

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РЕСПУБЛИКИ КАЗАХСТАН



**ҚазҰТЗУ ХАБАРШЫСЫ** \_\_\_\_\_

\_\_\_\_\_ **ВЕСТНИК КазНУ**

**VESTNIK KazNRTU** \_\_\_\_\_

**№ 5 (135)**

- [8] Анисимов Э.А. Анализ и состояние развития вендинг-бизнеса в республике Марий Эл . / Э. А. Анисимов, Е. Ю. Салдаева // Ученые записки Казанской государственной академии ветеринарной медицины им. Н.Э. Баумана. 2015. № 222 (2). С. 18-21.
- [9] Третьяков, М.М. Особенности организации автоматизированной розничной торговли . / М. М. Третьяков, А.В. Рудецкая // Вестник Тихоокеанского государственного университета, 2009. №2 (13).С. 139- 144.
- [10] Кахриманова, Д.Г. Вендинговый бизнес и сервис услуг . / Д. Г. Кахриманова, Г. Д. Магомедов // Инновационная наука : междунар. науч. журн. – 2015. - № 6. – С.86-88.
- [11] Гальперин, М.В. Электротехника и электроника: Учебник / М.В. Гальперин. - М.: Форум, НИЦ ИНФРА-М, 2013. - 480 с..
- [12] Баранов, В.Н. Применение микроконтроллеров AVR: схемы, алгоритмы, программы. Применение микроконтроллеров AVR: схемы, алгоритмы, программы / В.Н. Баранов. - М.: Додэка-XXI, 2006. - 288 с.
- [13] Бродин, В.Б. Микроконтроллеры. Архитектура, программирование, интерфейс / В.Б. Бродин, М.И. Шагурин. - М.: ЭКОМ, 1999. - 400 с.
- [14] Водовозов, А.М. Микроконтроллеры для систем автоматики: учебное пособие / А.М. Водовозов. - Вологда: ВоГТУ, 2002. - 123 с.
- [15] Мазиди, М.А. Микроконтроллеры PIC и встроенные системы. Применение ассемблера и С для PIC18 / М.А. Мазиди, Р.Д. МакКинли, Д. Кусэй; Пер. с англ. В.В. Литвина. - СПб.: КОРОНА-принт, МК-Пресс, 2009. - 784 с.

Леспеков Д.А., Налибаев Е.Д., Пахридин С.К., Қараман А.Д., Жамбыл А.

#### **Проектирование и разработка вендингового аппарата на базе микроконтроллера**

**Аннотация.** В статье описана разработка системы вендингового аппарата для зарядки мобильных устройств на базе микроконтроллера Atmega2560. Рассмотрена и изучена элементная база, соответствующая поставленным требованиям. Также была рассмотрена среда программного обеспечения. Объектом исследования работы являются анализ вендинговых автоматов и создания вендингового автомата, питающего мобильные устройства. Целью исследования является создание вендинговых автоматов для питания мобильных телефонов и рассмотрение его экономической выгоды.

**Ключевые слова.** Вендинговый аппарат, зарядное устройство, мобильный телефон, кофейные автоматы, микроконтроллер.

УДК 004.912; 004.62

**M.Y. Mansurova<sup>1</sup>, V.B. Barakhnin<sup>2</sup>, I.S. Pastushkov<sup>2</sup>,**

**M.Y. Kyrgyzbayeva<sup>1</sup>, Y. Khibatkhanuly<sup>1</sup>**

(1 Al-Farabi Kazakh National University, Almaty, Republic of Kazakhstan, 2 Institute of Computational Technologies SB RAS, Novosibirsk State University, Novosibirsk, Russian Federation, marzhan.kyrgyzbaeva@gmail.com)

#### **USING OF NEURAL NETWORKS TO EXTRACT NAMED ENTITIES FROM WEAKLY STRUCTURED TEXTS**

**Abstract.** This article is designed to provide all interested sides with intellectual tools to support decision making by automatically extracting knowledge from heterogeneous data sources, including the Internet. In the first part of the chapter, we considered the preprocessing and morphological analysis of texts. Then we considered a model of text documents in the form of a hypergraph and implemented the random walk method for extraction of semantically related words. As a result of calculations, we had a matrix with coefficients of affinity of words as well as a vocabulary that associates the word with the component of the vector. The second part of the chapter describes a neural network trained to extract linguistic constructions which include the possible values of descriptors of named entities of the texts. A neural network allows to retrieve information on one preselected descriptor, for example, location, in the form of the final result of the name of geographical objects. In a general case, a neural network can retrieve information on several descriptors simultaneously.

**Keywords:** machine learning, named entities, neural networks, ontology

**М.Е. Мансурова<sup>1</sup>, В.Б. Барахнин<sup>2</sup>, И.С. Пастушков<sup>2</sup>,**

**М.Е. Қырғызбаева<sup>1</sup>, Е. Хибатханулы<sup>1</sup>**

<sup>1</sup>Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан Республикасы,

<sup>2</sup>РҒА СБ Есептеу технологиялары институты, Новосибирск мемлекеттік университеті,

Новосибирск, Ресей Федерациясы, marzhan.kyrgyzbaeva@gmail.com)

## **ӘЛСІЗ ҚҰРЫЛЫМДАЛҒАН МӘТІНДЕРДЕН АТАУЛЫ МӘНДЕРДІ АЛУ ҮШІН НЕЙРОНДЫҚ ЖЕЛІЛЕРДІ ҚОЛДАНУ**

**Аннотация.** Бұл тарау барлық қызығушылық танытушы жақтарға Интернетті қоса алғанда әртүрлі дерек көздерінен автоматты түрде білімді алу арқылы шешім қабылдауды қолдау үшін интеллектуалды құралдарды беруге арналған. Мақаланың бірінші бөлімінде мәтіндерді бастапқы өңдеу және морфологиялық талдау қарастырылды. Содан кейін мәтіндік құжаттың үлгісі гиперграф ретінде қарастырылды және семантикалық байланысты сөздерді алу үшін кездейсоқ көшу әдісі жүзеге асырылды. Есептеулердің нәтижесінде біз сөздердің жақындық коэффициенттерімен матрицаны, сондай-ақ вектор компонентімен сөзді байланыстыратын сөздікті алдық. Екінші бөлім лингвистикалық құрылымдарды шығаруға оқытылған нейрондық желіні сипаттайды, ол мәтіндегі атаулы мәндер үшін дескрипторлардың ықтимал мәндерін қамтиды. Нейрондық желі алдын-ала таңдалған дескриптор бойынша ақпаратты, мысалы, орналасу жерін географиялық объектілердің атауының соңғы нәтижесі түрінде алуға мүмкіндік береді. Жалпы алғанда, нейрондық желі бір мезгілде бірнеше дескриптор туралы ақпаратты ала алады.

**Кілттік сөздер:** машиналық оқыту, атаулы мәндер, нейрондық желі, онтология

### **Кіріспе**

Қазіргі уақытта жергілікті және ғаламдық желілерде көптеген ақпарат жинақталған. Бұл ақпараттың басым бөлігі табиғи тілдегі мәтіндер түрінде көрсетілген. Оларда қамтылған пайдалы ақпарат құрылымды емес, тиісінше, классикалық есептеу әдістері мен құралдары арқылы өңдеу және талдау мүмкін емес. Сонымен қатар осы ақпарат көлемі тез өсуде. Көлемді мәтіндерді оқу және мәтіндік деректердің үлкен массивтерінен іздеу жүргізу тиімді емес, сондықтан құрылымдық емес мәтіндік массивтерден іздеу жүргізу және олардан жаңа білім алу тәсілдері көбірек танымал болып келеді. Компьютерлік лингвистикасының ғылыми саласы осы мәселелерді шешуге арналған [1].

Компьютерлік лингвистиканың міндеттері құрылымды деректерді құрылымдық емес немесе жартылай құрылымдық деректерден автоматты түрде алуда жатыр, яғни ақпарат алу (Information extraction). Ақпаратты алу жүйелері мәтіннің жекелеген бөліктерінен ақпарат жинауға, релевантты ақпараттың бөлінген қатынас, білім қоры және т.б. түрдегі құрылымдалған көрінісін құруға мүмкіндік береді. Мұндай жүйелерді пайдалана отырып, ақпаратты тиісті алгоритмдер көмегімен қосымша тұжырымдар жасауға мүмкіндік беретін семантикалық түрдегі нақты формаға келтіре аламыз.

Ақпаратты алудың маңызды міндеттерінің бірі - адам аттары, ұйым атаулары, орналасуы, уақыт және т.б. сол сияқты алдын ала анықталған категорияларға жіктеу мен атаулы мәндерді іздеу болып табылатын атаулы мәндерді тану (Named entity recognition). Бұл ішкі міндеттің шешімі белгілі бір компания немесе өнім туралы пікірлердің тоналдылығын анықтауға, атаулы мәндер арасындағы көптеген қатынастарды белгілеуге, сұрақтарды қабылдауға және оларға табиғи тілде жауап беруге мүмкіндік береді.

Бұл жұмыс мәтіндік құжаттардан фактілерді шығару бойынша зерттеулердің жалғасы болып табылады [2]. Жұмыстың мақсаты жаңалықтардағы хабарламалардан атаулы мәндерді алу технологиясын құру. Осы мақсатқа жету үшін келесідей міндеттер қойылды:

1. Мәтіндерге морфологиялық талдау жүргізу.
2. Семантикалық түрде байланысқан сөздерді алу үшін кездейсоқ көшу әдісін қолдану және мәтіндерді гиперграф түрінде моделдеу.
3. Нақты кілттік сөздердің дескрипторларға сәйкестілігіне оқытылған нейрондық желіні құру.

### **2 Қолданыстағы ақпаратты алу әдістеріне шолу**

Табиғи тілдерді өңдеу саласында ақпаратты шығару жаңа міндет емес. Ақпаратты алудың негізгі әдістері келесі түрде жіктеледі:

*Белгілер негізінде ақпарат алу әдістері.* Бұл әдіс белгіленген белгілер жиынының бар болуымен және мәтіннің алынатын элементтер белгілерінің салмағын қолданумен түсіндіріледі. Осылайша, алынатын элемент үшін оның белгілер векторы құрылады. Бұл класта барынша кең таралған Байестың ықтималдық классификаторы [3] мен жасырын марковтық моделдер [4-6] болып табылады. Қажет

элементті алу осы сегмент маңында анықталған белгілерге ықтималдық талдау жүргізу негізінде мәтіннің қандайда бір сегментін тануға жақындайды. Осы жұмыс барысында зерттеліп отырған фактілерді алу әдісі осы класқа жатады. Сонымен қатар жұмыста сипатталатын алгоритмнің ерекшелігі кілттік сөздерді алу үшін «Кездейсоқ көшу» әдісін және фактілерді дескрипторлар бойынша орналастыру үшін нейрондық желіні қолдану болып табылады.

*Ядроны қолданатын ақпарат алу әдістері.* Әдістің мәні танылатын элементтердің белгілік көрінісін сипаттайтын векторлардың скалярлық көбейтіндісін ядро деп аталатын қандай да бір функциямен алмастыру болып табылады. Бұл функция алгоритмдік түрде анықталады және мәтіндік сегмент құрылымын сипаттайтын танылатын элементтер мен олардың контекстін барынша күрделі көрінісін қарастырады. Мұндай әдістің кемшілігі – бұл сегмент құрылымын анықтау мен ядроларды есептеу кезінде есептеудің күрделілігі.

*Үлгілерді салыстыруға негізделген әдістер.* Бұл әдіс «үлгі» және оларды мәтін бөліктерімен салыстыру ережелері негізінде жұмыс жасайды [8, 9]. Мұнда үлгілер шектеулер тізбегін білдіреді және бұл тізбек қандай да бір үлгілік тіркестер болып табылады.

*Онтологияға негізделген әдістер.* Көрсетілген жұмыстарда ақпарат алу онтологияға негізделіп семантикалық торларды қолдану арқылы жүзеге асырылады [10,11]. Бұл жұмыстарда авторлар онтологиялық білім қоры негізінде құрылатын алгоритмді ұсынады және StanfordCoreNLP синтаксистік үштігінің мәтіндік парсер нәтижелерінің жиілік талдауын жүргізеді.

### **3 Ақпаратты алу мәселесін шешу жолы**

Ақпаратты алу үшін бірінші кезеңде семантикалық байланысты (кілттік) сөздерді алу қажет. Бұл жұмыста семантикалық байланысты сөздер (сөз тіркестері) мәтіндерді морфологиялық талдау және кездейсоқ көшу әдісі негізінде шығарылады.

#### **3.1 Мәтінді морфологиялық белгілеу**

Мәтінді морфологиялық белгілеу сөз формаларын анықтаудан және әр сөз формасына лексематикалық мінездеме мен грамматикалық сипаттама белгілеуден тұрады.

Мәтіндегі әр сөз үшін құрылатын морфологиялық ақпарат 4 "жолдан" немесе белгілеу тобынан тұрады:

1. Лексеманың сөзформасы (лексеманың «сөздік жазбасын» көрсетеді және қай сөз табына жататынын көрсетеді).
2. Лексеманың грамматикалық белгілер жиыны немесе сөз классификаторының мінездемесі
3. Сөзформаның грамматикалық белгілерінің жиыны немесе сөз түрлендіруші мінездемелер (мысалы, зат есім септелуі, етістік түрі, яғни көпше немесе жекеше түр )
4. Грамматикалық форманың стандартты еместігі, орфографиялық өзгертулер туралы ақпарат.

#### **3.2 Кездейсоқ көшу әдісі**

*n* құжаттан тұратын мәтін берілсін делік:

$$D = (d_1, d_2, \dots, d_n) \quad (1)$$

Бұл жұмыста құжат ретінде бөлек сөйлемдерді қарастырамыз. Мәтін сөздігі үшін келесі белгілеуді енгіземіз:

$$W = (w_1, w_2, \dots, w_n) \quad (2)$$

Талданатын мәтінді гиперграф түрінде моделдеу ыңғайлы және жиі қолданылатын әдістердің бірі болып табылады. Талданатын мәтінді HG (V, E) гиперграфы түрінде көрсетейік, мұндағы V – төбелер жиыны және E – гиперқабырғалар жиыны. Сонымен қатар  $\bigcup_{e \in E} e = V$ . Гиперграфта  $v \in V$  төбелері – бұл мәтін сөздері,  $e \in E$  гиперқабырғалары – мәтін құжаттары.

$HG = R^{|V| \times |E|}$  көмегімен гиперграфтың сыбайластық матрицасын белгілейміз:

$$h(v, e) = \begin{cases} 1, & \text{егер } v \in e \\ 0, & \text{егер } v \notin e \end{cases} \quad (3)$$

$HG(V, E, w)$  - өлшенген гиперграф болсын, мұндағы  $w: E \rightarrow R^+$  - гиперқабырға салмағы.

Біздің жағдайда гиперқабырға мен гиперграф төбелерінің дәрежесі келесі формуламен анықталады:

$$d(v) = \sum_{e \in E} w(e)h(v, e) \quad (4)$$

$$\delta(e) = \sum_{v \in V} h(v, e) = |e| \quad (5)$$

Қарастырылып отырған есепте граф төбелерінің салмағын, басқаша айтқанда өңделетін мәтін сөздерінің салмағын TF-IDF әдісінің негізінде келесі формула бойынша есептеуге болады:

$$w(v_i)_{tf-idf} = \frac{tf(v_i)}{N_w} * \log \frac{N}{df(v_i)} \quad (6)$$

Мұнда:

- $tf(v_i)$  - сөздің құжатта кездесу жиілігі;
- $N_w$  - құжатқа кіретін барлық сөздердің қосындысы;
- $N$  - D құжаттар жинағындағы құжаттар саны;
- $df(v_i)$  -  $v_i$  сөздері кездесетін D-дағы құжаттар жиыны.

Кездейсоқ көшу – кездейсоқ өзгерістер – уақыттың дискретті сәттеріндегі қадамдар процесінің математикалық моделі. Сонымен қатар әр қадамдағы өзгеріс алдыңғы қадамдар мен уақыттан тәуелсіз деп есептеледі. Талдаудың қарапайымдылығына байланысты бұл модель әр түрлі салаларда жиі пайдаланылады және мұндай модель нақты процестің маңызды жеңілдетілуі болып табылады.

Бірөлшемді дискретті кездейсоқ көшу – бұл  $\{Y_n\}_{n \geq 0}$  дискретті уақытқа ие кездейсоқ процесс және ол келесі түрге ие:

$$Y_n = Y_0 + \sum_{i=1}^n X_i \quad (7)$$

Мұнда:

- $Y_0$  – бастапқы күй.
- $X_i = \begin{cases} 1, & p_i \\ -1, & q_i = 1 - p_i \end{cases}, 0 < p_i < 1, i \in \mathbb{N}$
- $Y_0, X_i, i=1, 2, \dots$  тәуелсіз кездейсоқ шамалар.

Бірөлшемді дискретті кездейсоқ көшу бүтін күйлі Марков тізбегі болып табылады. Оның бастапқы үлестірімі  $X_0$  кездейсоқ шамасының ықтималдық функциясы арқылы беріледі, ал өтпелі ықтималдық матрицасы 1-суреттегідей болады:

$$P \equiv (p_{ij})_{i,j \in \mathbb{Z}} = \begin{pmatrix} \ddots & \ddots & \ddots & & & & \\ & q_{-1} & 0 & p_{-1} & & & \\ & & q_0 & 0 & p_0 & & \\ & & & q_1 & 0 & p_1 & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

1-сурет. Өтпелі ықтималдық матрицасы

Яғни:

$$p_{i,i+1} \equiv P(X_{n+1} = i + 1 | X_n = i) = p_i,$$

$$p_{i,i-1} \equiv P(X_{n+1} = i - 1 | X_n = i) = q_i, \quad i \in Z,$$

$$p_{ij} \equiv P(X_{n+1} = j | X_n = i) = 0, \quad |i - j| \neq 1.$$

$\{X_n\}_{n \geq 0}$  кездейсоқ дискретті шамалар тізбегі Марковтың қарапайым тізбегі (дискретті уақытпен) деп аталады, егер:

$$P(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n)$$

Осылайша, қарапайым жағдайда Марков тізбегінің кезекті күйін шартты үлестіру тек ағымдағы күйден ғана тәуелді және барлық алдыңғы күйлерден тәуелсіз.

$\{X_n\}$  кездейсоқ шамаларының мәндер облысы тізбек күйінің кеңістігі деп, ал  $n$  нөмірі – қадам нөмірі деп аталады.

$P(n)$   $n$ -ші қадамдағы өтулердің ықтималдық матрицасы деп аталады:

$$P_{ij}(n) = P(X_{n+1} = j | X_n = i) \quad (8)$$

Ал  $p = (p_1, p_2, \dots, p_n)^T$  векторы, мұндағы  $p_i = P(X_0 = i)$  - Марков тізбегінің бастапқы үлестірімі. Бұдан өтулердің ықтималдық матрицасы стохастикалық болатыны айқын, яғни:

$$\sum_j P_{ij}(n) = 1, \quad \forall n \in N \quad (9)$$

Егер өтулердің ықтималдық матрицасы қадам нөмірінен тәуелсіз болса, онда Марков тізбегі біртекті деп аталады, яғни

$$P_{ij}(n) = P_{ij}, \quad \forall n \in N \quad (10)$$

Гиперграф төбелерін ранжирлеу үшін кездейсоқ көшу процесін гиперграфқа жинақтаймыз. Графтағы төбелер арасындағы өтулер бұл кездейсоқ көшу процесі, яғни көрсетілген төбеден көршілес төбеге  $t$  дискретті уақытының әр қадамы арқылы өту. Біз төбелерді  $\{s_1, s_2, \dots, s_n\}$  күйлер жиыны ретінде, ал өтулерді осы күйлердегі Марковтың ақырлы тізбегі ретінде қарастыра аламыз. Өту ықтималдығы  $P(u, v) = \text{Pr ob}(s_{t+1} = v | s_t = u)$  есептеледі, бұл Марков тізбегі  $v$  төбесінде  $t+1$  уақытында, ал  $u$  -да  $t$  уақытында болатындығын білдіреді. Біздің жағдайда Марков тізбегі біртекті болып табылады, өту ықтималдығы  $t$  уақыттан тәуелсіз. Әр төбе үшін  $\sum_v P(u, v) = 1$ .

$M$  тек бір өту үшін есептелген мүмкіндікке ие, біртекті болып табылады. Өту матрицасы барлық өтулер үшін  $P \in \mathbb{R}^{|V| \times |V|}$  есептеліне алады.  $P$  өту матрицасы төбелер арасындағы өтулерді толығымен қамтиды.

Қарапайым графта кездейсоқ көшу процесі түсінікті, қандай да бір ықтималдықпен тек мақсатты төбеге баратын қабырғаны таңдау керек. Дегенмен бұл жағдайда гиперграф аздап ерекшеленеді. Бұл гиперграф құрылымының ерекшелігімен түсіндіріледі. Мысалы,  $\delta(e) \geq 2$  гиперқабырғасы үшін төбенің екіден көп нүктелері болуы мүмкін.

Гиперграфта кездейсоқ көшу процесін жалпылау үшін гиперқабырғада бір-біріне инцидентті екі төбенің арасындағы өтуді көшу түрінде модельдейміз. Жалпы кездейсоқ көшу процесі бір қадамнан емес, екі қадамнан тұратын процес болып табылады: біріншіден, кездейсоқ серфер кезекті  $u$  төбесімен инцидентті  $e$  гиперқабырғасын таңдайды. Екіншіден, таңдалған гиперграфта серфер  $u, v \in e$  шартын қанағаттандыратын мақсатты  $v$  төбесін таңдайды.

Гиперграфтағы кездейсоқ көшу жалпылау деп аталады, ал қарапайым графта кездейсоқ көшуде ерекше жағдай бар. Бұл қабырғаның тек бір төбесінің болуы, ал гиперграфта біз көптеген төбелер арасынан таңдай аламыз. Егер біз Марков тізбегі арқылы гиперграфта кездейсоқ көшу процесін анықтасак, мұнда төбелер жиыны күйлер жиынын құрайтын болады. Әрбір  $t$  уақыт қадамында серфер инцидентті гиперқабырғада орнын басқа төбеге өзгертеді.

Өлшенген гиперграфта гиперқабырғалармен бірге төбелер салмағын ескере отырып кездейсоқ көшудің жалпы анықтамасын берейік. Бұл жағдайда кездейсоқ көшу процесі гиперқабырғалар мен төбелер салмағын пайдалана отырып кеңейтіледі. Төбелер салмағы барлық инцидентті гиперқабырғалар көмегімен анықталады, бұл ерекшеліктер векторы:

$$\vec{v}_w = \{w(v_{e_1}), w(v_{e_2}), \dots, w(v_{d(v)})\} \quad (11)$$

Яғни  $u$  төбесі кездесетін әрбір  $e$  гиперқабырғасы үшін әртүрлі төбелер салмағы болады. Кездейсоқ көшудің болжамды процесі келесі түрде сипатталуы мүмкін:  $u$  төбесінен бастап серфер  $u$  төбесімен инцидентті және  $w(e)$  гиперқабырғасының салмағына пропорционал болатын  $e$  гиперқабырғасын таңдайды. Содан кейін серфер дәл осылай төбе салмағына пропорционал болатын  $v$  төбесін таңдайды, яғни қарастырылып отырған кезекті гиперқабырға.

$H_w \in R^{V \times E}$  инциденттік матрицасын өлшенетін гиперграфта келесі түрде анықтайық:

$$h_w(v, e) = \begin{cases} w(v_e), & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (12)$$

Осылайша, гиперқабырға градусын қайта анықтайтын болсақ:

$$\delta(e_w) = \sum_{v \in V} h_w(v, e) \quad (13)$$

Жоғарыда көрсетілген формулаларды пайдалана отырып, өту матрицаларын есептеуге болады:

$$P(u, v) = \sum_{e \in E} w(e) \frac{h(u, e)}{\sum_{\hat{e} \in E(u)} w(\hat{e})} \frac{h_w(v, e)}{\sum_{\hat{v} \in e} h_w(\hat{v}, e)} \quad (14)$$

Немесе матрицалық анықтау:

$$P = D_v^{-1} H W_e D_{ve}^{-1} H_w^T \quad (15)$$

Мұнда:

- $D_v$  – өлшенген төбе дәрежесінің диагональдық матрицасы;
- $H$  – гиперграф төбелерінің инциденттік матрицасы;
- $W_e$  – гиперқабырғалар салмағының диагональдық матрицасы;
- $D_{ve}$  – өлшенген гиперқабырға дәрежесінің диагональдық матрицасы;
- $H_w$  – өлшенген графтың диагональдық матрицасы;

Мұндағы  $P$  өту матрицасы стохастикалық болып табылады және жолдардың әрбір қосындысы 1-ге тең.

$P$  өту матрицасын есептегеннен кейін, кездейсоқ көшудің  $\pi$  стационарлық үлестірімін түсіну керек болады. Стационарлық үлестірім  $\vec{v}_0 \in R^{V \times 1}$  векторынан бастап есептеліне алады. Бұл  $1/|V|$  ықтималдығының қосындысы 1-ге тең. Біріншіден,  $P^T$  өту матрицасы  $\vec{v}_0 = P^T * \vec{v}$  түрінде  $\vec{v}_0$  векторбағанын көбейту арқылы алынады. Осылайша итерацияны  $\vec{v}$  векторының өзгерісі тоқтағанға дейін жалғастырамыз. Өту матрицасын ықтималдықтарды үлесіру векторына көбейту  $\vec{x} = P^T * \vec{v}$  келесі

үлестірім қадамын береді.  $x_i$  – бұл  $i$  төбесінде табу ықтималдығы болсын. Онда  $x_i = \sum_j p_{ij} v_j$ ,  $v_j - j$

түйінінде серфердің мерзімінен бұрын табылу ықтималдығы және  $p_{ij} - j$ -дан  $i$  – ға өту ықтималдығы.

Егер кездейсоқ көшу эргодикалық болса,  $\vec{v}$  векторының үлестірім ықтималдығының өзгерісі  $n$  кадамнан кейін тоқтайды.

### 3.3 Нейрондық желіні құру және оқыту

Келесі міндет өңделген мәтіндердің атаулы мәндерінің атрибуттарының ықтимал мәндерін қамтитын лингвистикалық құрылымдарды алуға үйретілген нейрондық желіні құру. Лингвистикалық құрылымдар жинағы және семантикалық тұрғыдан байланысты сөздердің жиынтығы нейрондық желіні оқытуға мүмкіндік береді.

Құрылған нейрондық желі сізге алдын-ала таңдалған бір дескриптор бойынша ақпарат алуға мүмкіндік береді, мысалы, географиялық объектілердің атауларын түпкілікті нәтиже ретінде көрсететін орналасқан орны. Жалпы жағдайда нейрондық желі бірнеше дескриптор бойынша бір уақытта ақпаратты шығара алады.

Нейрондық желіні оқыту үшін белгілердің векторларынан тұратын оқытушы жиын құрылған. Бір дескриптор үшін белгілер векторы келесі түрде құрылды: мақала мәтінінде қызығушылық тудыратын элементке дейін бес сөзден және одан кейін екі сөзден тұратын бос орын алынды. Сонымен қатар, әрбір дескриптор үшін көрсетілген сөздің сөздікте бар екендігіне жауап беретін сөздік құрылады. Әрбір дескриптордың барлық белгілері бір «bag of words-та» жиналады және белгілер векторы құрылады.

Желі әрбір кіріс деректер жиынтығын және келесі қатенің таралуын көрсету арқылы оқытылады. Нейрондық желіні оқыту алгоритмі қатені кері тарату әдісіне негізделген.

### 4 Есептеу тәжірибелерінің нәтижелері

Деректер қоймасы ретінде MongoDB 3 деректер қоры NoSQL пайдаланылды. Деректер жиынтығы қазақстандық жаңалықтар порталдарынан және төтенше жағдайлар порталдарынан жиналған 9723 жазбадан тұрады. Деректер келесі түрде белгіленеді: 1-класс – елді мекен, 0 – басқа жағдайда.

Елді мекендер бойынша жазба мысалы: “азаматтық қорғаныс туралы азаматтық кодекске сәйкес азаматтардың мүдделерін қорғайтын мемлекеттік және жергілікті атқарушы органдармен бірлесе отырып **алматы** қаласының табиғи және техногендік сипаттағы төтенше жағдайларды мемлекеттік есепке алудың негізгі функционалдық міндеттері азаматтық қорғау саласындағы ақпараттық-анықтамалық қолдауды қамтамасыз ету саласындағы азаматтық қорғаныс саласындағы қызмет туралы талдамалық материалдар дайындау және үкімет пайдаланатын ақпараттық жүйелердің жұмыс істеуін қамтамасыз ету азаматтық қорғаныс басқару әкімшілік практика саласындағы азаматтық қорғау жүйесі бақылау және қадағалау”.

Бастапқыда деректер өңделді, өйткені онда белгілеуде қателер болды. Белгілеу үшін ҚР елді мекендерінің тізімі алынды және айқын емес іздеу (difflib кітапханасы) мен корпустың қалыпты сөздерден толық сәйкестікте іздеу (morphy2 кітапханасы бойынша қалыпқа келтіру, [1])алгоритмдерінің көмегімен атаулары бар және жоқ мәтіндер табылды. Атаулардың бар немесе жоқтығы екі алгоритм бойынша да расталған мәтіндер тестіленетін және оқытылатын таңдамаға кірді.

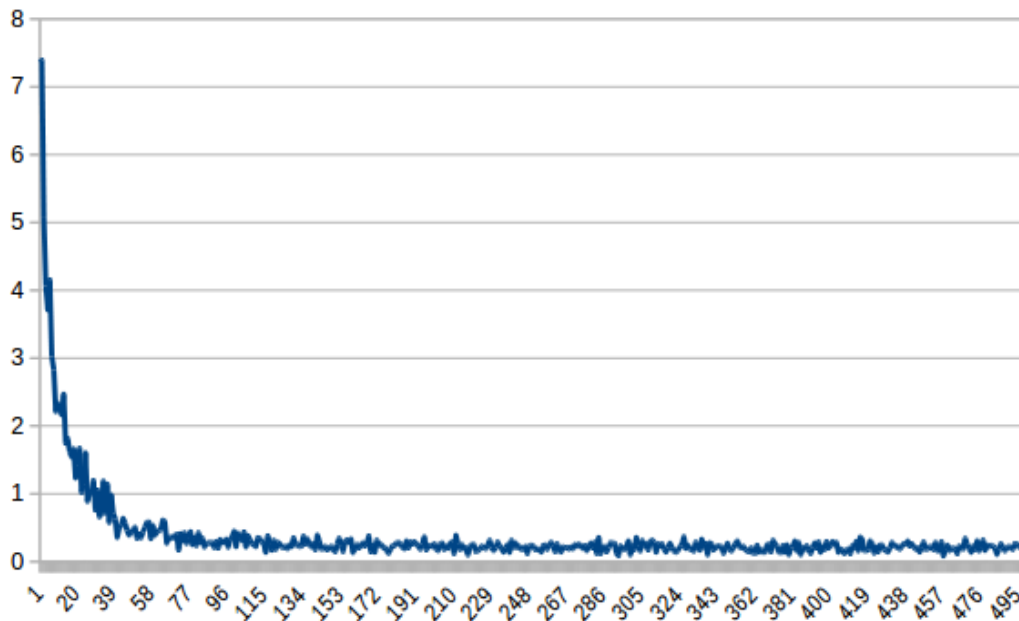
Осыдан кейін TensorFlow кітапханасын пайдалана отырып, TF-IDF матрицалары есептелді, олар кездейсоқ көшу әдісінде пайдаланылды. Нәтижесінде сөздердің жақындық коэффициенттері бар матрица, сондай-ақ вектор компонентіне сөзді сәйкестендіретін сөздік анықталды. Барлық мәтіндер алынған матрицаның мәндерімен кодталды және олардың әрқайсысы елді мекеннің атауы бар-жоқтығына байланысты. 1 немесе 0 -мен белгіленді.

Эксперименттік деректер 80% -20%, 20% - соңғы сынақ үшін болып бөлінді. Ал 80% -ы сәйкесінше 70% - 30% болып бөлінді, 70%-оқыту үшін және 30%-дұрыс тексеру үшін. Тексеру қайта оқытылған үлгісін анықтауға көмектеседі.

Барлық эксперименттер 2 жасырын қабатты нейрондық желіде өткізілді. Бірінші жасырын қабат 32 нейронды, екінші қабат 10 нейроннан тұрады. Жасырын қабат нейрондары үшін белсендіру функциясы ретінде relu және tanh әр түрлі вариациялары таңдалады, ал шығу қабаты - sigmoid. Желі оптимизациялары SGD, RMSPROP, Adam.



Төменде оқыту дәуіріне байланысты орташа квадратты жоғалту функциясының кестесі берілген. 500-ші дәуірден кейін жоғалту функциясының мәні  $\sim 0.168314$  құрайды, бұл – қарапайым желілік архитектура үшін қолайлы нәтиже.



2-сурет. Оқыту дәуіріне байланысты орташа квадратты жоғалту функциясының графигі

## 5 Қорытынды

Осы жұмыс нәтижесінде сәйкес тақырыптық мәтіннің кеңейтілген корпусында фактографиялық іздеуді ұйымдастыруға арналған, пәндік аймақтың құрылатын онтологиясында сипатталатын, алынған кілттік сөздерді кейін мән атрибуттарының мүмкін болатын мәндері ретінде қолдану үшін біртекті тақырыптардың мәтін корпустарынан кілттік сөздерді (сөз тіркестерін) алуға мүмкіндік беретін алгоритм алынды.

Есеп сөз таптарының мәтіндерінде белгілеу үшін морфологиялық парсерді қолдану арқылы кейін семантикалық байланысты кілттік сөздерді (сөз тіркестерін) алу үшін кездейсоқ көшу әдісін қолдану арқылы сәтті шешілді. Бұл сөз тіркестерінің жиынына нақты сөз тіркесін мән мәтінінде сипатталған қандай да бір атрибутқа сәйкес қою үшін жабық қабатты оқытылған нейрондық желі қолданылады. Осылайша семантикалық байланысты сөздер жұбы арқылы нейрондық желі жұмысы кезінде қалыптасатын нақты бір құжат үшін онтология құрылады.

## ӘДЕБИЕТТЕР

- [1] Шокин Ю.И., Федотов А.М., Барахнин В.Б. Проблемы поиска информации - Новосибирск: Наука, 2010 - 196 с.
- [2] Барахнин В. Б., Федотов А. М. Построение модели фактографического поиска // Вестник НГУ. Серия: Информационные технологии. - 2013. - Том 11, Выпуск № 4. - С. 16–27.
- [3] Pedersen T. A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation. ACM'2000.
- [4] Borkar V, Sarawahi S. Automatic segmentation of text into structured records, ACM, 2001.
- [5] Agichtein E., Ganti V. "Mining reference tables for automatic text segmentation," in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, -Seattle, USA, 2004.
- [6] Seymore K., McCallum A., Rosenfeld R. "Learning Hidden Markov Model structure for information extraction," in Papers from the AAAI- 99 Workshop on Machine Learning for Information Extraction, -1999, pp. 37–42.
- [7] Zelenko D., Chinatsu Aone. Kernel Methods for Relation Extraction//Journal of Machine Learning Research 3, - 2003.
- [8] Mary Califf, Raymond J. Moony, Bottom-Up Relational Learning of Matching Rules for Information Extraction, Journal of Machine Learning Research 4, 2003.
- [9] Herve Dejean. Learning Rules and Their Exceptions, Journal of Machine Learning Research 2, 2002.
- [10] Amy Aung, May Phy Thwal. Ontology Based Hotel Information Extraction From Unstructured Text // International Conference on Advances in Engineering and Technology (ICAET'2014) March 29-30, -Singapore, 2014.

[11] Raghu Anantharangachar, Srinivasan Ramani, S Rajagopalan. Ontology Guided Information Extraction from Unstructured Text // International Journal of Web & Semantic Technology (IJWesT) Vol.4, No.1, January 2013.

Мансурова М.Е., Барахнин В.Б., Пастушков И.С., Қырғызбаева М.Е., Хибатханулы Е.

**Применение нейронных сетей для извлечения именованных сущностей из слабоструктурированных текстов**

**Резюме.** Эта статья предназначена для предоставления всем заинтересованным сторонам интеллектуальных инструментов для поддержки принятия решений путем автоматического извлечения знаний из разнородных источников данных, включая Интернет. В первой части главы мы рассмотрели первичную обработку и морфологический анализ текстов. Затем мы рассмотрели модель текстовых документов в виде гиперграфа и реализовали метод случайного блуждания для извлечения семантически связанных слов. В результате вычислений мы получили матрицу с коэффициентами сродства слов, а также словарь, который связывает слово с компонентом вектора. Вторая часть главы описывает нейронную сеть, обученную извлекать лингвистические конструкции, которые включают в себя возможные значения дескрипторов именованных сущностей текстов. Нейронная сеть позволяет извлекать информацию по одному заранее выбранному дескриптору, например, местоположению, в виде конечного результата названия географических объектов. В общем случае нейронная сеть может извлекать информацию о нескольких дескрипторах одновременно.

**Ключевые слова:** машинное обучение, именованные сущности, нейронные сети, онтология

УДК 621.311

**D.K. Tursyn, N.R. Mazhrenova**

**HEAT ENGINEERING CALCULATIONS OF ENCLOSING STRUCTURES OF BUILDINGS.**

**Summary.** Building fences must have the required thermal protection properties and be sufficiently air-and moisture-proof. As regards heat-engineering aspect, the relation of external enclosing structures of buildings should have sufficient thermal properties to keep a room warm in cold weather and to protect the premises from overheating in summer. The temperature on the inner surfaces, the air permeability and the humidity of the outer barriers shall not exceed the limits permitted by the norms. The article presents the results of determining the design resistance to heat transfer of the main part of the fence structure. The evaluation of the following conditions: it is necessary that the total resistance to heat transfer was equal to or greater than the minimum permissible for sanitary and hygienic reasons, the resistance to heat transfer. To calculate the heat loss and thermal conditions in the room, the evaluation of the reduced resistance to heat transfer of the fence is given. The results can be used in the field of construction thermal physics.

**Key words:** energy, rationing, energy resource, SNiP, humidity, density, heat transfer coefficient

**Д.К. Тұрсын, Н.Р. Мажренова**

(Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан Республикасы.  
E-mail: dauren\_97.kkz@mail.ru)

**ҒИМАРАТТАРДЫҢ СЫРТҚЫ ҚОРШАУ КОНСТРУКЦИЯЛАРЫНЫҢ ЖЫЛУТЕХНИКАЛЫҚ ЕСЕБІ.**

**Түйін.** Ғимараттың қоршаулары қажетті жылу қорғау қасиеттеріне ие болуы және жеткілікті дәрежеде ауа және ылғал өткізбейтін болуы тиіс. Жылу техникалық жағынан ғимараттардың сыртқы қоршау конструкциялары үй-жайда жылуды суық уақытта сақтау және үй-жайды жазда қызып кетуден қорғау үшін жеткілікті жылу қорғау қасиеттеріне ие болуы тиіс. Ішкі беттердегі температура, сыртқы қоршаулардың ауа өткізгіштігі мен ылғалдылығы рұқсат етілген нормалардан асып кетпеуі тиіс. Мақалада қоршау конструкциясының негізгі бөлігінің жылу берілісінің есептік кедергісін анықтау нәтижелері берілген. Келесі шарттарды орындауға баға берілді: жылу берілісінің толық(жалпы) кедергісі жылу берілісінің санитарлық-гигиеналық пайымдаулары бойынша ең аз рұқсат етілген жылу берілісіне тең немесе одан көп болуы қажет. Жылу шығынын есептеу үшін және бөлмедегі жылу жағдайын есептеу үшін қоршаудың жылу берілісінің келтірілген кедергісіне баға берілді. Алынған нәтижелер құрылыс жылуфизикасы саласында пайдаланылуы мүмкін.

**Кілтті сөздер:** энергия, мөлшерлеу, энергоресурс, ҚНЖЕ, ылғалдылық, тығыздық, жылу беру коэффициенті

<i>Мансурова М.Е., Баракхнин В.Б., Пастушков И.С., Кыргызбаева М.Е., Хибатханулы Е.</i> ЭЛСІЗ ҚҰРЫЛЫМДАЛҒАН МӨТІНДЕРДЕН АТАУЛЫ МӨНДЕРДІ АЛУ ҮШІН НЕЙРОНДЫҚ ЖЕЛІЛЕРДІ ҚОЛДАНУ.....	223
<i>Тұрсын Д.К., Мажренова Н.Р.</i> ҒИМАРАТТАРДЫҢ СЫРТҚЫ ҚОРШАУ КОНСТРУКЦИЯЛАРЫНЫҢ ЖЫЛУТЕХНИКАЛЫҚ ЕСЕБІ.....	231
<i>Савинкин В.В., Ратушная Т.Ю., Кузнецова В.Н.</i> ПЛАЗМАЛЫҚ ЭНЕРГИЯ КӨЗІ ҚАЛПЫНА КЕЛТІРІЛГЕН ЖЭО ТУРБИНАЛАРЫ ҚАЛАҚТАРЫНЫҢ МЕХАНИКАЛЫҚ ПАРАМЕТРЛЕРІН ЗЕРТТЕУ.....	235
<i>Баимбетов Д., Шакир Е., Беляев Е., Калтаев А.</i> СУ ЖЫЛЫТУҒА АРНАЛҒАН КҮН ФОТОВОЛЬТИКАЛЫҚ ҚАЙТА ОСМОС ЖҮЙЕСІНДЕГІ ПРОЦЕССТЕРДІ МОДЕЛЬДЕУ.....	243
<i>Төрбек А.Ф., Мажренова Н.Р.</i> ЖЕКЕ ТҰРҒЫН ҮЙДІҢ ЭНЕРГИЯ ТИІМДІЛІГІ ПАРАМЕТРЛЕРІН ЕСЕПТЕУ.....	249
<i>Бостанбеков К.А., Юничева Н.Р., Алимова А.Н.</i> СМIP5 (COUPLED MODEL INTERCOMPARISON PROJECT) ТІЗІМІНЕ ЖАТАТЫН ЖАҒАНДЫҚ КЛИМАТТЫҚ МОДЕЛЬДЕРДІҢ ТАРИХИ ДЕРЕКТЕР ҚОРЫН ЖАСАУ.....	255
<i>Брянцев А.А.</i> ДӨҢГЕЛЕК САҢЫЛАУЛАРЫ БАР ПІСІРІНДІ ГОРФЛЕНГЕН АРҚАЛЫҚТАРДЫ БОЛМЫСТЫҚ СЫНАУ.....	261
<i>Генбач А.А., Джаманкулова Н.О.</i> ЖЫЛУ ЭНЕРГЕТИКАЛЫҚ ҚОНДЫРҒЫЛАР ПАЙДАЛАНУҒА КАПИЛЛЯРЛЫҚ-КЕУЕКТІК ҚҰРЫЛЫМНЫҢ ЖӘШІКТЕРІНДЕ БІРЛІК БУ КӨПІРШІКТЕРІНІҢ СИПАТТАМАСЫ.....	268
<i>Савинкин В.В., Ратушная Т.Ю., Кузнецова В.Н., Шакирова М.А.</i> ІСТЕН ШЫҒУ САЛДАРЫН ҚАЛПЫНА КЕЛТІРУГЕ ҮЛЕСТІК ШЫҒЫНДАРДЫ ЕСКЕРЕ ОТЫРЫП, ЖЭО ТУРБИНАСЫН ЭНЕРГИЯ ТИІМДІ ПАЙДАЛАНУ КРИТЕРИЙЛЕРІН НЕГІЗДЕУ..	276
<i>Отаров Е.Ж., Бегайдаров Б.А., Рахатов Т.Д., Исмаилов Ч.У.</i> МЕТАЛЛІ ҚҰРЫЛЫМЫН ЖАСАУ ЗАУЫТЫНДАҒЫ ЖҰМЫСШЫЛАРДЫҢ ЕҢБЕК ЖАҒДАЙЫНЫҢ КӘСІБИ ҚАУІПІН БАҒАЛАУ АСПЕКТІЛЕРІ (ӘДЕБИЕТКЕ ШОЛУ).....	285
<i>Мукашева А., Сапарходжаев Н.</i> РЕГРЕССИЯЛЫҚ МОДЕЛЬДЕР НЕГІЗІНДЕ ҚАНТ ДИАБЕТІНІҢ ӨСУІН БОЛЖАУ МОДЕЛІ .....	290
<i>Калимолдаев М.Н., Дрозденко А.А., Коплык И.В., Маринич Т.А., Абдилдаева А.А., Жукабаева Т.К., Галиева Ф.М.</i> ЭЛЕКТР ЭНЕРГИЯСЫН ТҰТЫНУДЫ БОЛЖАУДЫҢ ҚАЗІРГІ ЗАМАНҒЫ ТӘСІЛДЕРІН ТАЛДАУ.....	297
<i>Муканова Б.Г., Ахметжанов М.А., Азимова Д. Н., Намазбаев Б.Д.</i> ПАРАЛЛЕЛЬ ҚҰРЫЛЫМДЫ РОБОТТЫҢ МЫСАЛЫНДА БІРҚАЛЫПСЫЗ ЖАБУ АЛГОРИТМЫН ІСКЕ АСЫРУ (ANUCUBIC KOSSEL 3D ПРИНТЕРІ).....	307
<i>Баймаханов Г.А., Қалтай Т.Д., Қасымқұлқызы А.</i> ЖЫЛУ ЭНЕРГЕТИКАЛЫҚ ПАРАМЕТРЛЕРІН ОҢТАЙЛАНДЫРУ ЖӘНЕ ТҰТЫНУ ЖҮЙЕЛЕРІН ЖЕТІЛДІРУ.....	310
<i>Имансакипова Н.Б., Иргібаев Т.И.</i> МАГИСТРАЛЬДЫҚ МҰНАЙ ҚҰБЫРЛАРДАҒЫ ГАЗ-АУА ЖИНАҚТАРЫНЫҢ ҚАЛЫПТАСУЫ МЕН ТҰРАҚТЫЛЫҒЫНА БЕТТІК ӘСЕР ЕТУІН БАҒАЛАУДЫҢ ҒЫЛЫМИ-ӘДІСТЕМЕЛІК НЕГІЗІ.....	315
<i>Цой А.П., Алимкешова А.Х.</i> РАДИАЦИОНДЫ САЛҚЫНДАТҚЫШ ЭКСПЕРИМЕНТАЛЬДЫ СҮТ САЛҚЫНДАТҚЫШ ҚОНДЫРҒЫСЫН ЗЕРТТЕУ.....	320
<i>Жаникулов Н., Борисов И., Таймасов Б., Джанмулдаева Ж.</i> АЗ ЭНЕРГИЯ ТҰТЫНАТЫН КЛИНКЕРДІ КҮЙДІРУДІҢ ЖЫЛУ ТЕХНИКАЛЫҚ КӨРСЕТКІШТЕРІН ЕСЕПТЕУ НӘТИЖЕЛЕРІ.....	325
<i>Батыргалиев а.б.</i> ЖАНАМА ЭЛЕКТРОМАГНИТТІК СӘУЛЕЛЕНУ ЖӘНЕ НЫСАНАЛАУ БОЙЫНША АҚПАРАТ ТАРАЛУЫНЫҢ ТЕХНИКАЛЫҚ АРНАЛАРЫ АРҚЫЛЫ АҚПАРАТ ТАРАЛУЫНАН ТЕХНИКАЛЫҚ ҚҰРАЛДАРДЫҢ ҚОРҒАЛҒАНДЫҒЫН БАҒАЛАУДЫҢ КЕШЕНДЕРІ МЕН ЖҮЙЕЛЕРІНЕ ШОЛУ.....	330
<i>Аманжолов Т., Тунгатарова М., Төлеуханов А., Ахметов Б.</i> ГЕОТЕРМАЛДЫ ЖЫЛУ САҚТАУ АККУМУЛЯТОРЫ: ЖЕРДІҢ ЖЫЛУ ӨТКІЗГІШТІК КОЭФФИЦИЕНТІН ТАБУҒА АРНАЛҒАН ТӘЖІРИБЕ.....	337