



INTERNATIONAL SCIENCE REVIEWS



No. 4 (2) 2021

Natural Sciences and
Technologies series



INTERNATIONAL SCIENCE REVIEWS

Natural Sciences and Technologies series

Has been published since 2020

№4 (2) 2021

Nur-Sultan

EDITOR-IN-CHIEF:

Doctor of Physical and Mathematical Sciences, Academician of NAS RK, Professor
Kalimoldayev M. N.

DEPUTY EDITOR-IN-CHIEF:

Doctor of Biological Sciences, Professor
Myrzagaliyeva A. B.

EDITORIAL BOARD:

- | | |
|----------------------------|--|
| Akiyanova F. Zh. | - Doctor of Geographical Sciences, Professor (Kazakhstan) |
| Seitkan A. | - PhD, (Kazakhstan) |
| Baysholanov S. S | - Candidate of Geographical Sciences, Associate professor (Kazakhstan) |
| Zayadan B. K. | - Doctor of Biological Sciences, Professor (Kazakhstan) |
| Salnikov V. G. | - Doctor of Geographical Sciences, Professor (Kazakhstan) |
| Tasbolatuly N. | - PhD, (Kazakhstan) |
| Urmashev B.A | - Candidate of Physical and Mathematical Sciences, (Kazakhstan) |
| Abdildayeva A. A. | - PhD, (Kazakhstan) |
| Chlachula J. | - Professor, Adam Mickiewicz University (Poland) |
| Redfern S.A.T. | - PhD, Professor, (Singapore) |
| Cheryomushkina V.A. | - Doctor of Biological Sciences, Professor (Russia) |
| Bazarnova N. G. | - Doctor Chemical Sciences, Professor (Russia) |
| Mohamed Othman | - Dr. Professor (Malaysia) |
| Sherzod Turaev | - Dr. Associate Professor (United Arab Emirates) |

Editorial address: 8, Kabanbay Batyr avenue, of.316, Nur-Sultan,
Kazakhstan, 010000
Tel.: (7172) 24-18-52 (ext. 316)
E-mail: natural-sciences@aiu.kz

International Science Reviews NST - 76153

International Science Reviews

Natural Sciences and Technologies series

Owner: Astana International University

Periodicity: quarterly

Circulation: 500 copies

CONTENT

Mardenov Y., Zhukabayeva T, Abdildaeva A., Sultangazieva A. DETECTING AND PREVENTING BLACK HOLE ATTACKS ON WIRELESS SENSOR NETWORKS.....	5
Серік Фараби, Тасболатұлы Нұрболат МЕТОДЫ ОЦЕНИВАНИЯ ВНЕДРЕНИЯ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ БИЗНЕС-ПРОЦЕССАМИ ПРЕДПРИЯТИЯ	11
Леспекова А.А, Муканова А.С. МӘТІНДІ ТОНАЛДЫЛЫҚҚА АНЫҚТАУ ӘДІСТЕРІНЕ ШОЛУ.....	21
Мажит А., Муканова А.С. ИССЛЕДОВАНИЕ И АНАЛИЗ МЕТОДОЛОГИЙ УПРАВЛЕНИЯ ИТ-ПРОЕКТАМИ В ТЕЛЕКОММУНИКАЦИОННОЙ ОТРАСЛИ	31

УДК 004

МӘТІНДІ ТОНАЛДЫЛЫҚҚА АНЫҚТАУ ӘДІСТЕРІНЕ ТАЛДАУ

Леспекова А.А.¹, Муканова А.С.².

azizalespekova1998@gmail.com, assel.mukanova@aiu.edu.kz

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Кәсіпорындардағы компьютерлік жүйелер мен желілерді әкімшілендіру, бақылау және қорғау мамандығының 2-курс магистранты, Нұр-Сұлтан қаласы, Қазақстан.

²Астана халықаралық университеті

Аңдатпа. Мақалада мәтіннің тоналдылығын анықтау үшін қолданылатын бірнеше тәсілдерге талдау жүргізілді. Олардың артықшылықтары мен кемшіліктері көрсетілген. Сонымен қатар машиналық оқыту әдісі мен лексикаға негізделген әдістердің жұмыс жасау принциптері қарастырылды. Лексикаға негізделген әдіске бағытталған программаның жұмыс істеу мүмкіндіктері көрсетілді.

Түйін сөздер: тоналдылық, лемматизация, корпус, мәтін, талдау алгоритмдері

КІРІСПЕ

Мәтіннің тоналдылығын талдау (sentiment analysis) - табиғи тілді өңдеу (natural language processing, NLP) әдістерінің бірі, оның мақсаты мәтіннен эмоционалды мазмұнды алу болып табылады. Жасанды интеллекттің бұл бағыты табиғи тілде мәтін түрінде болатын әртүрлі ақпаратты алуға мүмкіндік береді.

Мәтіннің тоналдылығы немесе тоны — бұл автордың кейбір оқиғаларға, кейіпкерлерге, объектілерге немесе тікелей аудиторияға деген эмоционалды қатынасының көрінісі. Бұл мәтіннің эмоционалды компоненті. Бұл сипаттама өте маңызды. Өйткені, эмоцияларсыз мәтін тартымсыз және монотонды болады. Тоналдылық бір өлшемді немесе көп өлшемді эмоционалды кеңістікке жіктелуі мүмкін. Бір өлшемді кеңістікте тек бір шкала болады, онда бірнеше мәндер болуы мүмкін. Көп өлшемді кеңістікте бірнеше ортогональды өлшемдер бар, негізгі эмоцияларды (қуаныш, бақыт, қорқыныш) жатқызамыз. Осылайша, тоналдылықты талдауды кейбір көңіл-күй бағаларын тағайындау арқылы жүзеге асырылатын сапалы деректерді сандық сипаттау әдісі ретінде қарастыруға болады.

Тоналдылықты талдаудың мақсаты - мәтіндегі пікірлерді табу және олардың қасиеттерін анықтау. Аудитория мүшелерінің ойлау тәсілін зерттеу және өнімнің жағдайын қарама-қарсы тұрғыдан зерттеу мүмкіндігі.

Тоналдылықты анықтау үшін (Taboada, 2011жылы) [7] пікір терминдерін көңіл-күй сөздігімен салыстыратын лексикаға негізделген әдісті ұсынды. Ол жұмыста көңіл-күйді қаншалықты оң немесе теріс екенін бағалайды. Пікірлер сөздіктегі сөздер объективті.

(Hu, 2004жылы) [5] Сөздікке негізделген тәсілдің негізгі стратегиясы ұсынылған болатын. Онда белгілі бағдармен пікір білдіретін сөздердің шағын жиынтығы қолмен жинақталады.

Жарнама контекстінде (Qiu, 2010) [6] Сезімдерге қатысты сөйлемдерді анықтау үшін сөздікке негізделген әдісті қолданды. Олар жарнаманың өзектілігі мен пайдаланушылармен жұмыс істеу ыңғайлылығын арттыратын жарнамалық тәсілді ұсынды. Олар тақырыптық сөздерді шығару мәселесін шешуге және синтаксистік талдау мен көңіл-күй сөздіктерін қолдана отырып, жарнаманың кілт сөздерін алу кезінде тұтынушылардың көзқарасын анықтауға арналған ережеге негізделген стратегияны ұсынды. Олар автомобиль веб-форумдарына өз үлестерін қосты.

(Mandal Das, 2018) [8] Автор екі тілдің, ағылшын және бенгал тілдерінің көңіл-күйін анықтау үшін фильмдердің шолу мәліметтерімен жүргізілген әртүрлі эксперименттердің нәтижелерін талдады. Олар әр түрлі Машиналық оқыту алгоритмдерін сынақтан өткізді кодпен араласқан және ең жоғары дәлдікке қол жеткізген деректер 59,00% *аңғал Байес* моделін (NB) және 72,50% тірек векторлық машина модельдеріндегі (SVM) нәтижені алды. Олар мынадай тұжырымға келеді: оқу және тест деректері бірдей болған кезде SVM жоғарғы нәтижені көрсетеді, ал NB оқыту және тестілеу деректері әртүрлі болған кезде жақсы жұмыс істейді.

НЕГІЗГІ БӨЛІМ

Мәтінді тоналдылыққа талдау алгоритмдері. Ережеге негізделген тәсіл табиғи тілді өңдеудің негізгі процедурасын қамтиды. Ол келесі мәтіндік әрекеттерді қамтиды:

- Стемминг
- Токенизация
- Тотау сөздер
- Талдау Лексиканы талдау (тиісті контекстке байланысты)

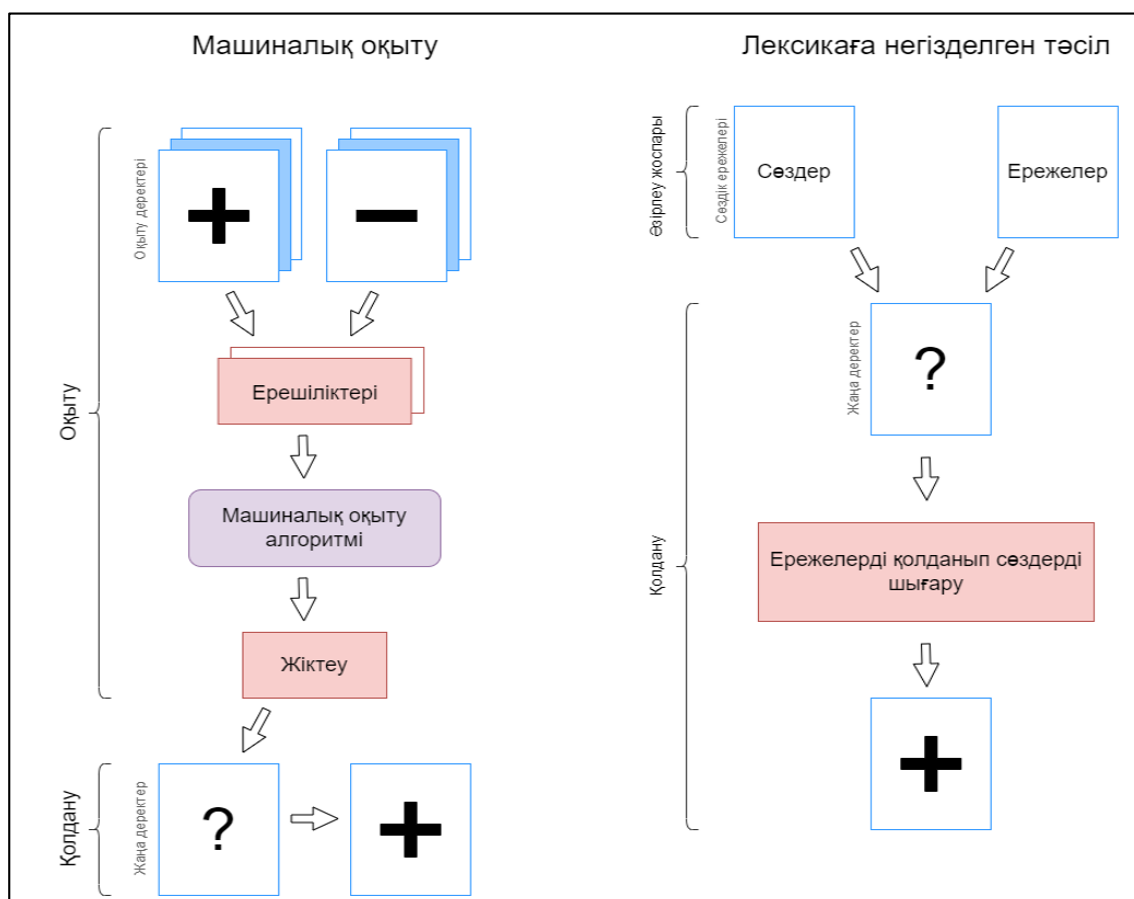
Мәтінді алдын-ала өңдеу мәтінді тоналдылыққа талдаудың бірінші кезеңі болып табылады. Бұл "шулы" мәтіннен тиісті ақпаратты бөліп көрсету үшін қажет. Мәтінді алдын-ала өңдеу барлық сөздерді бір регистрге келтіруді, тыныс белгілерін жоюды, тоқтату сөздерін жоюды, токенизацияны, сөздерді қалыпқа келтіруді және

қажет болған жағдайда басқа операцияларды қамтиды. Мәтінді талдау кезінде барлық сөздерді қалыпқа келтіру ұсынылады. Яғни. сөзді оның бастапқы түрінде ұсыну. Қалыпқа келтіруді екі жолмен жүргізуге болады: лемматизация және стемминг.

Лемматизация-сөзді оның бастапқы формасына (лемма) түрлендіру. Лемматизация морфологиялық сөздікке негізделген. Егер сөз сөздікте болмаса, онда сөзді өзгерту және оған лемма алу жолдары туралы гипотеза жасалады.

Стемминг-сөздің негізін алу, ал сөздердің соңы, жұрнақтары, префикстері алынып тасталады. Осылайша, мәтіндегі барлық сөздер бір нысанға келтіріледі. Стемминг негізделген морфологиялық ережелердің және сөздіктің болуын талап етпейді.

Оның жұмыс жасау принципі: Сөздердің екі тізімі бар. Олардың біреуіне тек оң, екіншісіне теріс кіреді. Алгоритм мәтінді қарайды, өлшемдерге сәйкес келетін сөздерді табады. Осыдан кейін алгоритм мәтінде қандай сөз басым екенін есептейді. Егер оң сөздер көп болса, онда мәтіннің оң полярлығы бар деп саналады(1-сурет). Төменгі суретті машиналық оқыту әдісі мен лексикаға негізделген тәсілдердің жұмыс жасау принциптері көрсетілген. Машиналық оқыту әдісінде тоналдылық автоматты түрде машиналық оқыту алгоритмі бойынша анықталады. Лексикаға негізделген тәсілде сөздерді реттеп отыратын мұғалім қажет болады. Ережеге негізделген алгоритмдердің мәні-олар қандай да бір нәтиже берсе де, оларды шынымен пайдалы ететін икемділік пен дәлдік жетіспейді. Мысалы, ережеге негізделген тәсіл контекстті ескермейді. Алайда, оны клиенттерге қызмет көрсету үшін пайдалы болуы мүмкін хабарламалардың үнін анықтау үшін жалпы мақсатта пайдалануға болады.



Сурет 1. Тоналдылыққа талдау әдістерінің жұмыс жасау архитектурасы

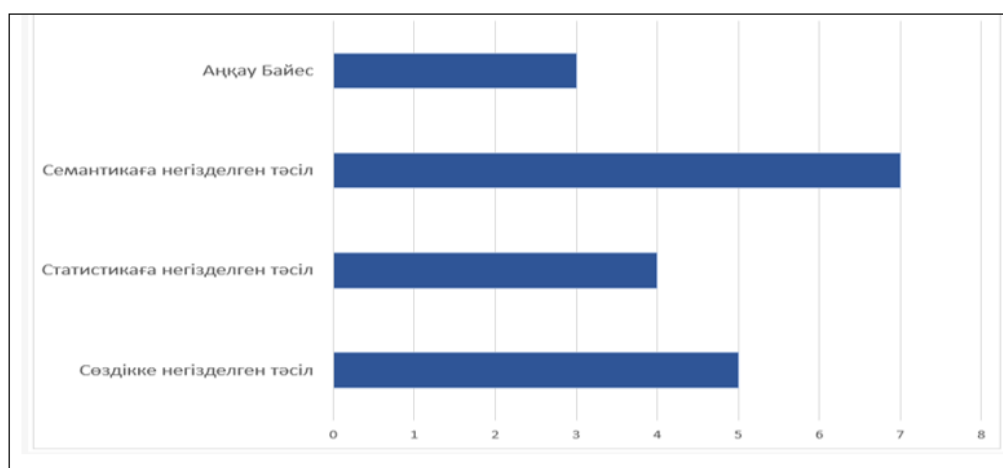
Тоналдылықты талдаудың қарапайым түрі "жақсы" және "жаман" сөздер сөздігінің көмегімен жүзеге асырылады. Сөйлемдегі әр сөз мағынасына қарай сандық сипатқа ие бола алады: әдетте +1 оң тоналдылық жағдайында, ал -1 – теріс тоналдылық жағдайында. Содан кейін біз сөйлемнің жалпы тоналдылығын білу үшін барлық сөздердің бағаларын қорытындылаймыз. Әлбетте, бұл тәсіл көптеген шектеулерге ие болады. Контекст пен жақын сөздерді елемей мұмкін. Мысалы, біздің қарапайым модельде "жақсы емес" деген сөзді 0 тоналды бағалаумен жіктеуге болады, өйткені "жоқ" -1 баллға ие, ал "жақсы" +1 баллға ие. Адам "жақсы емес" сөзін "жақсы" сөзінің болуына қарамастан теріс деп жіктеуі мүмкін, сондықтан да анықталмаған жағдайда 0 балл шкаласымен есептейміз.

Лексикаға негізделген тәсілдер жоғары сапалы көңіл-күй лексикасын қажет етеді, ал машиналық оқыту тәсілдері тиімді лингвистикалық функцияларды қажет етеді. Көңіл-күй лексикасы әртүрлі көңіл-күй полярлықтары бар көптеген сөздерден тұрады. Лексикаға негізделген тәсілдерге сөздікке негізделген тәсілдер,

корпус және қолмен жасалатын тәсілдер кіреді, бірақ көптеген зерттеушілер корпусқа негізделген әдістерге көбірек көңіл бөледі.

Сөздікке негізделген тәсіл. Ол полярлығы қолмен жинақталған бастапқы сөздердің аз санынан басталады. Содан кейін сентименталды сөздердің саны WordNet, SentiWordNe және тезаурус сияқты танымал лексикондарға негізделіп көбейеді. Бұл әдіс лингвистикалық талдауды қолдана отырып, алдын-ала құрастырылған тоналды сөздіктер мен ережелер бойынша мәтіндегі эмоционалды лексиканы іздеуге негізделген.

Корпустық тәсіл сентименталды сөздердің аз санынан басталуы мүмкін, содан кейін кейбір нақты ережелерге немесе формулаларға сәйкес үлкен корпус арасында сентименталды сөздердің саны артады. Корпус тәсілдері контекстке бағытталып оның мазмұнын таба алады, ал сөздікке негізделген тәсілдер мұндай сөздерді таба алмайды. Корпустың көптеген тәсілдері деректерге негізделген көңіл-күй лексикасын құру әдістеріне негізделген. Ол белгілі бір пәндік аймаққа тән көңіл-күй лексикондарын құруда артықшылықтарға ие. Корпусқа негізделген тәсілдерді нақты әдістер тұрғысынан статистикаға негізделген тәсілдер мен семантикаға негізделген тәсілдер деп бөлуге болады.



Сурет 2. Мәтінді тоналдылыққа таңдау әдістерінің қолданылу жиілігі

Статистикаға негізделген тәсілдің негізгі идеясы—көңіл -күй сөздері арасындағы қатынасты көрсететін тиісті статистикалық ережелерді әзірлеу. Бұл бастапқы көңіл-күй сөздерінің аз ғана бөлігінен басталады, ал статистикаға негізделген кейбір басқа тәсілдер бастапқы көңіл-күй сөздерін қажет етпейді. Статистикалық ережелер зерттеушілердің назарын аударды және олар бақылау, талдау және жалпылау арқылы статистикаға негізделген көптеген тиімді ережелерді анықтады. Мәтінді талдау үшін біз келесі алгоритмді қолдана аламыз: алдымен мәтіндегі әр сөзге сөздіктен оның тоналдық мәнін тіркейміз (егер ол

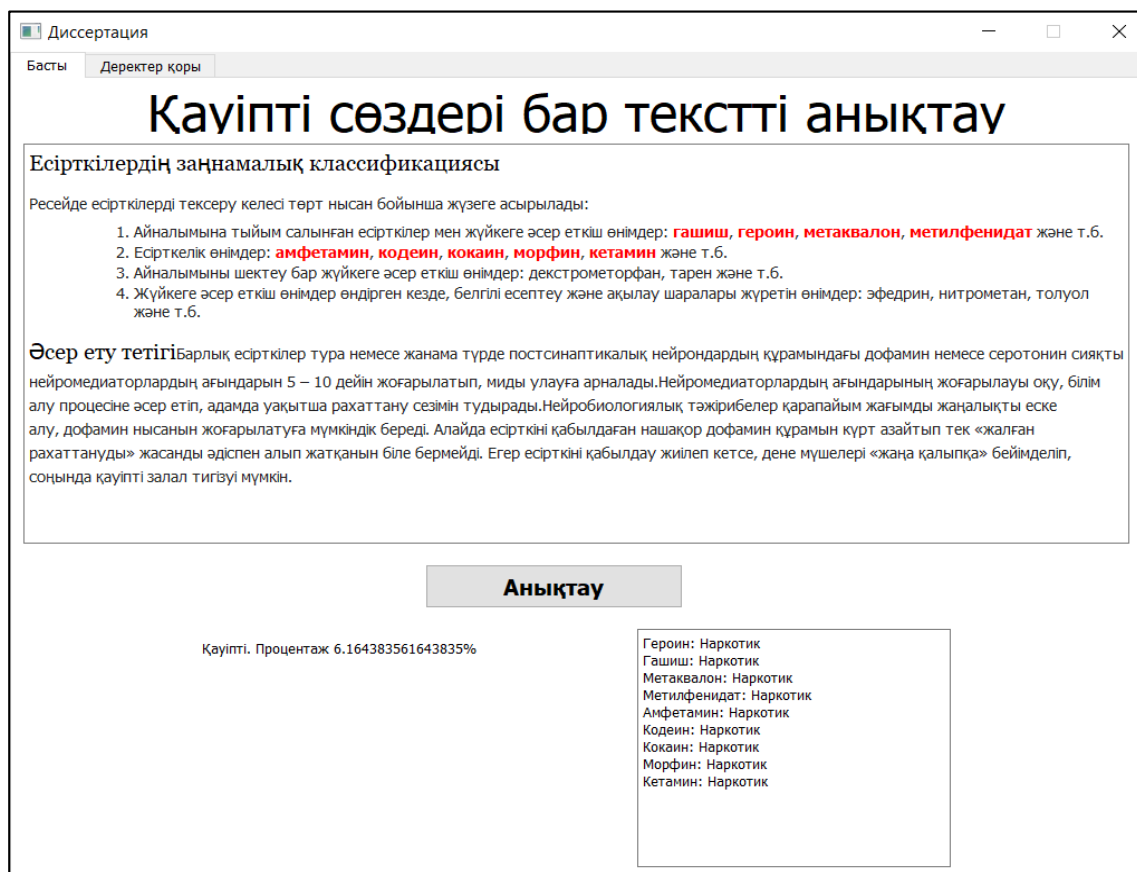
сөздікте болса), содан кейін әр жеке сөйлемнің тоналдық мәндерін қосу арқылы бүкіл мәтіннің жалпы тоналдылығын есептейміз. Қарапайым арифметикалық ортасын есептей қоямыз.

Семантикаға негізделген тәсіл сентименталды сөздердің мағыналарын белгілі бір принциптерге сәйкес анықтайды. Семантикалық мағынасы ұқсас сөздер көңіл-күйді ұқсас бағалауға ие болады. Семантикаға негізделген тәсілдің негізгі идеясы-көңіл-күйдің бастапқы сөздері, эмотикондар немесе көңіл-күйдің белгілі бір полярлығын білдіретін кез-келген басқа белгілер сияқты парадигмалық сөздерді анықтау. Әр сөзді дұрыс есептеу формуласын қолдана отырып, сөз парадигмасы арқылы өлшеуге болады. Бұл жағдайда көңіл-күйдің қарқындылығын көрсететін көңіл-күйді бағалау тиісті көңіл-күй сөзіне тағайындалады. Сонымен қатар, бұл тәсілді итеративті түрде жүзеге асыруға болады, яғни көңіл-күй лексикасының ауқымы біртіндеп артып келеді. 2004 жылы көңіл-күйдің тұқымдық сөздері мен бағалау факторы - "жақсы" және "жаман" деп аталатын *сын есімдер арасындағы* семантикалық қашықтықты есептейтін тәсілді ұсынды. Семантикалық қашықтық WordNet-пен есептеледі, ал соңғы нәтижелер бұл тәсіл әдеттегі сұраушыға қарағанда нақты мәселелер үшін жеткілікті жұмыс істейтінін көрсетеді.

Категориялар	Әдістер	Артықшылықтары	Кемшіліктері
Лексикалық тәсіл	Сөздікке негізделген тәсіл	Сөздік қоры тез жасалады және оның сапасы жоғары	Ол алдыңғы лексиканы немесе синонимдерді қажет етеді және оны қамту шектеулі
	Статистикаға негізделген тәсіл	Статистикаға негізделгендіктен оның қамтуы үлкен. Сонымен қатар, лексикондар әртүрлі салаларға жарайды	Лексиканың сапасы қол лексикондарына қарағанда нашар. Параметрлерден тым шекті мәндерді анықтау оңай емес
	Семантикаға негізделген тәсіл	Ол семантикалық ақпаратты енгізу арқылы жоғары қол жетімділікке және сапаға ие	Кейбір жағдайларда бұл ұзақ уақыт алуы мүмкін және оның параметрлері тәжірибеге сәйкес беріледі
Дәстүрлі машиналық оқыту тәсілі	Аңқау Байес	Онымен модель құру оңай және ол тұрақты жіктеу дәлдігіне ие	Әр түрлі тәуелсіз деген болжам нақты жағдайларда орындалмауы мүмкін

Сурет 3. Мәтінді тоналдылыққа талдау әдістемесінің қысқаша мазмұны

Жоғарыда көрсеткен әдістерді талдай отырып, мәтіннің тоналдылығын анықтай отырып, тыйым салынған сөздерді анықтауға бағытталған программа жүзеге асырылды. Программада мәтінді тоналдылыққа талдаудың лексикаға негізделген әдісі қолданылады және алдағы уақытта толықтырылады. Ол екі беттен тұрады. Бірінші бетте енгізілген мәтіннен корпуста бар сөздер белгіленеді. Мұнда анықтау батырмасы арқылы мәтінде кезіккен қауіпті барлық сөздерді және олардың қандай категорияға жататынын көрсетеді.



Сурет 4. Мәтіннен қауіпті сөздердің анықтау

Екінші бетте мәтін корпусымен жұмыс жасалады. Мұнда біз қауіпті сөздерді категорияларға бөліп, қоса аламыз. Алдағы уақытта бұл жұмыс өңделіп толықтырылады.

Диссертация

Басты Деректер қоры

ID	Қауіпті сөз	Қатысы
1	Анаша	Наркотик
2	Митинг	Политика
4	Наркотик	Наркотик
5	Героин	Наркотик
6	Гашиш	Наркотик
7	Метаквалон	Наркотик
8	Метилфенидат	Наркотик
9	Амфетамин	Наркотик
10	Кодеин	Наркотик
11	Кокаин	Наркотик
12	Морфин	Наркотик
13	Кетамин	Наркотик
14	Экстази	Наркотик

Қауіпті сөз:

Қатысы:

Сурет 5. Корпустағы сөздер жиыны

ҚОРЫТЫНДЫ

Бұл жұмыста мәтіннің тоналдылығын анықтаудың бірнеше тәсілдеріне талдау жүргізілді. Артықшылықтары мен кемшіліктері көрсетіліп өтілді.

Сонымен қатар машиналық оқыту әдісі мен лексикаға негізделген әдістердің жұмыс жасау принциптері қарастырылды. Тақырып бойынша жүргізілген жұмыстар сипатталды. Таңдалған әдістердің ішінде ең оңтайлысы корпуста негізделген әдіс. Лексикаға негізделген әдіске бағытталған программаның жұмыс істеу мүмкіндіктері көрсетілді.

ҚОЛДАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

1. Б. Лю и Л. Чжан, «Пікірлерді зерттеуге және көңіл-күйді талдауға шолу» 2012жыл, 412-425бет.
2. Кумбрия Эрик, "Аффективті есептеу және көңіл-күйді талдау", IEEE Intelligent Systems. 2016жыл, 102-107 бет.
3. Поляков Е. В., Восков Л. С., Абрамов П. С., Поляков С. В. Қысқа мәтіннің көңіл-күйін талдау мәселелерін шешуге жалпыланған көзқарасты зерттеу,

- табиғи тілді өңдеу тапсырмаларындағы хабарламалар. Ақпараттық-басқару жүйелері, 12-14бет. 2020ж.
4. Ениколопов С. Н., Кузнецова Ю. М., Смирнов И. В., Станкевич М. А., Чудова Н. В. мәтінді автоматты түрде талдау құралын жасау әлеуметтік-гуманитарлық зерттеулер. 1 бөлім. Әдістемелік және әдіснамалық аспектілер. Жасанды интеллект және шешім қабылдау, 2019жыл, 28-38 бет.
 5. Hu, M., Liu, B., & Street, S. M. (2004). Пайдалы қазбаларды өндіру және клиенттердің пікірлерін жалпылау.
 6. Кю, Г., Он, Х., Чжан, Ф., Ши, Ю., Бу, Дж., и Чен, с.((2010). Қолданба деректері бар сараптамалық жүйелер : Көңіл-күйге негізделген наразылыққа бағытталған жарнаманы талдау Q, қосымшалары бар сараптамалық жүйелер, 37 (9), 6182-6191. <https://doi.org/10.1016/j.eswa.2010.02.109>
 7. Табоада, М., Брук, Дж., және Фолл, к. (2011). Лексикаға негізделген көңіл-күйді талдау әдістері. 2010 жылғы қыркүйек
 8. Mandal, S., & Das, D. (2018). Көңіл-күйді анықтау үшін жіктеуіштер мен кодпен араласқан факторлардың рөлін талдау.

АНАЛИЗ МЕТОДОВ ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ ТЕКСТА

Аннотация. В статье проведен анализ нескольких подходов, используемых для определения тональности текста. Показаны их достоинства и недостатки. Также были рассмотрены принципы работы метода машинного обучения и методов, основанных на лексике. Продемонстрированы возможности функционирования программ, ориентированных на лексический метод.

Ключевые слова: тональность, лемматизация, корпус, текст, алгоритмы анализа

Сведения об авторе:

Леспекова А. А.¹, Муканова А. С.².

¹Евразийский национальный университет имени Л.Н. Гумилева, магистрант 2 курса, г. Нур-Султан, Казахстан.

²Международный университет Астана

ANALYSIS OF METHODS FOR DETERMINING THE TONALITY OF THE TEXT

Annotation. The article analyzes several approaches used to determine the tonality of the text. Their advantages and disadvantages are shown. The principles of the machine

learning method and methods based on vocabulary were also considered. The possibilities of functioning of programs focused on the lexical method are demonstrated.

Keywords: tonality, lemmatization, corpus, text, analysis algorithms

About the author:

¹*Lespekova A. A*, ²*Mukanova A. S.*

¹*L.N. Gumilyov Eurasian National University, 2nd year master's student, Nursultan, Kazakhstan.*

²*Astana International University*