EURASIAN JOURNAL OF TECHNOLOGY AND INNOVATION
Innovative Academy Research Support Center
Open access journal www.in-academy.uz

METHOD AND ALGORITHMS FOR CORRECTING ERRORS IN THE TRANSMISSION OF TEXT INFORMATION

[1]Tishlikov Sultonjon Abduraimovich
Gulistan state university
[2]Ismatillayev Avazbek
Gulistan state university
[3]Mavlonov Sherzod Hazratqulovich
Gulistan state university

ARTICLE INFO

Received: 25th May 2023
Accepted: 30th May 2023
Online: 31th May 2023

KEY WORDS

ABSTRACT

The study of information reliability control is reduced to the development of a computer system for processing texts and the use of modern information technologies that make it possible to effectively correct errors in texts based on the semantic and statistical properties of natural language. We have studied methods and algorithms aimed at interactive modes of text information processing, especially in widespread electronic document management systems. To analyze the efficiency of the system in the mode of interactive information processing, a criterion for the reliability of the system functioning is proposed, which takes into account the conditions for detecting errors.

Formulation of the problem. In computer word processing systems, a letter string obtained by word form transformations from the existing dictionary base is accepted as an erroneous word. And the solution to the problem of correcting a system error is to establish the original word form.

It should be noted that the main indicator of error control and correction systems is computational complexity, which, in turn, is determined by the methods and algorithms for word form transformations depending on the class and nature of text information distortions, such as replacing one letter with another, rearranging a vowel through a consonant, shift hands one position when typing part of a word on the keyboard. Moreover, the fight against combined distortions is a particularly difficult task and requires the development of specific methods of transformations.

It should also be emphasized that due to the high computational complexity, currently used systems for checking the spelling of natural languages, based on the use of morphological analysis models, are not used in interactive modes of text information processing, especially in widespread electronic document management systems.

In this regard, the development of spelling control methods and algorithms that do not require deep knowledge of the language, involve small volumes of dictionaries and allow you to quickly and without additional costs correct errors based on the use of the characteristics of the morphological model and features of the Uzbek language, are in great demand.

Method and algorithms for correcting errors out of context. In systems used in batch mode, the correctness of the variant for correcting detected errors is established by a human expert by analyzing the context, and the task of correcting errors is to find all the variant corrections. However, in practice, there may be cases when the correct variant of the word form being searched for in the dictionary of word forms may not always exist, since the true letter chain may either not be in the dictionary itself or contain an error that is not part of the analyzed context.

In this regard, we propose the construction of cyclic algorithms that represent options for correcting errors without taking into account the context. The functioning of the algorithm is carried out according to the following scheme:

Step 1. At each step of information control, a chain close to the one being checked is formed and it is presented to the dictionary matching procedure;

Step 2. If the presented string belongs to the class of admissible words, then it is added to the output set of generated correction options;

Step 3. The operation of the error control and correction algorithm ends when the formation of the next new desired chain is impossible.

Note that an important point of the algorithm is the construction of procedures for representing a chain and matching them with a dictionary, which consist in matching a controlled word with a subset of chains based on various hypotheses.

To make decisions about the reliability of information, we propose a hypothesis that consists in using a word form that is close to the original one, according to which the information is considered correct based on the use of additional information. In addition, to reduce the cost of searching and processing information, the proposed algorithm is assigned softer decision rules that allow for an optional match of the word form found in the dictionary with the original one.

The error correction algorithm also implements enumeration mechanism procedures included in the morphological dictionary management subsystem.

Another feature of the developed algorithm is the adaptation of the control rules when constructing the next hypothesis by using the information obtained from the results of the previous hypothesis.

Ways to optimize control processes and error correction. We have studied two main indicators of the effectiveness of error control and correction algorithms:

1. the number of calls to the procedure for matching a word with a dictionary;
2. is the number of disk memory accesses during the algorithm's operation.

Note that the calculation of these performance indicators of the developed algorithm directly affects the time of error control and correction. Four options for calculating performance indicators were considered for analysis:

1. the number of operations in the control procedure is calculated from the start of the algorithm to the receipt of the first version of the correction;
2. the number of operations is counted until the correct variant is obtained;
3. the number of operations is counted until the last version of the correction is received;
4. the number of operations is counted until the completion of the algorithm.

Moreover, in the batch mode of text processing, the most important task is to develop methods that lead to a reduction in the time from the start of the algorithm to its completion.

And in interactive mode, the waiting time for the first correct option is more important, where the user has the opportunity to stop the algorithm immediately after receiving it, if necessary.

At the same time, it is necessary to take into account the fact that the correct variant cannot be obtained before the first variant, and therefore it is necessary to develop additional procedures to minimize the waiting time of the first variant. In addition, we have solved the problem associated with minimizing the average waiting time for the last version of error correction, which allows us to terminate the algorithm forcibly after the average time for finding the last version, provided that the error is completely corrected.

It should be noted that the above indicators for evaluating the effectiveness of a software system for detecting and correcting errors serve as the basic foundations for calculating the criteria for the complexity of monitoring the reliability of information and the cost of implementing the system. In addition, the analysis of the effectiveness of such systems can be expanded by research on the criteria for the probability of not detecting errors and the reliability of the system in the mode of batch processing of information.

In this regard, below are the results of studies on assessing the reliability of the software system for monitoring and correcting errors, which takes into account the conditions for detecting errors at the channel and physical levels of information transmission. Models for estimating the complexity of the reliability control and information processing and the patterns of error detection are also proposed. The efficiency of the software system has been tested in batch and interactive modes.

Evaluation of the reliability of the system for detecting and correcting errors in texts. Note that the initial task of evaluating the effectiveness of a software system for detecting and correcting errors according to the reliability criterion is to determine the complexity of its creation ***B***, which is characterized by the complexity of programming ***E*** of the implemented information control algorithms and depends on the volume of programs being created ***V***

$$B = C \cdot E \cdot V$$

where ***C*** is a proportionality factor depending on the accepted metric, for example, the Halsted metric.

An estimate of the potential expected volume of software system (SS) components during design can be written as:

$$V^* = (n_2^* + 2)\log_2(n_2^* + 2), \qquad (1)$$

where n_2^* - is the minimum number of different operands that form the basic elements of the software system, and the value of n_2^* is detected at the stage of algorithmization and design of the software components.

Taking into account the links between the elements of the software system, the volume of the entire SS ***V*** after programming can be legitimately defined as:

$$V = (V^*)^2 / \gamma, \quad (2)$$

where γ - is the programming language level (metric). For the machine-level language according to the Halsted metric γ=0.88, for structural programming languages γ=1.53, for the English language (higher-level language) γ=2.16.

Entering $L = V^* / V$ designation, which means the level of the program, the complexity of programming can be expressed as follows:

$$E = V^2 / V = (V^*)^3 / \gamma^2 = V / L. \quad (3)$$

We have found that for the Java language E=3000 operas/sec.

Conclusion. The effectiveness of a software system for detecting and correcting errors was also studied in terms of the complexity of control and the cost of implementing the system. To analyze the efficiency of the system in the mode of interactive information processing, a criterion for the reliability of the system functioning is proposed, which takes into account the conditions for detecting errors.

References:

1. Jumanov, I. I., Karshiev, K. B., & Tishlikov, S. (2019). A Examination of the Efficiency of Algorithms for Increasing the Reliability of Information on Criteria of Harness and the Cost of Processing Electronic Documents. International Journal of Recent Technology and Engineering (IJRTE), 28(2).
2. Jumanov, I. I. (2019). Karshiev Kh. B., Tishlikov SA Examination of the Efficiency of Algorithms for Increasing the Reliability of Information on Criteria of Harness and the Cost of Processing Electronic Documents. International Journal of Recent Technology and Engineering (IJRTE) ISSN, 2277-3878.
3. Ахатов, А. Р., & Тишликов, С. А. (2016). Методы повышения достоверности передачи и обработки информации в системах электронного документооборота на основе нечеткой семантической гиперсети. Проблемы вычислительной и прикладной математики, (3), 10-19.
4. Ахатов, А. Р., & Тишликов, С. А. (2015). Оптимизация достоверности передачи и обработки электронных документов на основе свойств моделей переходного процесса. Химическая технология. Контроль и управление», ТГТУ, (1), 61.
5. Тишликов, С. А., Жуманов, И. И., & Юлдошев, О. (2012). Контроль достоверности информации электронных документов на основе нейронечеткой семантической гиперсети. «Химическая технология. Контроль и управление», (6), 38.
6. Жуманов, И. И., & Тишликов, С. А. (2011). Интерактивное обнаружение и исправление ошибок в текстах систем электронного документооборота. «Илмий тадқиқотлар ахборот-номаси», СамДУ, Самарқанд, (1).
7. Тишликов, С. А. (2013). ОТНОШЕНИЯ АВТОМАТИЗИРОВАННОГО ДЕЛОПРОИЗВОДСТВА НА ОСНОВЕ СТРУКТУРНО-ТЕХНОЛОГИЧЕСКОЙ ИЗБЫТОЧНОСТИ. In Современные инструментальные системы, информационные технологии и инновации (pp. 247-252).

8. Ахатов, А. Р., & Тишликов, С. А. (2009). МОДЕЛИРОВАНИЕ ПРОЦЕССОВ КОНТРОЛЯ ОРФОГРАФИИ НА ОСНОВЕ ПАРСИНГОВОГО ПРЕДСТАВЛЕНИЯ N-ГРАММНОЙ ЯЗЫКОВОЙ МОДЕЛИ. Техника и технология, (6), 43-47.
9. Ахатов, А. Р., & Тишликов, С. А. (2009). Парсинговое моделирование компонентов и оценка качества системы контроля и коррекции орфографии. Актуальные проблемы современной науки, (6), 180-184.
10. Sherzod Xazratqulovich Mavlonov, & Botir Baxtiyor O'G'Li Abdullayev (2021). TA'LIM JARAYONIDA CROCODILE ICT DASTURIDAN FOYDALANISH. Science and Education, 2 (3), 323-327.
11. Mavlonov, S. ., Monasipova , R. ., & Qilichev, S. . (2023). ELEKTRON TA'LIM MUHITLARIDA AXBOROT XAVFSIZLIGI TA'LIMINING SAMARALIGINI OSHIRISH TAMOILLARI. Евразийский журнал технологий и инноваций, 1(5 Part 2), 37–40. извлечено от https://in-academy.uz/index.php/ejti/article/view/15204
12. Sh. Mavlonov (2022). UZLUKSIZ TA'LIM TIZIMIDA AXBOROT XAVFSIZLIGI VA KIBERXAVFSIZLIKNI O'QITISH ZARURIYATI. Science and innovation, 1 (B7), 1198-1201. doi: 10.5281/zenodo.7306523
13. Mavlonov, S. (2023). ELEKTRON TA'LIM MUXITIDA "AXBOROT XAVFSIZLIGI" FANINI O'QITISHDA INTERFAOL KURS DIZAYNLARI ORQALI TAKOMILLASHTIRISH. Евразийский журнал технологий и инноваций, 1(5 Part 2), 15-18.
14. Mavlonov , . S. . (2023). ONLAYN TA'LIM MUHITLARIDA AXBOROT XAVFSIZLIGI TA'LIMLARINING SAMARALIGINI BAHOLASH. Инновационные исследования в современном мире: теория и практика, 2(17), 50–52. извлечено от https://in-academy.uz/index.php/zdit/article/view/15125
15. Hazratkulovich, M. S. (2023). Effective Strategies for Teaching Information Security in Online Learning Environments. Web of Synergy: International Interdisciplinary Research Journal, 2(5), 412-418.
16. Toshmatov, S., Yarashov, I., Otakhonov, A., & Ismatillayev, A. (2022). Designing an algorithmic formalization of threat actions based on a Functioning table. In 2022 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5).
17. Abduraximov, D., Taniberdiyev, A., Monasipova, R., & Ismatillayev, A. (2023). O 'QITUVCHINING INNOVATSION FAOLIYATGA TAYYORLASH METODLARI. Евразийский журнал технологий и инноваций, 1(5 Part 2), 123-127.
18. Abduraximov, D., Taniberdiyev, A., Monasipova, R., & Ismatillayev, A. (2023). AXBOROT TEXNOLOGIYALARI FANLARIDA DARSDAN TASHQARI MASHG 'ULOTLARNI TASHKIL QILISH. Евразийский журнал технологий и инноваций, 1(5 Part 2), 96-103.
19. Ismatillayev, A. (2022). Biometric information security systems. Science and Education, 3(8), 17-22.