



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ



Студенттер мен жас ғалымдардың
«ҒЫЛЫМ ЖӘНЕ БІЛІМ - 2014» атты
IX халықаралық ғылыми конференциясы

IX Международная научная конференция
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ - 2014»

The IX International Scientific Conference for
students and young scholars
«SCIENCE AND EDUCATION-2014»

2014 жыл 11 сәуір
11 апреля 2014 года
April 11, 2014



**ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ**

**Студенттер мен жас ғалымдардың
«Ғылым және білім - 2014»
атты IX Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
IX Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2014»**

**PROCEEDINGS
of the IX International Scientific Conference
for students and young scholars
«Science and education - 2014»**

2014 жыл 11 сәуір

Астана

УДК 001(063)
ББК 72
Ғ 96

Ғ 96

«Ғылым және білім – 2014» атты студенттер мен жас ғалымдардың IX Халықаралық ғылыми конференциясы = IX Международная научная конференция студентов и молодых ученых «Наука и образование - 2014» = The IX International Scientific Conference for students and young scholars «Science and education - 2014». – Астана: <http://www.eni.kz/ru/nauka/nauka-i-obrazovanie/>, 2014. – 5831 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-610-4

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 001(063)
ББК 72

ISBN 978-9965-31-610-4

©Л.Н. Гумилев атындағы Еуразия ұлттық университеті, 2014

ҚАЗАҚ ТІЛІ МОРФОЛОГИЯСЫН ЖӘНЕ АҚПАРАТТЫҚ-ІЗДЕУ ТЕЗАУРУС ҚҰРЫЛЫМЫН АВТОМАТТАНДЫРУ МОДЕЛІ

Самбетбаева Мадина Аралбаевна

Sambetbaeva_ma@enu.com

Л.Н.Гумилев атындағы ЕҰУ «Ақпараттық жүйелер» кафедрасының докторанты,
Астана қаласы, Қазақстан Республикасы
Ғылыми жетекшісі – Д.А.Тусупов

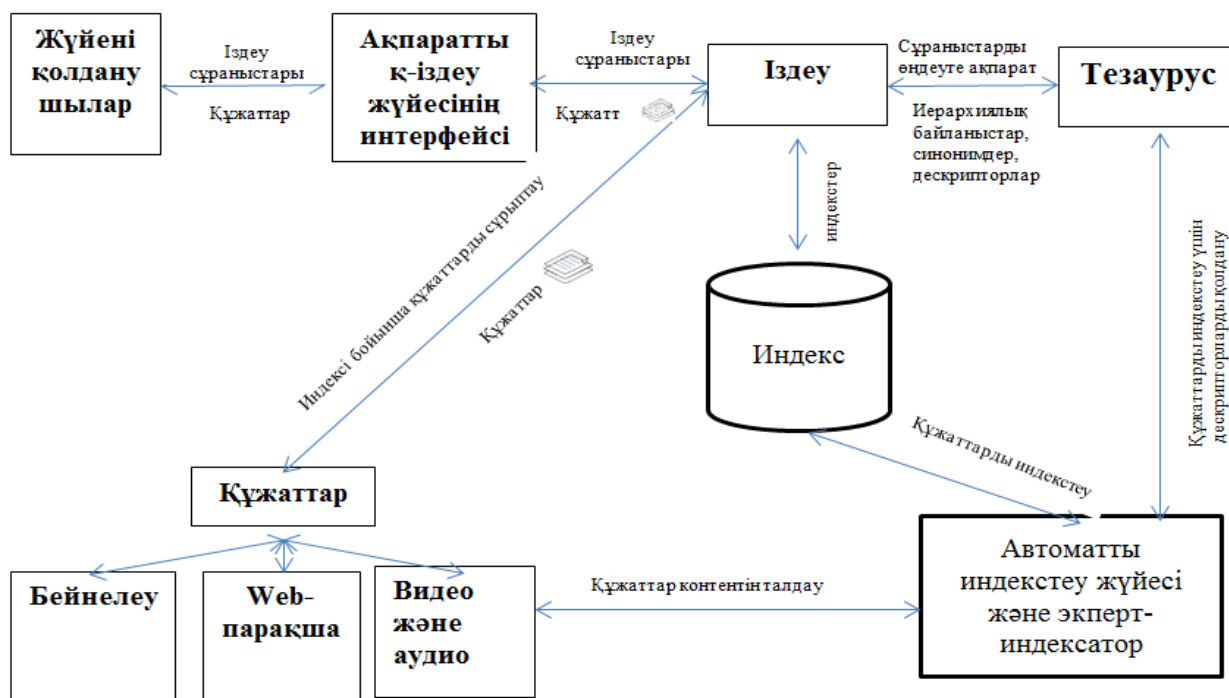
1. Кіріспе

Ақпаратты-іздігіру тезаурусы – үлкен көлемдегі мәліметтерді ұйымдастыру, индекстеу және басқаруға арналған қуатты құрал. Тезаурусты қолданудың негізгі мақсаты - құжаттарды индекстеу үшін қажетті сөздік қордың стандартты жиынтығын ұсыну және тезаурус концепттерінің өзара семантикалық байланыстары сатыларын іздеу сауалын қайта құру және басқа да тапсырмалар [1, 2].

Тезаурусты-бағытталған амалдардың күрделілігі мен семантикалық қордың қолдан жасау құнының жоғарылығы оның ақпараттық жүйеде кеңінен қолданылуына кедергі болады. Жұмыста осындай сөздікті құруды автоматтандыруға көмектесетін тәсілдер мен алгоритмдері, тезаурусты қолдан құру процесінің негізгі кезеңдері мен технологиясы, одан соң тезаурус құрылымының ақпараттық адами-машиналық технологиясын жасау арқылы осы процесті автоматтандыру моделі қарастырылған. Технологияны дайындау жұмыстың негізгі бөлімі және пән саласы мәтіндер кешені мен эксперттің қол жұмысының машиналық тәсілдері қиысуына негізделетін сараптамасы болып табылады.

2. Ақпараттық-іздеу тезаурусы

Келесі суретте тезаурустың ақпараттық-іздеу жүйесінде пайдаланылуына мысал көрсетілген:



1-сурет. Тезаурустың ақпараттық-іздеу жүйесінде қолданылу моделі

Ақпараттық-іздеу тезаурусы белгілі пән саласына арналған нақты ережеге сай құрастырылған терминдер мен сөз тіркестерінің сөздігі болып табылады, ол осы пән саласы

бойынша ақпараттық іздеу сапасын жақсартуға жасалады. Ақпараттық-іздеу тезаурусын негізінен екі бағытта қолданады, алдымен семантикалық ресурстағы концепттерді пайдаланып құжаттарды индекстеу үшін және пайдаланушының іздеу сауалын өңдеу кезінде сатылық, ассоциативтік пен синонимдік байланыстарын пайдалануда. Одан басқа, дескрипторлар арасындағы семантикалық байланыс құжаттарды топтастыру мен айдарлау, іздестірілетін сауалға байланысты тізімді жасауға және басқа да ақпараттық іздеу тапсырмалары үшін қолданылады.

Тезаурус бір-бірімен синонимдік немесе квазисинонимдік қатынаспен байланысқан бір немесе бірнеше дескрипторлары бар концепттерден тұрады. Дескрипторлар өз алдына пән саласының негізгі ұғымдарын бейнелейтін сөздер мен сөз тіркестерінен тұрады. Сонымен бірге, концепттер ассоциативтік, сатылық және басқа да қатынас түрлерімен байланысуы мүмкін. Концепт ретінде дескриптор жиынтығы бір ұғымға қатысты: «Жүйе», «Ақпараттық жүйе», «Жүйелік сараптама», «Жүйелендіру», «Білім жүйесі» мысал бола алады.

Ақпараттық-іздеу тезаурусы – сөздік, оны сөздік пен семантикалық ресурстар құрылымы саласындағы маман, сарапшы-тіл маманы қолмен жасайды. Мұндай сөздікті құруда бір немесе бірнеше пән саласының тезаурустық сипаттамасын алу міндеті тұрады, сондай-ақ, сөздік жасауға негіз болатын мәтін кешені жиі кездеседі. Сарапшы мәтін кешенін сараптайды және стандартта сипатталған [1] тезаурус құрылымын қолмен жасау технологиясын ұстана отырып, берілген пән саласын сипаттайтын термин тізімдерін жасайды және оларды дескриптор ретінде тезаурусқа қосады. Осыдан кейін терминдер концепттерге топтастырылады және олардың арасында сатылық және ассоциативтік қатынас бекітіледі.

Тезаурусты қолдан құру процесіне ресурс жасаудың жоғары құндылығы мен ұзақтығы, нәтиженің сарапшы біліктілігіне байланыстылығы, мәтін кешенінің барлығын сараптау мүмкіндігінің жоқтығы мен басқа да белгілер тән. Тезаурусты жасау процесін автоматтандыруды тезаурусты автоматты құру технологиясын пайдалана отырып жасау ұсынылды, ол Тезаурус Құрылымын Автоматтандыру жүйесінде жүзеге асырылады. Зерттеу нәтижесінде ақпараттық-іздеу тезаурустарымен жұмыс істеудің анықталған бағдарламалық шешімдері айқындалды: Oracle Text / Thesaurus Management System, Amicus Thesaurus, Livelink Collections Server Thesaurus Manager, Thesaurus Master және тағы басқалары.

Алайда, табылған өнімнің бірде-бірі тезаурус құрылымын автоматтандыру процесін іске асыра алмайды, бағдарламалар тек алдын-ала қолмен жасалған терминологияларды сақтау мен басқару мүмкіндіктерін жетілдіре алады.

3. Тезаурус құрылымын автоматтандыру технологиясы

Тезаурусты автоматты жасау талпынысы бұрын да болған, осылайша Филлипович [4] Вилкоксон, Грегори [5] статистикасы негізінде тезаурус құрылымын автоматтандыру үшін бірнеше тәсілдерді, нақтылап айтқанда осы мақаланың 4-бөлімінде қолданылған тәсілдерді сипаттайды, мәтіндердің тезаурус құрылымы техникасын ұсынған. Сондай-ақ, тезаурус құрылымының жеке кезеңдерін автоматтандыруға арналған жұмыстарды атап өтуге болады, мәселен, мәтіндер кешенінен терминологияны шығару. Осы мақалада тезаурус құрылымын автоматтандырудың жаңа технологиясы ұсынылады, онда тек алгоритм бөлігі ғана емес, сонымен қатар сөздік жасауда адам мен жүйе арасындағы жұмыстың бөліну ерекшелігі міндеттері қамтылған.

Ұсынылып отырған технология тезаурусты дедуктивті құру амалдарына негізделген, ол тезаурустың барлық дескрипторлары мәтін кешенінен алынатынын білдіреді. Қолмен жасауда сарапшы жиі кешенде кездеспейтін терминдерді қосады, алайда ондай сөздер көлемі 20-25% аспайды. Ұсынылған технология ерекшелігі сол, ол қажет болған жағдайда сарапшының терминді қолмен жазып қосып отыруына мүмкіндік береді. Қолдан жасаудан негізгі айырмашылығы сол, мітін кешенін алдын-ала машиналық өңдеу жүзеге асырылады және сарапшыға тезаурусты құрастырудың әр кезеңінде көптеген сөздер, сөз тіркестері мен олардың арасындағы байланыстар – тезаурусқа енгізуге лайықтылары ұсынылады. Технологияға сай адам сөздік элементтерінің аяққы тізімін өзі анықтайды, жүйе

ұсынған тізімнен керектісін алып немесе қосып және түрлендіруге мүмкіндігі болады. Тезаурус құрылымын автоматтандыру технологиясы бірнеше кезеңнен тұрады:

1) Мәтін кешенін алдын-ала өңдеу: мәтіндік ақпаратты машиналыю сараптау тәсілдерімен өңдеуге мүмкіндік беретін түрге жеткізу мақсатында құжат топтамасын түрлендіруге қажетті, қарапайым өңдеулер жасалады.

2) Көптеген қажетті дескрипторлар құру. Тезаурусқа енетін сөздер мен сөз тіркестерінің жиынтығын құрастыру. Сарапшы жиынтықты негізге ала отырып пән саласы негізгі ұғымдарының тізімін жасайды.

3) Тезаурус дескрипторлары арасындағы байланыс синонимдерін сөздіктерден іздеу, автоматты табылған қатынастарды сарапшының редакторлауы және дескрипторларды аққы концепттерге топтастыру.

4) Ассоциативтік және сатылық қатынастар жиынтығын құру. Тезаурус дескрипторлары арасындағы өкілдерінің типтендірілмеген қатынасының жиынтығын құру. Сарапшы, құрылған жиынтықты пайдалана отырып, тезаурус концепттері арасындағы сатылық пен ассоциативтік байланыстың аяққы жиынтығын анықтайды.

Көрсетілген төрт кезеңнен басқа өңдеу процесіне тағы бір өңдеу енуі мүмкін – кешен құжаттарын кластерлеу, ол мәтін кешені нашар құрастырылғанда және әр басқа пән салаларына жататын құжаттар кездескенде талап етіледі. Бұл кезең мақалада терең қарастырылмайды.

Мақаланың төртінші бөлімінде мәтін кешенінің алдын-ала өңделуі жан-жақты сипатталады. 2-суретте тезаурус құрылымын автоматтандыру технологиясының негізгі кезеңдерін көрсететін жалпы жобасы ұсынылған.

4. Мәтін кешендерін алдын-ала өңдеу

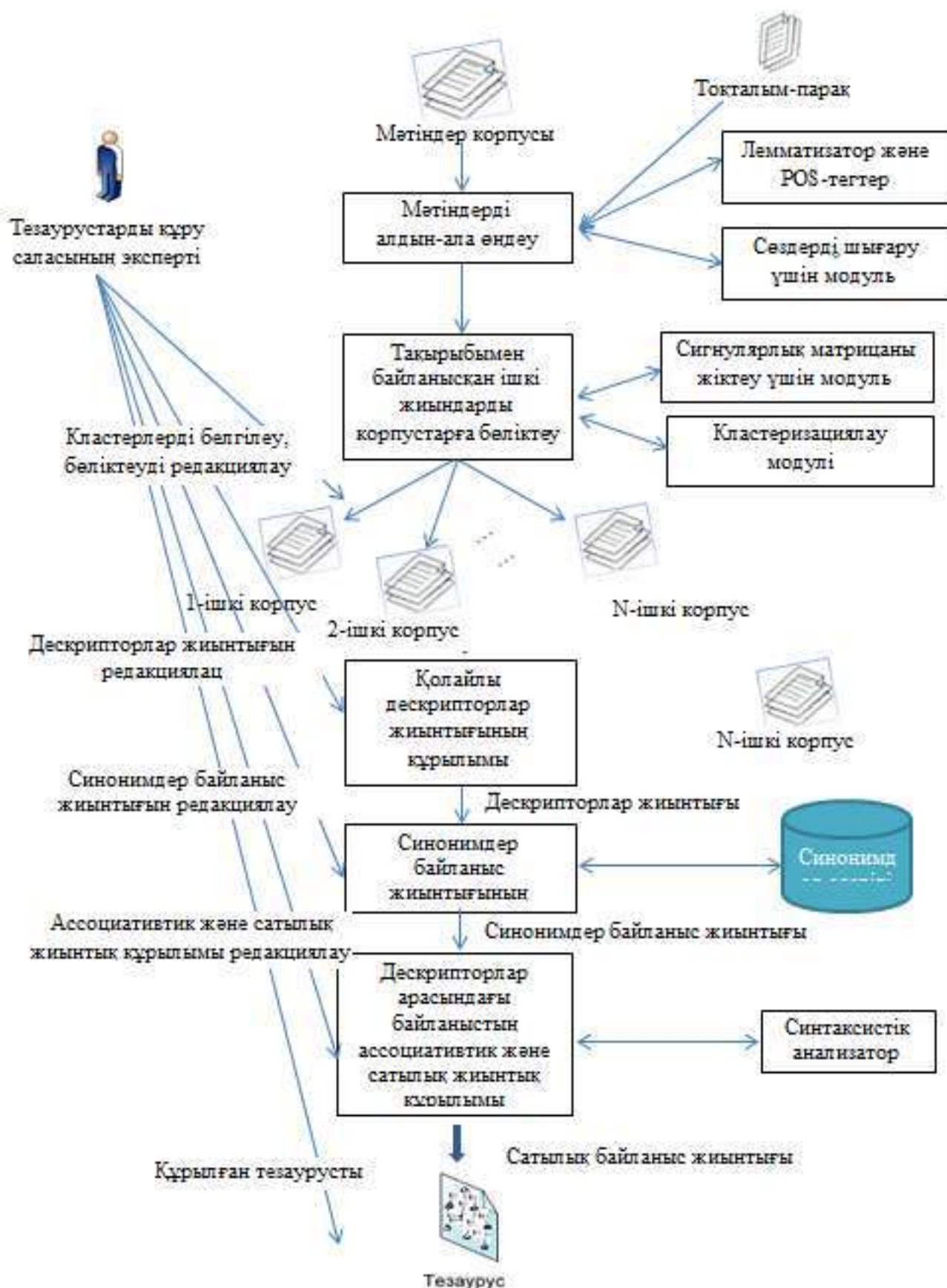
Үлкен мәтін кешені ондап, жүз мыңдаған, кейде миллиондаған құжаттармен есептелуі мүмкін. Мұндай мәтін кешендеріндегі сөз саны ондаған миллионға жетеді. Әрі қарапайым тілде жазылған мәтін құрылымы әлсіз және қатесі көп болуы мүмкін. Мұндай кешендегі сөздер мен сөз тіркестерінің маңыздысын айқындау және мәтінді сараптау алгоритмін қолдану кешенді алдын-ала өңдеуді қажет етеді. Сипатталып отырған технологияға мәтінді алдын-ала өңдеудің келесі кезеңдері таңдап алынды:

4.1. Мәтін кешенінен барлық сөздер мен сөз тіркестерін шығару.

ТОК кешенінде бар ерекше сөздердің барлық жиынтығы токенизация рәсімі көмегімен жасалады. Алгоритм сөздердің шекарасын тоқтатым-белгі сөздер жиынтығы – белгілер жиынтығы көмегімен анықтайды, олар мәтіндегі сөздерді [4], сондай-ақ, кейбір ережелерді бір-бірінен ажыратуға мүмкіндік береді, одан кейін кешендегі айрықша сөздер тізімі жасалады (кейде лемматизделінген – келесі параграфты қараңыз). ТОК жиынтығының көлемі көбіне 50000 сөзден аспайды. Сондай-ақ, ТОК бос жолдар бар деп есептейік.

Мәтіннен тек жеке сөздер ғана емес, сөз тіркестерін де шығара алу керек, өйткені пән саласының негізгі ұғымдары көбінесе құрамды сөздермен беріледі, мәселен «ақпараттық» және «жүйе» сөздерінің жеке қолданысынан гөрі «Ақпараттық жүйелер» сөз тіркесі «ақпараттық жүйе» дескрипторына ақпараттық тезаурус ретінде енуі анық.

Мақсаттылық түсінігіне сай бастапқы дескрипторлардың ұзындығын бес сөзбен шектейміз. Әрі, жеке дескриптор d_5 компонентті ретті кортежбен беруге болады $\langle w_1, w_2, w_3, w_4, w_5 \rangle \in TOK^5$. Егер әр сөзге $w \in TOK$ сәйкесінше 0 - бос жол болатын $\langle w, 0, 0, 0, 0 \rangle$ кортеж түрін қойсақ, $w \in TOK^5$ кортеж жиынтығы ретінде ТОК жеке сөздерінің барлық жиынтығын алуға болады. Бұл кезеңде ұзындығы екіден бес сөзге дейін болатын $MWE \in TOK^5$ кешеніндегі маңызды сөз тіркестерінің жиынтығы қалыптасады. Осындай сөз тіркестерінің әрқайсы $\langle w_1, w_2, w_3, w_4, w_5 \rangle$ түріндегі кортеж сипатында ұсынылады, ондағы үш сөзден тұратын сөз тіркесіне сай кортеж $\langle w_1, w_2, w_3, 0, 0 \rangle$ түрінде болады.



2-сурет. Тезаурус құрылымын автоматтандыру технологиясы

Маңызды сөз тіркестері жиынтығын табу үшін, сөз тіркесі құрастыруға қажетті, арнайы даярланған термин сөздігін пайдаланған аякқы өңдеушілер (finite state transducers) қолданылды[1].

Осы кезең нәтижесінде $D = W \cup MWE$ кешенінің дескрипторлар жиынтығы құрастырылады. Келешектегі сараптау мақсаты үшін кешендегі сөздер мен сөз тіркестері жиынтығын анықтау ғана маңызды емес, сондай-ақ, ақпаратты дескриптордағы қалпында сақтау да мәнді. Өңделетін кешенде сөйлемдер бар десек, онда сөйлеммен санды $SNT = \{1, 2, \dots, k\}$ бірізділігіне сай ұқсатамыз. SNT кешенінің сөйлем жиынтығынан D кешені дескрипторындағы жиынтығына $snt_d \subseteq SNT \times d$ сәйкестігін барлық реттелген жұптардың $(x, d), x \in SNT, d \in D$ жиынтық тапсырмасы арқылы анықтаймыз.

Жұмыста пән саласының мәтін кешеніне негізделген ақпараттық-іздеу тезаурус құрылымын автоматтандыру технологиясы ұсынылған. Жасалған технологияда мәтін кешенін алдын-ала машиналық өңдеу жүзеге асырылады және сарапшыға тезаурус құрастырудың әр кезеңінде сөздер, сөз тіркестері мен тезаурусқа енетін сөздер арасындағы байланыстар жиынтығы ұсынылады.

Қолданылған әдебиеттер

1. Лукашевич, Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во МГУ, 2011. – 495 с.
2. Baeza-Yates R., Ribeiro-Neto B. (1999). «Modern Information Retrieval». Addison Wesley Longman Publishing Co. Inc. 163-173.
3. Frakes W., Baeza-Yates R. (1992), «Information Retrieval. Data Structures & Algorithms». Prentice Hall PTR; Facsimile edition. pp.: 161-197.
4. Филиппович Ю., Прохоров А. (2002) «Семантика информационных технологий: опыты словарно-тезаурусного описания». Изд-во МГУП, ISBN 5-8122-0367-9.
5. Gregory G. (1994). «Explorations in Automatic Thesaurus Discovery». The Springer International Series in Engineering and Computer Science.

АВТОМАТИЗИРОВАННАЯ СИСТЕМА УЧЕТА ПОСЕТИТЕЛЕЙ ОРГАНИЗАЦИИ И УСЛУГ НА ПРИМЕРЕ ИП «APARTS.KZ»

Сейтзадина Г.Ж

j_gulshar@mail.ru

Студентка группы ИС-10 СКГУ им.М.Козыбаева

Научный руководитель - преподаватель кафедры ИС Гуляева А.Л

Сейчас во многих структурах используются специальное программное обеспечение, для автоматизации организационной и хозяйственной деятельности, для поддержки оперативного управления, ведения бухгалтерского и налогового учета. Перед большинством работающих сегодня организаций возникает одна и та же проблема – это возрастающий поток данных, вследствие чего, возникают проблемы получения быстрого доступа к информации. Помимо этого, также стоят задачи, связанные с усовершенствованием способов ведения архивов, усовершенствованием оборудования, программного обеспечения, и самое главное это – автоматизация ручного труда, следствием которого является повышение эффективности управления.

В современных условиях сравнительно быстро развивается рынок новых технологий организации автоматизированного учета, который используются на предприятиях разного профиля, в том числе и в организации учета ведения аренды квартир.

На сегодняшний день в организации «Aparts.kz» данные заносятся в обычный журнал. Поэтому создание автоматизированной системы учета посетителей и услуг является актуальным на данный момент, так как сейчас все предприятия переходят на автоматизацию своей деятельности. Это очень удобно и регистрация займет намного меньше времени.

На сегодняшний день процесс учета клиентов в деятельности фирмы не автоматизирован, что важно для времени обслуживания клиентов, поэтому возникает