tr>
<tr>
<td>How</td>
<td>Abstraction, Concretizing, Rule</td>
</tr>
</tbody>
</table>
In addition to the top-level terms as such, there must also be the definition of the relationships among them. The semantics of these relationships, as defined by the author, are presented in Table 2.

Table 2. The Semantics of Relationships in the Onto6 Meta-ontology

<table>
<thead>
<tr>
<th>Relationship between terms</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>[What, Where]</td>
<td>Relationship [T1, T2] defines specific attributes characterizing the cell in two-dimensional informatization grid. The dimensions of the grid are T1 and T2, where T1=What and T2= one of (Where, When, How, Why, Who)</td>
</tr>
<tr>
<td>[What, When]</td>
<td></td>
</tr>
<tr>
<td>[What, How]</td>
<td></td>
</tr>
<tr>
<td>[What, Why]</td>
<td></td>
</tr>
<tr>
<td>[What, Who]</td>
<td></td>
</tr>
</tbody>
</table>

Relationships are symmetrical, i.e., the relationship [T1, T2] is the same as the relationship [T2, T1].
3.2 A Meta-ontology Instance

In order to provide an informatization-based description of a particular domain, an instance of the meta-ontology must be created. This meta-ontology instance reflects the characteristic aspects of a particular domain. The initial ontology of the domain is created on the basis of this meta-ontology instance.

A meta-ontology instance is created by perhaps eliminating some of the top-level terms from the meta-ontology.

3.2.1 Input for creating a metaontology instance

The domain of informatization must have a description of the existing situation – usually in terms of a document that is prepared by a working group. This description serves as the main input for creating the meta-ontology. Another input is a documented statement of goals – the vision of the desirable situation in the future. In situations when there is lack of documents describing the situation and stating the goals the input needed can be obtained by organizing interviews or brainstorming sessions with representatives of interested organizations.

3.2.2 Stages in the creation of a metaontology instance

There are several stages in the creation of a meta-ontology instance:

1) Analysis of input documents. At this stage, the matches between the entities of input documents and the corresponding terms in the meta-ontology are determined;
2) Marking of insignificant terms in the meta-ontology. The threshold parameters hw and ht are applied to eliminate insignificant words and terms. The parameter hw is defined as the percentage of appearance vis-à-vis the appearance of the most frequent word. The word w is marked as insignificant if it occurs less often than hw does; The parameter ht is defined as the percentage of appearance vis-à-vis the appearance of the most frequent term. The term t is marked as insignificant if it occurs less often than ht does;
3) Creation of the meta-ontology instance. The terms which remain in the meta-ontology after the elimination of insignificant terms in the second stage are chosen for the meta-ontology instance.

The resulting meta-ontology instances can differ, depending on the values of hw and ht.

3.3 The Initial Ontology

The root metaphors of the initial ontology and the relationship among those metaphors must be defined by the informatization planner. The definition process uses the top level terms included in the meta-ontology instance. The attributes of the terms are analyzed – the corresponding instances significance is evaluated. The process of attributes evaluation is similar by its structure to the process of elimination of insignificant terms during the creation of the meta-ontology instance. For the attribute evaluation the same threshold parameter criterion can be used as the parameter used for the elimination of insignificant terms during the creation of the
meta-ontology instance. Usually some weights are applied to the attribute evaluation. Most significant instances are included in the initial ontology. The visual representation of the initial ontology usually is chosen by the domain experts taking into account their previous representations of facts concerning that particular domain.

3.4 Refinements of the Initial Ontology

After the initial ontology is created further work has to be done by informatization implementers refining the initial ontology. The refinement process uses some instance included in the initial ontology. The instance is considered itself as a term with possible attributes. The process of adding the relevant attributes with their instances resembles the process of creating the meta-ontology instance but takes place at more detailed level. The refinement process usually is being applied to several terms leading to several refinements of the initial ontology. In this sense, an ontology cluster which consists of a meta-ontology, a meta-ontology instance, an initial ontology, and the refinements of that initial ontology – these come together to reflect the conceptualisation of informatization in a particular domain. [SB06] describes the creation of initial ontologies and their refinement, as conducted with the participation of this author for several different domains.

The essential phases of the Onto6 methodology proposed by the author are seen in Figure 2.

![Fig. 2. The Onto6 Methodology](image-url)
4 Analysis of Several Domains for Informatization

Next we can review the ontologies which the author has defined for three domains which have not changed during the informatization of the corresponding domain.

4.1 Application of Onto6 Methodology to Education Informatization

According to the general Onto6 methodology the informatization conceptualization for a particular domain should be created in several stages:

- Creation of the meta-ontology instance,
  - Analysis of input documents,
  - Marking of significant terms in meta-ontology,
  - Building the meta-ontology instance;
- Creation of the initial ontology,
- Refinement of the initial ontology.

4.1.1 Creation of the meta-ontology instance

Education informatization is a significant subprogram of the national “Informatics” program [BBB98]. As the main input document for the meta-ontology instance creation the “Informatics” program document was used. The document consists of 239 pages. For the analysis of the document a full text indexing is required. The indexing can take place either manually or automatically. The author has tried both approaches. Manual indexing is very time consuming. A convenient way of an automatic indexing is to use a Word macro counting the frequency of words in Microsoft Word document. The author has modified such a macro [GenC07] adding the specifics of processing Latvian language words and taking into account the large size of documents to be processed.

The national program “Informatics” document contains 9753 words. Author has taken the decision to consider only the words occurring in the text at least three times. The number of such words is 3303. From the remaining words the redundant words were stripped, such as: a, the, and, at etc. The number of remaining words was 2985. The threshold parameter \( \text{hw} = 10 \) was applied to the remaining words allowing to obtain the significant words in the input document. There were 28 significant words left. After determining the significant words the matches between the words in the input document and the corresponding terms in the meta-ontology were calculated. In the Table 3 the result of the analysis is presented.
From the Table 3 it can be seen that the number of the significant words in term “What” substantially exceeds the numbers of the significant words in other terms. If we apply the threshold parameter $ht = 40$, we obtain very sparse meta-ontology instance consisting only of the term “What”. Taking into account that the initial document describes the whole national program “Informatics” but our task is specifically aimed at a subprogram “Education”, a further analysis was required. Author has analyzed the chapter 6 “Education subprogram” of the national “Informatics” program. The document contains 2190 words. Only 426 words occur in the text at least three times. Stripping the redundant words, such as: a, the, and, at etc. 337 words were left. Again the threshold parameter $hw = 10$ was applied to the remaining words allowing to obtain the significant words in the input document. There were 73 significant words left. After determining the significant words the matches between the words in the input document and the corresponding terms in the meta-ontology were calculated again. In the Table 4 the result of the analysis is presented.

From the Table 4 it can be seen that the numbers of the significant words in two terms “What” and “Where” substantially exceed the numbers of the significant words in other terms. Therefore we can apply a restricting threshold parameter $ht = 40$ thus retaining only the most significant terms for the meta-ontology instance. The resulting meta-ontology instance is seen in Fig 3.
4.1.2 Initial ontology creation
The next step was to develop the initial ontology for the informatization of education, basing this on the meta-ontology instance. The requirements of the national “Informatics” programme were taken into account [BBB98]. The resulting ontology is seen in Figure 4.

The informatization of education was co-ordinated under the framework of the Latvian Education Informatization System (LIIS) project and on the basis of the initial ontology.

4.1.3 Refinements of the initial ontology
The initial ontology was refined several times during the implementation of the project to reflect more concrete aspects of the informatization grid that was presented at first. Figure 5 shows a typical example of how the initial ontology was refined in the context of information services.
Another refinement is seen in Figure 6. A structure of the education system is presented on the basis of the initial ontology, taking into account the fact that the same structure of the education informatization system is being implemented at all stages and levels of education.
Many different activities are necessary to maintain the structure of the education system. Certain goals must be achieved – measurable improvements to infrastructure, enhancement of management services, supply of teaching materials, etc. Existing resources and routines must be used effectively, and activities aimed at the creation of new resources must be introduced and scheduled. The organisational aspects of planning the use of educational materials can be seen in Figure 7.

**Fig. 7. Refinement of the Initial LIIS Ontology for the Organizational Aspects of Informatization**

The proposed refinements of the initial ontology differ in their form of representation and in their content. The authors recognise, however, that they served as the leading motif during the implementation of the LIIS project between 1997 and 2005. No significant changes were made to the initial ontology instead the fact that the implementation of the informatization permanently creates the needs for further refinements of the initial ontology.
4.2 Application of the Onto6 Methodology to the Unified State Library Information System (VVBIS) Project

The goal of the VVBIS project is to establish a harmonised information system for national and public libraries, one which allows them to make use of the opportunities that are offered by modern information technologies [AVGK01]. These are the main functions of the VVBIS:

- Searching for information. The system allows users to seek out information of interest, and it must ensure access to sources of information such as books, documents, publications, government registers, statistics, and transnational sources of information.

- Ordering of information: Once search results are obtained, the user must be able to order the necessary book, report or document, and alternative means for delivering the documents must be established.

- Delivery of information: The system for delivering books, documents, publications, multimedia items, and other sources of information must ensure not only that users can use the information at a library, but also that copies of publications can be supplied in print or electronic form. The world’s leading document delivery centres must be used, and supply centres must also be established in Latvia so as to ensure that information sources can be supplied to users in Latvia and beyond. This must also involve the services of interlibrary systems.

- Services: The library must be able to provide information services to users, offering information about all kinds of subjects – the law, culture, education, etc. Librarians must become information brokers who are aware of sources for all kinds of information and who can evaluate both the status of the service (basic, value-added, fee-based, limited, etc.) and the scope of the service. Libraries must offer public access to universal information services.

- Creating resources: Libraries must be aware of local information resources which are of lasting value and can be offered to users (including remote users). This is because libraries are a component in the national heritage. The resources must be digitalised, with databases concerning the history of the relevant region, tourism destinations, etc. Unique documents must be copied in electronic form.

- Training of librarians and users: Librarians must be aware of their mission as intermediaries between users and information. They can greatly enhance the process of obtaining information, thus becoming very important members of the Information System. It is vitally important, therefore, to improve the system under which librarians are trained. Librarians must learn how to do the various things that are a part of their job. Libraries must establish user training systems, too, so that users can work more independently.

- Marketing of information: Information is increasingly becoming a product. Research must be conducted to determine the need for value added services, while existing services must be tested and advertised.

- Unified user registration. Users can receive information from many different libraries, and the existing system of user registration must be changed to
reflect this. People must have individual codes which allow for clear identification, examination of their relationship to other libraries, and specification of their right to receive information. Libraries must accept anyone who is seeking universal information services.

- Establishment of a national bibliography and conversion of retrospective data: Because Latvia’s informational resources are part of an international framework, they must be catalogued and made available through internationally accepted formats for the exchange of metadata. All materials which have been published in Latvia must be catalogued, with retrospective cataloguing of information resources. This must be done in accordance with demand for such resources.

The meta-ontology instance which was created for the VVBIS on the basis of the aforementioned requirements is shown in Figure 8.

![Fig. 8. The Meta-ontology Instance for the Unified State Library Information System](image)

Refinements to the initial ontology for data flows in the Unified State Library Information System project are shown in Figure 9. This schema remained unchanged during the creation of the VVBIS concept.
4.3 Application of the Onto6 Methodology to the Informatics Curriculum Standard at Latvia’s General Education Schools

The standard for the informatics curriculum for Latvia’s general education schools was prepared in 2002 by a team which included this author [VBS03]. The standard was based on a questionnaire which determined the frequency of ICT usage at that time and the relevant existing needs. The respondents were divided into three groups – infrequent users of ICT, everyday users of ICT, and ICT professionals. The results were used to propose new content for the teaching of information. The standard for the curriculum is in line with the requirements of the European Computer Driving Licence (ECDL), with a few departures that are based on the needs and traditions which exist in Latvia. Because of the need for activities in support of the introduction of the new curriculum, the development team also created a timetable for a transitional period to run from 2003 to 2005. Special attention was focused on the training of existing and future informatics teachers. Textbooks were written, lesson plans were drawn up and tested at schools, and the ECDL certification system was introduced for teachers.
The initial ontology for the informatics curriculum standard includes curriculum topics, their relationship to ECDL topics, as well as links to the relevant textbooks and lesson plans.

5 Common Methodology for Informatization

This author has dealt with many projects related to informatization, with the scope of domains ranging from the national level to an individual educational institution. The projects have been successful in creating positive changes in the planned timeline. The author’s approach is to analyse the nature of properties in a semi-formal way. The resulting ontology cluster is then checked for comprehensibility and relevance. This approach combines the general methodology for ontological analysis [WG01] with the specific requirements for information systems [2290, DR02, Gre05]. The essence of the proposed Onto6 methodology is to use the meta-ontology to create a meta-ontology instance which is relevant to the problematic domain. Then the initial ontology is developed on the basis of situational analysis and agreement among team members. An incremental plan is set up to refine the various aspects of informatization which are presented in the initial ontology. The result is a cluster of ontologies. Levels of formalisation and unification can differ during the development process, that depends on the potential end user.

Conclusion

The Onto6 methodology which is proposed in this paper identifies objects, determines their interaction, and specifies their functionality. The methodology is based on a
meta-ontology, and it involves the creation of an instance in accordance with the relevant domain. The initial ontology is created on the basis of the instance ontology, and it is then extended during the process of informatization. The planner of informatization must define the root metaphors of the initial ontology and the relationship among them. The implementers of informatization then do more work in refining the initial ontology. This creates an ontology cluster which consists of the meta-ontology, the meta-ontology instance, the initial ontology, and the refinements of the initial ontology. These reflect the conceptualisation and informatization of a particular domain.

References


[Fon02] Fonseca, Frederico, Martin, James, Rodriguez, Andrea. From Geo to Eco-Ontologies. 2002, 16p


