Development of a knowledge base on the topics of theoretical electrical engineering using software tools for analysis and generation of collections of WIKI articles

Authors:
Roman Alekhin
Pavel Varshavskiy
Anton Kozhevnikov

Department of Applied Mathematics and Artificial Intelligence
Pavel Butyrin
Felix Shakirzianov
Maria Karpunina

Department of Theoretical Fundamentals of Electrical Engineering

National Research University «MPEI»

Speaker:
Roman Alekhin
NRU «MPEI»
Improving modern computer-based instruction (CBI) tools, digital libraries (DL) and knowledge bases (KB) has led computers to become an integral part of the student learning system. The use of CBI plays an important role in the learning process and positively affects its performance.

This paper discusses various aspects related to the development of software for analysis and automatic generation of collections of wiki articles from the KB on the theoretical foundations of electrical engineering (TFEE) for use in the educational process of the Department of TFEE of NRU "MPEI".
Task of creation of CBI on TFEE

The discipline of TFEE provides the basis of a professional language, fundamental knowledge, a general methodology for solving research problems for students of all specialties, studying in the field of training "Power Engineering and Electrical Engineering". This discipline also forms a moving disposition of sociocultural-historical, cultural, philosophical, sociological information related to this area.

Thus, TFEE is the basic and at the same time the most complex, voluminous and multifaceted discipline in this area of training. In this regard, the task of providing a good gestalt (holistic vision, perception) of knowledge of TFEE with the possibility of effective independent development and continuous replenishment is extremely important. The task of creating a knowledge base on TFEE as an effective, modern, educational resource was first announced at the UIE open international congress.

CBI, Internet and Web-technologies provide the technological basis for creating such a knowledge base. The new habitus (lifestyle), generated by the gadget-tablet revolution, determines the mental and behavioral predisposition of students, and with them engineers and scientists, to perceive this knowledge base as the main repository of knowledge of TFEE and an effective educational tool for their development and replenishment. Such a knowledge base meets the needs of intellectualization and "digitalization" of the electric power industry and electrical engineering.
WIKI technology

To implement electronic knowledge base, a promising Internet technology \textit{wiki} was chosen.

The term \textit{wiki} was first used in 1995 by Ward Cunningham, the developer of the first Portland sample repository WikiWikiWeb wiki system code. Cunningham borrowed the word Hawaiian, meaning “fast,” and later the English backronym “What I Know Is ...” was coined.

\textit{Wiki} is an open Internet technology, the fundamental principle of which is that any visitor to a wiki resource can independently change its structure and content using the tools provided by such a resource.

\textit{Key features of wiki technology}: collaboration and version control, wiki markup language, use of hypertext.
Semantic WIKI

*Semantic wiki* is a web application using machine-processed data with strictly defined semantics in order to expand the functionality of a wiki system.

*Semantic wiki* allows you to specify the type of links between articles, the data type within articles, as well as page information.

Features of semantic wiki systems:
- semantic annotations (for example, to indicate the type of link links between articles);
- contextual presentation of data (for example, displaying related articles, context-sensitive links);
- support for Semantic Web standards, support for RDF, OWL, SPARQL query language;
- semantic search based on queries;
- support for logical inference.

Implementations of *semantic wiki* systems can be divided into two types according to the method of storing metadata. The first type (Semantic MediaWiki, etc.) implies the inclusion of semantic annotations directly in the page text using advanced wiki markup, in the second type of systems (OntoWiki, etc.) structured data is stored separately and entered using a special input interface.
Implementation of WIKI systems

There are a number of freeware implementations of wiki systems. The most common are MediaWiki, DokuWiki, OpenWiki, etc. The MediaWiki wiki engine was chosen for this work, since MediaWiki allows working with the Russian language, is actively supported by the WikiMedia Foundation and has great potential for expanding basic functionality through an extensive plug-in library.

MediaWiki implements the display of wiki-documents, provides an interface to work with the database of pages, flexible differentiation of access rights and user roles via ACL (Access Control List), the ability to process text in its own wiki-markup format, as well as in HTML and TeX, the ability to download and work with media files, full-text search and keyword search tools, messaging tools, support for many languages, etc. A flexible extension system allows users to add their own new opportunities and programming interfaces.

There are several approaches to displaying mathematical formulas in wiki projects using the MediaWiki engine:

- embedded mathematical formulas using TeX. To display the formulas, you need to wrap the LaTeX formula in the math tags or it is also possible to use external services that generate images from LaTeX formulas.

- JavaScript (MathJax and others). Such extensions require the installation of JavaScript on the client machine and generate images of formulas on the fly, which can affect the performance of the system as a whole.

Fig. 1. A fragment of a page of a graphic file
Development of knowledge base for TFEE

It is advisable to represent the structuring of knowledge on TFEE in the form of a semantic network, the main block of which is the block of basic categories <TFEE language; material objects; phenomena, effects and processes; laws; equations; tasks; methods ...>.

Each of these categories represents a certain ontology - a semantic network of a special type, the elements of which contain an ordered description of certain basic concepts. So a fragment of the ontology of some blocks of the TFEE language <definition → alphabet → dictionary of basic concepts → physical quantities in electrical engineering and their units → multiple and fractional units of physical quantities → sign designations → frequently used alphabetic abbreviations (acronyms) → Latin and Greek alphabets.

Other structures are also presented in the form of an ontology, for example, a dictionary of basic concepts implemented in GOST R52003-2003.
Analysis and automatic generation of collections of WIKI articles

With the growth of information available to each user, the problem of its analysis and selection arises. To solve this problem, a person needs quite a lot of time, a significant part of which is spent on analyzing information for its relevance.

Therefore, to reduce the cost of searching for information and forming the necessary set of knowledge in the form of a selection of wiki articles for CBI, an application was developed that collects information on a separate resource using MediaWiki technology and automatically generates a selection of wiki articles on a specific topic related to a user-defined concept.

For the application to work, you must select the wiki resource of interest, select the concept of interest to the user, and enter sections to categorize the articles. Educational wiki resources are often referenced outside their borders, for example, on the official Wikipedia website, so the application is limited only by the scope of the selected resource.
Development of a knowledge base on the topics of theoretical electrical engineering using software tools for analysis and generation of collections of WIKI articles

Analysis and automatic generation of collections of WIKI articles

Fig. 2. The algorithm of the application for analyzing wiki articles
Analysis and automatic generation of collections of WIKI articles

Each article stored in memory goes through several stages of analysis:

• highlighting keywords in a brief description;
• allocation of subsections (table of contents);
• highlighting keywords in the main text, taking into account statistical characteristics;
• highlighting keywords in the main text, taking into account the author's emphasis in font;
• highlighting links;
• selection of categories.
Analysis and automatic generation of collections of WIKI articles

In the case of the main article, the article is added to the set, and its selected links go through the procedure of updating mentions: in the list of all articles for articles corresponding to the selected links, the number of references increases.

In the case of a non-main article, the information received goes through a comparison procedure with articles in the application’s memory. To determine the similarity of the two articles, the Jaccard coefficient is used:

\[ K_i = \frac{c}{a + b - c}, \]

where \( a \) is the number of characteristics of the new article; \( b \) is the number of characteristics of the article stored in the application memory; \( c \) - the number of common characteristics for the two compared articles.
Development of a knowledge base on the topics of theoretical electrical engineering using software tools for analysis and generation of collections of WIKI articles

Analysis and automatic generation of collections of WIKI articles

Comparison of two articles includes the following steps:

1) checking the similarity of the content of articles;
2) determining the similarity of articles by keywords;
3) determining the similarity of articles by links to other articles;
4) determining the similarity of articles by category;
5) the decision to add a new article to the collection, taking into account the calculated degree of similarity and given coefficients (80% for non-main articles, 60% for the main article).

If, as a result of comparing the new article with the ones in the collection, more than half of the articles returned the consent agreement to add it to the set, then the article is added to the set, and its links are sent to the procedure for updating mentions.

This article is deleted from the list of all articles, and then the next article with the most references is selected for analysis.
Analysis and automatic generation of collections of WIKI articles

After receiving the entire set of articles, the articles are categorized into selected sections in order to structure the presentation of the material. The user enters the desired sections and their keywords.

As a result of the application’s work, a selection of articles on one topic around a concept specified by the user is stored in memory, this collection can be saved on the hard drive and can be viewed at any time.
Analysis and automatic generation of collections of WIKI articles

Fig. 3. An example of the result of the application
Conclusion

Today, the Internet and Web technologies are an integral part of the educational process and the main tool for the development of CBI and distance learning systems. The indicated tools and technologies are actively used in the implementation of modern electronic, corporate and private knowledge bases.

The main attention is paid to the description of the capabilities and features of wiki technology for developing CBI and the implemented software application for the analysis and automatic generation of wiki article collections, which can reduce the cost of developing informational and knowledge base for collecting and analyzing information from the Internet.
References

Thank you for attention!

Speaker’s contacts:

Roman Alekhin
NRU «MPEI»
AlekhinRV@mpeи.ru