



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.enu.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

межд. конф. «Actual problems of computer sciences», Алматы, КазНУ, 2003, 143-145

4. Л.Л. Ла, М. Нугманов Об одной многоцветовой схеме разделения секрета Вестник КазНПУ им.Абая, серия «Физико-математические науки», N (29), Алматы, 2010, 163-166.

5. М. Tompa, H. Woll How to Share a Secret with Cheaters\\ Proceedings on Advances in cryptology – CRYPTO '86, Santa Barbara, California, United States, pp. 261 – 265, 1987.

ҚАЗАҚ ТІЛІНДЕГІ МӘТІНДІ СЫН ЕСІМДЕР НЕГІЗІНДЕ СЕНТИМЕНТ ТАЛДАУ

Қажымұхан Дина Асқарқызы

askarovna0105@mail.ru

Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Ақпараттық технологиялар факультеті,
Информатика және ақпараттық қауіпсіздік кафедрасы, 5B060200 – Информатика
мамандығының 4 курс студенті, Астана, Қазақстан
Ғылыми жетекшісі - Бурибаева А.К.

Жалпы сентимент талдау деп табиғи тілдегі мәтінді өңдеу әдістері, статистика, машиналық оқыту көмегімен пікірдің эмоционалдық реңкін анықтауды айтамыз [1]. Бұл технология жаңа жағдайлар, өнімдер, ұйым, мемлекет және т.б. жағдайларды автоматты түрде бағалау үшін қолданылады [2].

Бүгінгі күні әлеуметтік желі, онда пікір қалдыру әр адамның күнделікті ісіне айналды. Әлеуметтік желі қолданушылары жаңалықтарға байланысты посттарға пікір қалдырып, өз ойларын білдіріп жатады. Әлеуметтік желі кең қолданыста болмаған кезде көптеген жылдар бойы «Адамдар не ойлайды?» деген сұраққа әлеуметтік сауалнамалар көмегімен жауап алу бір топ адамдарын ойын зерттеу тәсілі болып келді [4]. Интернетке кең қолданыс және әлеуметтік медиа, әлеуметтік топтардың ойын зерттеудің жаңа тәсілі – сентимент талдауды ұсынады. Алынған ақпаратты және оның қолданылуын талдау маркетингтік, саясат зерттеулерінде, жаңалықтарға әлеуметтік желі қолданушыларының әсерін қадағалау артықшылығын береді. Динамикалық түрде өсіп отыратын Интернеттегі ақпаратты өңдеу, әлбетте, автоматтандырылған ақпараттық технологиялар көмегімен іске аспайды.

Сентимент-талдаудың басқа тапсырмасы, пікірлердігі спамдарды, салыстыруларды іздеу үшін, аспектерді анықтау ретінде қолданылуы болып табылады [3].

Мәтіннің реңкін автоматты түрде анықтау үшін келесі тәсілдер бар:

1. Ережелер негізіндегі үлгілерді (rule-based with patterns) қолдану тәсілі. Ол мәтіннің реңкін анықтайтын ережелерді құрудан тұрады. Ол үшін мәтін сөз немесе сөздер тіркесіне (N-grams) бөлінеді. Одан кейін, алынған мәліметтер, жағымды немесе жағымсыз бағалау берілетін үлгілерді белгілеу үшін қолданылады. Белгіленген үлгілер «ЕГЕР шарт, ОНДА қорытынды» түріндегі ережені құру кезінде қолданылады.

2. Оқытушысыз машиналық үйрету (unsupervised learning). Бұл тәсіл мәтіннің көп бөлігін терминдер алатынды деген идеяға сүйенеді. Терминдерді реңкін анықтау арқылы, барлық мәтіннің реңкі туралы қорытынды шығаруға болады.

3. Оқытушымен машиналық үйрету (supervised learning). Бұл тәсілге мәтіннің эмоционалды кеңістік шеңберінде белгіленген, базасында статистикалық немесе ықтималдық жіктеуші (мысалы, байесовский) құрылатын оқыту жиынтығының болуы талап етіледі.

4. Гибридті тәсіл (hybrid method). Осы тәсіл жоғарыда қарастырылған барлық немесе бірнеше принциптерді үйлестіреді және анықталған тізбек негізіндегі классификаторларды қолданудан тұрады [5];

5. Теориялық-графтық модельдерге негізделген әдіс. Құжаттың мәтіндік корпусындағы барлық сөздер тең мағыналы еместігі жайлы болжам осы әдіс негізінде қолданылады. Кейбір сөздер салмағы үлкен болады және мәтін реңкіне қаттырақ әсер етеді. Бұл әдісті қолдану кезінде реңкті анықтау бірнеше кезеңдерге бөлінеді. 1) зерттелетін

мәтінге граф құру; 2) олардың төбелерін ранжирлеу; 3) табылған сөздерді классификациялау; 4) нәтижені есептеу [6].

Сентимент талдаудың тексерістен өткен әдістері:

а) жағымсыз сөздерді есептеу сөздігіне негізделген әдіс, яғни жағымды және жағымсыз сөздер сөздігі құрылған. Оларды толтыру үшін 6800 екі категориядағы аударылған ағылшын сөздер тізбегі қолданылған [7]. Сонымен қатар сөздіктерге тағы да түзетулер жүргізілген. Пікірді талдауда жағымды және жағымсыз сөздер саны есептеледі, сондай-ақ сөз алдында «не» жалғауы болуы да есептеледі. Егер сөз алдында «не» жалғауы кездеспесе ол қарама қарсы эмоционалды бейнеге ауысады. Сөздерді салыстыру кезінде Портердің стеммері қолданылған [8]. Стемминг - жалғау және суффиксті алып тастағанда қалған бөлік stem деп аталады, ол барлық грамматикалық формалар үшін бірдей болып келеді [9]. Өңдеу алдында жалғау және көмекші сөздер анықталмаған болып алынып тасталады. Б) Стэнфорд реңк анықтау кітапханасын және Яндекс аударудың API сервистерін қолдану әдісі – бұл әдіс алдын-ала аудару үшін арналған YandexTranslate API және Стэнфордпен әзірленген өзіне теориялық-графтық әдісті қосатын, табиғи тілді ағылшын үшін өңдеудің кітапханасын қолдануды көрсетеді. Тексеру нәтижесінде келесі нәтиже алынған: реңк сөздіктер негізінде жақсы анықталды – 71% пікір үшін бағалау дұрыс берілген. Екінші әдіс бойынша реңк тек 56% пікір үшін анықталған [10].

Сентимент талдау әдістерін қарастыру кезінде реңкті анықтау үшін тілдің құрылымына сүйену қажеттілігі көрсетіледі. Пікірді сентимент талдау барысында, сөздіктер негізіндегі әдісті қолдану және тілдің құрылымына байланысты ережелерді құру басты тапсырма болып табылады.

Қазақ тілі – құрылымы мен құрылысы кемелденіп жетік дамыған тіл. Әдеби тіл, жергілікті тіл (диалект), сөйлеу тілі, қарапайым сөйлеу тілі тәрізді қазақ тілінің түрлері бар. Яғни, осындай түрлерде ғаламторда пікірлер жазылатын болады. Пікірлердің реңкін анықтау үшін қазақ тілінің грамматикалық және морфологиялық ережелерін пайдаланып, бірқатар ережелер құру қажет.

Пікірлердің эмоционалдық реңкін анықтауда қазақ тіліндегі сын есімдер сөздігін және сын есімдердің әртүрлі сөз таптарынан туындауын қарастыра отыра құрылған ережелерді негізге алуды қарап көрелік. Сын есім дегеніміз, заттың сындық, сапалық қасиетін білдіретін, «қандай?», «қай?» сұрақтарына жауап беретін сөз табы. Сын есім дара және күрделі болып екіге бөлінеді. Дара сын есімдер ешқандай жалғаусыз тек түбір сөз болып келеді. Ал күрделі сын есімдердің туындауы ережелерді қамтиды. Сол ережелерге сүйене отыра талдау жүргізетін реңкті анықтау ережелерін құрамыз.

Сентимент талдауды қолдана отыра мәтіннің реңкін анықтау алгоритмі:

- Мәтінді сөздерге бөліп, оларды алдын-ала дайындалған деректер қорынан іздейді;
- Егер сөз деректер қорында болса, сөздің жағымды, жағымсыз екендігін анықтайды. Ізделінді сөз деректер қорында болмаса, оны деректер қорына қосып, оның реңкін анықтауды сұрайды.
- Реңкі анықталған сөздерді төменде келтірілетін формула бойынша есептеп, орташа есебін шығарады.

Сөздердің жағымды немесе жағымсыз екенін анықтау үшін келесі 1-кестедегідей деректер қорын құру қажеттігі туындайды, бұл сөздік негізінде талдау әдісін қолданатындығымызды көрсетеді.

1-кесте Сын есімдердің деректер қорында белгіленуі

Сөз	Сөз реңкі
жақсы	1
жаман	2
дұрыс	1
қате	2

Мен	0
адам	0

Мұндағы: «0» - нейтралды сөз

«1» - жағымды сөз

«2» - жағымсыз сөз

Дұрыс, нақты есептеу жүргізу үшін сын есімнің салмағы ν шамасын енгізу қажеттілігі туындайды. Бастапқы дара сын есім салмағының мәні жағымды 1-ге, жағымсыз -1-ге тең болып келеді. Қазақ тіліндегі пікірлер мен жарияланымдарда көп жағдайда сын есімнің асырмалы шырайы, яғни, *өте*, *тым*, *аса* сияқты көмекші сөздердің және сын есімді терістейтін, *емес*, *жоқ* тәрізді шылаулардың көп қолданылатынын байқауға болады. Осы сөздерді деректер қорына қосып, асырмалы шырай дара сын есім алдында кездесе, терістеу шылаулары дара сын есімнен кейін кездескен жағдайда сын есімнің салмағы өзгеретінін белгелей аламыз. Асырмалы шырайдың көмекші сөздері және терістеу шылаулары жеке деректер қоры ретінде енгізіледі.

1-кесте Сын есімдердің салмағы ν шамасы

Сөз	ν
жақсы	1
жаман	-1
өте	2
аса	2
Емес	-1

Ол шаманың мәні келесі ережелерге байланысты -1, 1,-2,2 сандарына тең болады:

1) Егер дара сын есімнен кейін терістеу шылаулары орналасса, онда сын есім салмағы жағымды -1, жағымсыз 1 мәндеріне ауысады.

Мысалы, жаман емес = жақсы: *жаман* жағымсыз сөз, *емес* шылауы жалғанғандықтан ол жағымды сөзге айналады.

2) Егер дара сын есім алдында асырмалы шырайдың көмекші сөздері кездесе, онда сын есім салмағы жағымды 2, жағымсыз -2 мәндеріне ауысады.

Мысалы, *өте жақсы*, *аса ұтымды*, *һәм жаман* бұл жерде сөздің мағынасы асырылып айтылып тұрғандықтан оның салмағы өседі.

3) Егер дара сын есімнен кейін терістеу шылауы, алдында асырмалы шырай көмекші сөзі орналасса, онда жағымды -2, жағымсыз 2 мәніне ауысады.

Мысалы, *өте жаман емес* = *өте жақсы*; *жаман* жағымсыз сөз, *емес* шылауы жалғанғандықтан ол жағымды сөзге айналады, алдында *өте* сөзі бар болғандықтан салмағы өседі.

Осы ережелерді ескере отыра пікірдің, жарияланымның жағымды немесе жағымсыз мәтін екенін анықтау формуласын құрамыз.

$$P = \nu p * k_p / k_p + k_n$$

Мұндағы: P – жағымды мәні, νp – жағымды сөздердің салмағы коэффициенті, k_p – мәтіндегі жағымды сөздердің саны, k_n – мәтіндегі жағымсыз сөздер саны.

$$N = \nu n * k_n / k_p + k_n$$

Мұндағы: N – жағымсыз мәні, νn – жағымды сөздердің салмағы коэффициенті, k_n – мәтіндегі жағымсыз сөздер саны, k_p – мәтіндегі жағымды сөздердің саны.

$$\nu p = \nu k_{p1} + \nu k_{p2} + \nu k_{p3} + \dots + \nu k_{pz}$$

νp – жағымды сөздердің салмағы коэффициенті мәтіндегі барлық жағымды сөздердің салмақтарының қосындысына тең.

$$\nu n = \nu k_{n1} + \nu k_{n2} + \nu k_{n3} + \dots + \nu k_{nz}$$

νn – жағымсыз сөздердің салмағы коэффициенті мәтіндегі барлық жағымсыз

сөздердің салмақтарының қосындысына тең.

Жекелей мәтіндегі жағымды және жағымсыз сөздердің мәнін анықтау жоғарыдағы формулалармен жүзеге асады. Енді, мәтіннің тональностын анықтау үшін салыстыру әдісін қолданамыз. Анықталған жағымды және жағымсыз мәндерді салыстырып, қайсысының сандық мәні үлкен болса, сол мәтіннің тональсты болып табылады. Салыстыру нәтиженсінде:

1) Егер $P > N$ болса, онда мәтін жағымды болып табылады.

2) Егер $N > P$ болса, онда мәтін жағымсыз болып табылады.

Қорыта келе, қазақ тіліндегі мәтінді сентимент талдаудың сөздіктер негізіндегі, оқытушымен үйрету әдістерін қолдану арқылы әлеуметтік желі қолданушыларының ойын бағалауды жүргізе алатындығымызды көреміз. Алдағы уақытта қазақ тіліндегі басқа сөз таптарының ерекшеліктерін ереже ретінде қосып, талдаудың сапасын жоғарылату жоспарлануда.

Қолданылған әдебиеттер тізімі:

1. Introduction to Sentiment Analysis, <http://www.lct-master.org/files/MullenSentimentCourseSlides.pdf>
2. Александр Прохоров, Александр Керимов, Сентимент-анализ и продвижение в социальных медиа <http://compress.ru/article.aspx?id=23115>
3. Александр Уланов, Обработка текстов на естественном языке (Лекция 10. Анализ мнений) https://compscicenter.ru/media/slides/nlp_2014_spring/2014_04_28_nlp_2014_spring.pdf
4. Yelena Aleksandrovna Mejova, SENTIMENT ANALYSIS WITHIN AND ACROSS SOCIAL MEDIA STREAMS, <http://ir.uiowa.edu/cgi/viewcontent.cgi?article=3090&context=etd>
5. М. В. Клековкина, Е. В. Котельников, Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики, <http://ceur-ws.org/Vol-934/paper15.pdf>
6. Анализ тональности текста, <https://ru.wikipedia.org/wiki/>
7. Hu and Liu, A list of positive and negative opinion words or sentiment words for English, <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>
8. Стеммер Портера для русского языка <http://www.algorithmist.ru/2010/12/porter-stemmer-russian.html>
9. Стеммер, http://www.solarix.ru/for_developers/api/stemmer.shtml
10. Mavljutov R. R., Ostapuk N. A., Using basic syntactic relations for sentiment analysis, <http://www.dialog21.ru/digests/dialog2013/materials/pdf/MavljutovRR.pdf>

УДК 004.658.6

АҚПАРАТТЫҚ РЕСУРСТАРДЫ ҚҰРУДА INTERNET/ИНТРАНЕТ ТЕХНОЛОГИЯЛАРДЫ ҚОЛДАНУ

Махаматиллаева А.

Студент, ЕНУ им. Л.Н. Гумилева, Астана, Қазақстан
Научный руководитель, к.т.н, доцент - К.Сагиндыков

Соңғы жылдары ақпараттық ресурстарды жасау мен әр түрлі бағыттағы ақпараттық жүйелерді құруда Интранет технологияларын қолдану мынадай себептер бойынша жаһандық ақпараттық кеңістікте басым болып отыр:

Қажетті ақпаратты іздеу жүйелерінде пайдаланушы үшін жеткілікті қарапайымдылықты ұйымдастыруға мүмкіндік береді.

Клиенттің жұмыс орнына техникалық жағынан да, бағдарламалық қамтама жағынан да ең төменгі талаптарды қояды (клиент стандартты бағдарламалық қамтамасыз етумен жұмыс істейді және тек WWW көрерменнің- соңғы версиялы браузерлердің бірінің жұмысын