

Afruz M. Gurbanova

DOI: 10.25045/jpis.v07.i2.07

Institute of Information Technology of ANAS, Baku, Azerbaijan
afruz1961@gmail.com

DEVELOPMENT OF TERMINOLOGICAL INFORMATION SYSTEM IN THE SUBJECT FIELD

The article provides an architecture and conceptual scheme of decision support information system for a wide range of experts of a certain subject in the field of terminological activity. It defines the opportunities created by the special knowledge database, which is oriented at term analysis of terminological information system.

Keywords: *terminological dictionary, information systems, ontology, semantic network.*

Introduction

In modern times, one of the methods to increase the level of training of highly qualified personnel in science and education system is the development and use of knowledge and data mining systems in various subject areas.

In recent years, significant changes and developments occurred in various fields of science, and technology has led to the enrichment of the terminological databases of separate subject areas and to the generation of new terms in the language. The growth of research and education centers in the world and Azerbaijan, the development dynamics of various fields of science and technology, meeting the international standards and ever-increasing demands for the terms, and other factors require the development and use of the terminological information system, which will provide the structuring of terms in a variety of subject areas, their concentration, processing and application.

Problem statement

The development of the ontology of separate subject fields serves to the training of the specialists of these fields both in the educational process and practical work. It also increases their level of knowledge.

The terminology database of the given subject field can be presented as a system aimed at certain goals [1].

$S = \langle M, R, P \rangle$ (1)

M - denotes the set of system elements: terms in the subject field and their definitions, R - the set of relationship among the terms in the majority, and P - the set of system properties to reach the goal.

As the elements of the terminology system, the terms should have the following properties:

- To be free of subjectivity of the life experience. Imagining the various objects about the same term is unacceptable.
- To be definite. That is the same term should not describe various objects of the same branch of science in different cases.
- To have fixed values, in other words, the set of fixed objects, which are described by this term.

The main feature of the system elements is that there is an exact definition of each term. To understand the term, it should have its definition, and the definition of the terms is to be used in its explanation.

The relationship among the elements of the terminology system reflects the hierarchy of the concepts, interacting with each other. The relationship among the terms is determined by the linguists and experts of the subject field, likewise the terms. The nature of the established relations may differ.

It should be noted that the terminological system is public: on the one hand, the system elements transform into the word in general use - determination, on the other, new elements – term flow (entry), are always included into the system.

Each separate property ($p_1, p_2, p_3, \dots, p_n$) of the set P characterizes the local functional quality (for example, p_1 - integrity, p_2 - publicity, p_3 - accuracy, etc.). Altogether, these properties characterize the system as a whole.

The information system, here, is reviewed as software and hardware systems, and as the realized data bank (knowledge bank). The system automatically and unequivocally supports the collection, search, identification, acquisition, storage, preservation, processing, and transmission of data in accordance with the users' requests within the framework of the certain methods, methodical and normative documents (standards).

The development of the architecture of terminological information system

The system architecture implies the information environment that covers the interaction of all its components, functions, structure, model, including the system itself. It is set on three levels (Figure 1).

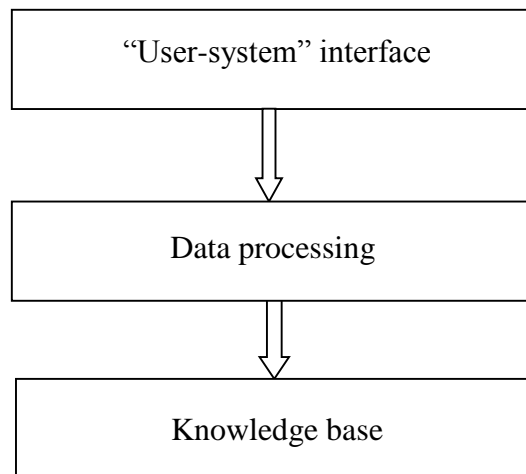


Figure 1. The architecture of information system

The first structural units of the system are the information objects, namely, the articles of the dictionaries. Here, a part of the object, for example, the title of the article, can be physically stored in the low level database (file). The text or a fragment of the article can be located or found in another database (file). However, the article is given to the request of the external interface, with the title, as a whole and in its original form.

The 3rd level of the system architecture is performed as follows:

- conceptual model of the knowledge base of the information system is created on the basis of the presentation of (1).
- ontology of the subject fields is established in the form of hierarchy of common notions (M – denotes the set of terms and definitions), in order to provide the knowledge space.

Ontology of the subject field – implies a formal description of the subject field. It is usually applied to identify the common terminology database of the subject field. In information technology and computer science, ontology means the set of the objects and the relations among their descriptions. Formally, ontology includes the terms, their description and access rules. Thus, the ontology is the model of the subject field, and accordingly, it is the basis for the knowledge bank and knowledge base.

The main purpose of the ontology is to standardize the knowledge model and to maximally (mathematically) describe it [2].

The advantages of establishing the ontology are:

- To facilitate the mutual access to the data structure about the subject field;
- To facilitate the compliance of various parts of the subject field models;
- To provide the availability and replacement of the models for new users;
- To separate the knowledge structure of the subject field from the specific data.

Ontology is developed in several phases. After defining the subject field, the search and reuse of existing ontology can be reviewed. Later, a complete list of the terms of the subject should be acquired, and the hierarchy of the classes should be developed and their features should be described. The following step includes describing the relationship between the ontology elements. At this phase, it is determined at which stage of the hierarchy the notion is, whether the object is classified or not, and so on. Separate examples of the classes are created at the last phase.

The ontological approach to the modeling of the subject field provides the development of new information systems and the interoperability of the preliminary information systems. At present, a paradigm of two-level information systems is growing.

In computer science, ontology is a detailed description of the certain knowledge fields with a conceptual scheme (semantic network of mutual relations of the notions and concepts within certain rules). Actually, this scheme reflects the structure of the data. It includes all relevant classes of the object, the links between them and adopted rules in this field (theorems, limitations) [3].

In the process of programming, ontology is used as a form of presentation of knowledge about the real world and its parts. It is mainly applied in the business process modeling, semantic web and artificial intelligence. Description language on the ontology is a formal language, which is used for its coding. Several similar formal languages (*OWL - Web Ontology Language*, *KIF - Knowledge Interchange Format*) are available.

The analysis of different models for the data description has justified the selection of multi-component semantic networks as the tool for ontology description.

The semantic network is an oriented graphics. Its tops denote the notions, and its edges show the relationships between the notions. The notion expresses any abstract or certain objects, whereas the relationship – the links between the objects. Semantic Web relations can be “partial-full” type (class-group, element-set), functional (“occur”, “affect” and so on.), quantitative (more, less, equal, etc.), spatial (far, near, under, above, in, etc.), logical (and, or, not), linguistic and so on [4].

In the knowledge base of a semantic network type, the search of solution is reduced to the search of network fragment, which reflects the query sent to the database in accordance with any sub-network. This model was proposed by an American psychologist Quillian M. [5].

The semantic network is implemented with the establishment of the thesaurus (an active tool for the description of the separate subject fields), where the hierarchy of the definitions is reflected. Unlike the explanatory dictionary, the thesaurus provides the notion not only by its definition, but also its understanding through the coordination with other words and groups.

The thesaurus defines five types of relations: a broader term - top; more narrow term - bottom; associated (linked) term - association; whole for the term - part; part for the term - whole.

It should be noted that, in fact, establishing the ontology in knowledge field and professional field of activity has a tendency to grow from field standard level up to technical materials.

The establishment of the 2nd level of the system architecture is based on the model of text type data search. This model refers to the presentation of documents and requests, the criteria for logical compliance, the mechanism for the survey results ranking (by its importance), and feedback mechanism for the evaluation of the document relevance (survey compliance rate).

The implementation of this model is associated with the search for the full text out of all documents. The user request rate (relevancy), search indexes relevance, reliability, service flexibility is determined by the search efficiency criteria.

The first level of the system architecture is established using *HTML* programming language through the data availability, which is based on hypertext information presentation.

It can be noted that the technical requirements of the software implementation are the ease of use of the information system, design compatibility, integrity, publicity, the presence of search system, the presence of query system, the availability of data protection tools, access and review of legislative documents and materials (the books related to terminology, terminological dictionaries, etc.).

The system has been established in accordance with the standards of the International Organization for Standardization [6, 7].

The conceptual model of terminological information system on a given subject is presented in Figure 2. Thus, the development of the terminological information system has created great opportunities for the development of the specialized knowledge base aimed at term analysis and for conducting studies in the field of terminology in Azerbaijan [8].

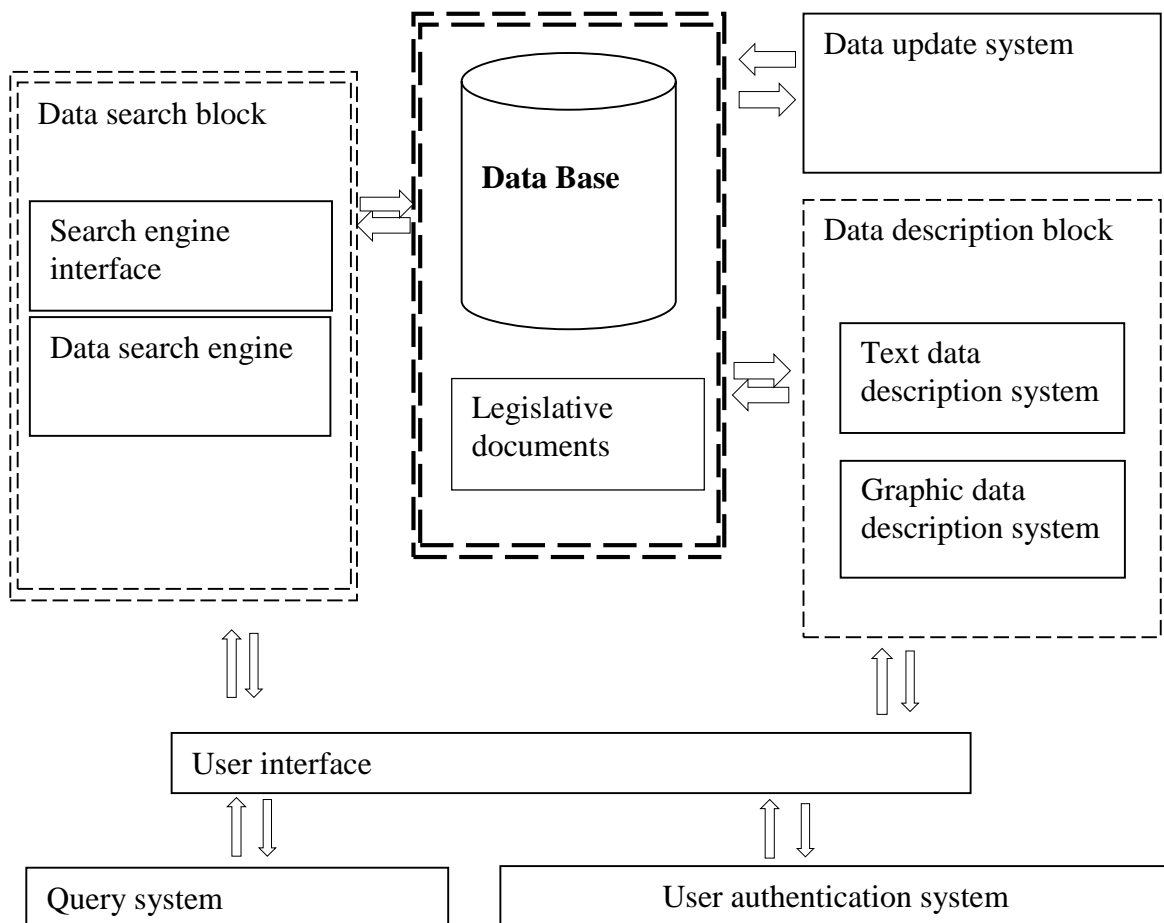


Figure 2. Conceptual model of the information system

Conclusion

The development of such a decision support system and software for a wide range of specialists in terminology in Azerbaijan is the first step.

From above mentioned, we can conclude that the terminological information system is a semantic network used on the set of terms. It is implemented through the thematic Internet portal that provides various interactive services to the users, where the key point is the search engine of the portal.

It should be noted that certain implementations have been done on the software of the terminological information system. The software has been developed with *PHP* program language running under the *Linux* operating system. The term sought in any subject area, definition of the term, its translation into the possible languages and other metadata are reflected here.

This work is supported by the Science Development Fund under the President of the Republic of Azerbaijan - Grant No SDF-2014-9 (24) KETPL-14/02/1

References

1. Novozhilova M.V., Usherov-Marshak A.V., Latorets E.V., Mikheev I.A. Development of terminological information systems in the field of concrete study. Kharkiv State Technical University of Construction and Architecture, Thematic collection of “Data processing systems”, 2010, edition 6 (87), pp. 139-142.
2. Konstantinova N.S., Mitrofanova O.A., Ontology as a knowledge storage system, Saint Petersburg State University, St. Petersburg, <http://www.ict.edu.ru/ft/005706/68352e2-st08.pdf>
3. [https://ru.wikipedia.org/wiki/Ontology_\(information_science\)](https://ru.wikipedia.org/wiki/Ontology_(information_science))
4. https://ru.wikipedia.org/wiki/The_semantic_web
5. Quillian M.R. Semantic memory. Semantic information processing, MIT Press; reprinted in Collins & Smith (eds.), Readings in Cognitive Science, section 2.1, 1968, pp.227–270.
6. ISO / TR 12618: 1994. Automated terminological data processing tools. Creation and use of terminological databases and text collections, <http://www.iso.org>
7. ISO 12207. “The processes of software life cycle”, <http://www.klubok.net/pageid313.html>
8. Aliguliyev R.M., Gurbanova A.M. Conceptual framework of developing the terminological information system. Problems of Information Society, No1, Baku, 2011, pp.3–8.