



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ТҰҢҒЫШ ПРЕЗИДЕНТІ - ЕЛБАСЫНЫҢ ҚОРЫ

«ҒЫЛЫМ ЖӘНЕ БІЛІМ – 2017»

студенттер мен жас ғалымдардың
XII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ

XII Международной научной конференции
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ – 2017»

PROCEEDINGS

of the XII International Scientific Conference
for students and young scholars
«SCIENCE AND EDUCATION - 2017»



14th April 2017, Astana



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

**«Ғылым және білім - 2017»
студенттер мен жас ғалымдардың
XII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XII Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2017»**

**PROCEEDINGS
of the XII International Scientific Conference
for students and young scholars
«Science and education - 2017»**

2017 жыл 14 сәуір

Астана

УДК 378

ББК 74.58

Ғ 96

Ғ 96

«Ғылым және білім – 2017» студенттер мен жас ғалымдардың XII Халықаралық ғылыми конференциясы = The XII International Scientific Conference for students and young scholars «Science and education - 2017» = XII Международная научная конференция студентов и молодых ученых «Наука и образование - 2017». – Астана: <http://www.eni.kz/ru/nauka/nauka-i-obrazovanie/>, 2017. – 7466 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-827-6

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 378

ББК 74.58

ISBN 978-9965-31-827-6

©Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2017

ҚАЗАҚША МӘТІНДЕГІ ЖАЛҚЫ ЕСІМДЕРДІ АНЫҚТАУ ЕРЕЖЕЛЕРІ

Бектурсынова Рыскул Серикқызы

ryskul_1996@mail.ru

5B060200 Информатика мамандығының 4-курс студенті

Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

Ғылыми жетекші – Б.Ж. Ергеш

Өндірілетін және тұтынылатын ақпараттың көлемінің экспоненциалды өсуі – бұл адамзат қызметінің барлық салаларын сипаттайтын қазіргі дамудың жаһандық тренді. Өзінің динамикасы мен масштабтарына қарай бұл тренд ақпараттық кеңістіктің екпінді түрде ретсіз өсуі болғандықтан, осы кеңістікте ақпаратты өңдеу үдерісі барған сайын күрделеніп, көп күшті қажет ететін және тиімсіз болып барады.

Үлкен деректерге жақсы бір мысал ретінде Интернет желісінде берілген табиғи тілдер мәтіндерінің жиынтығын қарастыруға болады. Бұл жиынтықтарда адам өзінің ғана күшімен өңдей алмайтын өте үлкен көлемдегі ақпарат бар. Ол үшін табиғи тілдегі құжаттардың мазмұнын талдап, олардан білімді бөліп алатын мүмкіндіктері бар электрондық интеллектуалдық ассистенттердің көмегі қажет. Табиғи тілді түсінетін интеллектуалдық көмекшілер (жүйелер, агенттер, қосымшалар) және олардың жұмыстарының негізі болатын әдістерді және үлгілерді жобалау және құру соңғы бірнеше ондаған жылдардағы компьютерлік ғылымның өзекті мәселесі болып табылады [1].

Қазіргі кезде қазақ тілін автоматтандыратын программалар аз, қазақ тілін автоматтандыру жеке сөздерді сөз таптарына бөлу арқылы орындау оңайға түседі. Сондай табиғи тілді өңдеудің қызықты есептерінің бірі жалқы есімдерді анықтау (ағылшынша NER-name entity recognizer).

Жалқы есімдерді анықтау – бұл «тексттегі атом элементтері», яғни адам аттары, мекеме атаулары, жер-су атаулары, уақыт, ақша атаулары, пайыздар және т.б. анықталған категорияларды мәтінде тану. Бұл text mining саласының бір есебі, ол құрылымдалмаған мәтіннің ішінен жалқы есімдердің мағынасын автоматты түрде анықтайды.

Жалқы есімдерге адам есімдерін, компаниялардың атауларын, географиялық атаулар, мекен жайлар, күндер, сонымен қатар электронды адресстер, телефон нөмірлері және олардың өзара байланысын жатқызуға болады. Қазіргі кезде жалқы есімдерді анықтау бизнестің әр түрлі салаларында маркетингтік зерттеулерден бастап, интеллектуалды жартылай және автоматты шешім қабылдау жүйелеріне дейін пайдаланылады. Жалқы есімдерді анықтау қайда және қалай қолдануға болады? Бірінші айқын пайдалану бұл - кейбір нарықтық зерттеулер, ұйымдар немесе олардың брендтері болып табылады. Осы зерттеулердің мәні кейбір брендтер немесе ұйымдар жайлы әртүрлі көздерден (әлеуметтік желілер, блогтар, сайтта түсіндірмелер, әріптес сайттар және т.б.) автоматты түрде пайдаланушыларының пікірлері жайлы ақпарат алу болып табылады. Жалқы есімдерді анықтау қалай жұмыс жасайды? Тұлғалар мен мәтіндерден бастап олардың арасындағы қарым-қатынастардың өндіру мәселесі оңай жұмыс емес. Сонымен қатар, ол белгілі бір тілге байланысты болады[2]. Мысалға, қазақ тілі үшін жазылған программа украин немесе орыс тілі үшін жұмыс істей бермейді.

Жалқы есімдерді анықтаудың бірнеше типтері бар: бірінші тип сөздіктерге негізделген, түпнұсқасында іздейтін жалқы есімді анықтау; екінші типі сөйлемдегі сөздердің лексикалық және грамматикалық талдау арқылы жалқы есім болу ықтималдылығына сүйене анықтау. Бұл алгоритм объектерді бірнеше ережелерді пайдалана отырып табады. Бұл ережелер түрлі сипатта білім базасынан бастап морфологиялық ерекшеліктеріне дейін болуы мүмкін. Тілдік талдау арқылы сөз құрамындағы морфемалар (түбір, аффикстер) мен синтагматикалық тіркестердің семантикасын анықтауға болады.

Жалқы есімдерді анықтаудың – негізгі мәселесін мына мысалда қарастырайық. Бізде

мынандай мәтін бар:

Президент Нұрсултан Назарбаев 6 шілдеде Алматы қаласында туылған.

Осы сөйлемнен бірнеше объекттерді белгілеп алуға болады (оларды арнайы тэгтермен белгілеміз):

Президент [Name: Нұрсултан Назарбаев] [Date: 6 шілде күні] [Location: Алматы қаласында] туылған.

Осылайша, мәтінді талдау жасау арқылы жалқы есімдерді анықтауға болады. Жалқы есімдерді анықтаудың бірінші типіне сәйкес, қазақ тіліндігі жалқы есімдердің білімдер базасы құрылды. Білімдер базасында сөздерді түпнұсқасын қарастыра іздеу тиімді әдіс болып табылады.

Деректер өз алдына мәселені шешу барысында өзгеруі мүмкін қысқа түпнұсқа ақпарат болып табылады. Ережелер болса, қолда бар мәліметтен жаңа деректер мен гипотезалар негізінде шығатын ұзақ, бірнеше лексемадан тұратын ақпараттарды анықтауға болады. Білім базасының мәліметтер базасынан негізгі айырмашылығы – үлкен шығармашылық мүмкіндіктердің болуы. Мәліметтер базасындағы деректер әдетте белсенді емес болып келеді: олар бар болады немесе жоқ болады. Білім базасы болса белсенді түрде жетпеген ақпаратты толықтыруға тырысады. Егер ... , онда ... форматындағы ережелер кең таралған, бірақ білімді көрсетудің жалғыз амалы емес. Осы мақсатта сараптамалық жүйелерде семантикалық желілер, фреймдар, нейрондық желілер және басқа да амалдар қолданылады. Кейбір терең деңгейлерде эквиваленттер анық көрінеді. Тандалған әдіс мәселе түрімен және пәннің спецификасымен анықталады. Сараптамалық жүйедегі көптеген ережелер эвристикалық, берілген пәндегі сарапшылардың эксперименттеріне негізделеді. Егер алгоритмдік әдіс нақты әрі оптималды шешімге кепілдік берсе, эвристикалық әдіс тек мүмкін шешімді қарастырады. Білім базасы логикалық шешім механизмі үшін кіру ағыны болып табылады [3], [4], [5].

Қазақ тіліндегі мәтіннен жалқы есімдерді анықтау есебін шешу үшін келесі категорияларды мәтін ішінде анықтауды жөн көрдік: name(есім), geo(жер-су), org (өндіріс, мекеме, ұйым), animals (үй жануарлары).

Осы категориялар бойынша білімдер базасы құрастырылды. Адам есімдері білімдер базасында 2230 сөз. Жер-су атаулары білімдер базасында 3099 сөз. Өндіріс, мекеме, ұйым атаулары базасында 900 сөз. Үй жануарлары білімдер базасында 132 сөз.

Келесі суреттерде (сурет 1- 3) сөздердің білімдер базасында көрсетілуі келтірілді.

	A	B	C	D	E	F	G
1	Қазақ қыздары қазақша есімдері мен есімдері мағынасы						
2	Сөз	Анықтамасы	Сурағы				
3	АБИРА	- тәжік тілінен алынған абера — шөбере, шөбере қыз.	Кім?				
4	АФИЛА	- араб және иран тілдерінен алынған агила — ақылды, парасатты сөзінен жасалған.	Кім?				
5	АҒЛИЯ	- грек тілінен — сұлулық, көркемдік, салтанат.	Кім?				
6	АДЛИЯ	- тәжік тілінен алынған адл — әділдік, ақиқаттық + ия қосымша.	Кім?				
7	АЖАР	- көрікті, келбет, нұрлы; түр, өң.	Кім?				
8	АЖАРГҮЛ	- гүл сияқты әдемі, сұлу, көрікті.	Кім?				
9	АЖАРХАН	- көрікті, ажарлы, әрлі, нұрлы.	Кім?				

Сурет 1. Адам есімдері білімдер базасы

2	АҒАРТУ	Таң бозарып, жарық тұсу	ет
3	АҒЫН	Су және басқа сұйықтық заттардың белгілі бір бағытқа жылжып ойысуы, ағыс.2. Ағып жатқан ағыс	зат
4	АҒЫНДАЙ	Ағын тәрізді	сын
5	АҒЫҢДЫ	Ағып жатқан, ағысты	сын
6	АҒЫНСЫЗ	Ақпайтын, ағыны жоқ	сын
7	АҒЫС	Су және басқа сұйықтық заттардың белгілі бір бағытқа жылжып ойысуы, ағын	зат
8	АҒЫСТЫ	Ағып жатқан, ағынды	сын

Сурет 2. Жер, су, қала атаулары білімдер базасы

	A	B	C	D	E
112	Құралай, қорықпас	не?			
113	Бұзау	не?			
114	Абадан	не?			
115	Теңбілше, шөнжік, сарымақ	не?			
116	Алан	не?			
117	Қонжық, апанақ	не?			
118	Інсін	не?			
119	Бөлтірік, көксерек	не?			
120	Күшік, итақай	не?			
121	Жаутаң, мыртот, түлкішек	не?			
122	Өзборақ	не?			
123	Соқыр, мәулен, марғау, безек	не?			
124	Мөндік	не?			
125	Торай	не?			
126	Жіптей, иіңпе	не?			
үй жануарлардың аттары		географиялық атаулар	газет, журнал, фильм, кітап	өндіріс, і	

Сурет 3. Жануар атаулары білімдер базасы

Қазақ тіліндегі атау сөздерге байланысты ережелерді толықтай қарастыру нәтижесінде келесі жалқы есемідерді анықтау ережелері туындады:

– Егер сөз бас әріптен басталса және кім? сұрағына жауап беретін сөзді деректер қорынан қарастырамыз, ол деректер қорында бар болса ол жалқы, жанды зат есім болса онда ол адамның есімі(name) болып табылады.

Мысалы төменгі суретте көрсетілген деректер қорында кім? сұрағына жауап беретін жалқы, жанды зат есімдер (сурет 1).

– Егер сөз бас әріптен басталса, сөзді деректер қорынан қарастырамыз, ол деректер қорында бар болса, ол жалқы, жансыз зат есім болса, онда ол жер-су, мекеме, қала, көше, аң-құс, газет, журнал атаулары болып табылады. Келесі суреттерден деректер базасынан жалқы, жансыз зат есімдерді көре аласыз (сурет 2, 3).

Жалқы есімдерді анықтау алгоритмі:

- Берілген сөйлемдегі сөздерді қарастырады;
- Бас әріптен басталған сөздерді табады;
- Табылған сөздерді деректер базасынан іздейді;
- Деректер базасында табылған сөздерді мәліметтер шығару терезесіне шығарады.

Жоғарыдағы алгоритм бойынша жалқы есімдерді анықтаудың автоматтандырылған

жүйесін жүзеге асыру жоспарланып отыр.

Қорытынды

Бұл жұмыс аясында мәтіндегі жалқы есімдерді анықтау әдістері зерттелді. Басқа тілдер үшін жасалған жүйелер қазақ тіліне бейімделуі көбінесе жақсы нәтиже бермейді, себебі әр тілдің өзіндік ерекшеліктері, құрылымдық өзгешеліктері бар. Жұмыс барысында қазақ тіліндегі адам есімдері, жер-су аттары, мекеме атаулары, жан-жануарлар атаулары, ақша атауларының білімдер базасы құрылды. Әдістерді зерттей келе ережеге негізделген әдіс таңдалды. Жалқы есімдерді анықтау ережелері мен алгоритмі құрылды. Болашақта осы жасалған жұмыстарды қолданып жалқы есімдерді анықтау автоматтандырылған жүйесін жасауды мақсат етіп отырмыз.

Қолданылған әдебиеттер тізімі

1. http://www.kazntu.kz/sites/default/files/kazntu.kz.SUMMARY_Nugumanova
2. <https://kk.wikipedia.org/wiki>.
3. <http://appclub.im/archi ve/details/3073>
4. Лорьер Ж. Системы искусственного интеллекта. – М.: Мир, 1991.
5. Нильсон Н. Принципы искусственного интеллекта. - М.: Радио и связь, 1995.

ВАРИАТИВНОСТЬ АУГМЕНТАЦИИ КАК ФАКТОР ПОВЫШЕНИЯ КАЧЕСТВА ВЫДЕЛЕНИЯ ИНФОРМАТИВНЫХ СТРУКТУР НА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ СВЕРТОЧНЫМИ НЕЙРОННЫМИ СЕТЯМИ

Глинский Алексей Валерьевич

alexionon@mail.ru

Младший научный сотрудник Белорусского государственного медицинского университета,
Минск, Беларусь.

Далидович Александр Вадимович

ZedGraph@ya.ru

Младший научный сотрудник Белорусского государственного медицинского университета,
Минск, Беларусь.

Косик Иван Иванович

ivankosik91@gmail.com

Младший научный сотрудник Белорусского государственного медицинского университета,
Минск, Беларусь.

Научный руководитель – Г.М. Карапетян

1. Реферат

В последнее время большинство задач компьютерного зрения решается с помощью свёрточных нейронных сетей (СНС) [1]. Подобная востребованность обусловлена умением СНС самостоятельно, в процессе обучения, находить различные особенности и повторяющиеся паттерны на цифровых изображениях. Принцип работы СНС состоит в построении ансамблей из различного рода фильтров (признаков), на основе операции свертки, и последующего перебора различных их комбинаций. А затем, путем алгоритма обратного распространения ошибки, позволяет сохранить наиболее эффективные признаки и их комбинации, отбрасывая при этом неэффективные, то есть те, вклад которых в формирование информативной картины несущественен. Таким образом, создаётся оптимальный набор признаков, позволяющий решить поставленную задачу классификации объектов на исходном изображении. Такой классификатор не восприимчив к сдвигам изображения, изменениям в яркости и прочим искажениям, что заметно повышает качество результата. Однако СНС обладает существенным для проводимого в данной работе исследования недостатком: для получения необходимого уровня качества, требуется большой