





FTAMP 20.53.19

А.Н. Шормакова¹ – негізгі автор, | ©
У.А. Тукеев²


 ¹Магистр, ²Техн. ғылым. д-ры, профессор
ORCID ¹<https://orcid.org/0000-0002-1637-4643>; ²<https://orcid.org/0000-0001-9878-981X>
 ¹Әл-Фараби атындағы Қазақ Ұлттық университеті,
 Алматы қ, Қазақстан Республикасы
 ¹shormakovaassem@gmail.com

<https://doi.org/10.55956/AEQO3045>

ҚАЗАҚ ТІЛІНДЕГІ АВТОМАТТАНДЫРЫЛҒАН СИНОНИМДЕР ТІЗІМІН ҚҰРУ

Аңдатпа. Мақалада ағылшын-қазақ сөздерінің синоним сөздер мәселелері қарастырылған. Зерттеу нәтижесінде онлайн ағылшын синоним сөздік қорынан қазақша синонимдер құрылды. Бұл жұмыстың ерекшелігі ағылшын онлайн синоним сөздігін қолдана отырып қазақ тіліндегі синоним сөздер тізімін автоматты түрде құру болып табылады. Автоматтандырылған синонимдер тізімін құру үшін ағылшыннан қазақшаға аударылған сөздер қолданылады. Зерттеу нәтижесінде мыңнан аса синоним сөздер тізімі құрылды. Ағылшын тілдегі сөздер онлайн тезаурустан ізделініп, табылған синонимдері машиналық аударма көмегімен аударылып, қазақ тіліндегі синонимдер каталогына енгізіледі. Орындалатын үрдіс толық автоматты түрде жүзеге асырылады.

Тірек сөздер: тезаурус, синонимдер тізімі, қазақ тілі, ағылшын тілі, автоматтандырылған.

 Шормакова, А.Н. Қазақ тіліндегі автоматтандырылған синонимдер тізімін құру [Мәтін] / А.Н. Шормакова, У.А. Тукеев // Механика және технологиялар / Ғылыми журнал. – 2022. – №3(77). – Б.44-49. <https://doi.org/10.55956/AEQO3045>

Кіріспе. Қазіргі кезде заман талабына сай ақпараттық технологиялар қарқынды дамып жатқан уақыт, сондықтан да үлкен мәліметтер қорын қамтитын құралдардың маңызы артуда. Дәлірек айтсақ, бүгінде қазақ тіліндегі синонимдер сөздігі үлкен сұранысқа ие және қолданылып та жүр [1,2].

Онлайн түрде қолданылып жүрген құралдар келесідей: Луғат, Емле, Терминком, Сөздікқор. Луғат бұл – қазақ тіліне арналған сөздік. Бұл құрылғыны қолдану үшін кез келген пайдаланушы желі арқылы тіркеліп, жұмыс істей алады және оны өзгертуге де мүмкіндігі бар. Жалпы сөздік қоры 534 000 сөз. Сөздікте көптеген тіл жұптары қарастырылған, оның ішінде: орысша-қазақша, қазақша-орысша, түсіндірме сөздік, қазақша-түрікше, түрікше-қазақша, ағылшынша-қазақша, қазақша-ағылшынша, аймақтық сөздік, орфографиялық, синонимдік, арабша-қазақша, қазақша-арабша сөздіктер. Ал, Емле – орфографиялық сөздік. Оның қызметі сөздердің жазылуын тексереді және онда әртүрлі мақалалар мен кітаптар бар. Терминcom.kz әртүрлі салада: ғылым, білім, технология және экономика, қоғамдық және әлеуметтік сөздердің республикалық терминологиялық электрондық базасы. Сөздікқор – ақпараттық технология саласында қазақ тіл сөз ерекшелігі ескерілетін, сөздер мағынасын көруге мүмкіндік беретін

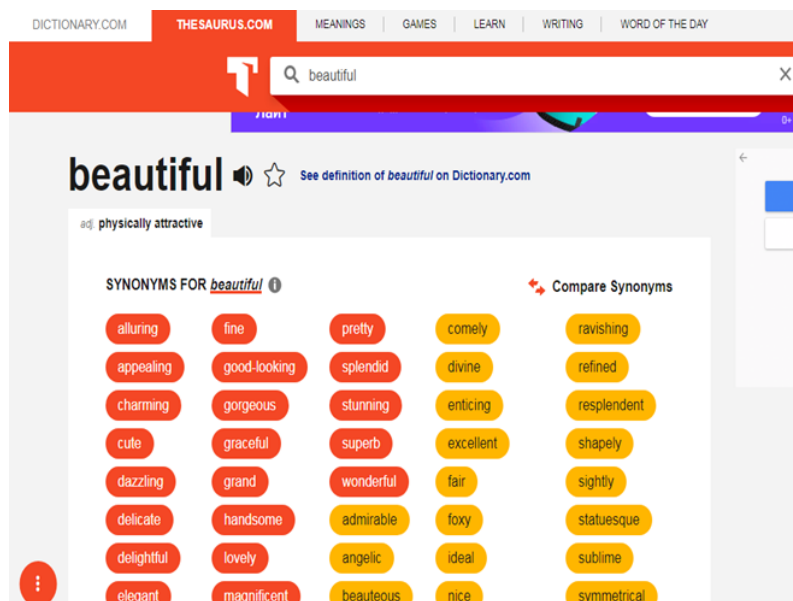
платформа. Тілдік сөздіктер Sozdik.kz болса бұл ең танымал қазақша-орысша (орысша-қазақша) сөздік. Жүйенің ішінде мәтіндерді аударуға болады және қазақша мәтінді кириллицадан латынға көшіруге арналған конвертермен де қамтылған. Бір ерекшелігі: пайдаланушы сөздікке жаңа сөз қоса алады [3]. Тек ескеретін жайт басылымдағы материалдарды онлайн түрінде қолдану өте ыңғайсыз, ал онлайн ресурстардың мәліметтер қоры толық қол жетімсіз болғандықтан жеке программалық жабдыққа қолдану ыңғайсыздық тудырады. Сондықтан ұсынылып отырған жұмыста онлайн ағылшын сөздігімен жұмыс істей отырып, синонимдер сөздік қорын ұлғайтамыз.

Зерттеу әдістері. Синоним сөздер тізімін құру мақсатында PyDictionary кітапханасы қолданылды. PyDictionary - мағыналар, аудармалар, сөздердің синонимдерін және антонимдерін алуға арналған Python 2/3 сөздік модулі. Ол synonym.com сайтынан синонимдер және антонимдер алу үшін, Google-ді аудармалар үшін және сөз мағыналарын алу үшін WordNet пайдаланады. WordNet® – ағылшын тіліндегі үлкен лексикалық дерекқор. Бұдан бөлек, тізім құру кезінде онлайн сөздік те қолданылды. Ол thesaurus.com сайтынан алынған dictionary.com ұсынған әлемдегі ең үлкен және ең сенімді тегін онлайн тезаурус.

Dictionary.com анықтамалығы сөздердің шығу тегі және тағы басқа құралдар үшін әлемдегі жетекші онлайн дереккөзі болып табылады. Dictionary.com миллиондаған адамдар үшін ағылшын тілінің құпияларын ашады. Dictionary.com бұл сөздермен жұмыс істейтін әлемдегі байланыстарды, қарым-қатынасты, оқуды, шығармашылықты және көптеген бағыттағы нәрселерді шабыттандыруға тырысатын құрал. Dictionary.com – әлемдегі жетекші цифрлық сөздік. Бұл сөздік миллиондаған ағылшынша анықтамаларды, емлелерді, дыбыстық айтылымдарды, үлгі сөйлемдерді және сөздің шығу тегін ұсынады. Бұл сайттың негізгі меншікті көзі – Random House Unabridged Dictionary, оны тәжірибелі лексикографтар тобы үнемі жаңартып отырады және ол әртүрлі тіл қажеттіліктерін қолдау үшін сенімді, танылған дереккөздермен, соның ішінде American Heritage және Harper Collins арқылы толықтырылады. Сонымен қатар, аударма қызметін, кроссворд шешімін және озық сөз әуесқойлары мен ағылшын тілін үйренушілерге пайдалы болатын көптеген редакциялық мазмұнды ұсынады. 20 жылдан астам уақыт бойы thesaurus.com миллиондаған адамдарға ағылшын тілін меңгеруді жақсартуға және 3,0 млн. астам синонимдер мен антонимдері бар нақты сөзді табуға көмектесіп келеді.

Ағылшын тіліндегі сөздердің синонимдерін thesaurus.com ресурсынан аламыз. Мысалы, beautiful сөзінің синонимдері келесі түрде шығады (1-сурет).

Синоним сөздерді thesaurus.com [4] сайтынан қолдану үшін Beautiful Soup кітапханасы пайдаланылды. Beautiful Soup – бұл HTML мен XML файлдарынан мәліметтерді алу үшін қолданылатын Python кітапханасы немесе Python бағдарламалау тілінде жазылған HTML/XML файлдарын синтаксистік талдауға арналған талдаушы (парсер) [5].



1-сурет. Thesaurus.com онлайн сөздігіндегі «beautiful» сөзінің синонимдер тізімі

Табылған барлық ағылшынша синоним сөздерді Google аудармашысы арқылы қазақшаға аударып, файлға сақтаймыз. Сонда аударылған қазақша синоним сөздер тізімі құрылады. Негізгі тізімнің құрылу сызбасын келесі суреттен көруге болады (2-сурет).



2-сурет. Синонимдер тізімін құру блок-схемасы

Зерттеу нәтижелері. Зерттеу нәтижесінде сөздік қоры 3,5 млн. ағылшын синонимдер сөздігі қолданылып, 1000 синоним сөздерден тұратын 6000 қазақ синонимдер тізімі құрылды. Құрылған қазақ тіліндегі синонимдер бір файлда сақталады. Нәтиженің бір бөлігін келесі тізімде көруге болады:

1. *Аз, жеткіліксіз, шамалы, шектеулі;*
2. *Карикатура, мақтан, түс, назар аудар, ұлғайт;*
3. *Сыпайы, жақсы, әділет, мейірімді, сүйкімді, қолайлы, тамаша, ерекие;*
4. *Бұрын, қара де, қанишалықты, кел;*

5. Денсаулық, пайда, әл-ауқат, қызық, прогресс;
6. Молшылық, өркенде, байлық, қаз;
7. Эволюция, аванс, гүлден, ынталандыр, дамыт;
8. Ис, пайда, бол, жылы, жағдай;
9. Мәртебе, әрекет;
10. Сақта, күшейт, көмек, қорға, қолда;
11. Іргетас, аяқ, тірек, негізі;
12. Тұрақ, мәңгі, үнемі, дәйекті түрде, әрқашан;
13. Клиент, дене, өмір, әйел, біреу, ада;
14. Кедей, қыс, жалған, кішкентай;
15. Жан, бойынша, бар;
16. Қол жетімді, қолайлы, бос, мүмкін;
17. Сызба, қорытынды, құрылым, түйіндеме, жоба;
18. Мутация, түрлендір, аударым, революция, жаңар;
19. Фирма, арнайы, тамыр, орнат;
20. Секір, көтер, арттыр, жоғары;
21. Жол, жолақ, рәсім;
22. Симпатикалық, пайдалы, достық;
23. Үлес, дәйексөз, қват;
24. Тілек, мақсат, бағыт, ұмтыл;
25. Ниет, амбиция, міндет;
26. Күш сал, ойна, пайдалан, ләззат, тәжірибе, жүзеге асыр;
27. Төсек, өріс, балабақша, терассалар, патио, бақша;
28. Алыс, сырт, жоқ, тауыс, өлі, шық;
29. Кадр, физикалық, конституция, сурет, құрастыр;
30. Жина, қап, кластер, жиналыс, кеңес;
31. Тұтқа, бер, қамтамасыз ет, тарат, қызмет ет;
32. Салыстырмалы, байланыс, танымал, сияқты;
33. Кеңсе, бизнес, жеткіз, қызмет көрсет;
34. Жоғарыда айтылған, ұқсас, осы;
35. Істе, салтанат, дайында, іске қос, бас;
36. Қатынас, нүкте, материал, тақырып;
37. Үлкен, зор, түршігерлік, тамаша;
38. Стерлинг үлгілі;
39. Қалыптастыр, өндіріс, мекеме, жобала, құр;
40. ...т.б.

Құрылған синонимдер сөздігі автоматты түрде жүзеге асырылды.

Ғылыми нәтижелерді талқылау. Мәліметтер қорын әлі де ұлғайту алдағы жоспарда бар. Ағылшын онлайн сөздігінде тек синонимдер ғана емес, омоним, антоним сөздерді қолдануға мүмкіндік бар. Синоним сөздер қоры көп болған сайын сөздік қоры да кеңейе түседі. Сөздік қор машиналық аударма мәтіндері үшін де маңызды [6, 7]. Тізім автоматты түрде және машиналық аударманың көмегімен құралғандықтан, сөздердің аударылу сапасын тексеру мен жақсартуды ескеру қажет.

Қорытынды. Жұмыстың негізі мақсаты ағылшын сөздерінен қазақ тіліндегі синонимдер сөздігін құру болып табылды. Жұмыс соңында қажет нәтиже алынды. Сөздік қоры үш жарым миллионнан астам сөздері бар ағылшын сөздігінен қазақ тілінде мың синоним сөздерден құралған қазақша синонимдер тізімі құрылды.

Әдебиеттер тізімі

1. Орысша-қазақша түсіндірме сөздік: Ғылымтану [Мәтін] / Жалпы редакциясын басқарған э.ғ.д., профессор Е. Арын. - Павлодар: ҒӨФ «ЭКО», 2006.
2. Бизақов, С. Синонимдер сөздігі [Мәтін]. - [?]: Арыс баспасы, 2007. – 640 б.
3. Ақпарат дереккөзі [Электрондық ресурс]. – Қол жеткізу режимі: <https://massaget.kz/layfstayl/zhadnama/59225/> Сұраныс күні: 09.09.22.
4. Питон кітапханасы [Электрондық ресурс]. – Қол жеткізу режимі: <https://pypi.org/project/PyDictionary/> Сұраныс күні: 09.07.22.
5. Синоним сөздігі [Электрондық ресурс]. – Қол жеткізу режимі: <https://www.synonym.com/> Сұраныс күні: 09.07.22.
6. Сөздер қоры [Электрондық ресурс]. - Қол жеткізу режимі: <https://wordnet.princeton.edu/> Сұраныс күні: 09.07.22.
7. Онлайн сөздік [Электрондық ресурс]. - Қол жеткізу режимі: <https://www.dictionary.com/> Сұраныс күні: 09.07.22.
8. Онлайн сөздік [Электрондық ресурс]. - Қол жеткізу режимі: <https://www.thesaurus.com/> Сұраныс күні: 09.07.22.
9. Питон модулі [Электрондық ресурс]. - Қол жеткізу режимі: <https://www.crummy.com/software/BeautifulSoup/> Сұраныс күні: ?
10. Абеустанова, А.Н. Қазақ тіліндегі көпмағыналы сөздердің бірін анықтаудың бір болжамы [Мәтін] / А.Н. Абеустанова // ҚазҰТУ хабаршысы. – 2015. - №4(110). - Б. 625-628.
11. Shormakova A., Zhumanov Z.H., Rakhimova D. Post-editing of words in Kazakh sentences for information retrieval // Journal of Theoretical and Applied Information Technology, 2019. No 97(6), PP. 1896–1908.

Материал редакцияға 14.09.22 түсті.

А.Н. Шормакова, У.А. Тукеев

Казахский Национальный Университет имени аль-Фараби, г. Алматы, Казахстан

ФОРМИРОВАНИЕ АВТОМАТИЗИРОВАННОГО СПИСКА СИНОНИМОВ В КАЗАХСКОМ ЯЗЫКЕ

Аннотация. В статье рассматриваются синонимы англо-казахских слов. В результате исследования были созданы казахские синонимы из онлайн-словаря английских синонимов. Особенностью данной работы является автоматическое создание списка казахских синонимов с использованием английского онлайн-словаря синонимов. Слова, переведенные с английского на казахский, используются для создания автоматизированного списка синонимов. В результате исследования был создан список из более чем тысячи синонимов. Поиск английских слов осуществляется в онлайн-тезаурусе, найденные синонимы переводятся с помощью машинного перевода и заносятся в каталог казахских синонимов. Процесс полностью автоматизирован.

Ключевые слова: тезаурус, список синонимов, казахский язык, английский язык, автоматизированный.

A.N.Shormakova, U.A.Tukeyev

Al-Farabi Kazakh National University, Almaty, Kazakhstan

**FORMATION OF AN AUTOMATED LIST OF SYNONYMS
IN THE KAZAKH LANGUAGE**

Abstract. The article deals with synonyms of English-Kazakh words. As a result of the study, Kazakh synonyms were created from the online dictionary of English synonyms. A feature of this work is the automatic creation of a list of Kazakh synonyms using an English online dictionary of synonyms. Words translated from English to Kazakh are used to create an automated list of synonyms. As a result of the research, a list of more than a thousand synonyms was created. The search for English words is carried out in the online thesaurus, the found synonyms are translated using machine translation and entered into the catalog of Kazakh synonyms. The process is fully automated.

Keywords: thesaurus, list of synonyms, Kazakh, English, automated.

References

1. Oryssha-kazaksha tysindirme sozdik: Gylymtanu [Russian-Kazakh explanatory dictionary: Science]. The editor-in-chief, Professor E. Aryn. - Pavlodar: GEF "ECO", 2006. [in Kazakh].
2. Bizakov S. Sinonimder sozdigi [Dictionary of synonyms]. - [?]: Arys baspasy, 2007. – 640 p. [in Kazakh].
3. Akparat derekkozi [Source of information]. [Electronic resource]. – Access mode: <https://massaget.kz/layfstayl/zhadnama/59225/>. Date of access: 09.07.22.
4. Piton kitaphanasy [Python library] [Electronic resource]. - Access mode: <https://pypi.org/project/PyDictionary/>. Date of access: 09.07.22.
5. Sinonim sozdigi [Synonym dictionary]. [Electronic resource]. - Access mode: <https://www.synonym.com/>. Date of access: 09.07.22.
6. Sozder qory [Vocabulary]. [Electronic resource]. - Access mode: <https://wordnet.princeton.edu/>. Date of access: 09.07.22.
7. Onlayn sozdik [Online dictionary]. [Electronic resource]. - Access mode: <https://www.dictionary.com/>. Date of access: 09.07.22.
8. Onlain sozdik [Online dictionary]. [Electronic resource]. - Access mode: <https://www.thesaurus.com/>. Date of access: 09.07.22.
9. Piton moduli [Python module]. [Electronic resource]. - Access mode: <https://www.crummy.com/software/BeautifulSoup/>. Date of access: [?].
10. Abeustanova A.N. Kazak tilindegi kopmagynaly sozderdin birin anyktaudyn bir bolzhamy [A prediction for determining one of the polysemous words in the Kazakh language] // Bulletin of KazNTU. - 2015. - No. 4(110). - PP. 625-628. [in Kazakh].
11. Shormakova A., Zhumanov Z.H., Rakhimova D. Post-editing of words in Kazakh sentences for information retrieval // Journal of Theoretical and Applied Information Technology. - 2019. – No. 97(6). - PP. 1896-1908.