

SUBSYSTEM SPELL CHECKING ELECTRONIC LIBRARY CATALOGBASED ON TECHNOLOGY HUNSPELL

R. Yarmoliuk

Khmelnytsk National University

System of an electronic catalog plays an important role in ensuring the basic functions of a modern library. The quality of the data contained in the electronic catalog directly affects the quality of information retrieval and integration capabilities of the library. When working with these electronic library catalog can be exceptional situations that lead to the emergence of various kinds of errors. Key attributes of an e-catalog (author, title, publisher, abstract source) are text type. Accordingly, the main errors in text attributes are symbolic distortion or spelling mistakes caused by both objective and subjective factors. On how efficiently searching and correction of spelling errors in text attributes of an e-catalog, depends on the quality of baseline information for search queries.

Special attention to theoretical and practical foundations of the problem of finding and correcting spelling errors in the records of the electronic catalog of his works had paid Vershynyn M.I., Karaush A.S., Nielsen R., Ballard T. and others. In his writings, and development data the authors pay attention to the development of theoretical methods and means of finding and correcting spelling errors in bibliographic records. However, practical implementation based on the finished software solutions remained unnoticed authors. The book Vershynyn M.I. pays great attention to the problems of electronic classification of various technical ways. In the matter of finding misspellings offered only general ideas based on theory of indistinct sets, however, the direction of practical implementation of these ideas have been proposed.

The main algorithmic problems arising in the development of software tools spellchecking and analysis of text strings include:

The main algorithms for matching text strings are:

- Knuth–Morris–Pratt algorithm
- Rabin–Karp algorithm
- Boyer–Moore algorithm and its variations.

The main algorithms for calculating the distance between text lines are:

- Wagner–Fischer algorithm
- Hirschberg's algorithm
- Khunt-Shymanskiy algorithm
- Ukkonen-Maers algorithm.

The main algorithm unclear comparison of text strings are:

- k – mismatch Landau-Vishkin algorithm
- k – differences Landau-Vishkin algorithm.

Classical approaches to the analysis of text strings highlighted in the papers: R. Hemming, and Lowenstein V. Analysis of the possibilities of modern automated library information system (AILS) showed that the developers of these systems, insufficient attention is paid search tools and correcting spelling errors in the records databases of the electronic catalog. The absence of the module to check the spelling of a common Automated Library Information System (AILS), as UFD/Library, MARC-SQL, UNILIB, LIBER, ALEPH, Koha, ISIS, CDS Invenio, OpenBiblio, Evergreen, causing many difficulties for electronic editor directory. Editor electronic catalog - a library employee responsible for the introduction of new data in the bibliographic database of the electronic catalog and their validation. In particular, when dealing with electronic catalog, there is a need for manual review records. Therefore, the development of approaches to the design of the subsystem spell checking e-catalog is an important and timely technical problem in the development of software systems verification of the electronic library catalog (SVELC). As a result of the study were asked to author approaches for designing subsystems spell check, formulated in the structure of articles and departments of the principal.

Analyzing functional requirements were selected Hunspell open technology to build on its core subsystems spell checking electronic library catalog. The proposed scheme of structural subsystems based on platform .NET. It has been proved in practice that on the basis of the scheme might build an effective spell checking subsystem catalog of the library. The proposed scheme is new and illustrates the workflow spell check on key technologies Hunspell.

Description and analysis of object-oriented model modified platform NHunspell showed that the proposed modified Nhunspell platform incorporates the necessary tools for spell checking, custom dictionaries based on words and affixes, morphological analysis, offers a mechanism for misspelled words, work on the thesaurus.

Experiments of assess the effectiveness of the upper time limit core system Hunspell showed a linear dependence on the characteristics and parameters of computer platforms dictionaries proving to effectively apply this technology in the development SVELC.

The main purpose of this research is to develop approaches to the design of the subsystem spell checking technology based on spell checking dictionary Hunspell. A number of requirements that apply to the subsystem spell checker had been considered. The choice of system Hunspell had been described. The structure of vocabulary words and affixes had been considered. The structural and object-oriented schema subsystem spell checker had been proposed and described. The analysis of the effectiveness of the proposed approaches based on experiment has been conducted.

Keywords - electronic catalog, system of spell checking Hunspell, dictionary, affix, AILS, object-oriented model, platform .NET.