Life Cycle of the Electronic Encyclopedia as the Information System

Pavlo Zhezhnych, Mariya Hirnyak

Information Systems and Networks Department,
Lviv Polytechnic National University, UKRAINE, Lviv,
S. Bandery street 12, E-mail: mhirnyak@ukr.net

Abstract – The paper dwells on the integral aspect of the electronic encyclopedia, namely its life cycle. There is a generalized system of the electronic encyclopedia formation in this article. It reveals the essence of the each life cycle stage. It is justified the choice of knowledge extraction from open texts for the encyclopedic conten formation.

Key words – electronic encyclopedia, life cycle, information system, knowledge extraction, open texts.

I. Introduction

Nowadays there is a lot of information concerning that or other topic in the Internet space. Search engine (Google, Yandex) receives an inquiry and generates data containing a lot of unnecessary and insufficient information. This information is given in the form of text documents, diagrams, graphs, audio- and videomaterials in web-portals, forums, encyclopedias, etc. In most cases to get a sound, terse and reliable information people search in the electronic encyclopedias to quickly get the answer to their inquiry. Moreover, during the last decade it is observed a rapid grow of the encyclopedias in the World Wide Web [6].

Electronic encyclopedias provide background information on various issues. Therefore, it should be considered the primary peculiarities of the electronic encyclopedia formation due to the adequate encyclopedia functioning and content. The quality of the electronic encyclopedia derives from the appropriateness of the software and a comprehensive and reliable content.

Regardless of the electronic encyclopedia type, it is analyzed a generalized system of its formation.

II. Life Cycle of the Electronic Encyclopedia

Electronic encyclopedia is the source of information in electronic form, which covers the most comprehensive and reliable background objective, scientific and verified material from one or more fields of knowledge in the form of articles (entries) in cmbination with audio and video materials to meet the information needs of people with access via a local (CD, DVD disks) and global (Internet) network. Online encyclopedia is getting more and more widespread; they are also referred to as web and online encyclopedias.

This article demonstrates the formation of the electronic encyclopedia to be presented in the World Wide Web.

Developing the electronic encyclopedia it should be considered the life cycle of the electronic encyclopedia as the information system. It is composed of the following stages:

1) requirements analysis;
2) design;
3) realization / implementation;
4) testing and verification;
5) maintenance.

Let us dwell on each of the stages, focusing on the structural features of the electronic encyclopedia and open texts. Open texts are processed by knowledge extraction and makes up the encyclopedic content.

Knowledge extraction is a methodology based on the triad «the object, subject and method of research». The object is electronic texts, the subject is the computer technologies of their intellectual processing, the basic methods are statistic analysis and the theory of fuzzy sets [2].

For the electronic encyclopedia content formation it should be applied knowledge extraction from open texts because most of the information resources contained in the web network (World Wide Web) is a semi-structured texts to be processed by the content. Knowledge extraction enables the comparison, analysis and synthesis of heterogeneous information from open texts that are available to the general public and are intended for the storage, transmission and conversion [3].

Requirements analysis. Requirements to the electronic encyclopedia content are the following:

• a source of the electronic encyclopedia content is a set of open texts that are processed using knowledge extraction;

• the content is to contain the logical, consistent, coherent presentation of the material with a reference nature from the scientific and verified sources.

Design. The design stage is characterized by:

• the formation of the thematic encyclopedia rubrics (classifiers);

• development of an information model of the terminological dictionary and encyclopedic articles (entries);

• development of the electronic encyclopedia architecture system taking into account the needs and interests of users of all ages and fields of activities.

The architecture of the electronic encyclopedia formation system is a complex of such subsystems [Fig. 1]:

• the subsystem of the key words formation on account of the target audience (in the form of the dictionary);

• the subsystem of the open sources analysis on account of the terminological dictionary of the encyclopedia (classifies the information blocks of the open texts by the keywords);

• the subsystem of knowledge extraction from open resources (identifies knowledge from classified information blocks);

• the subsystem of the encyclopedic article (entry) formation (on the basis of the extracted knowledge);

• the subsystem of the cross-references development (specifies connections with external and internal articles / resources).

In general, a system of the electronic encyclopedia formation is the establishment of a new (subsequent) versions of the articles (entries) using the above subsystems. A result of this system functioning is a new article or new (updated) version of an existing article (entry).
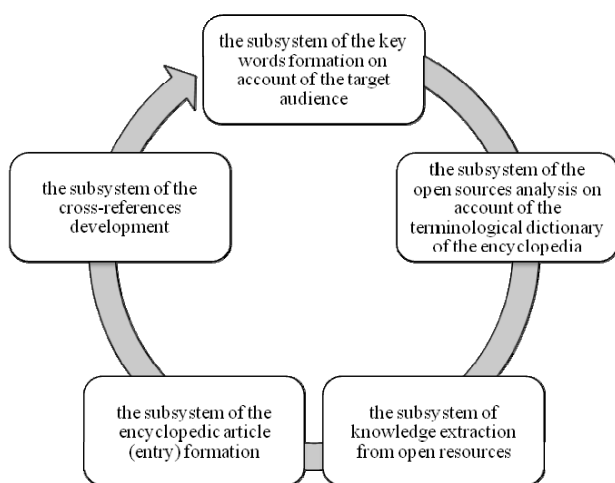
Fig. 1. System of the electronic encyclopedia formation

Realization / implementation. Electronic encyclopedia contains a detailed system of classifiers (rubrics) and the system of internal search which is realized by the dictionary of special terms. The development of the electronic encyclopedia begins with the formation of the dictionary of specialized terms, where each term is accompanied by the creation of a single encyclopedic article.

Electronic encyclopedia design is an automated process. A human expert deals with the development of the classifiers and the retrieval system, and the computer fetches the necessary material for the encyclopedic content formation distinguishing the basic elements of that material by knowledge extraction from open texts.

Considering the process of knowledge extraction from open texts it should be analyzed a system of knowledge extraction. The primary issue for the electronic encyclopedia content formation is the development of a knowledge extraction system or using the existing one.

Testing and verification. The testing stage is characterized by testing the quality of the electronic encyclopedia content.

To estimate the quality of the content it is defined such common evaluation criteria as the following:

• the presentation of artices (entries) within each classifier (rubric);

• the coherence of the articles (entries);

• the completeness and relevance of information.

Maintenance. Installation of the electronic encyclopedia into the web network, support and development through constant changes and updates for the completeness and relevance of the content.

To facilitate the electronic encyclopedia to be in use it should be done the above and should be taken into consideration the needs and interests of the users through the feedback and various questionnaires.

Conclusion

In this paper it is analyzed the main stages of the electronic encyclopedia life cycle. It should be taken into consideration all the requirements and peculiarities of these stages in order to develop a high quality functioning electronic encyclopedia.

Electronic encyclopedia formation requires the systematization of the material in accordance with the developed classifiers (rubrics), defining the key points and the appropriate cross-references and hyperlinks. The process of knowledge extraction makes the process of electronic encyclopedia formation automated.

Such issues as the elaboration of the approach to improve the quality of the electronic encyclopedia, elaboration of the electronic encyclopedia information model, detailed analysis of the subsystems (that in complex form a system of electronic encyclopedia formation) need further study.

References

[1] V.V. Osadchyi, "Peredumovy ta tekhnolohii stvorennia osvitnikh Internet-resursiv" ["Prerequisites and technologies of the educational Internet resources generation"], Problemy inzhenerno-pedahohichnoi osvity − Problems of engineering and pedagogical education, № 22-23, pp.162-170, 2009.

[2] V.I. Gor'kova, "Informetrija (kolichestvennye metody v nauchno-tehnicheskoj informacii)" ["Informetrics (quantitative methods of scientific and technical information)"], Itogi nauki i tehniki. Ser. Informatika − Outcomes of science and technology. Ser. Informatics, Vol. 10, 1988.

[3] P.I. Zhezhnych and M.H. Hirnyak, "Formuvannia elektronnoi entsyklopedii za dopomohoiu ekstraktsii znan z vidkrytykh tekstiv" ["Electronic encyclopedia formation through the knowledge extraction from open texts"], in Proceedings of the I[st] International Conference "Information, communication, society": ICS-2012, April 25-28, 2012, Lviv. Ukraine. Lviv: Vydavnytstvo Lvivskoi politekhniky Publ., 2012. pp. 32-33.

[4] A.P. Kurshev, E.A.Sulejmanova and I.V. Trofimov, "Rol' znanij v sistemah izvlechenija informacii iz tekstov" ["The role of knowledge in systems of information extraction from texts"], Programmnye sistemy: teorija i prilozhenija − Software systems: theory and applications, № 3 (12), pp. 57-70, 2012.

[5] T.A. Gavrilova and V.F. Horoshevskij, Bazy znanij intellektual'nyh sistem [Knowledge bases of intelligent systems]. St. Petersburg: Piter, 2000. (Russian)

[6] P.I. Zhezhnych and M.H. Hirnyak, "The architecture of the electronic encyclopedia formation system on the basis of open texts", Computer design systems. Theory and practice, №747, pp. 79-83, 2012.