METHOD AND SOFTWARE TO FIND SPELLING MISTAKES IN TEXT WRITTEN IN UZBEK LANGUAGE BASED ON THE LATIN ALPHABET

[1]**Alloyorov O.M.,** [2]**Allaberganov S.M.**

[1]Urgench branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi

[2]Data Training Center

Email: [1]alloyorovoybek1997@gmail.com, [2]alsimur@mail.ru

Abstract– This article is devoted to software development for Uzbek texts, which detects spelling errors and corrects them in some places. Due to the constant use of the Uzbek language based on the Latin script and some misunderstandings in the transition from Cyrillic to Latin, it is especially important to find and correct spelling errors. The article describes the importance and advantages of the software developed for finding spelling errors in texts written in Uzbek.

Key words– correction of spelling mistakes, roots, prefixes and suffixes, spelling of letters "o'", "g'", spelling of letters "x" and "h".

I INTRODUCTION

The relevance of spelling correction software is that it helps to eliminate spelling errors and omissions that may occur in various documents, texts, and articles. This, in turn, is important to prevent situations such as violations of language norms, an increase in the misuse of lexemes in languages, and the gradual normalization of these errors.

To date, the Uzbek Spelling Dictionary has identified more than 85,000 correct spelling words and word forms [1]. One of the most common mistakes is the inverted use of the upper right side of the letter "o'" in our alphabet, which represents the sound and the letter "o'", and the same confusion is used with the letter "g'", which represents "g'". There are also errors in the use of the letters "x" and "h". For this reason, it is important to check and correct the Uzbek text for spelling errors.

Analysis of the most common spelling mistakes in the Uzbek language.

Many spelling mistakes are made in the process of writing in Uzbek. If spelling mistakes are not corrected regularly, these words will be misspelled and the words will change altogether. There are different ways to detect errors [3]. They can be used in Uzbek spelling dictionary, annotated dictionary or in programs that help to find and edit mistakes [2].

One of the most common mistakes in the Uzbek language is given in the following table:

	wrong	wrong	wrong	right
The symbol used in the letters "o'" and "g'"	o', g'	o', g'	o', g'	o', g'
The modifier letter apostrophe	'	,	'	,

Another common mistake is to misuse the letters "x" and "h". Because these two letters represent different sounds, it is important to use each one correctly. For example, the words "xush" and "hush" are semantically different. For example: "Xush kelibsiz", "Hushidan ketmoq" [1].

It follows that it is important to know the norms of literary language when writing and using texts in Uzbek. Knowledge of literary language norms is also required in the process of software development. This, in turn, will greatly help to understand how mistakes are made and how to fix them.

Roots and suffixes

In Uzbek, as in other languages, words consist of root and suffixes. That is, a word may contain a root and one or more suffixes. The root word is the part of speech that can come into its own and is independent and meaningful. Adding prefixes and suffixes to word stems creates completely new words. There are many such suffixes in Uzbek language. For example, conjunctions: "-ning", "-ni", "-ga", "-dan", "-da" help to connect words by joining the root, and they are also

called word-changing suffixes.

The genitive case can be answered by questions "kimning?", "nimaning?", "qayerning?". The suffix of the genitive case is "-ning". "-ning" suffix connects a noun to another noun. The nouns of the genitive case are in the main role. Examples: "oʻquvchining kitobi", "koʻylakning yoqasi", "daraxtning kurtagi", "buloqning koʻzi".

Accusative case can be answered by questions "kimni?", "nimani?", "qayerni?". The suffix of accusative case is "-ni". Accusative case suffix connects a noun to verb in the sentence. The nouns of accusative case are main parts of sentence. Examples: "She'rni oʻqidi", "Multfilmni tomosha qildi".

Dative case can be answered by questions "kimga?", "nimaga?", "qayerga?". The suffix of dative case is "-ga", ("-ka", "-qa"). Accusative case suffix connects a noun to verb in the sentence. The nouns of accusative case are main parts of sentence. The "-ka" suffix is added to the nouns which end with the sound "k", the "-qa" suffix is added to the nouns with the last sound "-q" and in other cases "-ga" suffix is added to the nouns.

Ablative case can be answered by questions "kimda?", "nimada?", "qayerda?". The suffix of ablative case is "-da". Ablative case suffix connects a noun to verb in the sentence. The nouns of ablative case are main parts of sentence. Examples: "Saroyda ishladi", "Togʻda yashaydi".

In addition, there are word-forming, noun-forming, adjective-forming and verb-forming suffixes.

For instance:

- word-changing suffixes are "-ning", "-ni", "-ga", "-da", "-dan";

- word-forming suffixes are "-chi", "-la", "-li", "-kor", "-dosh";

- noun-forming suffixes are "-chi", "-zor", "-dosh", "-kor", "-k", "-q";

- adjective-forming suffixes are "-ser", "-be", "-siz", "-li", "-chan", "-dor", "-q";

- verb-forming suffixes are "-la", "-lan", "-sira", "-illa", "-(ulla)", "-lash".

Noun-forming suffixes are added to words and made nouns. Examples: "chi", "hasharchi"; "-zor", "gʻallazor"; "-dosh", "sinfdosh"; "-kor", "paxtakor"; "-k", "elak"; "-q", "taroq".

Verb-forming suffixes are added to words and made words which is in the category of verbs. Examples: "-la", "bogʻla"; "-lan", "tayyorlan"; "-sira", "suvsira"; "-illa", "-(-ulla)", "vizilla", "shivilla"; "-lash", "tiklash".

The present continuous tense verb ("hozirgi zamon fe'li") refers to an action that is performed (or not performed) while speaking. "-yap", "-moqda" are present continuous tense ("hozirgi zamon") suffixes. For example: "Oʻrik bechora oppoq, nozik gullarini qayoqqa yashirishni bilmayapti".

The future tense verb ("kelasi zamon fe'li") refers to an action that can be performed (or will not be performed) after the speech. "-moqchi" is a future tense ("hozirgi zamon") suffix. Example: "Sevara oʻzi yozgan she'rini oʻqib bermoqchi".

The past tense verb ("oʻtgan zamon fe'li") refers to an action performed (or not performed) before the present tense. "-di", "-gan" are past tense ("hozirgi zamon") suffixes. For example: "Dilnozaning uyiga Sarvi xola chiqdi".

The past tense suffix "-kan" for verbs ending in "k"; for verbs ending in "q", "-qan"; added to other verbs in the form "-gan". Examples: "qirq" – "qirqqan", "qoʻrq" – "qoʻrqqan", "choʻk" – "choʻkkan", "chop" – "chopgan".

Software development process

In the process of developing the program, an attempt was made to find a solution by studying the norms of the Uzbek language. From the above information, it follows that words in the Uzbek language are formed from suffixes added to the roots, and by adding many suffixes to one root, completely different words are formed. So, it can be a bit tricky to check every word in the text to see if it's incorrect. For example, the word "kitoblarimizning" consists of parts such as "kitob", "-lar", "-imiz", and "-ning".

The solution is to examine the roots and suffixes separately. This makes the software less cluttered and easier to find errors.

How the program works

With the help of the Uzbek dictionary based on the Latin script, a separate database for stems and suffixes was created. Suffixes are also divided into prefixes and suffixes. The program is installed in Microsoft Word in macro mode, the desired text is selected, and macro is run to detect spelling errors in the text. When the macro starts, it first checks to see if the selected word has prefixes. All prefixes are compared, and if there is a prefix, it is removed and re-checked. The process continues until there are no prefixes left. The next step is to check for the presence of suffixes by comparing them through a database of suffixes. This process continues until there are no more suffixes in the word. The word without suffixes is checked from the base database. If the word does not exist in the database, it is considered an error. Even if there is an error in the suffixes, it will not be found when comparing through the databases of suffixes and the database of roots, and will be automatically recognized as an error. Words found to be incorrect are highlighted in red. All words in the text are checked in the same way.
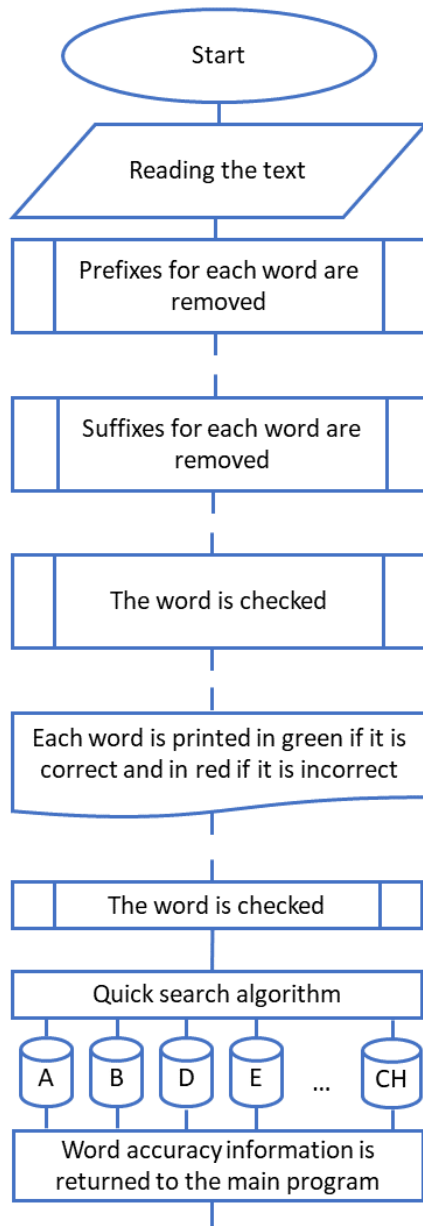
the modifier letter apostrophe " ' ".

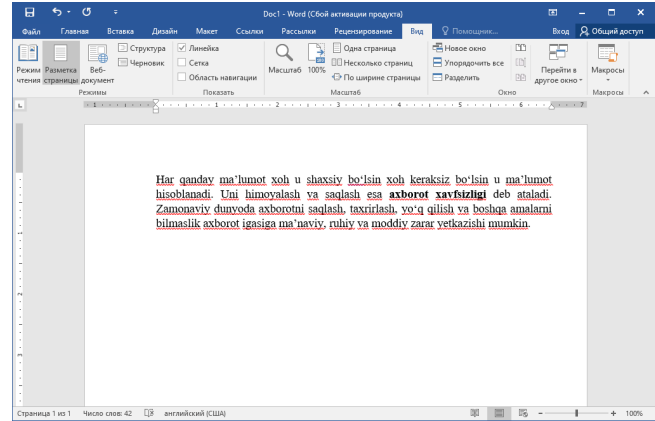The following is a view of the text before the scan and after the scan is completed:

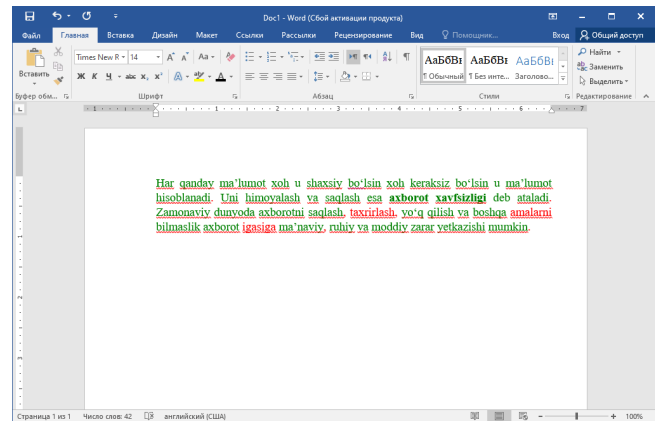**Fig. 2:** View of the text before the program starts

Fig. 3: The appearance of the text after the start of the program

There are three misspelled words in the given text, and as shown in Figure b, the program marked the correct words in green and the incorrect words in red.

II CONCLUSION

This software, which detects errors in Uzbek texts based on Latin script, can detect word errors relatively quickly using the above methods. This is due to the fact that the bases and suffixes in the words are placed in separate databases, the front and back suffixes of the word, and the base is checked separately. Dividing the base database into sections by letter, determining which letter the word starts with, and checking the word base that only starts with that letter can significantly increase performance. This is because there is no need to check for incorrect databases in this case, which saves time.

Fig. 1: Block scheme of program

The program also includes the function of correcting the special character of the letters "o'" and "g'". The function searches for the symbols "o" and "g", if one of the symbols " ' ", " ' ", " ' " is used after them, it replaces the characters with the symbol " ' ". If the special character " ' " of the letters "o'" and "g'" is dropped, the program will recognize the word as an error. In the same way, if one of the " ' ", " ' " or " ' " characters is used instead of the modifier letter apostrophe " ' " in a word, the program will replace it with

This software detects text errors, highlights correct words in green and incorrect words in red.

III   REFERENCES

[1] "Explanatory dictionary of the Uzbek language". - Tashkent: National Encyclopedia of Uzbekistan, 2006. 1-2-3 years.

[2] A software tool designed to popularize Uzbek spelling rules and improve literacy. Makes it possible to detect spelling errors in the text and improve the quality of your content. https://savodxon.uz/

[3] Google Docs Editor Google's browser-based word processor for creating, editing, and sharing documents online. Can be accessed from any computer with an Internet connection. https://support.google.com/docs#topic=1382883

[4] Grammarly is an artificial intelligence-based online platform to help you communicate in English. To effectively create text in English, it provides recommendations on the correctness, clarity, engagingness and tone of the message. https://www.grammarly.com/grammar-check

[5] Prüfer, D & Kawchuk, Lawrence & Rohde, W. (2006). Polerovirus ORF0 genes induce a host-specific response resembling viral infection. Canadian Journal of Plant Pathology - Revue Canadienne de Phytopathologie, v.28, 302-309 (2006). 28. 10.1080/07060660609507299.

[6] Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language. In Proceedings of the 11th Global Wordnet Conference, pages 8–19, University of South Africa (UNISA). Global Wordnet Association.

[7] "Explanatory dictionary of the Uzbek language". - Moscow: Russian, 1981.

[8] Mirxanova Gulandom Rustamovna. (2020). Presenting synonyms in the explanatory dictionaries of the Uzbek language. Middle European Scientific Bulletin, 6, 117-120.