

ISSN 1563 – 0277
eISSN 2617 – 4871
Индекс 75872; 25872

ӘЛ-ФАРАБИ атындағы ҚАЗАҚ ҮЛТТЫҚ УНИВЕРСИТЕТІ

ХАБАРШЫ

Математика, механика, информатика сериясы

КАЗАХСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ имени АЛЬ-ФАРАБИ

ВЕСТНИК

Серия математика, механика, информатика

AL-FARABI KAZAKH NATIONAL UNIVERSITY

**Journal of Mathematics, Mechanics
and Computer Science**

№ 4 (100)

Алматы
«Қазақ университеті»
2018

*Зарегистрирован в Министерстве информации и коммуникаций Республики Казахстан,
свидетельство № 16508-Ж от 04.05.2017 г. (Время и номер первичной постановки на учет
№ 766 от 22.04.1992 г.). Язык издания: казахский, русский, английский. Выходит 4 раза в год.
Тематическая направленность: теоретическая и прикладная математика, механика, информатика.*

Редакционная коллегия

научный редактор – Б.Е. Кангүжин, д.ф.-м.н., профессор, КазНУ им. аль-Фараби,
заместитель научного редактора – Д.И. Борисов, д.ф.-м.н., профессор, Институт
математики с вычислительным центром Уфимского научного центра РАН,
Башкирский государственный педагогический университет им. М. Акмуллы, Россия,
ответственный секретарь – Г.М. Даирбаева, к. ф.-м. н., доцент, КазНУ им. аль-Фараби.

Айсагалиев С.А. – д.т.н., профессор, КазНУ им.аль-Фараби, Казахстан

Ахмед-Заки Д.Ж. – д.т.н., Университет международного бизнеса, Казахстан

Бадаев С.А. – д.ф.-м.н., профессор, КазНУ им.аль-Фараби, Казахстан

Бектемесов М.А. – д.ф.-м.н., профессор, Казахский национальный педагогический
университет имени Абая, Казахстан

Жакебаев Д.Б. – PhD доктор, КазНУ им.аль-Фараби, Казахстан

Кабанихин С.И. – д.ф.-м.н., профессор, чл.-корр. РАН, Институт вычислительной
математики и математической геофизики СО РАН, Россия

Кыдырбекулы А.Б. – д.т.н., профессор, КазНУ им.аль-Фараби, Казахстан

Майнке М. – профессор, Департамент Вычислительной гидродинамики Института
аэродинамики, Германия

Малышкин В.Э. – д.т.н., профессор, Новосибирский государственный технический
университет, Россия

Ракишева З.Б. – к.ф.-м.н., доцент, КазНУ им.аль-Фараби, Казахстан

Ружанский М. – д.ф.-м.н., профессор, Имперский колледж Лондона, Великобритания

Сагитов С.М. – д.ф.-м.н., профессор, Университет Гетеборга, Швеция

Сукачев Ф.А. – профессор, академик АН Австралии, Университет Нового Южного Уэльса

Тайманов И.А. – д.ф.-м.н., профессор, академик РАН, Институт математики им. С.Л.
Соболева СО РАН, Россия

Темляков В.Н. – д.ф.-м.н., профессор, Университет Южной Каролины, США

Токмагамбетов Н.Е. – PhD доктор, КазНУ им.аль-Фараби, Казахстан

Шиничи Накасука – PhD доктор, профессор, Университет Токио, Япония

Научное издание

Вестник. Серия математика, механика, информатика, № 4(100) 2018.

Редактор – Г.М. Даирбаева. Компьютерная верстка – Г.М. Даирбаева

ИБ № 12547

Подписано в печать 28.12.2018 г. Формат 60 × 84 1/8. Бумага офсетная. Печать цифровая. Объем 10 п.л.

Тираж 500 экз. Заказ N 164. Издательский дом “Қазақ университеті”

Казахского национального университета им. аль-Фараби. 050040, г. Алматы, пр.аль-Фараби, 71, КазНУ.

Отпечатано в типографии издательского дома “Қазақ университеті”.

1-бөлім	Раздел 1	Section 1
Математика	Математика	Mathematics

IRSTI 27.39

Arens Algebras and Matricial SpacesDenis Potapov¹ and Fedor Sukochev²¹School of Mathematics and Statistics, University of New South Wales, Kensington,
NSW 2052 Australia, E-mail:d.potapov@unsw.edu.au²School of Mathematics and Statistics, University of New South Wales, Kensington,
NSW 2052 Australia, E-mail: f.sukochev@unsw.edu.au

Let \mathcal{M} be a finite von Neumann algebra equipped with a finite faithful normal trace τ and let $L^p(\mathcal{M}, \tau)$ be the corresponding noncommutative L^p space of τ -measurable operators associated with the couple (\mathcal{M}, τ) , $1 \leq p < \infty$. Let M_N be the algebra of all complex $N \times N$ -matrices equipped with the standard trace Tr . In this note we study the properties of Arens “algebras” over finite dimensional matrix spaces, given by Trunov’s construction for noncommutative L^p -spaces. In this work we show that the Arens “algebras” built upon Trunov’s noncommutative L^p -spaces fails to form an algebra in general. We also show that the Arens space $L^\omega(\alpha, h)$, with $0 \leq \alpha \leq 1$, fails to form an algebra in general, even in the setting of finite algebras associated to a trace, in contrast to $L^\omega(\mathcal{M}, \tau)$. In particular, we provide an example of a finite von Neumann algebra, with an associated trace, such that $L^\omega(\alpha, h)$ is not an algebra, for any choice of $\alpha \in [0, 1]$.

Keywords: von Neumann algebra, finite trace, Arens “algebras”, Noncommutative L^p -spaces.

Алгебры Аренса и матричные пространстваДенис Потапов¹ и Федор Сукочев²Школа математики и статистики^{1,2}, Университет Нового Южного Уэльса, Кенсингтон,
НЮУ 2052 Австралия

Пусть \mathcal{M} - конечная алгебра фон Неймана, снабженная конечным точным нормальным следом τ и пусть $L^p(\mathcal{M}, \tau)$ - соответствующее некоммутативное пространство L^p τ -измеримых операторов, связанных с парой (\mathcal{M}, τ) , $1 \leq p < \infty$. Пусть M_N - алгебра всех комплексных $N \times N$ -матриц, снабженных со стандартным следом Tr . В этой заметке мы изучаем свойства “алгебр” Аренса над конечномерными матричными постстранствами, заданные конструкцией Трунова для некоммутативного L -пространства. В этой работе мы покажем, что “алгебра” Аренса построена на некоммутативном L -пространстве Трунова не образуют алгебру. Мы также показываем, что пространство Аренса $L^\omega(\alpha, h)$, с $0 \leq \alpha \leq 1$, не образует алгебру, даже в случае когда алгебра конечная, связанных со следом, в отличие от $L^\omega(\mathcal{M}, \tau)$. В частности, мы приводим пример конечной алгебры фон Неймана с связанный следом, такой, что $L^\omega(\alpha, h)$, не является алгеброй, для любого выбора $\alpha \in [0, 1]$.

Ключевые слова: алгебра фон Неймана, конечный след, “алгебры” Аренса, некоммутативные L^p -пространства.

1 Introduction

Given a von Neumann algebra with a suitable trace, we may consider the now well developed theory of noncommutative integration (see, for example, [4, 9]). If we do not have a trace, or wish to consider integration with respect to a state, then the theory becomes substantially more involved. We may define noncommutative L^p -spaces, however in the most common

approaches, such as those of Haagerup [5, 10], or Connes and Hilsum [3, 6], the intersection of any two L^p -spaces for different values of p is necessarily trivial.

Our goal is to better understand the properties of noncommutative Arens algebras for integration with respect to a state. Obviously these constructions with trivial intersection are not suitable, however there are alternative possibilities. In this note we study the properties of Arens “algebras” over finite dimensional matrix spaces, given by Trunov’s construction for noncommutative L^p -spaces [11, 12]. This extends our understanding of Arens’ construction [2] for different noncommutative spaces. This topic has also been investigated by Albeverio, Ayupov, and Abdullaev [1], who characterise when this construction forms an algebra, for von Neumann algebras of type I_n .

In this note we show that the Arens space $L^\omega(\alpha, h)$, with $0 \leq \alpha \leq 1$, fails to form an algebra in general, even in the setting of finite algebras associated to a trace, in contrast to $L^\omega(\mathcal{M}, \tau)$ [7]. In Section 2 we review the necessary background, and in Section 3 we detail some intermediate results for matrix algebras. Then in Section 4 we provide an example of a space which fails to form an algebra for $\alpha \in (0, 1)$. Finally, in Section 5 we show that the example also fails to form an algebra for $\alpha \in \{0, 1\}$.

2 Preliminaries

Let \mathcal{M} be a von Neumann algebra equipped with a finite trace τ and let $L^p(\mathcal{M}, \tau)$ be the noncommutative L^p space associated with the couple (\mathcal{M}, τ) , $1 \leq p < \infty$. The norm in the space $L^p(\mathcal{M}, \tau)$ is given by

$$\|x\|_p = (\tau |x|^p)^{\frac{1}{p}}.$$

Let $h \in \mathcal{M}$ be a positive operator such that $\tau(h) = 1$ and let $0 \leq \alpha \leq 1$. Introduce the Trunov L^p -spaces

$$L^p(\alpha, h) = \left\{ x \in \tilde{\mathcal{M}} : h^{\frac{1-\alpha}{p}} x h^{\frac{\alpha}{p}} \in L^p(\mathcal{M}, \tau) \right\},$$

where $\tilde{\mathcal{M}}$ is the algebra of all τ -measurable operators [4]. We equip the latter space with the norm

$$\|x\|_{p,\alpha,h} = \left\| h^{\frac{1-\alpha}{p}} x h^{\frac{\alpha}{p}} \right\|_p, \quad x \in L^p(\alpha, h).$$

See [11, 12] for further details.

It can be seen via the Hölder inequality that the space

$$L^\omega(\mathcal{M}, \tau) = \bigcap_{1 \leq p < \infty} L^p(\mathcal{M}, \tau) \subseteq L^1(\mathcal{M}, \tau)$$

is an algebra, i.e. $xy \in L^\omega(\mathcal{M}, \tau)$ provided $x, y \in L^\omega(\mathcal{M}, \tau)$. Here we will see that the latter observation fails in general for the space

$$L^\omega(\alpha, h) = \bigcap_{1 \leq p < \infty} L^p(\alpha, h) \subseteq L^1(\mathcal{M}, \tau).$$

In other words, the spaces $L^\omega(\alpha, h)$ is not an algebra in general.

3 Matricial case

Let M_N be the algebra of all complex $N \times N$ -matrices equipped with the standard trace Tr . Let us define the matrices x_N and h_N by setting

$$x_N = \sum_{k=1}^N e_{k1} \quad \text{and} \quad h_N = \sum_{k=1}^N 2^{-k} e_{kk},$$

where $\{e_{jk}\} \subseteq M_N$ are the matrix units. If $\|\cdot\|_p$ stands for the p -th Schatten-von Neumann norm on the algebra M_N (see, for example, [8]), then

$$\begin{aligned} \left\| h_N^{\frac{1-\alpha}{p}} x_N h_N^{\frac{\alpha}{p}} \right\|_p &\leq \left\| \sum_{k=1}^N 2^{-\frac{2}{p}(1-\alpha)k} e_{k1} \right\|_2 \leq \frac{1}{2} \left(\sum_{k=1}^N 2^{-\frac{2}{p}(1-\alpha)k} \right)^{\frac{1}{2}} \\ &\leq c_{\alpha,p} = \frac{1}{2 \left(2^{\frac{2}{p}(1-\alpha)} - 1 \right)^{\frac{1}{2}}}. \end{aligned} \tag{1}$$

The constant $c_{\alpha,p}$ is finite for every $1 \leq p < \infty$ and every $0 \leq \alpha < 1$. Furthermore, the latter constant does not depend on N . On the other hand, we also have that

$$\left\| h_N^{\frac{(1-\alpha)}{p}} x_N^* x_N h_N^{\frac{\alpha}{p}} \right\|_p \geq \frac{N}{2}, \tag{2}$$

which follows easily from calculation of $(h_N^{\frac{(1-\alpha)}{p}} x_N^* x_N h_N^{\frac{\alpha}{p}})_{1,1} = \frac{N}{2}$.

4 The example

Let us consider the algebra \mathcal{M} given by the following direct sum

$$\mathcal{M} = \bigoplus_{N=1}^{\infty} M_N.$$

We equip the latter algebra with the finite trace τ defined by

$$\tau = \sum_{N=1}^{\infty} \frac{\lambda_N}{N} Tr, \quad \lambda = (\lambda_N) \in \ell_+^1,$$

where ℓ_+^1 is the space of non-negative real valued sequences which are 1-summable. We define the elements $x, h \in \mathcal{M}$ by setting

$$x = \bigoplus_{N=1}^{\infty} x_N \quad \text{and} \quad h = \bigoplus_{N=1}^{\infty} h_N.$$

Proposition 4.1. *For $0 < \alpha < 1$, the space $L^\omega(\alpha, h)$ fails to form an algebra.*

Proof. It follows from (1) that $x, x^* \in L^\omega(\alpha, h)$ for every $0 < \alpha < 1$. Indeed,

$$\|x\|_{p,\alpha,h} \leq \left(\sum_{N=1}^{\infty} \frac{\lambda_N}{N} c_{\alpha,p}^p \right)^{\frac{1}{p}} < +\infty$$

and

$$\|x^*\|_{p,\alpha,h} \leq \left(\sum_{N=1}^{\infty} \frac{\lambda_N}{N} c_{1-\alpha,p}^p \right)^{\frac{1}{p}} < +\infty.$$

On the other hand, from (2) we also have that

$$\|x^*x\|_{p,\alpha,h} \geq \frac{1}{2} \left(\sum_{N=1}^{\infty} \lambda_N N^{p-1} \right)^{\frac{1}{p}}.$$

The latter quantity is infinite, subject to appropriate choice of $\lambda \in \ell^1$. Thus, we see that the space $L^\omega(\alpha, h)$ is not an algebra, if $0 < \alpha < 1$. \square

5 The case $\alpha = 0$ and $\alpha = 1$

Proposition 5.1. *For $\alpha \in \{0, 1\}$, the space $L^\omega(\alpha, h)$ fails to form an algebra.*

Proof. Let us return to the matricial setting of Section 3. According to (1) we clearly have that

$$\left\| h_N^{\frac{1}{p}} x_N \right\|_p \leq c_p$$

and the constant c_p is finite for every $1 \leq p < \infty$. Moreover, the constant c_p does not depend on N .

Observe that the operators x_N and x_N^* are rank one operators and hence they are equivalent, i.e. there is a unitary operator $v_N \in M_N$ such that

$$x_N^* = v_N x_N v_N^*.$$

Consequently, we obtain that

$$\left\| h_N^{\frac{1}{p}} v_N x_N \right\|_p = \left\| h_N^{\frac{1}{p}} x_N^* \right\|_p \geq \frac{1}{2} N^{\frac{1}{2}},$$

by unitary invariance.

Indeed, this inequality follows as

$$\left\| h_N^{\frac{1}{p}} x_N^* \right\|_p = \left\| \left(\sum_{k=1}^N 2^{-k} e_{kk} \right) \left(\sum_{k=1}^N e_{1k} \right) \right\|_p \geq \left\| \sum_{k=1}^N \frac{1}{2} e_{1k} \right\|_p = \frac{1}{2} N^{\frac{1}{2}}.$$

Repeating the construction of Section 4 with

$$x = \bigoplus_{N=1}^{\infty} x_N \quad \text{and} \quad v = \bigoplus_{N=1}^{\infty} v_N$$

shows that $L^\omega(\alpha, h)$ fails to be an algebra also for $\alpha = 0$. Considering x^* and v^* proves that this also fails for $\alpha = 1$. \square

References

- [1] Albeverio S., Ayupov Sh. A., and Abdullaev R. Z. "Arens spaces associated with von Neumann algebras and normal states,"*Positivity*. 14 (2010): no.1, 105-121.
- [2] Arens R. "The space L_w and convex topological rings,"*Bull. Amer. Math. Soc.* 52 (1946), 931-935.
- [3] Connes A. "On the spatial theory of von Neumann algebras,"*J. Funct. Anal.* 35 (1980): no.2, 153-164.
- [4] Fack T. and Kosaki H. "Generalized s-numbers of τ -measurable operators,"*Pacific J. Math.* 123 (1986): no.2, 269-300.
- [5] Haagerup U. " L_p -spaces associated with an arbitrary von Neumann algebra,"*Proc. Colloq. Internat. CNRS.* 274 (1979), 175-184.
- [6] Hilsum M. "Les espaces L_p d'une algèbre de von Neumann définies par la dérivée spatiale,"*J. Funct. Anal.* 40 (1981): no.2, 151-169.
- [7] Kunze W. "Noncommutative Orlicz spaces and generalized Arens algebras,"*Math. Nachr.* 147 (1990), 123-138.
- [8] S. Lord, F. Sukochev, and D. Zanin: *Singular traces. Theory and applications* (De Gruyter, Berlin, 2013).
- [9] M. Takesaki: *Theory of operator algebras. II.* (Springer-Verlag, Berlin, 2003).
- [10] M. Terp: *L_p spaces associated with von neumann algebras* (Notes, Math. Institute, Copenhagen Univ., 1981).
- [11] Trunov N. V. "Integration in von Neumann algebras and regular weights,"In *Constructive theory of functions and functional analysis*. Kazan. Gos. Univ., Kazan. (1981): no.3, 73-87.
- [12] Trunov N. V. " L_p spaces associated with a weight on a semifinite von Neumann algebra,"In *Constructive theory of functions and functional analysis*. Kazan. Gos. Univ., Kazan. (1981): no.3, 88-93.

IRSTI 50.07.05; 27.35.14

On the existence of a conditionally periodic solution of one quasilinear differential system in the critical case

Suleimenov Zh., al-Farabi Kazakh National University,
Almaty, Kazakhstan, E-mail: zh-suleimenov@mail.ru

In the theory of nonlinear oscillations one often encounters conditionally periodic oscillations resulting from the superposition of several oscillations with frequencies incommensurable with each other. When finding a solution to a resonant quasilinear differential system in the form of a conditionally periodic function, the problem of a small denominator arises. Consequently, the proof of the existence, and even more the construction of such a solution is not an easy task. In this article, drawing on the work of VI. Arnold, I. Moser, and other researchers proved the existence and constructed a conditionally periodic solution of a second-order quasilinear differential system in the critical case. Accelerated convergence method by N.N. Bogolyubova, Yu.A. Mitropolsky, A.M. Samoylenko. The result can be applied to construct a conditionally periodic solution of specific differential systems.

Key words: conditionally periodic, accelerated convergence, frequency, resonance.

Квазисызықтық дифференциалдық жүйенің сындарлы жағдайдағы шартты-периодты шешімінің бар болуы

Сулейменов Ж., әл-Фараби атындағы Қазақ ұлттық университеті,
Алматы қ., Қазақстан Республикасы, E-mail: zh-suleimenov@mail.ru

Сызықтық емес тербелістер теориясында жиілігі өлшемдес емес бірнеше тербелістердің қабаттасып келуі нәтижесінде пайда болатын шартты-периодты тербелістермен жиі кездесуге тұра келеді. Осындағы резонансстық жағдайдағы квазисызықтық жүйенің шартты-периодты шешімін табу үдерісі "кішкене белім" мәселеін туындалады. Бұл мәселе шешімінің бар болуын дәлелдеу мен оны құру есебін қынданат түседі. Біздің ұсыныш отырган мақаламызда В.И. Арнольдтің, И. Мозердің және басқа да зерттеушілердің жұмыстары негізінде екінші ретті бір сындарлық жағдайдағы квазисызықтық дифференциалдық жүйенің шартты периодты шешімінің бар болатыны дәлелденіп, оны құру жолы көрсетілді. Шешімді құру барысындағы жуықтау тізбегі Боголюбов, Ю.А. Митропольский, А.М. Самойленко ұсынған үдемелі әдіске сүйеніп берілді. Жұмыстың нәтижесін нақты дифференциалдық жүйелердің шартты-периодты шешімдерін құру үшін пайдалануға болады.

Түйін сөздер: шартты-периодты, үдемелі жинақтылық, жиілік, резонанс.

О существовании условно-периодического решения одной квазилинейной дифференциальной системы в критическом случае

Сулейменов Ж., Казахский национальный университет им. аль-Фараби,
г. Алматы, Республика Казахстан, E-mail: zh-suleimenov@mail.ru

В теории нелинейных колебаний приходится часто встречаться с условно-периодическими колебаниями, возникающими в результате наложения нескольких колебаний с несоизмеримыми между собой частотами. При отыскании решения резонансной квазилинейной дифференциальной системы в виде условно-периодической функции возникает проблема малого знаменателя. Вследствие этого, доказательство существования, а тем более построения такого решения является нелегкой задачей. В данной статье опираясь на работы В.И. Арнольда, И. Мозера и других исследователей доказано существование и построено условно-периодическое решение одной квазилинейной дифференциальной системы второго порядка в критическом случае. Методом построения последовательности приближения выбран метод ускоренный сходимости Н.Н. Боголюбова, Ю.А. Митропольского, А.М. Самойленко. Результат может быть применен для построения условно-периодического решения конкретных дифференциальных систем.

Ключевые слова: условно-периодическое, ускоренная сходимость, частота, резонанс.

1 Introduction

Conditionally periodic functions are functions that are representable by trigonometric polynomials or series of the form

$$\sum_{|k_1|+...+|k_n|\geq 0} A^{(k_1,...,k_n)} \cos(k_1 w_1 + \dots + k_n w_n) t + B^{(k_1,...,k_n)} \sin(k_1 w_1 + \dots + k_n w_n) t$$

where t – argument, the summation is over all possible integer values k_1, k_2, \dots, k_n ; numbers w_1, w_2, \dots, w_n – fixed real, and rationally incommensurable numbers i.e. such that for any integers k_1, k_2, \dots, k_n

$$k_1 w_1 + k_2 w_2 \dots, k_n w_n \neq 0$$

$A^{(k_1,...,k_n)}$, $B^{(k_1,...,k_n)}$ – constant coefficients, or k_1, k_2, \dots, k_n . Set of numbers w_1, w_2, \dots, w_n is called the frequency basis or the frequency spectrum of the conditionally periodic function. It is characteristic that the frequency basis of the conditionally periodic function is finite.

The name "conditionally - periodic" was introduced by O. Staude.

The conditionally periodic function is often written for the convenience of the operations used in a complex form

$$\sum A^{(k_1,...,k_n)} e^{i(k_1 w_1 + \dots + k_n w_n) t}$$

If termwise integrating a conditionally periodic function written in the form of the indicated series without a free term, then this series may converge or may diverge. With convergence, we obtain a conditionally periodic function. If the series diverges, then one cannot integrate term by term.

Based on the ideas of A.N. Kolmogorov, V.I. Arnold managed to beat this difficulty and achieve serious results. In the present paper, we have produced a proof of the existence of a conditionally periodic solution of one quasilinear system.

2 Literature review

As already noted in celestial mechanics, nonlinear oscillations are often encountered, expressed by conditionally periodic functions. But the problem of a small denominator, arising as a result of the integration of such functions, made it difficult to carry out deep and diverse studies of the existence and construction of the conditionally periodic solution of differential equations. At the same time, KL's estimates were obtained in the theory of real numbers. Seidel [2], under which integrated conditionally-periodic functions are admissible: for the majority (in the sense of Lebesgue measure) frequencies belonging to the region n of the dimensional number space Ω with a bounded sum of modules of any n numbers.

The original literature is the works of A.N. Kolmogorov [1], V.I. Arnold [3-4], who developed a method of proving the existence and built conditionally periodic solutions of Hamiltonian systems, I. Moser [5-7], who considered a common system characteristic of the problems

of the theory of nonlinear oscillations and proved the existence of conditionally periodic solutions with a smaller number of basic frequencies than the number of degrees of freedom of the system. G.A. Merman [9] investigated the divergence of conditionally periodic series in powers of a small parameter. Yu.A. Ryabov, E.A. Grebennikov, L.K. Lika [10-11] addressed the issues of constructing conditionally-periodic solutions of canonical systems.

The method of constructing a sequence of approximate solutions was chosen by the method of accelerated convergence N.N. Bogolyubova, Yu.A. Mitropolsky, A.M. Samoylenko [12].

In recent years, A. Bari, H. Brezis, E. Feireist, H.P. have been investigating the existence of periodic and conditionally periodic solutions of nonlinear differential equations. Pelyukh, Suvak O.A. [13-15].

3 Materials and methods

Take quasi-linear system of differential equations

$$\frac{dx}{dt} = Ax + \varepsilon f(t, x), \quad (1)$$

Where

$$x = \text{colon}(x_1, x_2), \quad A = (a_{jk}), \quad j = k = 1, 2, \quad f(t, x) = \text{colon}(f_1(t, x_1, x_2), f_2(t, x_1, x_2))$$

conditionally-periodic by t with frequency basis $\omega_1, \omega_2, \dots, \omega_n$; analytical by t and x in the domain $= \{(t, x) \subset C^3 : \|x\| \leq h, \|Im\omega t\| \leq q\}$ function, $\det |A - \lambda E| = 0$ has purely imaginary roots $i\sigma_1, i\sigma_2$, and rational numbers σ_1, σ_2 non-co-measurable with $\omega_1, \omega_2, \dots, \omega_n$, ε —is a small parameter.

Let S —be a matrix, making the matrix A to Jordan form:

$$J = \begin{pmatrix} \sigma_1 i & 0 \\ 0 & \sigma_2 i \end{pmatrix}.$$

With conversion $x = Sy$, of the system (1) reduces to the form

$$\frac{dy}{dt} = Jx + S^{-1}\varepsilon f(t, Sy) \quad (2)$$

Without loss of generality, can assume that the system (1) has the form (2), i.e. A has the form: $J := \text{diag}(\sigma_1 i, \sigma_2 i)$.

3.1 The method of successive approximations

In order to find a conditionally-periodic solutions of the method of accelerated convergence [12] is applied. As an initial approximate conditionally-periodic solutions of the system

(1) $x^{(0)}(t, \varepsilon) = 0 := \text{colon}(0; 0)$ is chosen. Its residual denoted by $x^{(1)}(t, \varepsilon)$ and take this function as a first approximation to the original conditionally-periodic solutions of the system (1). Then the system relatively to $x^{(1)}(t, \varepsilon)$ will look like:

$$\frac{dx^{(1)}(t, \varepsilon)}{dt} = (J + \varepsilon P^{(0)}(t))x^{(1)}(t, \varepsilon) + \varepsilon \chi^{(1)}(t, x^{(0)}),$$

where $P^{(0)}(t) := f'_x(t, 0) := \left(\frac{\partial f_j}{\partial x_k}\right)_{(t; 0)}$, $j, k = 1, 2$; $\chi^{(1)}(t, x^{(0)}) := f(t, 0)$. Amendment to $x^{(1)}(t, \varepsilon)$ denote as $y^{(1)}(t, \varepsilon) = \text{colon}(y_1^{(1)}(t, \varepsilon), y_2^{(1)}(t, \varepsilon))$. Then it has the system:

$$\frac{dy^{(1)}(t, \varepsilon)}{dt} = (J + \varepsilon f_x^1(t, x^{(1)}))y^{(1)}(t, \varepsilon) + \varepsilon Y^{(1)}(t, x^{(1)}), \quad (3)$$

where $Y^{(1)}(t, x^{(1)}) := f(t, x^{(1)}) - f(t, 0) - f'_x(t, 0)x^{(1)}$.

The second approximation is determined by the formula $x^{(2)}(t, \varepsilon) := x^{(1)}(t, \varepsilon) + y^{(1)}(t, \varepsilon)$, and

the amendment is denoted by $y^{(2)}(t, \varepsilon)$ etc. Then to determinex^(j)(t, ε) := colon(x^(j)₁(t, ε), x^(j)₂(t, ε)) and y^(j)(t, ε) := colon(y^(j)₁(t, ε), y^(j)₂(t, ε)), j = 1, 2, ..., the following system of equations is obtained

$$\frac{dx^{(j)}(t, \varepsilon)}{dt} = (J + \varepsilon P^{(j-1)}(t))x^{(j)} + \varepsilon \chi^{(j)}(t, x^{(j-1)}), \quad (4)$$

$$\frac{dy^{(j)}(t, \varepsilon)}{dt} = (J + \varepsilon P^{(j)}(t))y^{(j)} + \varepsilon Y^{(j)}(t, x^{(j-1)}, y^{(j-1)}), \quad (5)$$

where $P^{(j-1)}, \chi^{(j)}, Y^{(j)}, j = 2, 3, \dots$ – are defined similarly to $P^{(0)}, \chi^{(1)}, Y^{(1)}$.

3.2 Integrating Model Equation

The systems (4) and (5) have the same structure and are linear non-homogeneous systems of the form:

$$\frac{dz}{dt} = (J + \varepsilon P(t))z + \varepsilon q(t) \quad (6)$$

where $J = \text{diag}(\sigma_1, \sigma_2)$, $P(t) = (p_{jk}(t))$, $j, k = 1, 2$; $q(t) := \text{colon}(q_1(t), q_2(t))$. Matrix $P(t)$ and vector-function $q(t)$ are considered to be analytical and conditionally periodic by t with frequency basis $\omega_1, \omega_2, \dots, \omega_n$. Let $B = (b^1, b^2) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ be a constant, and $R(t) = (r^1(t) \ r^2(t))$ is purely conditionally-periodic parts of the matrix $P(t)$, satisfying the equation $P(t) = B + R(t)$. Denote $T(t) := \int R(t)dt$ and let $TR = RT$, $BT = TB$. Making a replacement in the equation (6) with $z = e^{\varepsilon T(t)}\vartheta$, have

$$\frac{d\vartheta}{dt} = (J + \varepsilon B)\vartheta + \varepsilon g(t), \quad (7)$$

where $g(t) := \exp(-\varepsilon T(t))q(t)$, $g(t) := \text{colon}(g_1(t), g_2(t))$.

Let $g(t)$ be the following

$$g(t) = \sum_{\|k\| \geq 0} C^k \exp(i(k, \omega)t), \quad (8)$$

where $k := (k_1, \dots, k_n)$, $\omega = (\omega_1, \dots, \omega_n)$, $\|k\| := |k_1| + \dots + |k_n|$, $C^k := \text{colon}(C_1^k, C_2^k)$, $k\omega := k_1\omega_1 + \dots + k_n\omega_n$.

Conditionally-periodic solution of the system (7) is defined by the same form of series

$$\vartheta(t, \varepsilon) = \sum_{\|k\| \geq 0} d^k \exp(i(k, \omega)t), \quad (9)$$

where $d^k = \text{colon}(d_1^k, d_2^k)$ —undetermined coefficients. Using (9) in the system (7) considering (8) the

formal solution of the system (7) is the following :

$$\vartheta(t, \varepsilon) = \sum_{\|k\| \geq 0} \varepsilon \begin{bmatrix} i((k, \omega) - \sigma_1) - \varepsilon b_{11}; & -\varepsilon b_{12} \\ -\varepsilon b_{21}; & i((k, \omega) - \sigma_2) - \varepsilon b_{22} \end{bmatrix}^{-1} C^k \exp(i(k, \omega)t). \quad (10)$$

It is actual conditionally-periodic solution if the series (10) converges uniformly by t .

Let: $k_{n+1} := -1$, $\omega_{n+1} := \sigma_1$, $k_{n+2} := -1$, $\omega_{n+2} := \sigma_2$, $k^* := (k_1, \dots, k_{n+2})$, $\omega^* := (\omega_1, \dots, \omega_{n+2})$, $(k^*, \omega^*) := k_1\omega_1 + \dots + k_n\omega_n + k_{n+2}\omega_{n+2}$.

Then for most of the frequencies (meaning Lebesgue) $\{\omega_1, \dots, \omega_n, \omega_{n+2}\} \subset Q^*$ and for integers k_1, \dots, k_{n+2} the following estimation is true:

$$|(k^*, \omega^*)| \geq K (\|k\| + 2)^{-(n+2)}, \quad (11)$$

where $K > 0$ —is some fixed number smaller than [1] :

In the strip $\|Jm\omega t\| \leq q$ lets define $M_0 := \|q(t)\|_0 = \sup_t \|q(t)\|$, $N_0 := \|R(t)\|_0 = \sup_t \|R(t)\|$, lets define $\|\cdot\|_0$ is the norm of the strip $\|Jm\omega t\| \leq q$.

Then for the coefficients of the series $r^m(t) = \sum_{\|k\| \geq 0} \rho^{mk} e^{i(k, \omega)t}$, $\rho^{mk} := \text{colon}(\rho_1^{mk}, \rho_2^{mk})$, $m = 1, 2$; and the following estimation is true: $\|\rho^{mk}\|_0 \leq N_0 e^{-\|k\|q}$, $m = 1, 2$.

Therefore, for $T(t)$ in the strip $\|Jm\omega t\| \leq q - 2\delta_1$; $2\delta_1 \leq q$ there is $T(t) = \sum_{\|k\| \geq 0} \frac{1}{i(k, \omega)} \rho^{mk} e^{i(k, \omega)t}$, $\|T(t)\|_1 \leq \frac{Q N_0}{\delta_1^{2n}}$, $Q := \frac{1}{K} \left(\frac{4n}{e} \right)^n$, $\forall \omega \in \bar{G}$, where \bar{G} — is a set where the

following satisfies $|k, \omega| \geq K\|k\|^{-n}$, where $\|\cdot\|_1$ is the norm of the strip $\|Jm\omega t\| \leq q - 2\delta_1$. In general, $\|\cdot\|_j$ is taken as the norm of the strip $\|Jm\omega t\| \leq q - 2(\delta_1 + \dots + \delta_j)$. Then $\|g(t)\|_1 \leq M_0 e^{\varepsilon N_0 Q \delta_1^{-2n}} =: M_1$, $\|C^k\| \leq M_1 e^{-\|k\|(q-2\delta_1)}$.

Therefore in the strip $\|Jm\omega t\| \leq q - 2\delta_1 - 2\delta_2$, $2\delta_2 \prec q - 2\delta_1$ for conditionally-periodic solution of the system (6) have

$$\|x(t, \varepsilon)\|_2 \leq \frac{\varepsilon M_1}{\sqrt{2}\delta_2^{2n+1}}(1 + Q_1) \exp(2\varepsilon N_0 Q \delta_1^{-2n}), \quad Q_1 := \frac{1}{4K} \left(\frac{4(n+2)}{e} \right)^{n+2}.$$

Lets prove convergence of the series of the estimated solutions $x^{(j)}(t, \varepsilon)$ to solution of system (1).

Return to equations (4) and (5). Let in equation (2)matrix $P^{(0)}(t) := f'_x(t, 0)$ has a form:

$$P^{(0)}(t) = B^{(0)} + iC^{(0)} + R^{(0)}(t), \quad B^{(0)} := (b_{lm}), \quad C^{(0)} := (c_{lm}), \quad l, m = 1, 2;$$

where $R^{(0)}(t)$ — pure conditionaly-periodic part of matrix $P^{(0)}(t)$. Then

$$\|x^{(1)}(t, \varepsilon)\| \leq \frac{\varepsilon M}{\sqrt{2}\delta_2^{2n+1}}(1 + Q_1) \exp(2\varepsilon N_0 Q \delta_1^{-2n})$$

where $N_0 := \|R^{(0)}(t)\|_0$. If $x^{(1)}(t, \varepsilon)$ stays in the domain D , then $f(t, x^{(1)})$ and $Y^{(1)}(t, x^{(1)})$ also

will be analytical by t and $x^{(1)}$. If $N_1 := \|R^{(1)}(t)\|$, $R^{(1)}(t)$ — is the pure conditionally- periodic part of matrix $P^{(1)}(t)$, then

$$\|y^{(1)}(t, \varepsilon)\| \leq \frac{\varepsilon M_2}{\sqrt{2}\delta_4^{2n+1}}(1 + Q_1) \exp(2\varepsilon N_1 Q \delta_3^{-2n})$$

For $\omega = (\omega_1, \dots, \omega_{n+1}, \omega_{n+2})$, $\omega_{n+i} := \sigma_i + \varepsilon(c_{11}^{(i)} + c_{22}^{(i)})$, $i = 1, 2$, satisfying the inequality (11), where $c_{lm}^{(1)}$ from the following expansion

$$P^{(1)}(t) = B^{(1)} + iC^{(1)} + R^{(1)}(t), \quad B^{(1)} := (b_{lm}^{(1)}), \quad C^{(1)} := (c_{lm}^{(1)}), \quad R^{(1)} := (r_{lm}^{(1)}), \quad l, m = 1, 2.$$

Similar estimation is obtained for $y^{(j)}(t, \varepsilon)$, $j \geq 2$.

Let in the region D the following conditions hold:

$$\|f'_x(t, x)\| \leq \frac{1}{2}L, \quad \|f''_{x^2}(t, x)\| \leq 2P, \quad f''_{x^2} := \left(\frac{\partial^2 f_l}{\partial x_m^2} \right), \quad m, l = 1, 2.$$

Then under the condition that all the approximations $x^{(j)}(t, \varepsilon)$ stay in the region $\|x\| \leq h$ we have

$$\|P^{(j)}(t)\| \leq L, \|Y^{(j)}(t, x^{(j-1)}, y^{(j-1)})\| \leq P \|y^{(j-1)}\|^2.$$

Therefore,

$$\|x^{(j)}(t, \varepsilon)\|_2 \leq \frac{\varepsilon Q_0}{\sqrt{2}\delta_2^{2n+1}} (1 + Q_1) \exp(2\varepsilon LQ\delta_1^{-2n}), \quad (12)$$

$$\|y^{(j)}(t, \varepsilon)\|_{2j+2} \leq \frac{\varepsilon P(1 + Q)}{\sqrt{2}\delta_2^{2n+1}} \exp(2\varepsilon LQ\delta_{2j+1}^{-2n}) \|y^{(j-1)}(t, \varepsilon)\|_{2j}^2, \quad j \geq 1, \quad (13)$$

where $Q_0 := \|f(t, 0)\|_0$.

3.3 Convergence of the sequence of approximation

Solution of the system (1) is the sum of the series

$$x(t, \varepsilon) = \sum_{j=1}^{\infty} y^{(j)}(t, \varepsilon). \quad (14)$$

Then using the inequalities (12), (13) we obtain that this series is majored by the series

$$m_0 + \sum_{j=1}^{\infty} m_j, \quad m_0 := \frac{\varepsilon Q_0}{\sqrt{2}\delta_2^{2n+1}} (1 + Q_1) \exp(\varepsilon Q^* \delta_1^{-2n}), \quad Q^* := 2LQ,$$

$$m_j := \frac{\varepsilon P}{\sqrt{2}\delta_{2j+2}^{2n+1}} (1 + Q_1) \exp(\varepsilon Q^* \delta_{2j+1}^{-2n}) m_{j-1}^2. \quad (15)$$

Consider the series

$$\sum_{j=0}^{\infty} \bar{m}_j = \sum_{j=0}^{\infty} \varepsilon_0 m_j, \quad \varepsilon_0 := \frac{\varepsilon P}{\sqrt{2}} (1 + Q_1). \quad (16)$$

instead of (15).

Choose $\delta_1 \prec 1$, $\delta_2 \prec 1$, so that

$$\delta_2^{r(2n+1)} \exp(-r\varepsilon Q^* \delta_1^{-2n}) =: \bar{m}_0,$$

$$2\delta_2^{r(2n+1)} \exp(-r\varepsilon Q^* \delta_1^{-2n}) = (\varepsilon(1 + Q_1))^2 P Q_0 \exp(\varepsilon Q^* \delta_1^{-2n}) \delta_2^{-(2n+1)}$$

$$2\delta_2^{(r+1)(2n+1)} = (\varepsilon(1+Q_1))^2 \exp((r+1)\varepsilon Q^* \delta_1^{-2n}).$$

Let $(2-\alpha)r = \alpha$, $1 \prec \alpha \prec 2$.

The other δ_j choose so that $\delta_{2j+2} = \delta_{2j}^\alpha$, $\delta_{2j+1} = \delta_{2j-1}\alpha^{-1} = \delta_1\alpha^{-j}$, $j = 1, 2, \dots$

Then the series (16) has the form $\sum_{j=0}^{\infty} \bar{m}_j = \sum_{j=0}^{\infty} \bar{m}_0^{\alpha^j}$. This series converges for $\bar{m}_0 \prec 1$. And

if $\bar{m}_0^{\alpha-1} \prec \frac{1}{2}$, then its sum S_1 satisfies the inequality $S_1 \leq 2\bar{m}_0$. The series $\sum_{j=0}^{\infty} \delta_j$ also converges due to the relationship between its members, its sum satisfies the inequality $2S_\delta \leq q$, if

$$2[\delta_2^{2n+1} \exp(-\varepsilon Q^* \delta_1^{-2n})]^{\alpha-1} \prec 1. \quad (17)$$

Thus we have derived the convergence conditions (15) and (16). Setting $2\alpha = 3$ for definiteness, we get $r = 3$. Then from (16) and $\bar{m}_0^{\alpha-1} \prec \frac{1}{2}$ it follows that

$$\varepsilon \prec \frac{1}{16\sqrt{PQ_0}(1+Q_1)} =: \varepsilon_1.$$

Besides from the inequality $2S_\delta \leq q$ we obtain

$$\varepsilon \prec \frac{q^2}{\sqrt{PQ_0}(1+Q_1)} =: \varepsilon_2.$$

The condition that the approximations $x^{(j)}(t, \varepsilon)$ stay in the region $\|x\| \leq h$ yields $S_2 \leq h$, here S_2 is a sum of the series dominating the series (14). By (15): $S_1 = \varepsilon_0 S_2$. Consequently, $\varepsilon_0 S_2 \leq 2\bar{m}_0$. Hence, under the assumption $2\bar{m}_0 \leq \varepsilon_0 h$ the approximations stay in the region $\|x\| \leq h$. This condition implies the inequality

$$\varepsilon \prec \frac{h^2\sqrt{P}}{Q_0^{3/2}(1+Q_1)} =: \varepsilon_3.$$

Thereby, with $\varepsilon \prec \min(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ the series (14) uniformly converges on the real axis t to the conditionally-periodic solution of the system (1).

4 Results and reasoning.

The existence of a conditionally periodic solution of system (1) is proved. For this purpose, a sequence has been constructed, rapidly converging to the conditionally periodic solution of system (1).

First, a linear inhomogeneous system is investigated in detail, which is a model equation for the terms of the sequence of approximations. The research results are used to build a

sequence of approximate solutions and in their evaluation. The conclusions of the work have scientific and practical value. Simultaneously with the proof of the existence of a solution, the construction of its construction is given. Unlike previous researchers, a critical case was considered and the method of accelerated convergence was applied.

5 Conclusion

The question of the existence and construction of a conditionally periodic solution is one of the most important in celestial mechanics. In problems of celestial mechanics, a conditionally periodic solution is a solution in which positional variables (semi-major axis, eccentricity, inclination, etc.) are expressed as conditionally-periodic functions, and angular variables (pericenter longitude, angle longitude, etc.) are expressed as a formula:

nt – conditionally periodic function,

where n – average angular variation for a given variable. Movement in an orbit corresponding to a conditionally periodic solution occurs in a limited region of space, and after a certain period of time, the celestial body returns to as close to any point within this region as you like. A characteristic feature of the series representing a conditionally periodic function is that their members are arranged according to increasing degrees of one or several small parameters, as in ordinary power series. However, the question of the convergence of these series in a strict mathematical sense has long remained open. Moreover, it was proved that a considerable number of these series diverges.

Thus, the assumption that conditionally periodic solutions in the problems of celestial mechanics exist and that the real motions of celestial bodies should be described by conventionally periodic functions required proof. G.A. Merman's paper first obtained a rigorous proof of the existence of a conditionally periodic solution in the plane bounded three-body problem.

V.I. Arnol'd developed a method for proving the existence and constructing conditionally periodic solutions of Hamiltonian systems of differential equations of a satisfied general form [3-4]. This method of A.N. Kosmogorov-V.I. Arnold was disclosed in terms of their practical application by Yu.A.Rabov and E.G. Grebennikov for the same Hamiltonian systems.

In recent years, A.Bari, H. Brezis, Feireist E., Pelyukh H.P., Syvak Q.A. have become involved in the problems described above or close to them [13-15]. In some international conferences, the author of this article made reports of particular details of the formulation of the question and the election of the method [16-17], theses of which were published.

References

- [1] Kolmogorov A.N., "On the conservation of conditionally periodic motions at small change in Hamilton function."Reports of the USSR Academy of Sciences, 98(4), (1954): 527-530.
- [2] Siegel C.L., "Vorlesungen über Himmelsmechanik,"Berlin:Gottingen: Heidelberg: Springer-Verlag, 1956.
- [3] Arnold V.I., "Small denominators. Proof of Theorem A.N. Kolmogorov on the conservation of conditionally-periodic motions with a small change in the Hamiltonian function,"SMS, 18(5), 1963:13-39.
- [4] Arnold V.I., "Small denominators and the problem of stability in classical and celestial mechanics,"SMS, 18(6), (1963): 92-191.
- [5] Moser J., "Convergent series expansion for quasi-periodic motions,"Mathem Ann, 169, (1967):136-176.

- [6] Moser J., "A rapidly convergent iteration method and nonlinear partial differential equations,"Ann, Scuola Nozm Super, de Pisa, ser111 20(2), (1966): 65-315.
- [7] Moser J., "Regularization of Kepler's problem and averaging method on a manifold,"Communs Pure and Appl. Math. 23, (1970): 609-636.
- [8] Jeffreys W., Moser J., "Quasi-periodic solutions three-body problem,"Actron.J., 7, (1966):568-578.
- [9] Merman G.A., "New class of periodic solutions in Hill's restricted problem,"Works of the Institute of Theoretical Astr. USSR Academy of Sciences 8, 5.
- [10] Lika L.K., Ryabov Yu.A., "Construction of conditionally periodic solutions of canonical systems of differential equations,"Izvestiya Academy of Sciences of the Moldavian Soviet Socialist Republic. Series of physical and technical and mathematical sciences, 2, (1971):204-311.
- [11] Grebennikov E.A., Ryabov Yu.A., "Resonances and small denominators in celestial mechanics,"M.Nauka, 1978. -pp. 126.
- [12] Bogolyubov N.N. Mitropolsky Yu.A., Samoylenko A.M., "Accelerated Method of Convergence in Nonlinear Mechanics,"Kiev: "Naukova dumka 1969.
- [13] Bari A., Brezis H., "Periodic solutions of nonlinear wave equations,"Comm. Pure Aple Math. 31(1), (1978):1-30.
- [14] Feireist E., "On the existence of period o Solutions of a semilinear wave equation with a super linear forcing term,"Chechosl. Math. J., 38(1), (1988): 78-87.
- [15] Pelyukh H.P., Syvak O.A., "Periodic Solutions of the Systems of Nonlinear Functional Equations,"Journal of Math. Sciences, 1, (2014): 92-95.
- [16] Suleimenov "On the existence and stability of quasi-periodic solutions of a quasilinear system of differential equations,"Actual problems of mathematics and mathematical modeling are dedicated to the 50th anniversary of the creation of the Institute of Mathematics and Mechanics Academy of Sciences of the Kazakh SSR Almaty June 1-5, 2015.
- [17] Suleimenov Zh., "The method of accelerated convergence for constructing conditional-periodical solutions,"Third International Conference on Analysis and Applied Mathematics ICAAM 2016.

2-бөлім	Раздел 2	Section 2
Информатика	Информатика	Computer Science
IRSTI 20.23.17		

Dynamic request distribution for enhanced Quality of Service

Joldasbayev S.K., Al-Farabi Kazakh National University,
Almaty, Kazakhstan, E-mail: serykjoldasbaev@mail.ru

Balakayeva G.T., Al-Farabi Kazakh National University,
Almaty, Kazakhstan, E-mail: gulnardtsa@gmail.com

Aidarov K.A., Al-Farabi Kazakh National University,
Almaty, Kazakhstan, +77013460160, E-mail: kanataidarov@yahoo.com

Chris Phillips, Newcastle University, Newcastle, Great Britain,
0191 208 7972, E-mail: chris.phillips@ncl.ac.uk

The emergence of Web-services as an open components supporting flexible and inexpensive set of distributed applications as well as exploiting them as a promising solution for integration with other application and providers of software-hardware resources is very much in demand. Using Web-services simplifies and enhances system functionality because of the availability of programs interaction with each other through Internet using open protocols. Thereby it is necessary to provide QoS provision issues such as distribution of request streams, enhanced efficiency of operations at peak loads. One of the ways to tackle these issues is to apply dynamic request distribution which ensures continuity of data transfer and processing ignoring crashes and equipment failures, redistribution of data flow to the available nodes in case of their loss. Abovementioned can be achieved using certain heuristics based on methods of Mathematical statistics and Probability modeling. Moreover, we provide analysis of methods and models developed for QoS provision.

Key words: Quality of Service, QoS, dynamic balancing, resource allocation, load balancing, queuing.

Қызмет көрсету тиімділігін арттыру үшін сұраныстарды үlestірудің динамикалық моделі

Жолдасбаев С.К., әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ.,
Казақстан Республикасы, E-mail: serykjoldasbaev@mail.ru

Балақаева Г.Т., әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ.,
Казақстан Республикасы, E-mail: gulnardtsa@gmail.com

Айдаров К.А., әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ.,
Казақстан Республикасы, E-mail: kanataidarov@yahoo.com

Chris Phillips, Newcastle University,

Newcastle, Great Britain, E-mail: chris.phillips@ncl.ac.uk

Веб-сервистердің пайда болғаны икемді және қымбат емес кеңейтілген қосымшаларды қолдайтын ашық компоненттер, сондай-ақ оларды басқа қолданбалармен және бағдарламалық және аппараттық ресурстармен жабдықтаушылармен интеграциялау үшін перспективалық шешім ретінде пайдалану үлкен сұранысқа ие. Веб-қызметтерді пайдалану ашық хаттамаларды қолданып Интернет арқылы бір-бірімен бағдарламалық өзара әрекеттесудің болуына байланысты жүйенің жұмысын женілдетеді және жетілдіреді. Осылайша, QoS қамтамасыз ету мәселелерін шешу қажет, мысалы, сұраныс ағындарын бөлу, ең жоғары жүктемелерде операциялардың тиімділігін арттыру. Бұл мәселелерді шешу жолдарының бірі деректерді беру және өндөудің үзіліссіздігін қамтамасыз ететін динамикалық сұранысты таратуды, сәтсіздіктер мен жабдықардың істен шығуын елемеуді, жоғалған жағдайда қол жетімді түйіндерге деректер ағынның қайта бөлінуін қамтамасыз ету болып табылады. Жоғарыда көрсетілген математикалық статистика мен ықтималдық модельдеу әдістеріне негізделген белгілі әвристиканы қолдана отырып қол жеткізуге болады.

Бұдан басқа, біз QoS-ті қамтамасыз ету үшін әзірленген әдістер мен модельдерді талдауды үсінамыз.

Түйін сөздер: Қызымет көрсету сапасы, QoS, динамикалық тендеңстіру, ресурстарды үлестіру, жүктемені теңестіру, тапсырыс.

Динамическое распределение запросов для повышения качества обслуживания

Джолдасбаев С.К., Казахский национальный университет имени аль-Фараби,

Алматы, Республика Казахстан, E-mail: serykjoldasbaev@mail.ru

Балакаева Г.Т., Казахский национальный университет имени аль-Фараби,

Алматы, Республика Казахстан, E-mail: gulnardtsa@gmail.com

Айдаров К.А., Казахский национальный университет имени аль-Фараби,

Алматы, Республика Казахстан, E-mail: kanataidarov@yahoo.com

Крис Филлипс, Университет Нюкасл, Нюкасл, Великобритания,

E-mail: chris.phillips@ncl.ac.uk

Появление Web-сервисов как открытых компонентов, поддерживающих гибкий и недорогой набор распределенных приложений, а также их использование в качестве перспективного решения для интеграции с другими приложениями и поставщиками программно-аппаратных ресурсов, очень востребовано. Использование Web-сервисов упрощает и улучшает функциональность системы из-за доступности взаимодействия программ друг с другом через Интернет с использованием открытых протоколов. Таким образом, необходимо решать вопросы обеспечения QoS, такие как распределение потоков запросов, повышение эффективности операций при пиковых нагрузках. Одним из способов решения этих проблем является применение динамического распределения запросов, которое обеспечивает непрерывность передачи и обработки данных, игнорируя сбои и выхода из строя оборудования, перераспределение потока данных к доступным узлам в случае их потери. Вышеупомянутые могут быть достигнуты с использованием определенных эвристик на основе методов математической статистики и моделирования вероятностей. Кроме того, мы предлагаем анализ методов и моделей, разработанных для обеспечения QoS.

Ключевые слова: Качество обслуживания, QoS, динамическая балансировка, распределение ресурсов, балансировка нагрузки, очередность.

1 Introduction

In modern telecommunication networks, the character and volume of traffic transmitted and, accordingly, the provided info communication services are constantly changing. The increasing dissemination of web services, whether as an internal organizational infrastructure, or as part of paid services provided by the organization, issues related to service quality become very relevant and will ultimately be a significant factor in the success of service providers. For example, when customers contact you through the customer service centre's of services, they usually interact with automated support systems and expect an immediate answer. However, if the issue requires human intervention, it is likely that consumers will hold on the line, waiting for a response operator. Similarly, when the stock market is unstable and a large number of brokers and traders is online, as a rule, the trading platform can become overloaded, inhibiting or causing significant difficulties in the operation of such systems. The inability to buy and sell in a timely manner can cause serious financial losses.

2 Review of literature

To date, a large number of well-known scientists, for example, Kleinrock, S. Blake, D. Grossman, Z. Wang, Steklov VK, Berkman LN, as well as research centers such as Mobile Ad-hoc

Networks, Internet Engineering Task Force, Center for Embedded Networked Sensing deals with the management and distribution of traffic. However, despite the huge number of publications and the efforts of manufacturers, the task of constructing traffic patterns that best reflects its functioning in real conditions is still not solved [20-25].

Specifically, because reinforcement learning offers the potential to develop optimal allocation policies without explicit model knowledge by learning from the consequences of each action, existing works on ML algorithms mainly focus on reinforcement learning [11], [12]. They require neither an explicit system model nor an explicit traffic model to learn.

RL refers to a learning process, where a learning agent can learn to make appropriate decisions through interactions with an external environment [13]. Specifically, beyond the learning agent and the environment, a reinforcement learning system consists of a policy, a reward function and a value function. Let S be the set of environment states and A be the set of actions, respectively.

Another popular machine learning algorithm is the support vector machine (SVM). It has been widely applied for different areas such as pattern recognition, classification and data mining. However, SVMs are not preferred in on-line applications since the training and testing complexity of standard SVM are $O(nm + m^3)$ and (m) respectively, where n is the data size and m denotes the number of support vectors. On the other hand, some approximated methods have been proposed to reduce the complexity [14]. For example, [15] reduces the complexity to $O(nd \max)$, where $d \max$ is the number of basis functions elected.

A few of works on machine learning algorithms have been proposed for the resource management problem [16-19]. For admission control, [8] derived a complex rule set that can be used to identify the optimal configuration for unobserved workload based on machine learning algorithms. [19] applied RL to configure parameters automatically in multi-tier Web systems, where eight parameters at web tier and application tier are selected to consist of the state space. For each parameter, there are three possible actions: increase, decrease and keep. The policy is based on the e-greedy method. In order to suppress the poor performance due to bad initialization, they proposed an algorithm to construct different initialization policies for different scenarios. For VM scaling, [16] proposed an iterative model training technique based on artificial neural network (ANN) to predict computing resource demand in virtual environments. [17] applied RL to train nonlinear approximators (e.g., multi-layer perceptrons) instead of the lookup table for VM horizontal scaling, where the state is defined as the request arrival rate and the action is to determine the number of servers allocated. Since the state space grows exponentially with the number of parameters in practice, the authors applied a nonlinear function approximator as an external policy to avoid poor performance that would be expected during online learning.

Recently, a few of works on fuzzy control for the resource management have been proposed in [10-12]. In [10], the admission control is conducted by fuzzy control in order to manage the QoS, where the turning parameter Maxclients in each interval is controlled by the fuzzy controller. For VM scaling, [11] attempted to capture the non-linear behaviors in VM resource usages by designing a fuzzy model estimator. The approach is divided into two steps. First, a fuzzy logic based modeling method is used to learn the system behaviors without requiring any priori knowledge. Then a predictive controller predicts the resource demand of all VMs and takes actions based on this model. [12] proposed a neural fuzzy controller for percentile-based end-to-end delay guarantee through a virtualized multi-tier server cluster, where Gaussian

membership functions are first used to fuzzify the average service time, s_i , and the variance of service time, σ_{2i} , distribution of requests at tier i , respectively. Then a fuzzy neural network is applied for online learning at the Inference stage. In addition, an output scaling factor is introduced to further enhance the performance. It is model-independent and capable of adapting control parameters through fast online learning. Compared with other supervised machine learning techniques, it does not require off-line training.

The urgency of the work on the material implementation of QoS in the Internet is a time-consuming task, but its solution will certainly affect the balancing of the load as a whole.

The quality of traffic management and distribution is characterized by such a notion as Quality of Service (QoS), which is the ability of a computer network to provide the required level of service to a given traffic within a certain technological framework [1].

The necessary service is characterized by many parameters, the main ones being the following:

- Bandwidth (bandwidth);
- priority;
- Delay of data transmission (delay);
- delay variation – jitter;
- loss of packets in the transmission of network data (packet losses).

In order to determine which simulation is most suitable and effective for managing and distributing traffic, typical problems that arise with bit transfer during periods of network congestion should be considered.

Developments in this area contribute to the effective operation of the IT service delivery system. Today there are huge resources for data processing, however, its effective application will contribute to the correct allocation of time and memory of the system, which are important factors in modern conditions. In addition, the information in the servers of the system, as well as requests for them, are steadily growing, which gives the demand for research in this direction relevant [2].

QoS (quality of service) – the ability of a network to provide the necessary service to a given traffic within a certain technological framework. "In this technical sense, this term means a set of methods for managing packet network resources. The quality of the connection is determined by four parameters:

- 1) Bitrate, describes the nominal bandwidth of the media. Depends on the bandwidth of the communication channel (Hz) and the signal-to-noise ratio;
- 2) Delay in the transmission of a packet (Delay), measured in milliseconds;
- 3) Fluctuation (jitter) of packet delay;
- 4) Packet loss. Determines the number of packets lost on the network during transmission [3].

To date, in modern telecommunications, one of the most urgent tasks is the transfer of traffic in compliance with a number of requirements for quality of service. Inefficient use of network resources, a large number of subscribers, as well as stringent requirements to the quality of service parameters may be the reason for the decline in the quality of services provided in multiservice data networks. Service quality technologies enable you to meet the workload or application maintenance requirements by measuring network bandwidth, monitoring network state changes (eg, used or available bandwidth), and prioritizing (or regulating) network

traffic [4]. Quality of service QoS refers to non-functional properties of interaction with Web services. These properties determine the degree of satisfaction in the query. QoS on customer terms can be grouped into three categories: performance, security and availability. These categories contain a set of various non-functional properties and metrics that determine the non-functional properties of the web service in quantitative forms in a certain order, linking it to unambiguous semantics. When measuring metrics, you should take into account the measurement indicators, facilities, organizations and the point in the network under which measurements are to be taken.

If several queue control algorithms are planned for wireless network design, for example, FIFO algorithm, priority maintenance algorithms, weighted queue algorithm, then in this case it is suggested to use the following order for comparison.

1. Calculation of the generation period

$$\tau = \frac{1}{f},$$

where τ is the generation period; f is the network frequency.

2. Determine the network speed

$$b = \frac{1}{m},$$

where b is the average service time; m is the data rate.

3. Bandwidth of channels

$$C = \sum_{k=1}^K \frac{1}{t_k}$$

where C is the throughput; t is the delay time; k is the number of channels.

Comparative analysis using the presented formulas will allow to judge, firstly, what equipment is necessary in this situation when building a computer network, which minimizes the delay in information transfer, and also effectively manages and distributes traffic. Secondly, what algorithm is better to use in that or other equipment.

In order to improve the quality of service, modern systems use intelligent service delivery systems that contribute to the improvement of the service delivery system.

For example, in [5-6] some solutions to the problems of availability and fault tolerance with the use of Apache Zoo Keeper technologies are proposed. The authors of this paper to solve the problem of distribution of flows, especially at high workloads, mainly, the transfer of some of the main loads to other servers (followers). of distribution of flows, especially at high workloads, mainly, the transfer of some of the main loads to other servers (followers). The proposed method is aimed at scenarios in which the Zab protocol is potentially a bottleneck in performance, this can happen for some reasons: (1) the intensity of work with intensive recording; (2) the number of clients is quite large; or (3) a large number of ensemble N , $N > 3$. In other words, the method reduces the traffic of messages, both incoming and outgoing, from the leader. This is achieved by changing the behavior of followers in two simple but important ways, while at the same time supporting the structure of Zab itself, which is well understood and implanting the structure. Thus, the costs associated with reaching an agreement were reduced, lower overheads and increased productivity.

3 Materials and methods

The urgency of the work on the material implementation of QoS in the Internet is a time-consuming task, but its solution will certainly affect the balancing of the load as a whole, both globally.

To describe the QoS work in the physical representation, that is, the definition of what it is, let's go down to the hardware implementation. It will be appropriate to determine that in the work we do not touch this level (the hardware part) and no change in its architecture or the like. solutions is not a task. At the same time, without describing the hardware structure, it will be difficult to state the task at the middleware & software level.

Implementation of dynamic QoS solutions for servers of large-scale clusters exists relatively long ago. But the use of the Internet of things in different approaches can bear other promising solutions. The use of fairly simple elements of the Internet things from improvised tools in turn can have a grandiose significance for QoS in general.

There are three models of implementing QoS: the best possible; integral and differentiated. In the hardware, we can define this as the following:

The best possible type of service is realized on the network, when everything is done to deliver the package, but nothing is guaranteed (for example, FTP or HTTP). The integrated type of services (RFC-1633, 1994) was developed first and implemented by point-to-point redundancy (RSVP protocol, 1997, section 4.4.9.6). RSVP provides a signaling mechanism for configuring remote routers to obtain the desired QoS [3], [8]. The protocol is oriented to work with three types of traffic: best efforts (normal transmission of IP-data without connection), sensitive to the transmission speed and sensitive to delays. Traffic sensitive to download requires the formation of a channel with guaranteed bandwidth. The application is forced to put up with certain delivery delays (a class of services with a guaranteed bit rate in seconds).

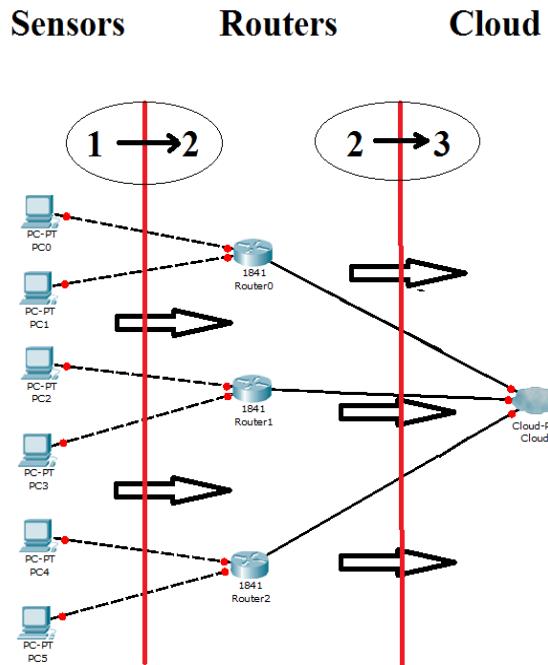


Figure 1: The general picture of network connection for dynamic model

3.1 Mathematical model

The third type of traffic (sensitive to delays) guarantees a minimum delay and a low variance in the delivery time. The capacity can vary. An example of this type of traffic can serve as voice or video. RSVP defines two types of services for this type of traffic: a service with controlled delays and a predictable service (for real-time applications (video conferences and telephone conversations) [6].

It is known that the transmission of information through Internet services occurs in a portioned, in the form of a session, containing many individual requests. These requests are expressed as a queue of atomic units that have a finite number of elements, called threads. Each request in the stream is a package, each of which has an application and a limited number of m. Therefore, one packet inside the stream at a certain point in time will be processed in only one of the limited sets of available N processors [4].

Streams can be m different types with different performance requirements. a stream of a i -type contains $k(i)$ applications sent at the frequency of $l(i)$ applications per second. These values will be needed later in the description of various QoS heuristics.

3.2 Development of algorithm of performance evaluation

Suppose that for each i -th stream, K_i - the total number of packets organized in the FIFO queue. The flow has important characteristic, λ_i is the rate of arrival of the i -th stream. Each processor has its own characteristic b_i - the average processing time, which means the average time of finding the packets in the processing processor. And each from this processor can process only one packet at a time. Each packet in the stream has a characteristic L_j^i - delay j -th packet in i -th stream. Delay means the total time that a package has been in the system, from the moment it hits the queue, until it leaves the processor after processing. On the basis of this we introduce the following variable: W_j^i is the waiting time of the packet and we define it as

$$W_j^i = L_j^i - b_i$$

where $j = 1, 2, \dots, K_i$ and i stream index. We introduce one more variable Q_i - the average waiting time for the i -th stream and express it as follows:

$$Q_i = \sum_{j=1}^{K_i} W_j^i$$

The application vendor leases server resources from the cloud and provides user work execution services, which is expressed in the form of a Service Level Agreement (SLA). This agreement contains a list of quality parameters for servicing methods and means of their control, the time of the supplier's response to the consumer's request, as well as penalties for violations of this agreement. It consists of three components:

1. A description of the quality of service guarantee that the cloud platform provides to the user;
2. A revenue-generating scheme that is used by the platform for charging for the services provided;
3. Penalties for violations of the agreed performance guarantee under the contract.

SLA in our case is defined as follows:

$$Q_i \leq \sigma_i, i = 1, 2, \dots, M$$

Here M is the total number of streams accepted for maintenance, σ_i is the profit received for completing the maintenance of the i -th stream (processing of all packets of this stream). P_i is penalties for violations of the SLA.

We define a new set of variables:

$$\text{profit}_i = \sigma_i - A_i \times \text{cost} \times \text{duration}_i$$

Which A_i - number of dedicated processors for the i -th stream, cost - a fixed number of units per unit time, duration_i - amount of time spent processing.

To determine system performance, we use the utility function. R - the average profit received by the service provider per unit time is the main metric considered in this work.

$$R = \sum_{i=1}^m a_i [c_i - r_i P(W_i > q_i)]$$

where $a(i)$ is the average number of i -type flows received per system per unit time, $P(W_i > q_i)$ is the probability of exceeding the observed waiting time for streams commitment q_i .

Heuristics with the same flows can be optimized by applying one or more steps to improve the heuristic, similar to another. The ultimate goal of resource allocation and heuristics of admission control is to maximize this value (R).

4 Results and discussion

To improve the efficiency of providing services at the software and middleware levels, it is proposed to use the dynamic model of an intelligent system at high loads without affecting the hardware of the system as a whole. The method of solution and the mathematical model of regulation of load balancing are presented in the article according to the statement of the problem.

The results of this work at this stage can be defined as the proposed effective method of solution, as well as the correct wording of the problem itself.

An algorithm for the functioning of a new model for carrying out extensive computational experiments is being developed.

In the future, emphasis will be placed on numerical solutions and simplifying the problem to the required level.

5 Conclusion

In a dynamic system, instead of the limited resources N , a non-constant number of processors with the total amount limitation will be taken into account: $I \in N$.

In this connection we use a dynamic solution to the problem of distribution of queries to solve the problems of distribution of flows, in order to increase the efficiency of operations, especially at high loads, which will ensure the continuity of data transmission and processing, ignoring equipment breakdowns and outages, distributing data on accessible nodes, in case of their loss.

References

- [1] Hyunyoung Kil, Reeseo Cha, Wonhong Nam. Transaction history-based web service composition for uncertain QoS. *International journal of web and grid services*, vol. 12 (2016): 42.
- [2] Balakayeva G.T., Aidarov K.A. Research of algorithms and methods of load balancing and construction of models for queuing networks. *Proceedings of the International Conference on Computational and Applied Mathematics "VPM'17" in the framework of the Marchuk Scientific Readings, Novosibirsk, June 25 – July 14 [Electron. resource]*, (2017): 17–21.
- [3] Goldstein B.S., Marshak MA, Mishin E.D., Sokolov N.A., Tum A.V. "Indicators of the functioning of the multiservice communication network of public use". *Journal of Communication Engineering*, no. 3–4 (2009): 17.
- [4] Balakayeva G., Aidarov K., Simulation of load balancing algorithms based on queuing networks. *Abstracts of VI Congress of the Turkic World Mathematical Society (TWMS 2017), Astana*, (2017): 313.
- [5] EL-Sanosi I. and Ezhilchelvan, P. Improving zookeeper atomic broadcast performance by coin tossing. In *European Workshop on Performance Engineering, Springer*, (2017): 249–265.
- [6] Flannagen E. Michael. Syngress (2001) "Administering CISCO QoS in IP-Networks", *Syngress Media, ISBN 1928994210, 9781928994213*, (2001): 519.
- [7] Goldstein B.S., Marshak M.A., Mishin E.D., Sokolov N.A., Tum A.V. Kontrol pokazatelei kachestva obsluzhivanya s uchotom perekhoda k seti svyazi sleduyushego pokoleniya [Control of quality of service indicators, taking into account the transition to a next-generation communication network]. *Tekhnika Svyazi*, no 1 (2009).
- [8] Andrzejak A., Arlitt M., Roila J. Bounding the Resource Savings of Utility Computing Models. *Technical Report HPL-2002, Internet Systems and Storage Laboratory, HP Laboratories*, (December 2002): 339.
- [9] Kharchenko V, Illiashenko O, Boyarchuk A, Sklyar V, Phillips C Emerging curriculum for industry and human applications in Internet of Things. In: *9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) Bucharest, Romania: Institute of Electrical and Electronics Engineers Inc.*, (2017): 918–922.
- [10] Chase, J. S., Anderson, D. C., Thakar, P. N., Vahdat, A. M., Doyle, R. P. Managing energy and server resources in hosting centers. *ACM SIGOPS Operating Systems Review*, no 35(5), (2001): 103.
- [11] Lee, Y. C., Zomaya, A. Y. . Energy efficient utilization of resources in cloud computing systems. *The Journal of Supercomputing*, 60(2), (2010): 268–280.
- [12] Enokido, T., Aikebaier, A., Takizawa, M. A Model for Reducing Power Consumption in Peer-to-Peer Systems. *IEEE Systems Journal*, 4(2),(2010): 221–229.
- [13] Liu, S., Ren, S., Quan, G., Zhao, M., Ren, S. Profit Aware Load Balancing for Distributed Cloud Data Centers. *2013 IEEE 27th International Symposium on Parallel and Distributed Processing*, (2013): 611–622.
- [14] Vakilinia, S., Heidarpour, B., Cheriet, M. Energy Efficient Resource Allocation in Cloud Computing Environments. *IEEE Access*, 4, (2016), 8544–8557.
- [15] Zhang, W., Zhang, Z., Chao, H.-C. . Cooperative Fog Computing for Dealing with Big Data in the Internet of Vehicles: Architecture and Hierarchical Resource Management. *IEEE Communications Magazine*, 55(12),(2017): 60–67.
- [16] Nagpure, M. B., Dahiwale, P., Marbate, P. An efficient dynamic resource allocation strategy for VM environment in cloud. *2015 International Conference on Pervasive Computing (ICPC)* (2015).
- [17] Mukherjee, M., Shu, L., Wang, D. Survey of Fog Computing: Fundamental, Network Applications, and Research Challenges. *IEEE Communications Surveys and Tutorials*, (2018): 1–1.
- [18] Bodenstein, C., Schryen, G., Neumann, D. Energy-aware workload management models for operation cost reduction in data centers. *European Journal of Operational Research*, 222(1), (2012): 157–167.
- [19] Mohammad Ali, H. M., El-Gorashi, T. E. H., Lawey, A. Q., Elmirghani, J. M. H. Future Energy Efficient Data Centers With Disaggregated Servers. *Journal of Lightwave Technology*, 35(24), (2017): 5361–5380.
- [20] Hameed, A., Khoshkbarforoushha, A., Ranjan, R., Jayaraman, P. P., Kolodziej, J., Balaji, P., ... Zomaya, A. A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. *Computing*, 98(7), (2014): 751–774.

- [21] Ge, Y., Zhang, Y., Qiu, Q., Lu, Y.-H. A game theoretic resource allocation for overall energy minimization in mobile cloud computing system. *Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design - ISLPED '12.*, (2012): 279-284.
- [22] Kliazovich, D., Arzo, S. T., Granelli, F., Bouvry, P., Khan, S. U. e-STAB: Energy-Efficient Scheduling for Cloud Computing Applications with Traffic Load Balancing. *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing,* (2013):7-13.
- [23] Aikebaier, A., Yang, Y., Enokido, T., Takizawa, M. Energy-Efficient Computation Models for Distributed Systems. *2009 International Conference on Network-Based Information Systems* (2009): 424-431.
- [24] Sharma, B., Wood, T., Das, C. R. HybridMR: A Hierarchical MapReduce Scheduler for Hybrid Data Centers. *2013 IEEE 33rd International Conference on Distributed Computing Systems* (2013): 102-111.
- [25] Gao, Y., Guan, H., Qi, Z., Song, T., Huan, F., Liu, L. Service level agreement based energy-efficient resource management in cloud data centers. *Computers and Electrical Engineering*, 40(5), (2014): 1621–1633.

Optical character recognition with neural networks

A. Shalakhmetov, University of International Business,
Almaty, Republic of Kazakhstan, e-mail: aidar.shalakhmetov@gmail.com

S. Aubakirov, University of International Business,
Almaty, Republic of Kazakhstan, e-mail: aubakirov.sanzhar@gmail.com

XXI century is the age of global automation and digitization. There is high demand for optical recognition software, including character recognition. There are different approaches in solution optical recognition problem. Some of them based on classical feature extraction methods. Other based on machine learning algorithms. In this work, we observed related works in machine learning field and propose the plan for further research. The work relies on two research studies that describe basics and fundamentals of machine learning. These research include various experiments in this field. We conducted several experiments to get acquainted with methods and techniques and to identify key features that are affecting on optical character recognition process. We analyzed two main architectural structures: multilayer perceptron and convolutional neural network. In conclusion, we identified key points of machine learning techniques and composed our own strategy for further researches. The result shows the difference in performance of different convolutional neural network models under the same conditions. Further work will cover researches and experiments on performance of several architectures. In addition, we observed latest tools, software programs and environments for the most convenient way to organize implementation process.

Keywords: OCR, neural network, convolutional neural networks.

Оптическое распознавание символов с помощью нейронных сетей

Шалахметов А.З., Международный Университет Бизнеса,
г. Алматы, Республика Казахстан, e-mail: aidar.shalakhmetov@gmail.com
Аубакиров С.С., Международный Университет Бизнеса,
г. Алматы, Республика Казахстан, e-mail: aubakirov.sanzhar@gmail.com

XXI век – это век глобальной автоматизации и оцифровки данных. В наше время имеется огромный спрос на системы оптического распознавания, включая системы распознавания символов. В сфере оптического распознавания используются различные подходы в решении поставленных задач. Некоторые из них основываются на классических методах выделения характерных признаков. Некоторые базируются на алгоритмах машинного обучения. В данной работе рассматриваются исследования в сфере машинного обучения и предложения для последующих исследований. Данная статья основывается на двух публикациях, которые описывают основы машинного обучения. Мы поставили несколько аналогичных экспериментов для ознакомления с методами и техниками данного подхода, а также для определения основных принципов, которые влияют на процесс оптического распознавания. Мы проанализировали две основные архитектуры: многослойный перцептрон и сверточные нейронные сети. В заключении, мы ознакомились с основами алгоритмов машинного обучения и составили стратегию для дальнейших исследований. Полученный результат отражает разницу в производительности между разными моделями сверточных нейронных сетей при одинаковых условиях. Последующие работы будут содержать исследования и эксперименты различных архитектур. В дополнении, мы рассмотрели различные утилиты, программное обеспечение и среды для создания оптимального процесса реализации системы.

Ключевые слова: OCR, нейронные сети, сверточные нейронные сети.

Символдарды нейрон желілерінің көмегімен оптикалық тану

Шалахметов А.З., Халықаралық бизнес университеті,
Алматы қ., Қазақстан, e-mail: aidar.shalakhmetov@gmail.com
Аубакиров С.С., Халықаралық бизнес университеті,
Алматы қ., Қазақстан, e-mail: aubakirov.sanzhar@gmail.com

Жиырма бірінші ғасыр – мәліметтерді жахандық автоматизация мен цифралау ғасыры. қазіргі уақытта оптикалық тану жүйелеріне, оның ішінде символдарды тану жүйелері үлкен сұранысқа ие. Оптикалық тану саласында қойылған мақсаттарға жету үшін түрлі амалдарды қолданады. Ол амалдардың кейбірі классикалық ерекше болу әдісінде негізделген. Басқалары машиналық үйрету алгоритмдарында негізделеді. Бұл әдісте машиналық үйрету саласының зерттемелері қарастырылады да, келешектегі зерттеу жұмыстарына ұсыныстар қарастырылады. Берілген мақала, машиналық үйретудің негізін сипаттайтын екі басылымға қараған негізделген. Біз берілген әдістің амалы және техникасымен танысу үшін көрсетілген тәжірибелердің қайта орындаудың Сонымен қатар оптикалық тануға әсерін тигизетін негізгі принциптарды анықтамақ болдық. Біз басты екі архитектураны талдадық; көп қатпарлы перцептон мен үйірткілі нейрон желілері. қортындысында біз машиналық үйрету алгоритмдерінің негізімен танысып, келешектегі зерттеулер стратегиясын құрдық. Келесі жұмыстарымызда зерттемелер мен түрлі архитектуралардың тәжірибелерін табуга болады. Нәтиже бірдей шарттарда конвективті нейрондық желілердің әртүрлі модельдері арасындағы әнімділіктің айырмашылығын көрсетеді. Оған қоса біз түрлі утилиталар, бағдарламалық жабдықтау мен жүйені жүзеге асыру процесін жасайтын салаларды қарастырық. Түйінді сөздер: символдарды оптикалық тану, нейрондық желілер, үйірткілі нейрон желілері.

Түйін сөздер: OCR, нейрондық желілер, конвективті нейрондық желілер

1 Introduction

Nowadays there is growing demand for the software system to recognize characters in a computer system when information is scanned through paper documents. This paper presents proposal work in the field of Optical Character Recognition that describes chosen methods of character recognition. In our days technological development growth increasing exponentially in many fields. Machine learning is one of the rapidly increasing and most interesting fields. An approach, that provides machine learning, can be applied to solution of almost any problem. Machine learning and deep learning[1] techniques allows to process a large amount of unstructured data, perform tasks much faster, unlike any person who might spend dozens of man-hours. Main advantage of these techniques is the ability to generalize and abstract data[2]. First prototype of neural network was made in 1943[3], which was conceived as a perception element. With the appearance of high definition cameras, big amount of different data in different areas, and computation capacities, development of machine learning is increasing faster and faster. Nowadays there are large amount of variations of neural networks as well as areas of usage[4]. The architecture and type of neural networks, as well as the approach, is selected depending on the main problem. Neural networks can be divided by their purposes on these categories namely: image recognitions, decision-making, clustering, prediction making, approximation, associative memory, data analysis and optimization[5]. Neural networks can have various shapes and sizes. Usually these shapes, sizes, and architecture in general, are chosen according to the problem. For instance, convolutional neural network greatly deals with the problem of object recognition due to saving the spatial structure of data. Neural networks have ability to compute almost any function. No matter what the function, it is guaranteed to be a neural network so that for every possible input value, there is some close approximation of output value from the network due to universal approximation theorem. Universal approximation theorem states that a single hidden layer feed-forward neural network can approximate continuous functions on compact subsets of real numbers. Neural networks take their basis from biology. The whole mechanism is copied from the per-

formance of a brain. Similarly, as the signals flow between neurons by synapses, the data flows between mathematical elements, also called neurons. First try of creation of mathematical model, which simulates brain processes, was made by Warren McCulloch and Walter Pitts in 1943. Since then neural networks become more complex and can perform various tasks. Most of the problems, which can be solved by neural networks falls under recognition, prediction, classification, approximation and data analysis problems[6]. For different problem there are different types and architectures of the neural network - with supervised and unsupervised training, fixed and dynamic weights, analog and digital input data, and with different propagation algorithms[7]. Basic neural networks are the mathematical models, which is built to compute mathematical operations in different variations. Therefore, neural network contains several main elements - nodes, which are called neurons, weights, activation functions. Neural networks usually are built step-by-step by grouping all elements in so called layers. Weights are represented as links between each neuron of each layer. Each weight has its own value, which represents “relation” between neurons in the form of certain value. This value is passed to the target neuron in activation phase. Activation function is a mathematical function, which takes sum of all weights as an arguments and, by applying certain function, computes the output. So neuron gathers all described elements in itself. Basically neuron decides, will it pass processed input data further or not. This approach is successfully introducing to various fields. One of the fields of interest is computer vision[8], [9] and artificial perception of the handwritten text. Using neural networks in optical character recognition and text recognition systems can make big contribution in data digitization.

2 Related works

The book, on which this work is partially based, is written by Michael Nielsen and called “Neural network and Deep learning”[10]. This online book describes basic principles of neural networks and its implementation on Python. Convolutional neural networks were described in paper entitled “Gradient-Based Learning Applied to Document Recognition”[11] by Y. LeCun, L. Bottou, Y. Bengio and P. Haffner. The paper describes the process they used to achieve up to a 99.1% accuracy on the MNIST dataset, using both a 3-layer convolutional network and a 5-layer network that failed to outperform the former. Another great example of appliance of machine learning techniques to text recognition problem is described in paper “Recognizing Handwritten Digits and Characters” by Vishnu Sundaresan and Jasper Lin, 2015[12]. Paper describes appliance of convolutional neural networks to text recognition problem.

2.1 “Neural network and Deep learning”, Michael Nielsen

The research was made on multilayer perceptron with MNIST dataset[13]. Neural networks consist of layers which contains some amount of neurons respectively[14]. These layers can be categorized as input layers, hidden layers and output layers. Input layers contains input neurons[15]. Those neurons do not process data with activation functions, they just pass them further. Those neurons can be also called “entry points” of neural networks. Neurons of hidden layers gather data from each neuron of input layers, sum outputs of applied activation functions on weights, and pass it further (or do not pass at all). Most of all computations are performed in these layers. And more model has hidden layers, more powerful the model

becomes, but it also takes more resources for computations in the other hand. Neurons of output layers do the same operations as neurons of hidden layers, but they have the only output weight (or “link”) due to absence of next layer. The output value of this type of neuron can be interpreted as the prediction and should be considered in the context. Figure 1 shows four-layer neural network. Activation function serves as the firing mechanism of the output.

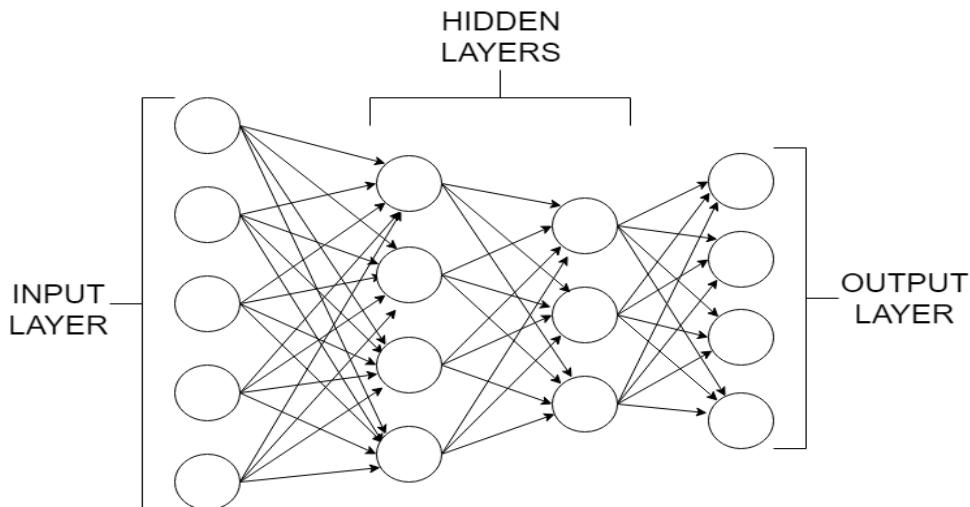


Figure 1: Layers of neural network

In other words, activation function shows whether is neuron activated or not. Activation function can be almost any function, which fits the mathematical model. It is obvious that activation functions have different behavior in the same domain, which affects the output. Usually sigmoid[16], hyperbolic tangent, Rectified Linear Unit and Softmax functions are used as activation functions in neural networks. Activation functions applied to the model based on assumed result. Each element of the neural network imitates a corresponding component of a brain cell. Neuron or perceptron imitates a neural cell itself, weights stand for axons, dendrites and synapses combined, and activation function can be interpreted as neurotransmitter. In reality organic brain is more complex and neural network should not be considered as exact digital analogy to organic brain. When the architecture of the model is defined and weights are randomly set, it can take input data and compute output value. Most likely the output would be different as it is expected, i.e. predict incorrectly. In order to get right predictions with high accuracy, the model must be trained. During the supervised training model takes input data, computes it and generates output value, which is compared to the actual result. The difference between the output and the actual result can be determined by different functions called loss functions. The choice of the loss function directly effects on the performance of the neural network. Loss functions can be divided into two groups - regressive loss functions and classification loss functions. Mean Absolute Error, Mean Square Error or Arctan could be used as regressive loss functions. Cross-Entropy loss, Hinge loss or logistic loss could be used in solution of classification problems. The main idea of the training process is changing the values of the weights and aspiring to reduce an error. One of the methods for defining weight such

that would reduce the loss function is taking the derivative of that loss function with respect to every parameter. This method comes down to finding such weights (or parameters in case of the function), that would minimize the outcome of the loss function. When derivative is found for each parameter, these parameters are being tuned according to the derivative. This process is called backpropagation. An initial experiment, described in this work, was performed on multilayer perceptron with MNIST dataset. MNIST stands for “Modified National Institute of Standards and Technology”. This dataset contains over 60,000 samples of handwritten digits. All samples are normalized, smoothed, resized to 28x28 pixels and converted into grayscale images. The model consists of 1 input layer, 1 hidden layer and 1 output layer. All neurons of each layer are connected to all neurons of next layer, except of output layer. An MLP architecture is shown on figure 2. Model in this experiment[17] is created with following architecture: input layer consists of 728 neurons, each for every pixel of 28x28 image[18]. Every neuron gets pixel intensity as input value. Sigmoid is applied as an activation function. Second layer has 30 neurons. Activation function is also sigmoid. Each neuron gathers all values from input connections, applies activation function and produces an output value. Output layer have 1 neuron for each classification class. It means that the output layer has 10 neurons for digits from 0 to 9. Activation function in last layer is softmax. After all, input data is passed forward and computed, model calculates loss with cross-entropy function. The last step of the iteration is backpropagation. This iteration shows the process of learning for one sample. One run with all samples is called epoch. To increase accuracy, training process can be repeated with several epochs. With this configuration Michael Nielsen has achieved accuracy of 96%.

2.2 “Recognizing Handwritten Digits and Characters”, V. Sundaresan and J. Lin

Another interesting research in this field was made in 2015 year by Vishnu Sundaresan and Justin Lin from Stanford University. Authors used convolutional neural network for digit and character recognition. Convolutional matrices are the main elements of this type of architecture. These matrices, also known as filters or kernels, compute dot product of its entries and the same sized input. In other words filter overlays an image, slides over it and searches for specific pattern. Each kernel learns to recognize a specific pattern like line, curve or corner. The figure 2 shows the convolution process. This approach allows to save spatial structure of the data. The result is stored in a so called feature map. Sometimes feature maps are compressed to reduce the number of parameters. This procedure is known as pooling. The figure 3 shows the CNN architecture. Authors used both MNIST and Chars74K[20] datasets in their experiments. They have built their own model along with LeNet and AlexNet. Figure 4 shows the architecture of custom model: With this configuration V. Sundaresan and J. Lin has achieved accuracy of 98% on MNIST dataset and 71% on Chars74K dataset comparing to the lower result of other networks.

3 Methods

3.1 Datasets

We conducted experiments with MNIST and Chars74K as datasets. MNIST dataset is a large set of thousands of handwritten digit samples. This dataset is used for basic calibration and

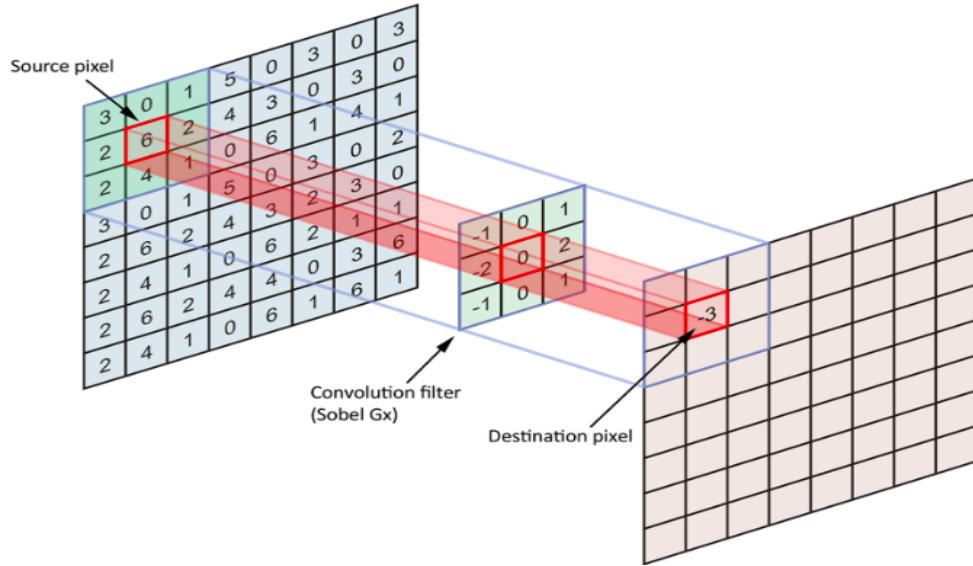


Figure 2: Convolution

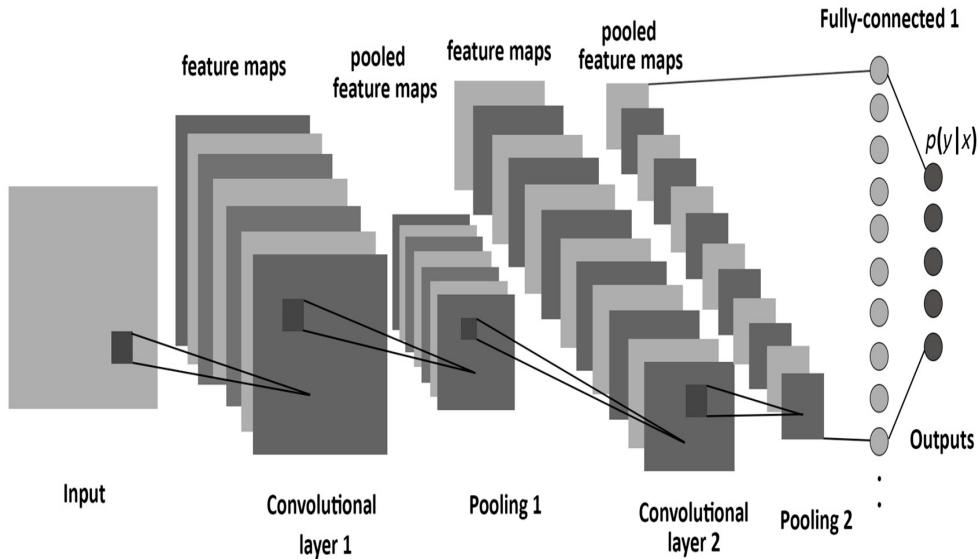


Figure 3: Architecture of Convolutional Neural Network[19]

tuning of different methods for image recognition. Each sample is preprocessed and unified. Preprocessing translates images into one common format. Unified format of images is the requirement for convolutional neural networks. In other words, architecture of convolutional neural networks based on structure of input data. Preprocessing also helps to generalize image features and, therefore, it directly affects on performance of convolutional neural networks[21], [22], [23]. The most common parameters of images are height, width, pixel intensity and number of channels. Typically, images have one or three color channels for grayscale or RGB respectively. Basing on these parameters, we can define architecture of neural network. To

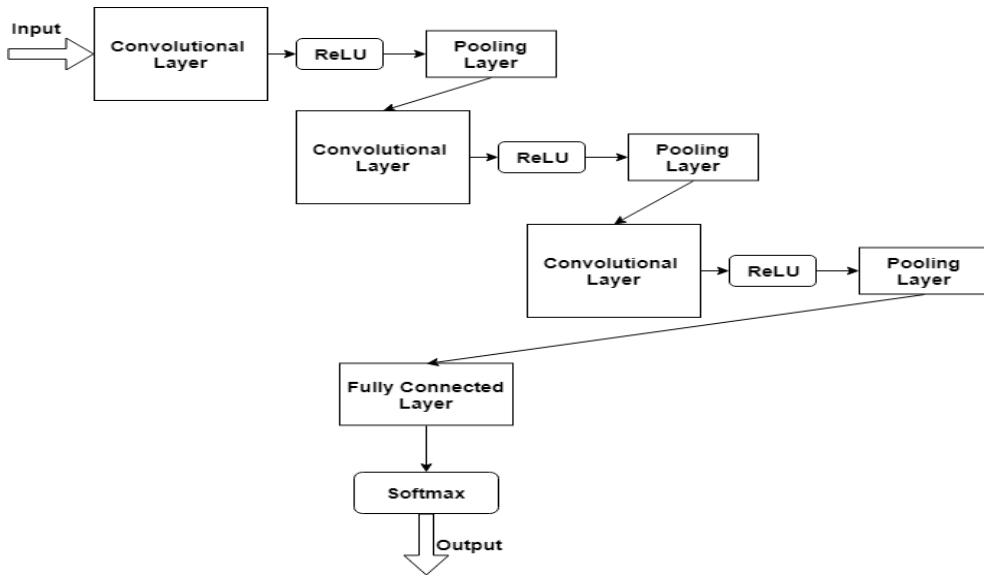


Figure 4: Pipeline of CNN

Table 1: Aggregation of Test Accuracies

	MNIST	Chars74K
LeNet	99%	45.3%
AlexNet	Na	63.4%
Custom network	98.1%	71.7%

increase accuracy of the model, we should perform further preprocessing. Usually it consists of resizing and rescaling, normalization and mean distribution of pixels of images. Convolutional neural network models are designed with fixed number of input nodes and kernels. Each node takes one pixel of image as input. Each kernel is intended for one of the channels. Therefore, for RGB input image should be defined three input kernels for each channel. First of all it is must be ensured that images have the same size and ratio. If some of the samples are not fitting to the specified size, they must be cropped and rescaled. Image should be cropped with the region of interest in center. Afterward image should be rescaled into defined size. There are plenty techniques to perform these transformations. Normalization of input data leads to similar distribution. One of the normalization techniques is to subtract from pixel mean value of all pixels and divide it by the result of the standard deviation. These actions lead to faster convergence during backpropagation. Another step of preprocessing is the data augmentation. This method transforms samples from dataset in different ways, such as rotations and scaling. Due to this method dataset is increases in dozens and helps to prevent recognition of unwanted features. MNIST is the dataset of handwritten digits, which contains about 60 000 of training data and 10 000 validation data. All of the samples are normalized, centered and resized to 28x28 pixels. Images in MNIST dataset grayscale, i.e. the have only

one channel representing gray tone intensity. Chars74K is the set of typed letters. Data in this dataset is raw, that means it was not preprocessed. Samples in dataset has different sizes, not normalized and augmented. All images in Chars74K in RGB color model. Chars74K contains around 7 000 samples, which we divided into training, validation and test sets in proportion of 60%, 20%, 20% respectively.

3.2 Neural network models

We conducted few experiments with two different convolutional neural networks. LeNet and AlexNet architectures was chosen for research to repeat the experiment, described in paper by Vishnu Sundaresan and Jasper Lin. Firstly LeNet architecture was introduced by Yann LeCun in 1989[24] and last modified version was introduced in 1998. LeNet model consist of two convolutional and three consecutive fully connected layers. There also pooling layers after each convolutional layer. AlexNet model was designed by Alex Krizhevsky and Ilya Sutskever in 2012[25]. This model contains five convolutional layers and 3 fully connected layers. Both models use ReLU loss for each layer. We expect the AlexNet model will show better results than LeNet model due to the depth of the model.

3.3 Tools and utilities

We used latest tools and utilities as an environment for experiments. We used latest version of PyTorch – machine-learning framework, along with Numpy, Pandas and Matplotlib frameworks, provide transparent and simple modeling process. Google provides free platform for research purposes. Google Colaboratory platform provides with free preconfigured Python environment and one GPU (in latest versions TPU – tensor-processing unit).

4 Results

4.1 Experiments

Firstly, we conducted experiments on MNIST dataset. We designed LeNet model with 2 convolutional layers and 3 fully-connected layers. We have chosen Cross-Entropy Loss as a loss function and Adam algorithm as an optimizer. We trained model 20 epochs long with batch size of 128 samples per batch. After conducting experiments, we got 99% accuracy – model correctly predicted 9904 sample out of 10 000. Same experiment we tried with AlexNet model with 5 convolutional layers and 3 fully-connected layers. Cross-Entropy Loss and Adam optimizer also had been chosen for this experiment. As the result we got 98.9% accuracy – model predicted 9896 samples out of 10 000. Concluding the experiment we can state that LeNet and AlexNet models showed almost the same result. We conducted experiments with Chars74K dataset. The dataset was not normalized. We rescaled images to the size of 128 pixels. We designed LeNet and AlexNet models with same architecture as in previous experiments. We used different learning rates increasing each by ten times per experiment. First experiment was made on LeNet model. As it is shown on table 2, different learning rates affect on results. With higher learning rate the model have not learned. Optimizer could not get to the global minimum during backpropagation. Learning rates with values equaled to 1e-3 and 1e-4 shown almost similar results – 74% and 71% respectively. Figure below shows loss

Table 2: LeNet learning results

Experiment №	Learning rate	Learning time	Accuracy	Loss
1	1e-2	0:35:40.46	8% (118/1541)	0.0320
2	1e-3	0:35:46.23	74% (1138/1541)	0.0093
3	1e-4	0:34:47.73	71% (1093/1541)	0.0100

reduction during training process. As you can see, the optimizer with learning rate equaled to 1e-3 optimizes weights faster than with learning rate equaled to 1e-4. Optimizer with 1e-2 learning rate could not minimize the loss at all.

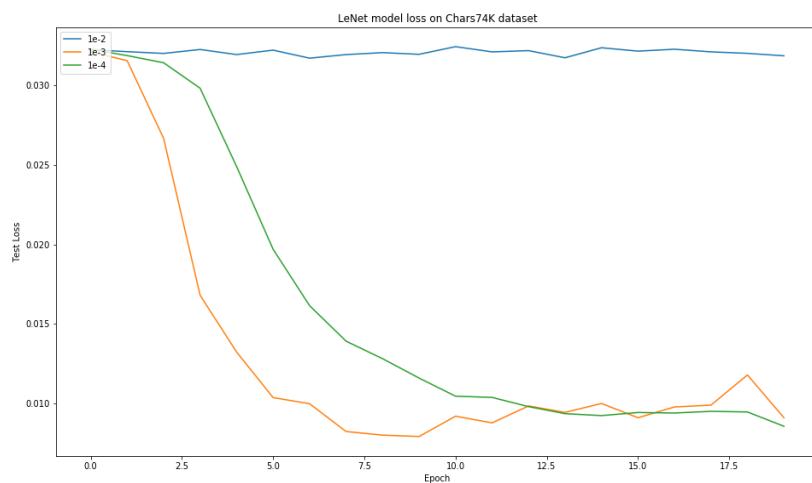


Figure 5: LeNet model loss

As it is shown, on figure 6, most accurate predictions were made while the optimizer was training the model with learning rate equaled to 1e-3. We can state that learning rate in the second experiment is the most optimal, because learning rate in the third experiment takes more time to learn to get same results. On the same dataset, we also trained AlexNet model. The result shown in table 3 below.

AlexNet had reached maximal accuracy with learning rate equaled to 1e-3 – 78% with 1197 correct predictions out of 1541. The third experiment showed much less accurate predictions than the experiment on LeNet model with the same learning rate. That could happen due to greater amount of convolutions in AlexNet model than in LeNet and it takes more computations to reach the same result. Graphical visualization of loss and accuracy is shown

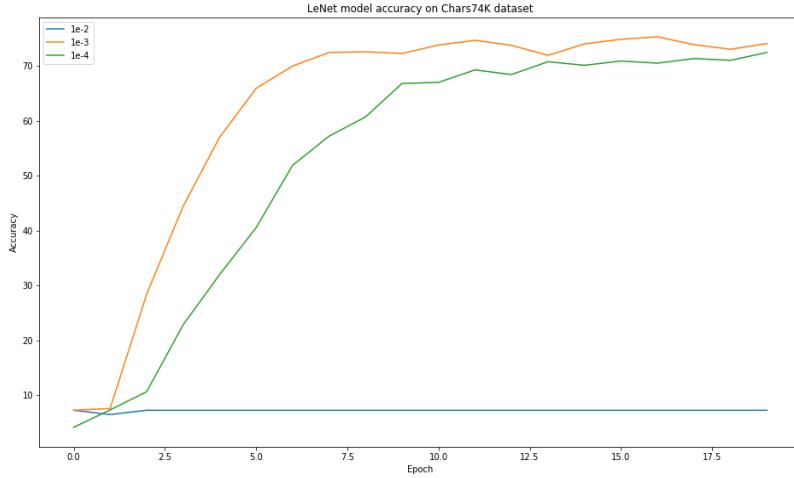


Figure 6: LeNet model accuracy

Table 3: AlexNet learning results

Experiment №	Learning rate	Learning time	Accuracy	Loss
1	1e-2	0:40:15.10	7% (107/1541)	0.0323
2	1e-3	0:41:06.41	78% (1197/1541)	0.0079
3	1e-4	0:41:45.23	64% (990/1541)	0.0108

on the figure 7 and 8.

Comparing two models within the same conditions, we can state that AlexNet model showed slightly better result than LeNet model. AlexNet model is more powerful than LeNet, and the result can be improved by preprocessing of dataset, but it takes more time to train this network. The figure below describes loss decrease of each model over learning epochs. LeNet started to decrease loss right from the beginning, while AlexNet model started to learn after ninth epoch. At the end of twentieth epoch, AlexNet showed slightly better results than LeNet.

Same result could be observed of accuracy comparison graphic. AlexNet started to improve prediction accuracy right from eighth epoch and reached its best result at the end of twenties epoch, meanwhile LeNet showed steady growth in accuracy, but could not beat the AlexNet at the end.

Comparing implementations of AlexNet and LeNet models, it is noticeable that there are great difference in total amount of trainable parameters.

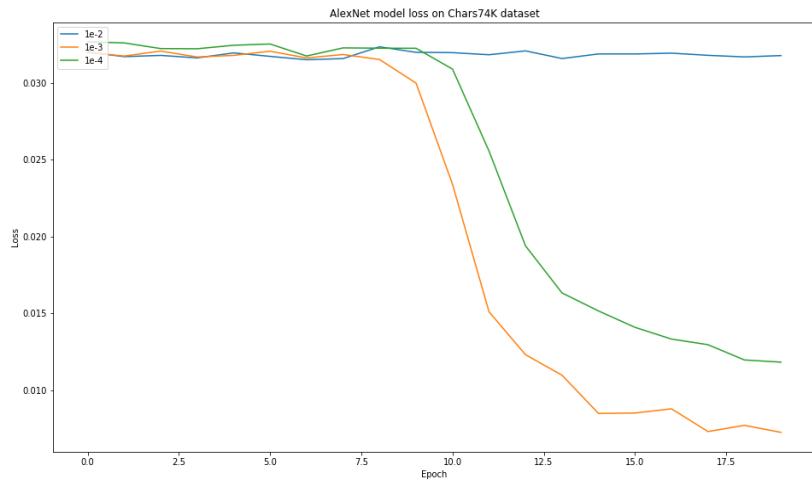


Figure 7: AlexNet model loss

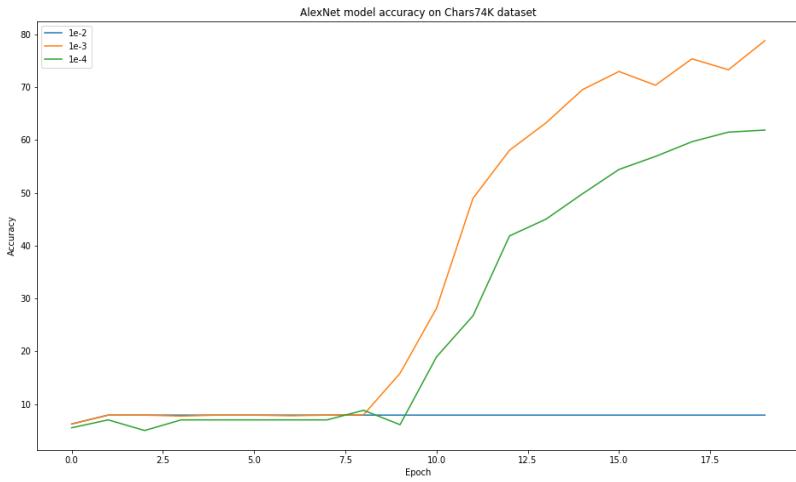


Figure 8: AlexNet model accuracy

Implementation of LeNet model contains 5 times more trainable parameters than AlexNet model. Transition from convolutional layer to fully-connected layer generate large number of trainable parameters. Basing on these numbers, AlexNet model uses 8 MB of memory to feedforward the signal and LeNet uses about 45 MB of memory to pass the same signal.

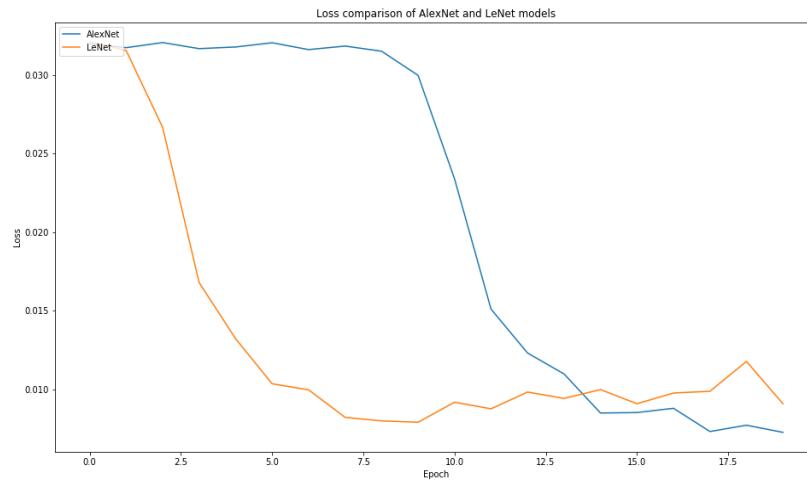


Figure 9: Loss comparison

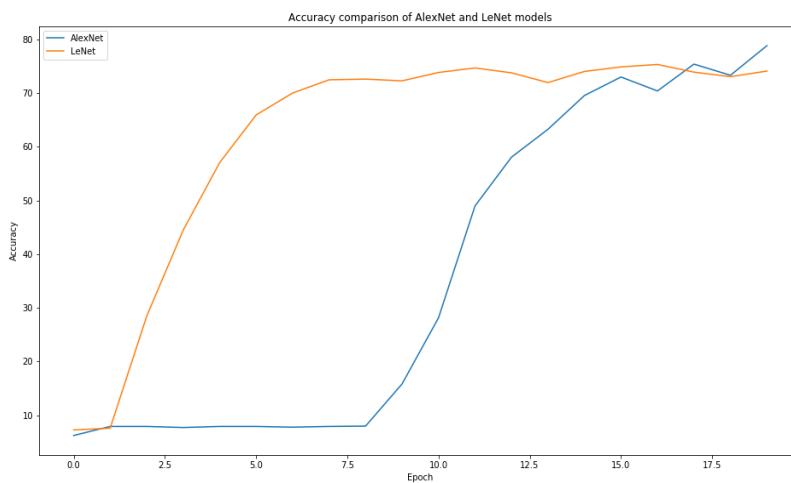


Figure 10: Accuracy comparison

5 Conclusion

In conclusion, we can state that it is obvious that convolutional neural networks could be used as solution of optical recognition problem. It is reasonable to study these fields due to the steady development of object recognition and machine learning in general. In this work, we observed related works in optical character recognition with neural networks area. Also, we conducted experiments with implementations of AlexNet and LeNet models on Chars74K dataset. As the result of experiments, implementation of AlexNet model has shown greater

Table 4: Aggregation of Test Accuracies

Model	Total number of parameters	Memory usage
LeNet	11,412,382	45 MB
AlexNet	2,021,918	8 MB

prediction accuracy than LeNet model. In the other hand AlexNet took longer time to learn and reach the same result as LeNet model. At the first sight, it could seem that implementation of AlexNet model has more trainable parameters than LeNet due to the greater number of layers in architecture. However, convolutional layers generate less parameters than fully connected layers. Taking in account all results of experiments, we can conclude that AlexNet model showed better result in general, than LeNet model. AlexNet could be used in optical character recognition and text recognition systems. Data from Chars74K dataset, which was used in experiments, was not preprocessed and augmented. Preprocessing of data is very important step; it can highly increase accuracy. In further researches, we should consider using different models and combination of architectures, such as recurrent neural networks and combination of convolutional and recurrent networks.

References

- [1] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep learning", *Nature* (2015): 436–444.
- [2] Behnam Neyshabur et al., "Exploring Generalization in Deep Learning", accessed October 14, 2018, <https://papers.nips.cc/paper/7176-exploring-generalization-in-deep-learning.pdf>.
- [3] Warren S. McCulloch and Walter H. Pitts, "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics* (Springer US, 1943): 115–133.
- [4] Vidushi Sharma, Sachin Rai and Anurag Dev, "A Comprehensive Study of Artificial Neural Networks", *International Journal of Advanced Research in Computer Science and Software Engineering* 10 (2012): 278-284.
- [5] Jayesh B. Ahire, "Real world Applications of Artificial Neural Networks", accessed October 14, 2018, <https://medium.com/@jayeshbahire/real-world-applications-of-artificial-neural-networks-a6a6bc17ad6a>
- [6] Sumit Das et al., "Applications of Artificial Intelligence in Machine Learning: Review and Prospect", *International Journal of Computer Applications* 115 (2015): 31-41.
- [7] Sonali B. Maind and Wankar Priyanka, "Research Paper on Basic of Artificial Neural Network", *International Journal on Recent and Innovation Trends in Computing and Communication* 2 (2014), accessed November 5, 2018, <http://www.ijritcc.org/download/Research%20Paper%20on%20Basic%20of%20Artificial%20Neural%20Network.pdf>.
- [8] Alex Krizhevsky, Ilya Sutskever and Geoffery E. Hinton, "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing systems* 25 (2012), accessed November 5, 2018, <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>.
- [9] Christian Szegedy et al., "Rethinking the inception architecture for computer vision", *Paper presented at the 29th IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, June 26 – July 1, 2016*.
- [10] Michael Nielsen, "Neural network and Deep learning", accessed October 20, 2018, <http://neuralnetworksanddeeplearning.com/index.html>.
- [11] Yann LeCun et al., "Gradient-based learning applied to document recognition", *Proceedings of the IEEE* 86 (1998):2278-2324.

- [12] Vishnu Sundaresan and Jasper Lin, "Recognizing Handwritten Digits and Characters", *accessed October 20, 2018*, http://cs231n.stanford.edu/reports/2015/pdfs/vishnu_final.pdf.
- [13] Michael Nielsen, "Neural Network and Deep Learning: Learning with gradient descent", *accessed October 21, 2018*, <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [14] Michael Nielsen, "Neural Network and Deep Learning: The architecture of neural networks", *accessed October 21, 2018*, <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [15] Michael Nielsen, "Neural Network and Deep Learning: Perceptron", *accessed October 21, 2018*, <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [16] Michael Nielsen, "Neural Network and Deep Learning: Sigmoid function", *accessed October 21, 2018*, <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [17] Michael Nielsen, "Neural Network and Deep Learning: Implementing our network to classify digits", *accessed October 21, 2018*, <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [18] Michael Nielsen, "Neural Network and Deep Learning: A simple network to classify handwritten digits", *accessed October 21, 2018*, <http://neuralnetworksanddeeplearning.com/chap1.html>.
- [19] Saleh Albelwi and Ausif Mahmood, "A Framework for Designing the Architectures of Deep Convolutional Neural Networks", *Entropy* 19(2017): 242-262.
- [20] Amit Choudhary, "A Review of Various Character Segmentation Techniques for Cursive Handwritten Words Recognition", *International Journal of Information & Computation Technology* 4 (2014): 559-564.
- [21] Shuang Wu et al., "L1-Norm Batch Normalization for Efficient Training of Deep Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems* (2018), *accessed November 28, 2018*, doi:10.1109/TNNLS.2018.2876179.
- [22] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *Paper presented at the 32nd International Conference on Machine Learning, Lille, France, July 06 – 11, 2015*.
- [23] Yann LeCun et al., "Efficient backprop, Neural Networks: Tricks of the Trade", *second edition* (Springer US, 1998): 9-48.
- [24] Yann LeCun et al., "A handwritten digit recognition: Applications of neural net chips and automatic learning", *Neurocomputing* (Springer US, 2005): 303-318.
- [25] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *accessed November 20, 2018*, <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

МРНТИ 20.23.25, 16.31.21, 28.23.39

Сравнение морфологических правил глагола казахского и турецкого языков

Жеткенбай Л., Евразийский национальный университет имени Л.Н.Гумилева,
г. Астана, Республика Казахстан, E-mail: jetlen7@gmail.com

Шарипбай А.А., Евразийский национальный университет имени Л. Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: sharalt@mail.ru

Бекманова Г.Т., Евразийский национальный университет имени Л. Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: gulmira-r@yandex.ru

Кажымухан Д., Евразийский национальный университет имени Л. Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: askarovna0105@mail.ru

Каманур У., Евразийский национальный университет имени Л. Н. Гумилева,
г. Астана, Республика Казахстан, E-mail: unzila.88@mail.ru

В данной статье приведено детальное сравнение глаголов казахского и турецкого языков. Для этого сначала строятся онтологические модели морфологических правил глаголов казахского и турецкого языков. В этих онтологических моделях формально и визуально описаны свойства и функции используемых лексических единиц для образования глаголов в указанных языках, а также отношений между этими единицами. Для обозначения грамматических категорий и понятий в предлагаемых онтологических моделях используются термины из унифицированной метаязыка UniTurk. После построения онтологических моделей глаголов указанных языков был построен алгоритм их сравнение. С помощью онтологических моделей формализованы морфологические правила тюркских (казахского, турецкого) языков. Результат этих работ могут быть использованы для решения задач NLP, например, для разметки корпусов, в системах извлечения знаний, системах информационного поиска, машинного перевода и т.д.

Ключевые слова: онтологические модели, морфологические правила, лексические единицы, глагол казахского и турецкого языков.

Қазақ және түрік тілдер етістігінің морфологиялық ережелерін салыстыру
Жеткенбай Л., Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана қаласы,

Қазақстан Республикасы, E-mail: jetlen7@gmail.com

Шәріпбай А.А., Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана қаласы,
Қазақстан Республикасы, E-mail: sharalt@mail.ru

Бекманова Г.Т., Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана қаласы,
Қазақстан Республикасы, E-mail: gulmira-r@yandex.ru

Кажымухан Д., Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана қаласы,
Қазақстан Республикасы, E-mail: askarovna0105@mail.ru

Каманур У., Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана қаласы,
Қазақстан Республикасы, E-mail: unzila.88@mail.ru

Мақалада қазақ және түрік тілінің етістіктерін толық салыстыру көрсетілген. Бұл үшін алдымен қазақ және түрік тілі етістіктерінің морфологиялық ережелерінің онтологиялық моделі құрылады. Бұл онтологиялық моделдерде аталған тілдердің етістік жасалу үшін қолданылатын лексикалық бірліктерінің қасиеттері сонымен қатар бұл бірліктер арасындағы қатынастары да формалды және визуалды сипатталады. Ұсынылатын онтологиялық моделдерде грамматикалық категориялар мен түсініктерді белгілеу үшін UniTurk метатілі терминдері қолданылады. Аталған тілдер етістіктерінің онтологиялық моделдері құрылған соң, оларды салыстыру алгоритмі құрылды. Онтологиялық моделдер арқылы түркі (қазақ, түрік) тілдерінің морфологиялық ережелері формалданған. Бұл жұмыстың нәтижелерін NLP есептерін шешу үшін қолдануға болады. Мысалы, корпустарды белгілеуде, білімдерді шығару жүйелері, ақпараттық іздеу жүйелері, машиналық аудары және басқалары.

Түйін сөздер: онтологиялық модельдер, морфологиялық ережелері, лексикалық бірліктер, қазақ және түрік тілі етістік.

Comparison of the morphological rules of the Kazakh and Turkish languages

Zhetkenbay L., L.N.Gumilyov Eurasian National University, Astana,

Republic of Kazakhstan, E-mail: jetlen7@gmail.com

Sharipbay A.A., L.N.Gumilyov Eurasian National University, Astana, Republic of Kazakhstan,

E-mail: sharalt@mail.ru

Bekmanova G.T., L.N.Gumilyov Eurasian National University, Astana, Republic of Kazakhstan,

E-mail: gulmira-r@yandex.ru

Kazhymukhan D., L.N.Gumilyov Eurasian National University, Astana, Republic of Kazakhstan,

E-mail: askarovna0105@mail.ru

Kamanur U., L.N.Gumilyov Eurasian National University, Astana, Republic of Kazakhstan,

E-mail: unzila.88@mail.ru

This article gives a detailed comparison of the verbs of Kazakh and Turkish. For this, ontological models of morphological rules of verbs of Kazakh and Turkish languages are first constructed. These ontological models formally and visually describe the properties and functions of the lexical units used to form verbs in the specified languages, as well as the relationships between these units. To denote grammatical categories and concepts in the proposed ontological models, the terms from the unified metalanguage UniTurk are used. After constructing ontological models of verbs of these languages, an algorithm was constructed to compare them. Using ontological models, the morphological rules of the Turkic (Kazakh, Turkish) languages are formalized. The result of these works can be used in the NLP applications, for example, for corpus tagging, in knowledge extraction systems, information retrieval systems, machine translation, etc.

Key words: ontological models, morphological rules, lexical units, verb of the Kazakh and Turkish languages.

1 Введение

В настоящее время в связи с резким увеличением объема информации на естественных языках в интернете и социальных сетях исследование и разработки в области компьютерной лингвистики становятся чрезвычайно актуальными. Как известно, компьютерная лингвистика является новым научным направлением и входит в состав искусственного интеллекта, который также является новым направлением информатики (вычислительной науки). Компьютерная лингвистика включает в себя компьютерную обработку естественных языков (ОЕЯ) – Natural Language Processing (NLP). Для компьютерной обработкой любых естественных языков требуются, во-первых, формализация их грамматических (морфологических и синтаксических) правил, во-вторых, разработка алгоритмов анализ и синтеза слов и предложений по этим правилам, в-третьих, программная реализация всех этих алгоритмов, в-четвертых, построение текстовых корпусов (база данных размеченных текстов) и аудиокорпусов (база данных размеченных аудиозаписей) и других программ для анализа и обработки текстов, например, сентимент анализ. Все языковые уровни характеризуются наличием базовых элементов. Изучение языка может идти с двух позиций – анализа и синтеза, ибо выявленные правила синтеза могут способствовать проведению анализа и наоборот. В данном случае казахский и турецкий языки изучаются с позиций, как анализа, так и синтеза. Именно такой комплексный подход позволяет детально изучить все закономерности и выявить такие нюансы, которые, при использовании только одного из подходов остались бы за пределами нашего внимания. Для исследования и максимальной формализации каждой языковой подсистемы необходимо создавать программный инструментарий, реализующий процесс изучения путем выявления и проверки правил анализа и синтеза, тем самым максимально

автоматизируя исследовательский процесс, освобождая при этом исследователя как от рутинного процесса накопления и сбора информации, так и снимая вопрос трудоемкости ее обработки. Моделирование морфологии имеет отношение ко всем приложениям, таким как обработка естественного языка и задач, включая поиск информации, анализ настроений, исправление орфографии, обнаружение генерируемых текстов, разметка частей речи. Морфология занимается изучением структуры и формирования слов. Агглютинативными языками (лат *agglutinare*: Склейивать) являются языки, морфологическая система которых характеризуется агглютинацией («склеиванием») различных формант. В качестве формант выступают либо префиксы либо суффиксы, каждый из которых имеет только один свой собственный смысл.

2 Обзор литературы

Поскольку казахский и турецкий языки относятся к группе тюркских языков, а языки этой группы могут быть классифицированы как агглютинативные языки. Эти языки богаты словоформами (флексиями). Флекции образуются путем добавления суффиксов. Суффиксы прикреплены в строгой последовательности и возникающие в результате новые слова могут иметь другую часть речи. Притяжательная форма на казахском языке подобна притяжательной форме в английском языке [1, 2]. Проведено много исследований в данном направлении по формализации морфологических правил и морфологическому анализу [3-9] тюркских языков. Первый морфологический анализатор казахского языка был разработан в 2009 году и имел в своей основе процедурный метод. Процедурный метод предполагает предварительную систематизацию морфологических знаний о естественном языке и разработку алгоритмов присвоения морфологической информации отдельной словоформе [10]. Вторая версия морфологического анализатора была разработана в 2012 году и была основана на формальных морфологических правилах [11]. Более поздние версии были основаны на использовании онтологических моделей и гиперграфа [12-17]. Другими исследовательскими группами были разработаны собственные морфологические анализаторы [18-20]. Для турецкого языка работы по созданию морфологического анализа ведутся достаточно давно и представлены в работах [21-25].

3 Материал и методы

Онтология является мощным и широко используемым инструментом для формализованного описания свойств и функций объектов заданной предметной области, а также отношений между ними. Можно классифицировать онтологии на основе степени зависимости от задачи или прикладной области, модели представления онтологических знаний и выразительности, а также других критериев. Основная часть формально представленных знаний основана на концептуализации: объекты, концепции и другие объекты, которые существуют в некоторой области интересов и отношения, определенные между ними. Концептуализации является абстрактным, упрощенным взглядом на мир, который мы хотим представить для какой-то цели [26]. Для того, чтобы разработать онтологию вначале нужно ответить на следующие вопросы:

1. Какие области охватывает онтология? Ответ: Глагол.
2. Для чего нам нужна онтология? Ответ: Нужно (необходимо) для создания сравнительной онтологической модели глагола казахского и турецкого языков.
3. На какие виды вопросов должна отвечать информация в онтологии? Ответ: Необходимо для определения спряжения глагола по составу и смыслу и категории глаголов.
4. Кто будет использовать и поддерживать онтологию? Ответ: Лингвисты и программисты.

Согласно вышеуказанным вопросам сравнительная онтологическая модель глагола будет выглядеть $O(X, R, I)$, где X – наименования входящие в структуру глагола (объекты и понятия), R – связи между наименованиями, а I – множество наименований этих структур и связей. Сравнительная онтологическая модель глагола разработана в среде Protege (<http://protege.stanford.edu>). Язык Protege OWL дает возможность описать не только понятия, но и конкретные объекты. Онтологическая модель глагола казахского языка разработанная в среде Protege отображен на 1-рисунке, а онтологическая модель глагола турецкого языка на 2-рисунке.

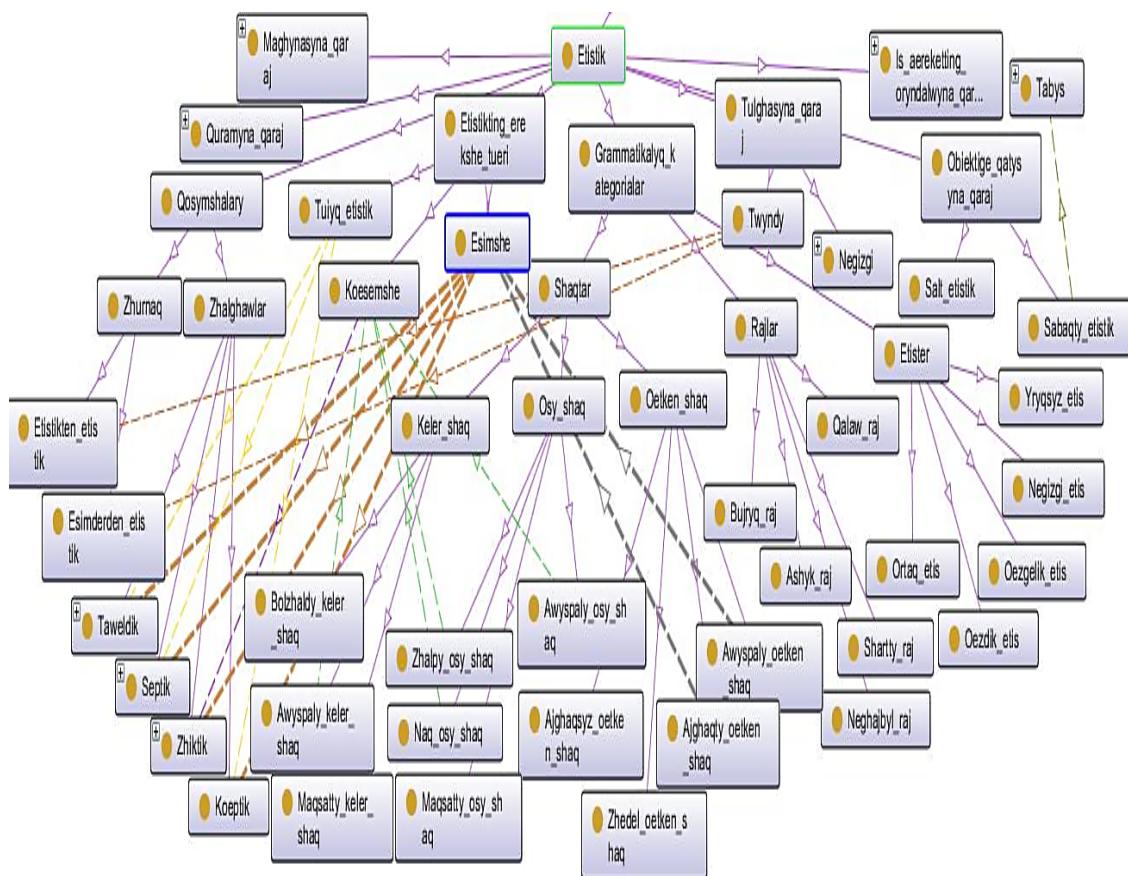


Рисунок 1: Онтологическая модель глагола казахского языка

Таким образом, сравнительная онтологическая модель глагола охватывает все

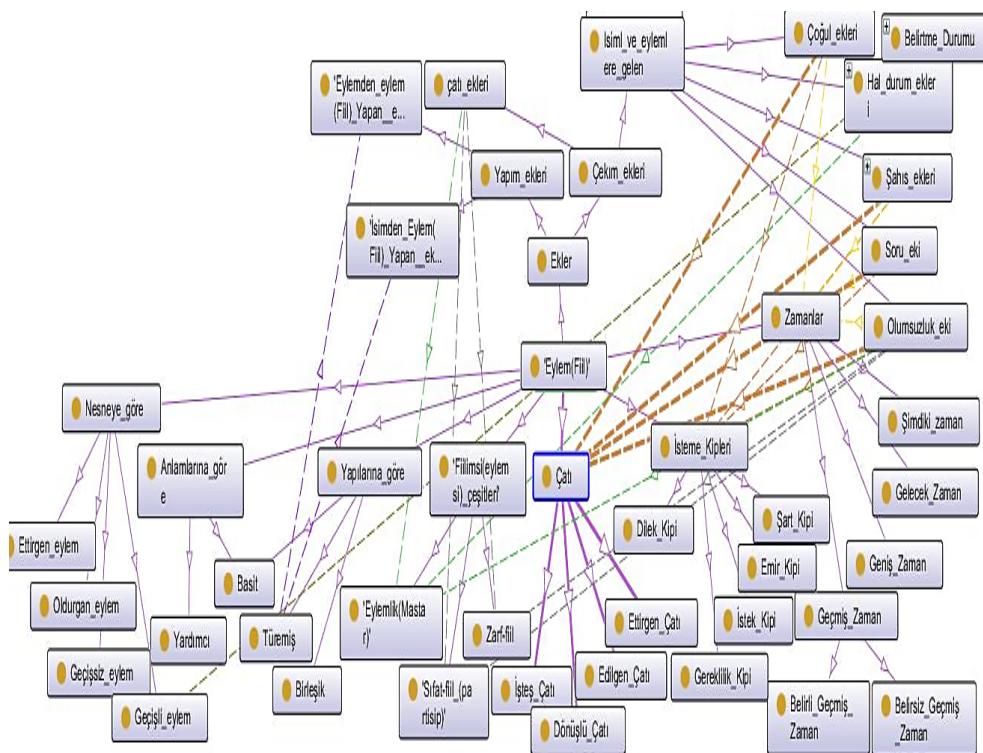


Рисунок 2: Онтологическая модель глагола турецкого языка

компоненты множества морфологического анализа. Сравнение онтологических моделей позволило создать общую систему обозначений морфологических признаков, которые используются в морфологическом анализаторе. Теперь остановимся на идентичности и различиях глагола казахского, турецкого языков. Сравнение по категориям залога и накопления непроизводного глагола «ки-giy» на казахском и турецком языках представлено на таблице 1. Приведение в соответствие правила по категориям залога и накопления на примере слова «ки-giy» на казахском и турецком языках представлено на таблице 1.

Таблица 1 – Сравнение по категориям залога и накопления непроизводного глагола «ки-giy»

<i>Казахский</i>	<i>Турецкий</i>	<i>Казахский</i>	<i>Турецкий</i>	<i>Казахский</i>	<i>Турецкий</i>				
Утвердительный	Отрисательный	Вопросительный							
Рай-Наклонение-Kipi									
Шартты рай (Условное наклонение) (аффиксы : -са, -се) -ps (ps- Личное окончание)									
Şart Kipi (аффиксы : -sa, -se) -ps									
ки-се-ps	giy-se-ps	ки-ме-се-ps	giy-me-se-ps	ки-се-ps бе, ме?	giy-me-se-ps mi?				
Gereklik Kipi (аффиксы : -mali, -meli)-ps									

	gel-meli-ps		gel-me-meli-ps		gel-meli mi-ps
Қалау рай (Желательное наклонение) (аффиксы : -гай, -гей, -қай, -кей) едi-ps					
Қалау рай (Желательное наклонение) (аффиксы : -ғы, -ті, -қы, -кі)-ps келеді					
қалау рай (Желательное наклонение) (аффиксы : -са, -се)-ps еken					
Dilek Kipi (аффиксы : -sa, -se) -ps					
ки- гi-ps келеді	giy-se-ps	ки- гi-ps келмейді	giy-me-se-ps	ки- гi-ps келеді	giy-me-se-ps mi?
ки-се-ps екен		ки- ме-се-ps екен		ме? ки-се-ps екен бе?	
Бұйрық рай (Повелительное наклонение) (аффиксы : -айын, -ейін, -иын, -иін, -айық, -ейік, -иық, -иік)					
Emir Kipi (аффиксы :) -ps					
ки-ейін	giy- ps	ки -мей- ps	giy-me-ps	ки - ейін бе?	
Ашық рай (Изъявительное наклонение) (аффиксы :) -ps					
Istek Kipi (аффиксы :) -ps					
Етіс-Залог- Çatı					
өздік етіс (Возвратный залог) (аффиксы : -ын, -ін, -н) -ps (түбір етістік: ки)					
Dönüslü Çatı (аффиксы : -n, -in, -in, -un, -ün) -ps (түбір етістік: gel) (ps-жіктік жалғауы)					
Dönüslü Çatı (аффиксы : -l, il, -il, -ul, -ül)					
ки-ін	giy-in	ки-ін-бе	giy-in-me		
өзгелік етіс (Понудительный залог) (аффиксы : -т; -тыр,-тір, -дыр, -дір)					
өзгелік етіс (Понудительный залог) (аффиксы : -қыз -ғыз, -кіз, -гіз)					
Ettirgen Çatı (аффиксы : -t, -tır, -tir, -tur, -tür, -dir, -dir, -dur, -dür)					
Ettirgen Catı (аффиксы : -r, -ır, -ir, -ur, -ür)					
ки- дір ки - гіз	giy- dir	ки- дір-ме ки- гіз-бе	giy- dir-me		
Ортақ етіс (Взаимный залог) (аффиксы : -с, -ыс, -ic)					
Işteş Çatı(аффиксы : -ş, -ış, -iş, -uş, -üş)					
ки-ic	giy-iş	ки-ic-пе	giy-iş-me		
Ырықсыз етіс (Страдательный залог) (аффиксы : -ыл, -іл, -л)					
Edilgen Çatı (аффиксы : -l, il, -il, -ul, -ül)					
Edilgen Çatı (аффиксы : -n, -in, -in, -un, -ün)					
ки-іл	giy-in	ки-іл-ме	giy- in-me		

Таблица 2 - Приведение в соответствие правила по категориям залога и накопления на примере слова «ки-giy»

Казахский	Турецкий
Шартты рай (Условное наклонение)	Sart Kipi
кисем: ки+Verb+ Cond+A1sg	giysem: giy+Verb+ Cond+A1sg
кисем: ки+Verb+Pos+Cond+A1sg	giysem: giy+Verb+Pos+Cond+A1sg

кимесем: ки+Verb+Neg+Cond+A1sg	giyimesem: giy+Verb+Neg+Cond+A1sg
кисе?: ки+Verb+ Cond+ A2sg	giysen: giy+Verb+ Cond+ A2sg
кисен: ки+Verb+Pos+Cond+ A2sg	giysen: giy+Verb+Pos+Cond+ A2sg
кимесен: ки+Verb+Neg+Cond+ A2sg	giyimesen: giy+Verb+Neg+Cond+ A2sg
кисеніз: ки+Verb+ Cond+ A2sgpol	
кисеніз: ки+Verb+Pos+Cond+ A2sgpol	
кимесеніз: ки+Verb+Neg+Cond+ A2sgpol	
кисе: ки+Verb+ Cond+ A3sg	giyse: giy+Verb+ Cond+ A3sg
кисе: ки+Verb+Pos+Cond+ A3sg	giyse: giy+Verb+Pos+Cond+ A3sg
кимесе: ки+Verb+Neg+Cond+ A3sg	giymese: giy+Verb+Neg+Cond+ A3sg
кисек: ки+Verb+ Cond+ A1pl	giysek: giy+Verb+ Cond+ A1pl
кисек: ки+Verb+Pos+Cond+ A1pl	giysek: giy+Verb+Pos+Cond+ A1pl
кимесек: ки+Verb+Neg+Cond+ A1pl	giyimesek: giy+Verb+Neg+Cond+ A1pl
кисендер: ки+Verb+ Cond+ A2pl	giyseniz: giy+Verb+ Cond+ A2pl
кисендер: ки+Verb+Pos+Cond+ A2pl	giyseniz: giy+Verb+Pos+Cond+ A2pl
кимесендер: ки+Verb+Neg+Cond+ A2pl	giyimeseniz: giy+Verb+Neg+Cond+ A2pl
кисеніздер: ки+Verb+ Cond+ A2plpol	
кисеніздер: ки+Verb+Pos+Cond+ A2plpol	
кимесеніздер: ки+Verb+Neg+Cond+ A2plpol	
кисе: ки+Verb+ Cond+ A3pl	giyse(ler): giy+Verb+ Cond+ A3pl
кисе: ки+Verb+Pos+Cond+ A3pl	giyse(ler): giy+Verb+Pos+Cond+ A3pl
кимесе: ки+Verb+Neg+Cond+ A3pl	giyimeseler: giy+Verb+Neg+Cond+ A3pl

4 Результаты и обсуждение

В ходе научного исследованы морфологические признаки казахского и турецкого языков. Проведено сопоставление онтологий, разработана единая система обозначений морфологических признаков, переписаны морфологические правила казахского и турецкого языков через новую систему обозначений. Разработан единый морфологический анализатор, основанный на общем алгоритме морфологического анализа.

5 Заключение

Разработанные онтологические модели для компьютерной обработки казахского и турецкого языков являются важным шагом при сравнительном исследовании двух тюрских языков. Поэтому исследование структуры и значений схожих глагола казахского и турецкого языков, и результаты их сравнения безусловно дает большую возможность для обработки естественных языков. Результаты будут применяться для создания систем семантического перевода с казахского языка на турецкий язык, и наоборот, и для электронного обучения указанным языкам через компьютеры или через Интернет.

6 Благодарность

Работа выполнена при поддержке грантового финансирования научно-технических программ и проектов Министерством науки и образования Республики Казахстан (грант № АР05132249, 2018-2020 годы)

Список литературы

- [1] *Қазақ, грамматикасы. Фонетика, сөзжасам, морфология, синтаксис.* – Астана, 2002. – 784 6.
- [2] *Lewis, Geoffrey.* Turkish Grammar. Oxford University Press, 2001.
- [3] Sharipbay A., Bekmanova G. The synthesis of word forms of Turkic language using semantic neural networks // Modern problems of applied mathematics and information technologies: abstracts – Al Khorezmy, 2009. – P.145.
- [4] Sharipbayev A. A., Bekmanova G. T. The building of logical semantics of the Kazakh words // The materials of the all-Russian conference with International participations «Knowledge-Ontology-Theory, ZONT-09». – Novosibirsk, 2009. – P. 246-249.
- [5] Altintas K., Cicekli I. A Morphological Analyser for Crimean Tatar // Proceedings of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks TAINN'2001, North Cyprus, 2001. – P.180-189.
- [6] Tantug, A. C., Adali, E., Oflazer, K. Computer Analysis of the Turkmen Language Morphology // Lecture Notes in Computer Science, Springer, 2006. – P. 186-193.
- [7] Orhun, M., Tantuğ, A. C., Adali, E. Rule Based Analysis of the Uyghur Nouns // International Journal of Asian Language Processing, 2009. – P.33-43.
- [8] Orhun, M., Tantuğ A., Adali, E. Rule Based Tagging of the Uyghur Verbs // Fourth International Conference on Intelligent Computing and Information Systems, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt, 19-22 March , 2009. – P.811-816.
- [9] Sulejmanov D.Sh., Nevzorova O.A., Gatiatullin A.R., Gilmullin R.A., Ayupov M.M., Pyatkin N.V. The main components of the application of the grammatical model Tatar (In Russian) // In Proc. Dialogue Conference, Computational linguistics and intelligent technologies. Russian State Humanitarian University, 2007. – P.525-530.
- [10] Bekmanova G.T. Some approaches to the problems of automatic inflection and morphological analysis in the Kazakh language // the newsletter of D. Serikbayev East Kazakhstan state technical university. – Ust-Kamenogorsk, 2009. – C. 192-197.
- [11] Sharipbayev, A., Bekmanova, G., Mukanova, A., Buribayeva, A., Yergesh, B., Kaliyev, A. Semantic neural network model of morphological rules of the agglutinative languages // The 6th International Conference on Soft Computing and Intelligent Systems The 13th International Symposium on Advanced Intelligent Systems. – Kobe, Japan, 20-24 November 2012. –P.1094-1099.
- [12] Yergesh, B., Mukanova, A., Bekmanova, G., Sharipbay, A., Razakhova, B. Semantic hyper-graph based representation of Verbs in the Kazakh language // Computacion y Sistemas; Volume 18, Issue 3, 1 July 2014. – P. 627-635.
- [13] Mukanova, A., Yergesh, B., Bekmanova G., Razakhova, B., Sharipbay, A. Formal models of Verbs in the Kazakh language // Leonardo Electronic Journal of Practices and Technologies; Issue 25, July-December, 2014. – P. 264-273.
- [14] Zetkenbay, L., Sharipbay, A., Bekmanova, G., Kamanur, U. Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages // Journal of Theoretical and Applied Information Technology, Vol. 91. No.2, 2016. – P. 257-263.
- [15] Kamanur U., Sharipbay A., Altenbek G., Bekmanova G., Zhetkenbay L. Investigation and use of methods for defining the extends of similarity of Kazakh language sentences // 15th China National Conference, CCL 2016, and 4th International Symposium, NLP-NABD 2016, Yantai, China, October 15-16, 2016. – P.153-161.
- [16] Bekmanova G., Sharipbay A., Altnbek G., Adali E., Zhetkenbay L., Kamanur U., Zulkhazhav A. The uniform morphological analyzer for the Kazakh and Turkish languages. Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017), Moscow, Russia, July 2017. – P. 20-30.

- [17] Sharipbay A., Mukanova A., Yergesh B., Zhetkenbay L., Zulkhazhav A., Yelibayeva G. Ontology modeling of morphological rules of the Kazakh and Turkish languages // Abstract of the VI international conference «Modern problems of applied mathematics and information technology - al-Khorezmiy» 2018. – P.51-52.
- [18] Tukeyev, U., Zhumanov, Zh., Rakimova, D., Kartbayev, A. Combinational Circuits Model of Kazakh and Russian Languages Morphology // Abstracts of International Conference «Computational and Informational Technologies in Science, Engineering and Education». – Almaty: Al-Farabi KazNU Press, 2015. – P.241-242.
- [19] Kessikbayeva, G., Cicekli I. Rule Based Morphological Analyzer of Kazakh Language // Linguistics and Literature Studies, 2016 . – P. 96-104.
- [20] Makhambetov O., Makazhanov A., Yessenbayev Zh., Sabyrgaliyev I., and Sharafudinov A. Towards a data-driven morphological analysis of Kazakh language // Türkiye bilisim vakfi Bilgisayar bilimleri ve mühendisliği dergisi. 2014. – P. 69-74.
- [21] Eryigit, G., Adalı, E. An affix stripping morphological analyzer for Turkish // In Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, 2004. – P. 299-304.
- [22] Hakkani-Tür, D. Z., Oflazer, K., Tür, G. Statistical morphological disambiguation for agglutinative languages // In Proceedings of COLING. ICCL, Staarburken, Germany, 2000. – P.285-291.
- [23] Sak H., Güngör T., and Saracclar M. A stochastic finite-state morphological parser for Turkish, // in Proceedings of the ACL-IJCNLP, 2009 Conference. Stroudsburg, PA, USA: ACL, 2009, – P.273-276.
- [24] Ilgen B., Adalı, E., Tantug A.C. Exploring feature sets for Turkish word sense disambiguation // Turkish Journal Of Electrical Engineering and Computer Sciences, 2015. – P.4391-4405.
- [25] Hankamer, J. Finite State Morphology and Left to Right Phonology // Proceedings of the West Coast Conference on Formal Linguistics Stanford University, 1986.– P.29-34.
- [26] Gruber, T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing // International Journal Human-Computer Studies Vol. 43, Issues 5-6, – P.907–928.

References

- [1] Kazakh grammar. Phonetics, word formation, morphology, syntax (in Kazakh). Astana (2002).
- [2] Lewis Geoffrey, Turkish Grammar: Oxford University Press, 2001.
- [3] Sharipbay A. and Bekmanova G., “The synthesis of word forms of Turkic language using semantic neural networks”, “*Modern problems of applied mathematics and information technologies: abstracts – Al Khorezmy*” (2009), 145.
- [4] Sharipbayev A. A. and Bekmanova G. T., “The building of logical semantics of the Kazakh words”, *The materials of the all-Russian conference with International participations “Knowledge-Ontology-Theory” (ZONT-09)* (Novosibirsk, 2009): 246-249.
- [5] Altntas K. and Cicekli I., “A Morphological Analyser for Crimean Tatar”, *Proceedings of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks* (TAINN’2001), (North Cyprus, 2001):180-189.
- [6] Tantug, A. C. and Adalı, E. and Oflazer, K., “Computer Analysis of the Turkmen Language Morphology” *Lecture Notes in Computer Science* (Springer, 2006): 186-193.
- [7] Orhun, M., Tantug, A. C., Adalı, E., “Rule Based Analysis of the Uyghur Nouns”, *International Journal of Assian Language Processing*, (2009):33-43.
- [8] Orhun, M. and Tantug A. and Adalı E., “Rule Based Tagging of the Uyghur Verbs”, *Fourth International Conference on Intelligent Computing and Information Systems, Faculty of Computer and Information Science*, Ain Shams University (Cairo, Egypt, 19-22 March , 2009).811-816.
- [9] Sulejmanov D.Sh., Nevezorova O.A, Gatiatullin A.R., Gilmullin R.A., Ayupov M.M., Pyatkin N.V., “The main components of the application of the grammatical model Tatar (In Russian)”, *In Proc. Dialogue Conference, Computational linguistics and intelligent technologies* (Russian State Humanitarian University, 2007): 525-530.
- [10] Bekmanova G. T., “Some approaches to the problems of automatic inflection and morphological analysis in the Kazakh language”, the newsletter of D. Serikbayev East Kazakhstan state technical university, (Ust-Kamenogorsk, 2009): 192-197.

- [11] Sharipbayev, A., Bekmanova, G., Mukanova, A., Buribayeva, A., Yergesh, B., Kaliyev, A., "Semantic neural network model of morphological rules of the agglutinative languages" (paper presented at The 6th International Conference on Soft Computing and Intelligent Systems The 13th International Symposium on Advanced Intelligent Systems, Kobe, Japan, November 20-24, 2012): 1094-1099.
- [12] Yergesh, B., Mukanova, A., Bekmanova, G., Sharipbay, A., Razakhova, B., "Semantic hyper-graph based representation of Verbs in the Kazakh language", *Computacion y Sistemas* (2014): 627-635.
- [13] Mukanova, A., Yergesh, B., Bekmanova G., Razakhova, B., Sharipbay, A., "Formal models of Verbs in the Kazakh language", *Leonardo Electronic Journal of Practices and Technologies* (2014): 264-273.
- [14] Zetkenbay, L., Sharipbay, A., Bekmanova, G., Kamanur, U., "Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages", *Journal of Theoretical and Applied Information Technology* (2016): 257-263.
- [15] Kamanur U., Sharipbay A., Altenbek G., Bekmanova G., Zhetkenbay L., "Investigation and use of methods for defining the extends of similarity of Kazakh language sentences", (paper presented at The 15th China National Conference, CCL 2016, and 4th International Symposium, NLP-NABD 2016, Yantai, China, October 15-16, 2016): 153-161.
- [16] Bekmanova G., Sharipbay A., Altnbek G., Adal? E., Zhetkenbay L., Kamanur U., Zulkhazhav A., "The uniform morphological analyzer for the Kazakh and Turkish languages", (paper presented at Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017), Moscow, Russia, July 2017): 20-30.
- [17] Sharipbay A., Mukanova A., Yergesh B., Zhetkenbay L., Zulkhazhav A., Yelibayeva G., "Ontology modeling of morphological rules of the Kazakh and Turkish languages", (paper presented at the Abstract of the VI international conference «Modern problems of applied mathematics and information technology - al-Khorezmiy» 2018): 51-52.
- [18] Tukeyev, U., Zhumanov, Zh., Rakhimova, D., Kartbayev, A. "Combinational Circuits Model of Kazakh and Russian Languages Morphology", (paper presented at the Abstracts of International Conference «Computational and Informational Technologies in Science, Engineering and Education». – Almaty: Al-Farabi KazNU Press, 2015): 241-242.
- [19] Kessikbayeva, G., Cicekli I., "Rule Based Morphological Analyzer of Kazakh Language", *Linguistics and Literature Studies* (2016): 96-104.,
- [20] Makhambetov O., Makazhanov A., Yessenbayev Zh., Sabyrqaliyev I., and Sharafudinov A. "Towards a data-driven morphological analysis of Kazakh language", *Türkiye bilisim vakfi Bilgisayar bilimleri ve mühendisliği dergisi*. (2014): 69-74.
- [21] Eryigit, G., Adali, E. "An affix stripping morphological analyzer for Turkish", (paper presented at the In Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, 2004): 299-304.
- [22] Hakkani-Tür, D. Z., Oflazer, K., Tür, G. "Statistical morphological disambiguation for agglutinative languages", *In Proceedings of COLING. ICCL*, (Staarburken, Germany, 2000): 285-291.
- [23] 23 Sak H., Gungor T., and Saracclar M. "A stochastic finite-state morphological parser for Turkish", *In Proceedings of the ACL-IJCNLP*, (Stroudsburg, PA, USA, 2009): 273-276.
- [24] Ilgen B., Adali E., Tantug A.C., "Exploring feature sets for Turkish word sense disambiguation", *Turkish Journal Of Electrical Engineering and Computer Sciences* (2015). 4391-4405. doi:10.3906/elk-1408-77.
- [25] Hankamer, J., " Finite State Morphology and Left to Right Phonology", (paper presented at the Proceedings of the West Coast Conference on Formal Linguistics Stanford University, 1986). 29-34.
- [26] Gruber, T.R., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", *International Journal Human-Computer Studies* (Vol. 43, Issues 5-6): 907-928.

МРНТИ 50.41.21

Разработка формата приложений для информационно-моделирующей системы экологического мониторинга

Искаков К.Т., Евразийский национальный университет имени Л.Н. Гумилева,
г.Астана, Республика Казахстан, E-mail: kazizat@mail.ru
 Хасенова З.Т., Евразийский национальный университет имени Л.Н. Гумилева,
г.Астана, Республика Казахстан, E-mail: zthasenova@mail.ru
 Мынбаева Б.Н., Казахский национальный педагогический университет имени Абая,
г.Алматы, Республика Казахстан, E-mail: bmynbayeva@gmail.com
 Оралбекова Ж.О., Евразийский национальный университет имени Л.Н. Гумилева,
г.Астана, Республика Казахстан, E-mail: oralbekova@bk.ru
 Мухтарова Ш.О., Казахский национальный педагогический университет имени Абая,
г.Алматы, Республика Казахстан, E-mail: shynarayum@gmail.com

Необходимость разработки формата приложений на основе современных технологий обусловлена актуальностью оптимального функционирования системы экологического мониторинга. Данная система состоит из несколько взаимодействующих модулей. Одна из них «модуль усвоения данных», которая моделирует процесс с применением алгоритма усвоения данных. Для оптимального моделирования необходимо исходные файлы форматом Network Common Data Form (NetCDF), который будет формироваться с разработанного приложения. В работе дается характеристика, особенности, структура формата NetCDF, представлены существующие утилиты и программные обеспечения, инструменты, программы визуализации по работе с данным форматом. Анализируются их назначение, функции. Приведены созданные файлы NetCDF с помощью математической программы Scilab, которая является альтернативой MATLAB. В работе представлена разработанная база данных по загрязнителям атмосферного воздуха. Описана структура основных таблиц : поля, тип данных, описание поля, свойства поля. Для создания базы данных использовалась система управления базами данных (СУБД) MS SQL Приведена реляционная модель созданной базы данных. Показано взаимодействие основных составляющих информационно-моделирующей системы.

Ключевые слова: данные, NetCDF, система управления базами данных, база данных, экологический мониторинг.

Экологиялық бақылаудың ақпараттық-моделдеу жүйесі үшін қосымшалар форматын әзірлеу

Ісқақов К.Т., Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail: kazizat@mail.ru
 Хасенова З.Т., Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail: zthasenova@mail.ru
 Мынбаева Б.Н., Абай атындағы Қазақ ұлттық педагогикалық университеті,
Алматы қ., Қазақстан Республикасы, E-mail: bmynbayeva@gmail.com
 Оралбекова Ж.О., Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail: oralbekova@bk.ru
 Мұхтарова Ш.Ө., Абай атындағы Қазақ ұлттық педагогикалық университеті,
Алматы қ., Қазақстан Республикасы, E-mail: shynarayum@gmail.com

Заманауи технологиялар негізінде қосымшалар форматын әзірлеу қажеттілігі экологиялық бақылау жүйесінің тиімді жұмыс істеуінің өзектілігімен анықталады. Бұл жүйе бірнеше өзара әрекеттесетін модулдерден тұрады. Оның бірі деректерді игеру алгоритмін қолдана отырып, процесті моделдейтін «деректерді игеру» модулі болып табылады. Тиімді моделдеу үшін әзірленген қосымшадан қалыптасатын Network Common Data Form (NetCDF) форматтағы бастапқы файлдар қажет. Бұл жұмыста NetCDF форматының сипаттамасы, ерекшеліктері, құрылымы сипатталды, берілген форматпен жұмыс істейтін қолданыстағы утилиттер және бағдарламалық қамтамалар, құралдар, визуализация бағдарламалары келтірілді. Олардың қызыметтері мен функциялары талданды. MATLAB-қа альтернатива болатын Scilab математикалық бағдарлама көмегімен құрылған NetCDF файлдары келтірілді. Жұмыста атмосфералық ауаның ластаушылары бойынша әзірленген деректер қоры көрсетілді. Негізгі кестелер құрылымы сипатталды: өріс, деректер типі, өріс сипаттамасы, өріс қасиеті. Деректер қорын құру үшін MS SQL деректер қорын басқару жүйесі (ДҚБЖ) қолданылды. Әзірленген деректер қорының реляциялық моделі келтірілді. Ақпараттық-моделдеу жүйесінің негізгі құрауыштарының өзара әрекеттесуі көрсетілді.

Түйін сөздер: деректер, NetCDF, деректер қорын басқару жүйелері, деректер қоры, экологиялық бақылау.

Development of format applications for information and modeling system environmental monitoring

Iskakov K.T., L.N.Gumilyov Eurasian national university, Astana, Kazakhstan,
E-mail: kazizat@mail.ru

Khassenova Z.T., L.N.Gumilyov Eurasian national university, Astana, Republic of Kazakhstan,
E-mail: zthasenova@mail.ru

Mynbayeva B.N., Abai Kazakh national pedagogical university, Almaty, Republic of Kazakhstan,
E-mail: bmynbayeva@gmail.com

Oralbekova Zh.O., L.N.Gumilyov Eurasian national university, Astana, Republic of Kazakhstan,
E-mail: oralbekova@bk.ru

Mukhtarova Sh.O., Abai Kazakh national pedagogical university, Almaty, Kazakhstan,
E-mail: shynaraym@gmail.com

The need to develop an application format based on modern technologies is determined by the relevance of the optimal functioning of the environmental monitoring system. This system consists of several interacting modules. One of them is the "data assimilation module", which simulates the process using the data assimilation algorithm. For optimal modeling, you need source files in the Network Common Data Form (NetCDF) format, which will be generated from the developed application. The paper describes the characteristics, features, structure of the NetCDF format, presents the existing utilities and software, tools, visualization programs for working with this format. Their purpose, functions are analyzed. The generated NetCDF files are shown using the Scilab math program, which is an alternative to MATLAB. The paper presents a developed database of air pollutants. The structure of the main tables is described: fields, data type, field description, field properties. Database management system (DBMS) MS SQL is used for creating the database. The relational model of created database is presented. The interaction of the main components of the information-modeling system is shown.

Key words: data, NetCDF, database management systems, database, environmental monitoring.

1 Введение

В настоящее время почти любая система осуществляет хранения и обработку огромного количества информации. В связи с обработкой этих данных, а также особенностями работы различного формата данных, соответственно требуется дополнительные разработки. В системе экологического мониторинга информационно-моделирующая подсистема занимает основное место, содержащая в себе хранения данных, расчетные

и аналитические блоки, где функциональные блоки реализованы в виде модулей, например: модуль «усвоения данных», модуль «метеопараметры», модуль «анализа данных» и т.д. Для оптимального расчета и моделирования необходимо создать приложение, которое сформирует файлы формата NetCDF.

Таким образом, одной из проблем, с которой сталкивается при моделировании является работа с пространственно-временными данными формата NetCDF. Данный формат разработан при поддержке объединения университетов в области исследований атмосферы University Corporation for Atmospheric Research. NetCDF – двоичный формат файлов, являющийся открытым стандартом для обмена научными данными [1].

Преимущества формата NetCDF: облегчение обмена массивами данных между разными приложениями и компьютерами, уменьшение программных сложностей и ошибок при интерпретации данных и вспомогательной информации, утверждение стандартного интерфейса [2]-[11].

2 Обзор литературы

В работах [12]-[13] приведены результаты экспериментальных работ с датчиками, которые обеспечивают многомерную и многопараметрическую информацию об осадках, для хранения данных ими используются универсальный формат NetCDF. На сегодняшний день формат NetCDF также широко используется для обмена данными и в атмосферных исследованиях, например, для хранения таких данных, как, температура, скорость ветра, атмосферное давление, направление ветра, влажность и т.п. [14]-[18], а также в задачах геофизики, океанологии [19]-[23]. Нами с помощью формата netCDF, все данные нашей информационно-моделирующей системы экологического мониторинга можно хранить в файле данного формата для дальнейших расчетов.

3 Материалы и методы

Структура файла netCDF состоит из следующих трех основных компонентов: размерность (dimensions); переменные (variables); атрибуты (attributes), которая схематично представлена на Рис. 1.

С подробным описанием об основных компонентах файла NetCDF можно ознакомиться в работе [24]-[25].

Для решения проблемы по обработке и управлению пространственно-временными данными формата NetCDF существуют разные утилиты и программные обеспечения, которые и анализируются и приводится в данной работе.

Так как NetCDF имеет двоичный код, в связи с этим для визуализации данного кода существует текстовое представление Common Data form Language (CDL) и программное обеспечение для конвертации исходных данных в текст.

Одна из самых простых программ для просмотра содержания файла является ncBrowse. Это бесплатная кроссплатформенная программа, которая дает возможность представить данные в графическом виде и перевести в формат ASCII на языке CDL.

При исследовании и решении разных тип задач, в том числе экологического характера необходимо работа с файлами, которые требуют осреднение и выборку по различным осям, интерполяцию полей, установку временной оси, объединение

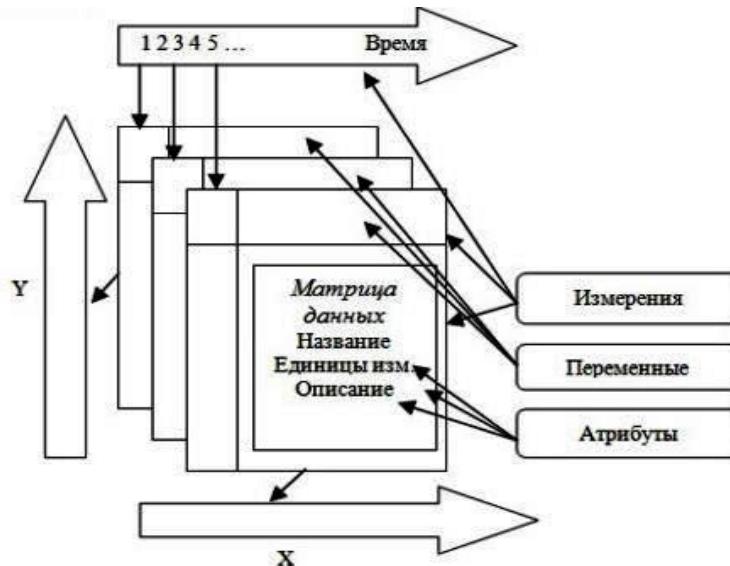


Рисунок 1: Схематичная структура формата файла NetCDF

и разделение файлов. Для этого можно использовать простую для понимания синтаксисом программу Climate Data Operators (CDO) на C++, позволяющая манипулировать файлами.

Для выдачи подробную информацию о файле, например, его типа, пользователя, статус защиты можно использовать инструмент анализа файлов FileViewPro. Также существуют инструменты NetCDF.view и NetCDF Reader, которые могут прочитать заголовки, извлекать из файла данные и сделать проекцию на любые измерения.

Кроме предложенных утилит и программ имеющие возможности работать с файлами NetCDF, среди созданных к настоящему времени программ визуализации и обработки данных формата можно отметить следующие: Panoply; GMT; NCL; NCO; GrADS; Open-GrADS.

В рамках исследования по моделированию переноса и трансформации примесей в атмосфере использовались файлы формата NetCDF. В ходе работы визуализация данных осуществлялась с помощью Panoply. Данная программа позволила увидеть все переменные, размерности и атрибуты, а также графики. Сформированный файл по данным города Алматы приведен на Рис.2.

В файле *cityCoeff_apriori_mean_Ref_traffic.nc* хранятся значения коэффициентов для моделирования переноса и трансформации примесей в атмосфере. Каждого коэффициента (переменных) можно посмотреть через построения графики. Пример визуализации коэффициента «U10» приведен на Рис. 3.

При формировании графика, например для «U10» – скорости ветра, можно выбрать ту или иную размерность. У этой переменной три размерности: скорость ветра по двум направлениям и время, то есть каждая переменная в NetCDF – это обычный многомерный массив.

Исходными файлами для получения результатов расчета являются файлы данного типа, следовательно, для создания файла была использована математическая

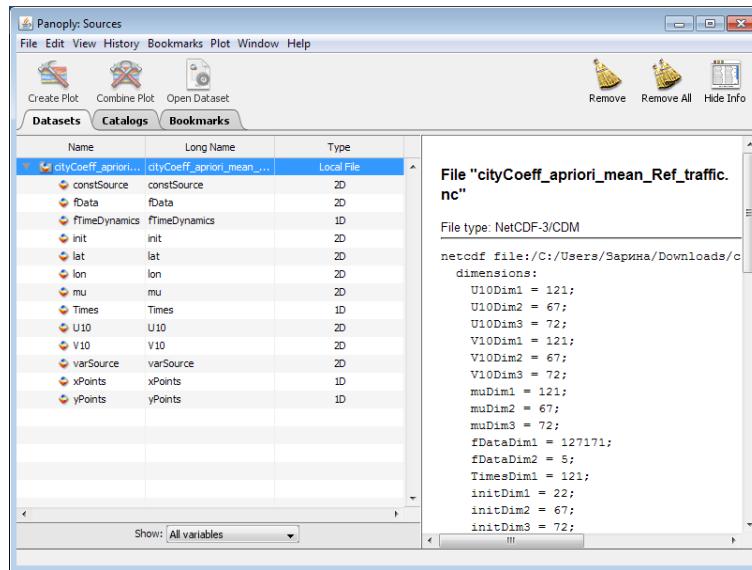


Рисунок 2: Сформированный файл NetCDF

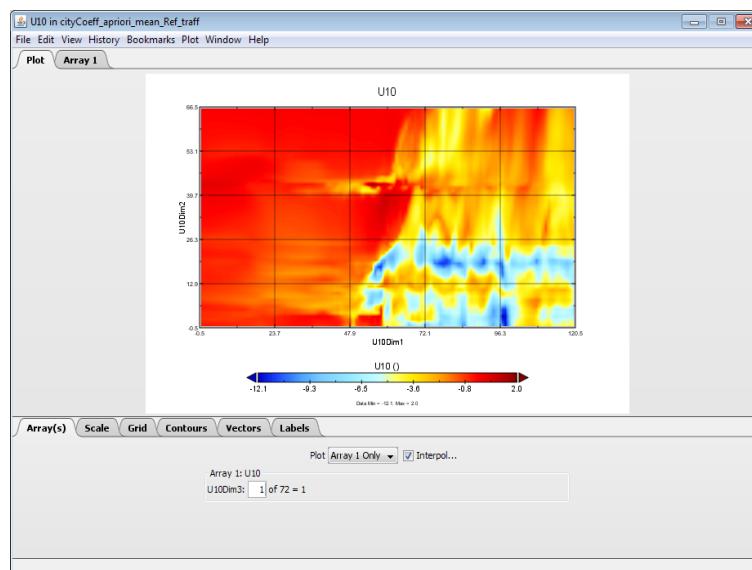


Рисунок 3: Пример визуализации коэффициента «скорость ветра»

программа Scilab. Scilab самая полная общедоступная альтернатива MATLAB для технических и научных расчетов. Для работы с форматом NetCDF подключается модуль sciNetCDF. Один из примеров результата созданного файла в Scilab представлен на Рис. 2.

Для создания приложения, формирующего файлов такого формата, создана база данных (БД) в качестве исходных данных. Всех данных можно условно классифицировать на четыре блока:

1. сведения об измерениях концентрации примеси конкретных тяжелых металлов в

атмосфере;

2. сведения о ветре;

3. информация о пользователях и их ролях в системе;

4. системная информация.

Для каждой таблицы базы данных определены название полей (атрибуты), типы данных, описание и свойства поля.

Приведем таблицы блока Pollution «сведения об измерениях концентрации примеси конкретных тяжелых металлов в атмосфере».

Блок Pollution состоит из 3 таблиц.

Таблица 1 содержит основную информацию о точках сбора данных.

Таблица 1: DataCollectionPoint

Поля:	Тип данных	Описание поля	Свойства поля
id	bigint	Идентификатор точки сбора данных	Ключевое поле
name	varchar	Наименование точки сбора данных	обязательное поле
lon	float	долгота	обязательное поле
lat	float	Широта	обязательное поле
cmt_ru	text	комментарии	необязательное поле

Таблица 2 содержит данных об измерениях в точках сбора информации.

Таблица 3 содержит данные о тяжелых металлах и других загрязнителях в атмосферном воздухе.

Другие таблицы остальных блоков определены аналогичным образом. Приведенные названия БД, таблиц и полей являются рабочими и могут быть изменены, а также дополняться другими таблицами. Таким образом, создана БД с целью сбора и обработки данных, реляционная модель которой приведена на Рис. 4. БД создана с помощью системой управления базами данных (СУБД) MS SQL. Во избежание проблем с переносом и совместимостью имен использованы английские названия всех объектов базы данных.

4 Результаты и обсуждение

Созданная база данных является источником для приложения, которое будет формировать файлы формата NetCDF. Сформировавшиеся файлы необходимы для модуля «усвоение данных» системы, где основой является математическое моделирование, то есть применение алгоритма усвоение данных в режиме реального времени. Фрагмент процесса работы системы можно представить следующим образом, как показано на Рис.5.

Таблица 2: Measurements

Поля:	Тип данных	Описание поля	Свойства поля
id	bigint	Идентификатор записи	Ключевое поле
point	bigint	Точки сбора данных	Обязательное, ключевое
dBeg	datetime	Время начала проведенного замера	обязательное
dEnd	datetime	Время конца проведенного замера	обязательное
element	bigint	Загрязнитель	Обязательное, ключевое
temperature	float	Температура воздуха во время измерения	обязательное
pressure	float	Давление воздуха во время измерения	обязательное
value	float	Значение	обязательное

Таблица 3: PollutionElements

Поля:	Тип данных	Описание поля	Свойства поля
id	bigint	Идентификатор записи	Ключевое поле таблицы
name	varchar	Наименование загрязнителя	Обязательное, внешний ключ
alias	varchar	Вторичное наименование	Необязательное
concentration	float	Концентрация загрязнителя в атмосфере	Обязательное, внешний ключ

5 Заключение

Для полностью функционирования системы для оптимизации мониторинга загрязнения атмосферного воздуха тяжелыми металлами полученные результаты являются основным и начальным этапом сложного процесса разработки целой системы. Использование современных технологий для работы с пространственно-временными данными повысит эффективность информационно-моделирующей системы в управлении качеством атмосферного воздуха.

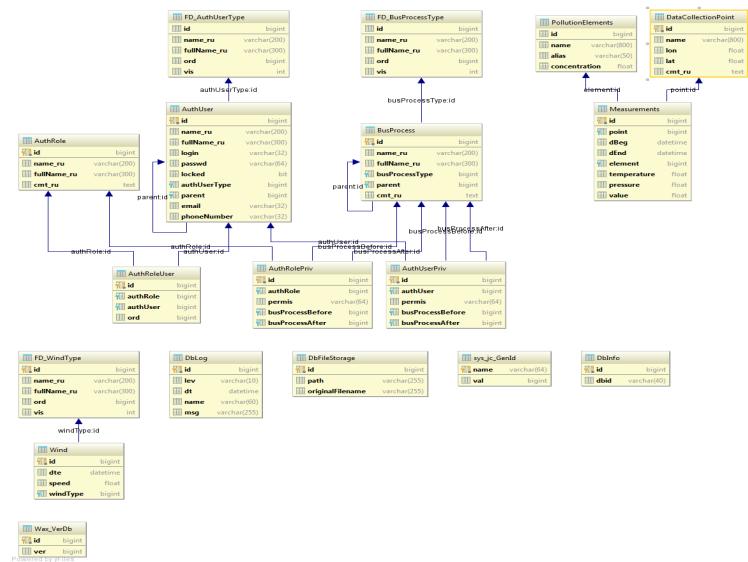


Рисунок 4: Реляционная модель базы данных

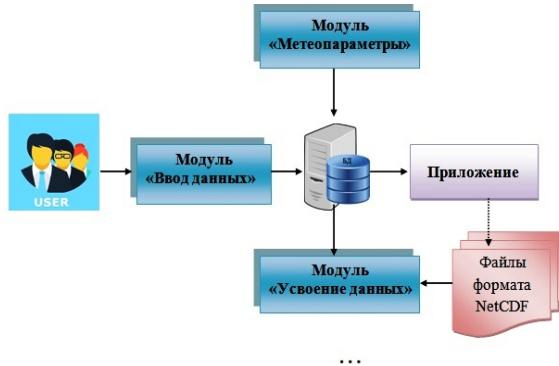


Рисунок 5: Фрагмент процесса работы системы

6 Благодарности

Работа выполнена при поддержке грантового финансирования научно-технических программ и проектов Министерством образования и науки Республики Казахстан (грант №AP05135992, 2018-2020 годы).

Список литературы

- [1] How to make a website. Unidata. NetCDF. Network Common Data Form NetCdf [Электрон. ресурс]. - URL: <http://website.fishes-aquarium.ru> (дата обращения: 20.09.2018г.)
- [2] Белоушико К.Е. Формат NetCDF как стандарт для обмена данными в атмосферных исследованиях [Электрон. ресурс]. - 2018. - URL: http://www.izmiran.ru/POLAR2012/REPORTS/POLAR_2012_Belouschko2.pdf (дата обращения: 20.08.2018г.)

- [3] *Bartz C., Chasapis K., Kuhn M., Nerge P., Ludwig T.* A best practice analysis of HDF5 and NETCDF-4 using luster // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – 2015. – №9137. – P.274-281.
- [4] *Yu J., Car N.J., Leadbetter A., Simons B.A., Cox, S.J.* Towards linked data conventions for delivery of environmental data using netCDF // IFIP Advances in Information and Communication Technology. – 2015. – №448 – P.102-112.
- [5] *Car N.J., Ip A., Druken K.* NetCDF-LD SKOS: Demonstrating linked data vocabulary use within NetCDF-compliant files // IFIP Advances in Information and Communication Technology. – 2017. – №507. – P.329-337.
- [6] *Zhuang K., North G.R., Stevens M.J.* A NetCDF version of the two-dimensional energy balance model based on the full multigrid algorithm // SoftwareX. – 2017. - №6. – P.198-202.
- [7] *Visheratin A.A., Rakitin S.V., Ageev D.V., Nasonov D., Boukhanovsky A.V.* Exarch -semantics-based modular distributed data storage // ACM International Conference Proceeding Series. – 2017. – P. 22-26.
- [8] *Scarpioni P., Coro G., Pagano P.* A collection of Aquamaps native layers in NetCDF format // Data in Brief. – 2018. - №17. – P.292-296.
- [9] *Hossain M.M., Wu R., Painumkal J.T., Kettouch M., Luca C., Dascalu S.M., Harris F.C.* Web-service framework for environmental models // Internet Technologies and Applications, ITA 2017: proceedings of the 7th International Conference. – 2017. – P.104-109.
- [10] *Rew R., Davis G.* NetCDF: An Interface for Scientific Data Access // IEEE Computer Graphics and Applications. – 1990. – V.10, №4. – P.76-82.
- [11] *Rew R.K., Hartnett E.J.* Merging NET CDF and HDF5 // Proceedings of the 21st International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology. – 2004. – P.1-11.
- [12] *Wingo S.M., Petersen W.A., Gatlin P.N., Pabla C.S., Marks, D.A.* The System for Integrating Multiplatform Data to Build the Atmospheric Column (SIMBA) precipitation observation fusion framework // Journal of Atmospheric and Oceanic Technology. – 2018. - №35. – P.1353-1374.
- [13] *Liu H., Van Oosterom P., Tijssen T., Commandeur T., Wang W.* Managing large multidimensional hydrologic datasets: A case study comparing NetCDF and SciDB // Journal of Hydroinformatics. – 2018. - №20. –P.1058-1070.
- [14] *Mbogo G.-K., Rakitin S.V., Visheratin A.* High-performance meteorological data processing framework for real-time analysis and visualization // Procedia Computer Science. – 2017. - №119. – P.334-340.
- [15] *Husar R.B., Hoijarvi K.* DataFed: Mediated web services for distributed air quality data access and processing // International Geoscience and Remote Sensing Symposium (IGARSS). – 2007. – P.4016-4019.
- [16] *Kim J., Kwon Y.C., Kim T.-H.* A Scalable High-Performance I/O System for a Numerical Weather Forecast Model on the Cubed-Sphere Grid // Asia-Pacific Journal of Atmospheric Sciences. – 2018. -№54. – P.403-412.
- [17] *Meng J., Ding G., Liu L., Zhang R.* A comparison and validation of atmosphere CO₂concentration OCO-2-based observations and TCCON-based observations // Communications in Computer and Information Science. – 2016. - №645. – P.356-363.
- [18] *Young A.H., Knapp K.R., Inamdar A., Hankins W., Rossow W.B.* The International Satellite Cloud Climatology Project H-Series climate data record product // Earth System Science Data. – 2018. – V.10, №1. –P.583-593.
- [19] *Loaiza J.M.C., Giuliani G., Fiameni G.* Big-data in climate change models - A novel approach with hadoop mapreduce // Proceedings - 2017 International Conference on High Performance Computing and Simulation, HPCS 2017. – 2017. - №8035056. – P.45-50.
- [20] *Iona A., Theodorou A., Sofianos S., Troupin C., Troupin C., Beckers J.-M.* Mediterranean Sea climatic indices: Monitoring long-term variability and climate changes // Earth System Science Data. – 2018. – V.10, №4. – P.1829-1842.
- [21] *Chiriaco M., Dupont J.-C., Bastin S., Badosa J., Lopez J., Haeffelin M., Chepfer H., Guzman R.* ReOBS: A new approach to synthesize long-term multi-variable dataset and application to the SIRTA supersite // Earth System Science Data. – 2018. – V.10, №2. – P.919-940.
- [22] *Shutler J.D., Land P.E., Piolle J.-F., Woolf D.K., Goddijn-Murphy L., Paul F., Girard-Ardhuin F., Chapron B., Donlon C.J.* FluxEngine: A flexible processing system for calculating atmosphere-ocean carbon dioxide gas fluxes and climatologies // Journal of Atmospheric and Oceanic Technology. – 2016. – V.33, № 4. – P.741-756.

- [23] Rodriges Zalipynis R.A. Array DBMS in environmental science: Satellite sea surface height data in the cloud // Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017. -2017. – V.2, №8095248. –P.1062-1065.
- [24] ArcGIS for Desktop. Базовые сведения о хранилище данных NetCdf [Электрон. ресурс].- 2016. - URL: <http://desktop.arcgis.com/ru/arcmap/10.3/manage-data/netcdf/what-is-netcdf-data.htm> (дата обращения: 21.08.2018 г.)
- [25] Rew R., Davis G., Emmerson S., Davies H., Hartne E. The NetCDF Users Guide [Электрон. ресурс]. -2009. - URL: https://www.unidata.ucar.edu/software/netcdf//old_docs/docs_4_0_1/netcdf.pdf (дата обращения: 21.08.2018 г.)

References

- [1] How to make a website. Unidata. NetCDF. Network Common Data Form NetCdf, Accessed August 20, 2018, <http://website.fishes-aquarium.ru>.
- [2] Beloushko K.E. Format NetCDF kak standart dlya obmena dannymi v atmosfernyh issledovaniyah [NetCDF format as a standard for data exchange in atmospheric research]. - 2018.- Accessed August 20, 2018, http://www.izmiran.ru/POLAR2012/REPORTS/POLAR_2012_Beloushko2.pdf.
- [3] Bartz C., Chaspis K., Kuhn M., Nerge P., Ludwig T. "A best practice analysis of HDF5 and NETCDF-4 using lustre,"Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9137 LNCS (2015): 274-281.
- [4] Yu J., Car N.J., Leadbetter A., Simons B.A., Cox, S.J. "Towards linked data conventions for delivery of environmental data using netCDF,"IFIP Advances in Information and Communication Technology, 448 (2015): 102-112.
- [5] Car N.J., Ip A., Druken K. "NetCDF-LD SKOS: Demonstrating linked data vocabulary use within NetCDF-compliant files,"IFIP Advances in Information and Communication Technology 507 (2017): 329-337.
- [6] Zhuang K., North G.R., Stevens M.J. "A NetCDF version of the two-dimensional energy balance model based on the full multigrid algorithm,"SoftwareX 6 (2017): 198-202.
- [7] Visheratin A.A., Rakitin S.V., Ageev D.V., Nasonov D., Boukhanovsky A.V. "Exarch -semantics-based modular distributed data storage,"ACM International Conference Proceeding Series (2017): 22-26.
- [8] Scarponi P., Coro G., Pagano P. "A collection of Aquamaps native layers in NetCDF format,"Data in Brief 17 (2018): 292-296.
- [9] Hossain M.M, Wu R., Painumkal J.T., Kettouch M., Luca C., Dascalu S.M., Harris F.C. "Web-service framework for environmental models,"Internet Technologies and Applications, ITA 2017 - Proceedings of the 7th International Conference (2017): 104-109.
- [10] Rew R., Davis G. "NetCDF: An Interface for Scientific Data Access,"IEEE Computer Graphics and Applications 10(4) (1990): 76-82.
- [11] Rew R.K., Hartnett E.J. "Merging NET CDF and HDF5,"Proceedings of the 21st International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology (2004): 1-11.
- [12] Wingo S.M., Petersen W.A., Gatlin P.N., Pabla C.S., Marks, D.A. "The System for Integrating Multiplatform Data to Build the Atmospheric Column (SIMBA) precipitation observation fusion framework,"Journal of Atmospheric and Oceanic Technology 35 (2018): 1353-1374.
- [13] Liu H., Van Oosterom P., Tijssen T., Commandeur T., Wang W. "Managing large multidimensional hydrologic datasets: A case study comparing NetCDF and SciDB,"Journal of Hydroinformatics 20 (2018): 1058-1070.
- [14] Mbogo G.-K., Rakitin S.V., Visheratin A. "High-performance meteorological data processing framework for real-time analysis and visualization,"Procedia Computer Science 119 (2017): 334-340.
- [15] Husar R.B., Hoijarvi K. "DataFed: Mediated web services for distributed air quality data access and processing,"International Geoscience and Remote Sensing Symposium (IGARSS) (2007): 4016-4019.
- [16] Kim J., Kwon Y.C., Kim T.-H. "A Scalable High-Performance I/O System for a Numerical Weather Forecast Model on the Cubed-Sphere Grid,"Asia-Pacific Journal of Atmospheric Sciences 54 (2018): 403-412.

-
- [17] Meng J., Ding G., Liu L., Zhang R. "A comparison and validation of atmosphere CO₂concentration OCO-2-based observations and TCCON-based observations,"Communications in Computer and Information Science 645 (2016): 356-363.
 - [18] Young A.H., Knapp K.R., Inamdar A., Hankins W., Rossow W.B. "The International Satellite Cloud Climatology Project H-Series climate data record product,"Earth System Science Data 10(1) (2018): 583-593.
 - [19] Loaiza J.M.C., Giuliani G., Fiameni G. "Big-data in climate change models - A novel approach with hadoop mapreduce,"Proceedings - 2017 International Conference on High Performance Computing and Simulation, HPCS 2017 8035056 (2017): 45-50.
 - [20] Iona A., Theodorou A., Sofianos S., Troupin C., Troupin C., Beckers J.-M. "Mediterranean Sea climatic indices: Monitoring long-term variability and climate changes,"Earth System Science Data 10(4) (2018): 1829-1842.
 - [21] Chiriaco M., Dupont J.-C., Bastin S., Badosa J., Lopez J., Haeffelin M., Chepfer H., Guzman R. "ReOBS: A new approach to synthesize long-term multi-variable dataset and application to the SIRTA supersite,"Earth System Science Data 10(2) (2018): 919-940.
 - [22] Shutler J.D., Land P.E., Piolle J.-F., Woolf D.K., Goddijn-Murphy L., Paul F., Girard-Ardhuin F., Chapron B., Donlon C.J. "FluxEngine: A flexible processing system for calculating atmosphere-ocean carbon dioxide gas fluxes and climatologies,"Journal of Atmospheric and Oceanic Technology 33(4) (2016): 741-756.
 - [23] Rodriges Zalipynis R.A. "Array DBMS in environmental science: Satellite sea surface height data in the cloud,"Proceedings of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2017 2(8095248) (2017): 1062-1065.
 - [24] ArcGIS for Desktop. Bazovye svedeniya o hraniashche dannyh NetCdf [Basic information about data storage NetCdf] (2016), Accessed August 21, 2018, <http://desktop.arcgis.com/ru/arcmap/10.3/manage-data/netcdf/what-is-netcdf-data.htm>.
 - [25] Rew R., Davis G., Emmerson S., Davies H., Hartne E. The NetCDF Users Guide (2009), Accessed August 21, 2018, https://www.unidata.ucar.edu/software/netcdf//old_docs/docs_4_0_1/netcdf.pdf.

МРНТИ 20.53.19

Подходы к разработке информационной системы мониторинга мнений и оценки социального самочувствия

Мамыкова Ж.Д., Казахский национальный университет им. аль-Фараби,

г. Алматы, Республика Казахстан, E-mail: Zhanl.Mamykova@kaznu.kz

Мутанов Г.М., Казахский национальный университет им. аль-Фараби,

г. Алматы, Республика Казахстан, E-mail: Rector@kaznu.kz

Сундетова Ж.Т., Казахский национальный университет им. аль-Фараби,

г. Алматы, Республика Казахстан, E-mail: Zhanerke.Sundetova@kaznu.kz

Торекул С.М., Казахский национальный университет им. аль-Фараби,

г. Алматы, Республика Казахстан, E-mail: Saule.Torekul@ kaznu.kz

Стремительное распространение социальных онлайн-сервисов и развитие технологий Больших Данных инициировали интерес к использованию сведений из социальных сетей в различных отраслях. Сегодня приобретают популярность технологии "мониторинга социальных сетей"(social listening) и контент-анализа. Данные сервисы в основном представлены зарубежными разработками, лингвистические словари, которых построены на английском языке, и они плохо адаптированы под русский и казахский языки.

В представленной статье описан процесс проектирования и разработки информационной системы для "мониторинга социальных сетей адаптированной под специфику казахского и русского языка, а также форм сленгового общения, с целью определения эмоционального окраса пользовательского восприятия контента, и оценки социального самочувствия в стране/регионе. В работе представлен алгоритм информационной системы с описанием функциональных возможностей, а также методы и модели. Кроме этого, рассматривается продукционная модель оценки социального самочувствия и модель семантического профиля восприятия обществом событий. Проведен сравнительный анализ разработок в вопросах мониторинга мнений для выявления сильных и слабых сторон. Информационная система дает возможность проводить мониторинг СМИ, социальных сетей, анализа управления репутацией, анализа пользовательского восприятия контента сети Интернет на региональном, республиканском и международном уровне.

Ключевые слова: мониторинг социальных сетей, анализ комментариев, социальное самочувствие, оценка пользовательского восприятия.

Пікірлерге мониторинг жүргізу және әлеуметтік көніл күйді бағалау ақпараттық жүйесін жобалау және құрастырыу

Мамыкова Ж.Д., әл-Фараби атындағы Қазақ ұлттық университеті,

Алматы қ., Қазақстан Республикасы, E-mail: Zhanl.Mamykova@kaznu.kz

Мутанов Г.М., әл-Фараби атындағы Қазақ ұлттық университеті,

Алматы қ., Қазақстан Республикасы, E-mail: Rector@kaznu.kz

Сундетова Ж.Т., әл-Фараби атындағы Қазақ ұлттық университеті,

Алматы қ., Қазақстан Республикасы, E-mail: Zhanerke.Sundetova@kaznu.kz

Төреқұл С.М., әл-Фараби атындағы Қазақ ұлттық университеті,

Алматы қ., Қазақстан Республикасы, E-mail: Saule.Torekul@ kaznu.kz

Әлеуметтік онлайн-сервистердің кең таралуы және Улken Деректер технологияларының дамуы әлеуметтік желілердегі деректерді әртүрлі салаларда қолдануға деген қызыгуышылық тудырды. Казіргі таңда контент анализ және "әлеуметтік желілерді мониторлау" технологиялары танымалдылыққа ие болуда. Мұндай сервистердің көпшілігі штедлік әзірлемелер болып табылады және лингвистикалық сөздіктер ағылшын тілінде құрастырылған, сондықтан қазақ және орыс тілдеріне бейімделген.

Мақалада елдің/аумақтың әлеуметтік көңіл күйін бағалау және пайдаланушының контентті қабылдауының эмоционалды түсін анықтауға арналған, қазақ және орыс тілдерінің, сонымен қатар ауызекі сөйлеу формаларының ерекшеліктеріне бейімделген "әлеуметтік желілерді мониторлау" ақпараттық жүйесін жобалау және әзірлеу процесі көрсетілген. Жұмыста ақпараттық жүйе алгоритмінің функционалдық мүмкіндіктерінің сипаттамасы және моделдері мен әдістері көрсетілген. Сонымен қатар, әлеуметтік көңіл күйіді бағалаудың өнімдік моделі және қоғамның оқиғаларды қабылдауының семантикалық моделі қарастырылған. Мықты және әлсіз тұстарын анықтау мақсатында пікірлерді мониторлау саласындағы әзірлемелерге салыстырмалы анализ жүргізілген. Ақпараттық жүйе аудандық, республикалық және халықаралық деңгейде БАҚ-ты және әлеуметтік желілерді мониторлауға, беделдерді басқаруға анализ жасауга, интернет контентті пайдаланушының қабылдауына анализ жасауга мүмкіндік береді.

Түйін сөздер: әлеуметтік желілерді мониторлау, комментарийлерге анализ жасау, әлеуметтік көңіл-куй, пайдаланушының қабылдауын бағалау.

Approaches of the development of information monitoring system of social wellness opinions and evaluation

Mamykova Zh.D., al-Farabi Kazakh National university,
Almaty, Republic of Kazakhstan, E-mail: Zhanl.Mamykova@kaznu.kz

Mutanov G.M., al-Farabi Kazakh National university,
Almaty, Republic of Kazakhstan, E-mail: Rector@ kaznu.kz

Sundetova Zh.T., al-Farabi Kazakh National university,
Almaty, Republic of Kazakhstan, E-mail: Zhanerke.sundetova@kaznu.kz

Torekul S.M., al-Farabi Kazakh National university,
Almaty, Republic of Kazakhstan, E-mail: Saule.Torekul@ kaznu.kz

A dynamic spreading of social on-line services and development of the Big Data technologies have caused interest to use information from social media in different spheres. Nowadays the "social listening" and content analysis technologies gain popularity. The given services are basically represented by foreign development works, where linguistic dictionaries are made in English and they are badly adapted to Russian and Kazakh.

The given article describes the process of designing and developing the information system to monitor "social media adapted to the specific character of the Kazakh and Russian languages and forms of slang communication as well to define the emotional coloring of a user's perception of the content and evaluation of the social well-being in the region and country. The work represents the algorithm of the information system describing functionalities and methods and models as well. Besides, a production model of social well-being evaluation and a semantic profile model for conceiving events by society has been considered. A comparative analysis of the development works has been made in the issues of opinions monitoring to reveal strengths and weaknesses. The information system gives an opportunity to monitor mass and social media, to analyze reputation management, to analyze a user's perception of the Internet content on the regional and a nationwide scale.

Key words: social media monitoring, comment analysis, social well-being, a user's perception evaluation.

1 Введение

В результате развития социальных сетей, СМИ, блогов, веб-ресурсов появилась острая необходимость, в инструменте определения оценки эмоционального окраса

тональности текстов, сообщений, возможность проведения репутационного аудита в режиме реального времени, выявления репутационных и информационных угроз, критик, негатива, дезинформации. Совместное использование неструктурированных и контентных данных потенциально позволяет применять социальные сети для решения широкого круга бизнес-задач: маркетинговая деятельность (мнение потребителей, анализ отзывов), управления брендом, мониторинг для бизнеса для отслеживания ключевых слов в режиме реального времени от основных новостных агентств и социальных медиа-платформ и др.

Проблемы систем мониторинга СМИ: слабое применение данных технологий для анализа интернет-пространства и оценки социального самочувствия в обществе; любая ситуация мониторится вручную; субъективизм в интерпретации данных событий; отсутствие практики применения методов Data Mining для обработки информационного массива потоков СМИ; отсутствие системы кодирования, позволяющей классифицировать текстовые сообщения в соответствии с категориями событий; сервисы мониторинга социальных сетей в основном представлены зарубежными разработками, лингвистические словари, которых построены на английском языке; отечественные системы мониторинга и анализа социальных сетей, плохо адаптированы под казахский язык.

В связи с чем, особую актуальность приобретает разработка Информационной системы мониторинга мнений и оценки социального самочувствия (далее - Система).

Основной целью разработки Системы является обеспечение сервиса мониторинга мнений интернет-пространства для проведения глубинного анализа неструктурированной информации, с целью извлечения знаний и получения важной информации из текстовых массивов по оценке пользовательского восприятия контента, и оценки социального самочувствия общества.

Для разработки системы анализа пользовательского восприятия контента социальных сетей, адаптированной под специфику казахского и русского языка, а также форм сленгового общения были поставлены следующие задачи: разработать модуль подключения к API-социальных сетей, разработать модуль управления тональным словарем, разработать модуль хранения данных для обработки и анализа результатов, разработать модуль моделирования социального самочувствия, программная реализация системы.

Реализация Системы позволит мониторить веб-ресурсы и социальные сети с подсистемой моделирования социального самочувствия общества, посредством оценки тональности сообщений/мнений.

2 Обзор литературы

Анализу проблем функционирования программных инструментов [1], предназначенных для мониторинга социальных сетей посвящены труды зарубежных ученых, которые позволили в полной степени рассмотреть характеристики программного обеспечения [2], типы социальных сетей, включая языковую поддержку [3], инструменты и сервисы, предназначенные для мониторинга социальной сети с использованием современных методов анализа синтаксиса и семантики на примере китайского языка [4]. В аспекте новых моделей мониторинга социальных сетей была рассмотрена модель для

мониторинга того, как социальные связи и социальная структура меняются внутри группы [5], абстрактная модель для сравнения метрик с одинаковым значением в разных социальных сетях [6] и концептуальные модели [7,8]. Проанализирован алгоритм мониторинга [9], который анализирует потоки из социальных сетей, сочетая в себе различные подходы машинного обучения с легкой степенью адаптированности и обогащен пользовательскими знаниями без необходимости сложной настройки, алгоритм сегментации по определению новых ключевых слов, где в качестве данных используются твиты, собранные из Twitter [10], алгоритм MMCRD (maximum monitoring contribution rate deployment) [11] и структура с открытым исходным кодом для мониторинга, анализа и поиска в нескольких социальных сетях [12].

Исследования в рамках разработки новых методов и методологии мониторинга социальных сетей представлены как методология информационного взаимодействия в социальных системах [13], методы измерения: пассивные, активные [14], с применением методов теории систем и кибернетики [15]. Методология, которая позволяет справиться с дефицитом данных при мониторинге реальных событий с использованием данных социальных сетей [16]. С точки зрения социального эффекта была изучена методология мониторинга социальных сетей, предназначенная для борьбы с кризисами (лесные пожары, вспышек заболеваний, преступления) [17].

Данные подходы позволили определить архитектуру разрабатываемой Системы с акцентом на построение лингвистического словаря для казахского языка и разработки алгоритмов чтения контента с "пониманием". Изучение зарубежного опыта нашло отражение в определении подходов в разработке алгоритмов машинного обучения для казахского языка и разработку интеллектуальной составляющей как платформы анализа данных.

3 Материал и методы

Происходящие глобальные изменения, связанные с интернет-технологиями и сетевыми инновациями, объясняют необходимость разработки инструментария и моделей для мониторинга общественного мнения и оценки социального самочувствия как одного из главных инструментов оценки по состоянию социально-экономической системы в контексте устойчивого развития.

На зарубежном рынке уже представлены инструменты по мониторингу социальных сетей и контент-анализу, как Hootsuite [18], Socialmention [19], BuzzLook [20], IQBuzz [21], Socialbakers [22], Keyhole [23], Brandwatch [24], YouScan [25] и другие аналоги, которые позволяют, применять социальные сети для решения широкого круга бизнес-задач: борьбы с мошенничеством, управления брендом, рекламы товаров и услуг, формирования новых каналов сбыта и др.

Рынок Казахстана не богато представлен отечественными разработками в вопросах мониторинга мнений (ИС "iMAS"[26], "Alem Media Monitoring"[27]) и разработка зарубежных компаний (ИС "Brand Analytics"[28], "Microsoft engagement"[29]).

Таблица 1: Функциональные возможности программных продуктов.

	iMAS	Alem Media Monitoring	Предлагаемая авторами решение	Brand Analytics	Microsoft engagement
1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
2	<input checked="" type="checkbox"/>				
3	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5	<input checked="" type="checkbox"/>				
6	<input checked="" type="checkbox"/>				
7			<input checked="" type="checkbox"/>		
8	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>
10			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>
11			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
12			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
13			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
14			<input checked="" type="checkbox"/>		
15			<input checked="" type="checkbox"/>		
16			<input checked="" type="checkbox"/>		

Примечание (модули систем): 1. Поддержка казахского языка; 2. Поддержка русского языка; 3. Поддержка английского языка; 4. Географическо-интерактивная карта мнений; 5. Мониторинг социальных сетей; 6. Мониторинг интернет-ресурсов; 7. Моделирование социального самочувствия; 8. Набор слов-исключений; 9. Сервис настройки каталога ресурсов; 10. Сервис добавления новых ресурсов; 11. Сервис настройки профиля правил для темы поиска; 12. Детализация анализа; 13. Сервис настройки выгрузки данных; 14. Отчеты по KPI системы; 15. Возможность приглашать "эксперта"; 16. Модуль управления тональным словарем.

Согласно данным Таблицы 1 все программные продукты поддерживают 3 языка (русский, английский и казахский) исключением является Microsoft engagement, у которого нет поддержки казахского языка. Для всех программных продуктов мониторинга мнений источниками являются социальные сети и интернет-ресурсы. Большинство программных решений позволяют увидеть географическо-интерактивную карту мнений, визуализацию результатов работы системы и дают возможность сузить круг поиска, с помощью добавления слов-исключений. Лишь отдельные программные продукты содержат в себе сервисы настройки каталога ресурсов, добавления новых ресурсов, настройки профиля правил для поиска, детализацию анализа и настройки выгрузки данных. В рассмотренных программных продуктах отсутствуют модули управления тональным словарем, моделирования социального самочувствия и возможность приглашения "эксперта".

Процесс проектирования подобных систем представляют собой платформенное решение. На основе использования функциональных возможностей аналогичных систем, нами была применена клиент-серверная архитектура программного

решения, использованы технологии проектирования информационных систем, технология программирования с применением языка Python, фреймворка Django, проектирование реляционной базы данных на основе PostgreSQL. Использованы методы продукционной модели, позволяющие построить гибкие правила оценки социального самочувствия, базирующиеся на статистических показателях социально-экономического развития страны, модель семантического профиля на основе стандарта для оценки эффективности контентной политики и коммуникации на площадке социальных сетей, разработанная сервисом SocialBakers [30]. Данные модели и методы позволили нам построить платформу количественного анализа и систему организации хранения этих данных. Предложенная система мониторинга социальных сетей и анализа настроения, предполагает использование алгоритмов машинного обучения. Анализ литературных источников [3,4,16] показал, что для создания полноценной системы необходимо применить алгоритмы машинного обучения (классификация, кластеризация, TF-IDF, Word2Vec и т.д.) для казахского языка. Тем самым позволив построить лингвистический словарь, который сформирован для казахского языка и решить задачу автоматической оценки тональности текста, с последующим прогнозированием настроения пользователей. Для следующего этапа развития системы реализовать интеллектуальный модуль путём формирования корпуса текстовых массивов казахского языка и его разметки по тональности; интегрировать модуль машинного обучения с платформой мониторинга социальных сетей.

3.1 Аспекты проектирования системы

Архитектура системы базируется на основе трехуровневой модели клиент-серверного приложения (графический интерфейс пользователя - бизнес-логика - СУБД), которая представлена на рис.1.

Система состоит из:

1. Модуля "Коннектор позволяющий производить настройки подключения к интернет-ресурсам и API известных социальных сетей из каталога ресурсов;
2. Модуля "Лингвистический конструктор позволяющий формировать тональный словарь на казахском и русском языках, с возможностью динамического наполнения и определения оценки тональности;
3. Модуля "Платформа обработки данных и анализа позволяющий моделировать социальное самочувствие по принципу "лакмусовой бумажки" и строить количественный анализ результатов мониторинга социальных сетей, производить предобработку и обработку данных с целью определения оценки тональности, с применением алгоритмов машинного обучения;
4. Модуля "Визуализации формирующий отчеты в формате PDF и позволяющий строить графические представления результатов количественного анализа.

Основное назначение Системы является обеспечение сервиса мониторинга мнений интернет-пространства для проведения глубинного анализа неструктурированной информации, с целью извлечения знаний и получения важной информации из текстовых

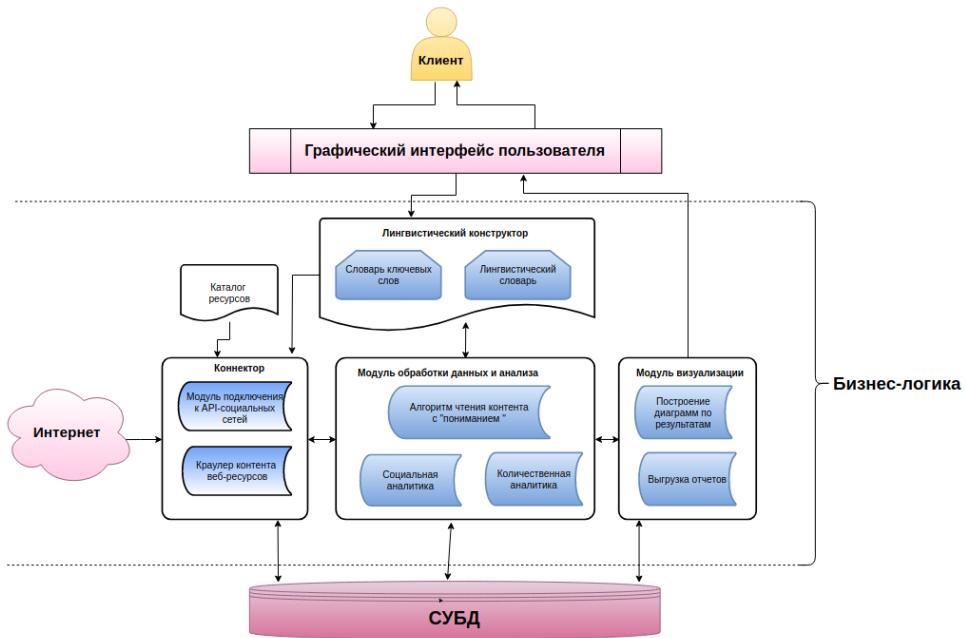


Рисунок 1: Архитектура Системы

массивов по оценке пользовательского восприятия контента, и оценки социального самочувствия общества.

Разработанная система имеет встроенные модули подключения к API социальных сетей Вконтакте, Facebook, Twitter, на этапе разработки подключение к Instagram, функциональные возможности позволяют:

- Работать с текстами на 2-х языках (казахский, русский) с автоматическим определением языка текста (казахский, русский, смайлы, символы);
- Определять тональность темы/текста (отрицательная, положительная, нейтральная);
- Поддерживать тональный словарь на казахском и русском языках;
- Формировать базу данных слов с ошибками на казахском и русском языках, с целью увеличить результаты процедуры оценки тональности темы (это связано с тем, что комментарии написаны с ошибками);
- Фиксировать диапазон времени при процедуре мониторинга социальных сетей (от года, 6 мес., 3 мес., месяц, неделя, день);
- Строить интегральную оценку тональности темы;
- Выявлять профиль участника социальных сетей, путем чтения данных по профилю участника; подсчета активности участника в теме (количество комментариев, лайков, репостов);

- Приглашать эксперта участвовать в дискуссии и наблюдать статистику мониторинга по изменению мнений под влиянием эксперта;
- Рассчитать индекс социального самочувствия общества;
- Работать с большим объемом данных (заложены принципы масштабирования и параллельной обработки);
- Создавать различные отчеты (по категории, ресурсам, оценки тональности, мультиязычности и др) в виде различных графиков и диаграмм с выгрузкой в формате PDF.

Разработанная архитектура системы базируется на ролевой политике доступа к системе, которая обеспечивает организацию алгоритмического подхода работы пользователя. В основе любого алгоритма действий пользователя лежит тема, которая представляет собой набор заданных параметров суперпользователем, такие как: название, категория, подкатегория темы, набор ключевых слов, язык, ресурсы и временной интервал поиска.

С учетом выше названных пунктов определен алгоритм работы пользователя с системой.

Таблица 2: Алгоритм Системы

Шаг 1. Запускаем систему (социальные сети, интернет-ресурсы новостных порталов)
Шаг 2. Создаем новую тему для анализа (тема, категория, подкатегория, период, язык поиска, область поиска, ключевые слова)
Например: Проблемы безопасности граждан (Категория: общество, безопасность; Диапазон поиска: 10.10.18-10.07.18; рус./каз., соц.сети: Вконтакте, facebook, Twitter; ключевые слова: потерпевший, убийство, задержан, следствие, МВД, полицейские)
Шаг 3. Запускаем поисковый механизм (поиск в режиме реального времени, подключение тонального словаря на рус/каз языках с целью определения тональности темы и комментариев; расчет социального самочувствия).
Шаг 4. Изучаем результаты поиска и анализа (отчеты по результатам поиска; аналитика по: оценки тональности, теме, записям, ресурсам, авторам, языку, работе эксперта) Например: Анализ по теме "Проблемы безопасности граждан". Общее количество результатов: 3594; из них записей: 973, комментариев: 2621 Результатов из "Вконтакте": 3090, из "Twitter": 504 на русском языке: 2632; на казахском языке: 63, на других языках: 899 Оценка тональности темы: отрицательная; "отрицательные": 1743, "положительные": 468, "нейтральные": 1383 Активные участники: Ф. И. О., Ф. И. О., Ф. И. О. Популярные слова: "жертва", "потерпевший", "убийство". Местоположение: определено 49, 14% (1766), из Казахстана: 52, 5% (927), из России: 47, 5% (839)
Шаг 5. Визуализация результатов анализа (диаграммы: оценка тональности, типы ресурсов, количество записей; оценка социального самочувствия)
Шаг 6. Сохранение результатов анализа (файл в формате pdf, разметка по датам поиска)

Длительность работы Системы по теме поиска занимает определенное время, которое зависит от периода поиска, количества ключевых слов и набора ресурсов для поиска, что в свою очередь влияет на количество возвращаемых Системой результатов мониторинга и на длительность обработки темы поиска. По завершению поиска по заданным параметрам можно получить "Сводный отчет по результатам темы". В отчете по результатам заданной темы отображаются все параметры, по которым производился поиск, количество найденных записей, количество найденных комментариев, общая тональность Темы.

Страница "Сводный отчет по результатам темы" представлена на рис.2.

Сводный отчет по результатам темы "Проблемы безопасности граждан"	
Тема	Проблемы безопасности граждан
Категория	Общество
Подкатегория	Происшествия
Ключевые слова	потерпевший, жертва, убийство, задержан, следствие, МВД, полицейские;
Ресурсы	Vk, Twitter,
Язык	KZ RU
Количество результатов	3594
Количество записей	973 Подробнее >>
Количество комментариев	2621 Подробнее >>
Диапазон времени	С 10 октября 2018 г. 3:08 по 17 июля 2018 г. 20:48
Оценка тональности темы	Отрицательный
Количество запросов модератору	
Участие эксперта	
Получить отчет в PDF	Посмотреть диаграмму по теме "Проблемы безопасности граждан"

Рисунок 2: Страница "Сводный отчет по результатам темы"

Кроме того, в Системе предусмотрен функционал визуализации различных отчетов по всем результатам (за весь период мониторинга, за квартал, за 1 месяц, за 1 неделю, за 1 день), диаграммы представлены на рис.3 а,б,в,г.

Модуль "Мониторинга социальных сетей" отвечает за выявление тональности пользовательских мнений, путем мониторинга аккаунтов новостных порталов в известных социальных сетях. Для этого используется тональный словарь на казахском и русском языках. При проектированиях тонального словаря были предусмотрены возможности создавать слова обычной формы, сленговые варианты слов, слова с ошибками, и предусмотрен функционал для добавления новых слов с целью динамического наполнения словаря.

Результативность работы разработанной системы можно определить путем оценки семантического профиля по результатам поиска. Для этого были рассмотрены модели стандартов оценки вовлеченности, которые основаны на способе измерения показателей



Рисунок 3: Диаграммы модуля "Визуализации мониторинга"

социальных сетей для SMM менеджмента, по формуле SocialBakers [31] на примере Facebook [32]. Исходя из которых, были определены следующие показатели для оценки семантического профиля восприятия обществом событий: уровень заинтересованности темой в обществе (R_{CT}), уровень активности обсуждения темы (R_{CE}), уровень настроения общества (R_{TS}).

Формула SocialBakers для расчета вовлечения аудитории имеет следующий вид:

$$\text{Daily ER} = \frac{\text{Total Actions Taken}}{\text{Daily coverage}} \times 100\% \quad (1)$$

где, Total Actions Taken – общее число действий, а именно – Лайки, Комментарии, Шэры, Вступления; а Daily Coverage – общий дневной охват контента, включая виральный и оплаченный [33]. На основании формулы 1 предлагаем расчет уровня активности обсуждения темы (R_{CE}), как:

$$R_{CE} = \frac{\frac{C}{CP}}{(CS)} \times 100\% = R_{CE} = \frac{\frac{L+R+C}{CP}}{(CS)} \times 100\%.$$

где, S – сумма количества подписчиков источников, CP – количество найденных постов по теме, CE – количество взаимодействий участниками по теме (комментарии (C), лайки (L), репосты (R)).

Уровень заинтересованности темой в обществе (RCT) определяется, как:

$$R = \frac{\times 100\%}{\max_{CT}}.$$

где, – количество найденных записей по данной теме, \max – максимальный объем записей для темы (должно устанавливаться экспертом за определенный период времени).

Уровень настроения общества (RTS) рассчитывается, как суммарная оценка результатов тем мониторинга социальных сетей (положительная, нейтральная, отрицательная).

Помимо семантического профиля в основе программного продукта лежит оценка социального самочувствия, которая позволит определить, что послужило формированию такого настроения пользователей в социальных сетях.

Проведенный анализ на тему "Проблемы безопасности граждан" показал, анализ настроения общества - отрицательный. Согласно оценке индекса социального самочувствия, проблема настроения в обществе лежит глубже. Например, "число зарегистрированных преступлений" по г.Алматы на 2016 году составило 68704 преступлений, а в 2017 году - 63797 [34], являются самыми высокими по всей республике, чем и объясняется причина отрицательного настроения общества. Предлагаемый в данной Системе подход к анализу социального самочувствия, путем оценки производственной модели ключевых показателей статистики, позволяет определить эмоциональное состояние общества по принципу "лакмусовой бумаги".

Для моделирования социального самочувствия использованы следующие статистические данные [34]: Занятое население; Безработное население; Уровень безработицы; Среднемесячная номинальная заработная плата; Среднемесячный прожиточный минимум населения; ВВП на душу населения; Доля населения с доходами ниже величины прожиточного минимума в общей численности; Число зарегистрированных преступлений. В основу расчета производственной модели взят алгоритмический подход, где сравниваются значения показателей за текущий и предыдущие годы, либо берется отношение значения показателя к среднему или минимуму. В результате были получены четыре выходных результата определяющих индекс социального самочувствия: социальное напряжение - критическая зона, не ярко выраженное социальное напряжение, удовлетворительное социальное самочувствие, стабильное социальное самочувствие. По индексу социального самочувствия формируется диаграмма оценки социального самочувствия общества (рис. 4).

4 Результаты и обсуждение

Сравнительный анализ аналогичных зарубежных программных продуктов позволил сделать вывод о том, что данные решения при введении входных данных на казахском и русском языках не возвращают результаты мониторинга, такие как, тексты и количественные данные записей и комментариев по теме поиска. Это объясняется тем, что информационное поле поиска изученных систем ориентировано на интернет-пространство только зарубежных стран, и не охватывает отечественное интернет-пространство, а алгоритмы обработки данных и алгоритм определения



Рисунок 4: Диаграмма по оценке социального самочувствия общества с 2013 по 2017 гг.

оценки тональности текста зарубежных программных продуктов не адаптированы для казахского языка. Однако, по результатам мониторинга разработанная Система выдает всю детальную информацию по теме поиска. Таким образом, сравнительно-сопоставительный анализ функциональных возможностей зарубежных систем с разработанной Системой является не рациональным.

Практическая значимость исследования состоит в том, что предлагаемая информационная система позволит не только мониторить социальные сети и интернет-ресурсы, но и выявлять причины пользовательского мнения, путем расчета индекса социального самочувствия и с помощью управления тональным словарем можно обучать систему для более точных результатов.

В результате проведенных работ информационная система мониторинга мнений является первой отечественной разработкой с поддержкой тонального словаря на казахском языке, построенного с применением алгоритмов машинного обучения.

5 Заключение

Потребность в Казахстане в подобной системе есть. Так как, важно выявлять мнение потребителей, например, образовательных, медицинских и государственных услуг, с целью выявления "узких мест" в работе соответствующих ведомств, оптимизации работы и сервисов, принятия оперативных решений по проблемным вопросам. Ранний доступ к анализу и информации обеспечивает лучшее принятие решений и более построения качественных информированных ответов, для немедленного реагирования.

Работа по мониторингу социальной сети и контент-анализу позволяет: осуществить детектирование информационного воздействия; определять круг интересов и распространения информации; осуществлять сбор расширенных социальных профилей для маркетинговой деятельности; отслеживать динамику вовлеченности пользователей в ту или иную тему; выявлять информационные поводы и отношения к ним; мониторить настроения интернет-пользователей на предмет восприятия конкретных тем обсуждения.

Данная система позволит развить новое направление рынка программных решений – разработка информационной системы мониторинга мнений и социальных сетей для обработки и анализа текстовых данных, с целью получения информации по оценке пользовательского восприятия контента и оценки социального самочувствия общества.

Список литературы

- [1] *Benedetto F., Tedeschi A.* Big Data Sentiment Analysis for Brand Monitoring in Social Media Streams by Cloud Computing // Sentiment Analysis and Ontology Engineering. – 2016. – № 639. – С. 341-377.
- [2] *Gu J., Xiao B., Gu Z.* Software Design for Network Monitoring System // International Conference on Information Computing and Applications. – Berlin: Springer, 2012. – С. 41-46.
- [3] *Laine M.O.J., Fröhwirth C.* Monitoring Social Media: Tools, Characteristics and Implications // International Conference of Software Business. – Berlin: Springer, 2010. – С. 193-198.
- [4] *Liu Y.* Social Media Monitoring // Social Media in China. – 2016. – С. 185-193
- [5] *Gesell, S.B., Barkin, S.L. & Valente, T.W.* Social network diagnostics: a tool for monitoring group interventions // Implementation Science. – 2013. – № 8(116). – С. 1-12.
- [6] *Tesconi M., Gazzè D., Duca A.L.* SocialTrends: A Web Application for Monitoring and Visualizing Users in Social Media // International Conference on Social Informatics. – Berlin: Springer. – 2012. – С. 535-538.
- [7] *Averchenkov V., Budylskii D., Podvesovskii A., Averchenkov A., Rytov M., Yakimov A.* Hierarchical Deep Learning: A Promising Technique for Opinion Monitoring and Sentiment Analysis in Russian-Language Social Networks"// Creativity in Intelligent Technologies and Data Science. – 2015. – № 535. – С. 583-592.
- [8] *Psallidas F., Ntoulas A., Delis A.* Soc Web: Efficient Monitoring of Social Network Activities // International Conference on Web Information Systems Engineering - Heidelberg: Springer. – 2013. – С. 118-136.
- [9] *Kirsch B., Giesselbach S., Knott D., Rüping S.* Robust End-User-Driven Social Media Monitoring for Law Enforcement and Emergency Monitoring"// Community-Oriented Policing and Technological Innovations. – 2018. – С. 29-36.
- [10] *Xiaoyan W., Kai X., Ying S., Jian-long T., Li G.* Research of New Words Identification in Social Network for Monitoring Public Opinion // International Conference on Trustworthy Computing and Services. – Heidelberg: Springer. – 2013. – С. 598-603.
- [11] *Guo Z., Wang Z., Zhang R.* MMCRD: An Effective Algorithm for Deploying Monitoring Point on Social Network // China Conference on Knowledge Graph and Semantic Computing. – Singapore: Springer. – 2018. – С. 40-51.
- [12] *Schinias M., Papadopoulos S., Apostolidis L., Kompatsiaris Y., Mitkas P.A.* Open-Source Monitoring, Search and Analytics Over Social Media // International Conference on Internet Science. – Greece: Springer. – 2017. – С. 361-369.
- [13] *Grübbauer J., Haric P.* Social Media Monitoring Tools as Instruments of Strategic Issues Management // Handbook of Social Media Management. – 2013. – С. 671-687.
- [14] *Gyarmati L., Trinh T.A.* Measurement Methods of User Behavior in Online Social Networks // Computational Social Networks. – 2012. – С. 157-175.
- [15] *Stiehm S., Welter F., Jeschke S.* Shaping the Future Through Cybernetic Approaches of Social Media Monitoring // Automation, Communication and Cybernetics in Science and Engineering. – 2016. – С. 179-191.
- [16] *Souza R.C.S.N.P., de Brito D.E.F., Cardoso R.L., de Oliveira D.M., Meira W., Pappa G.L.* An Evolutionary Methodology for Handling Data Scarcity and Noise in Monitoring Real Events from Social Media Data // Ibero-American Conference on Artificial Intelligence. – Chile: Springer. – 2014. – С. 295-306.
- [17] *Sparks R.* Social Network Monitoring: Aiming to Identify Periods of Unusually Increased Communications Between Parties of Interest // Frontiers in Statistical Quality Control 11. – 2015. – С. 3-13.
- [18] *Hootsuite* [Электрон. ресурс]. – URL: <https://hootsuite.com/> (дата обращения: 21.11.2018).
- [19] *Socialmention* [Электрон. ресурс]. – URL: <http://socialmention.com/about/> (дата обращения: 21.11.2018).

- [20] *BuzzLook* [Электрон. ресурс]. – URL: <https://br-analytics.ru/about/> (дата обращения: 21.11.2018).
- [21] *IQBuzz* [Электрон. ресурс]. – URL: <http://iqbuzz.pro/about.php> (дата обращения: 21.11.2018).
- [22] *Socialbakers* [Электрон. ресурс]. – URL: <https://www.socialbakers.com/company/> (дата обращения: 21.11.2018).
- [23] *Keyhole* [Электрон. ресурс]. – URL: <https://keyhole.co/about-us> (дата обращения: 23.11.2018).
- [24] *Brandwatch* [Электрон. ресурс]. – URL: <https://www.brandwatch.com/about/> (дата обращения: 23.11.2018).
- [25] *YouScan* [Электрон. ресурс]. – URL: <https://youscan.io/product/> (дата обращения: 23.11.2018).
- [26] *iMAS* [Электрон. ресурс]. – URL: <https://imas.kz> (дата обращения: 24.11.2018).
- [27] *Alem Media Monitoring* [Электрон. ресурс]. – URL: <https://alem.kz> (дата обращения: 24.11.2018).
- [28] *Brand Analytics* [Электрон. ресурс]. – URL: <https://br-analytics.ru> (дата обращения: 24.11.2018).
- [29] *Microsoft engagement* [Электрон. ресурс]. – URL: <https://dynamics.microsoft.com/ru-ru/microsoft-social-engagement> (дата обращения: 24.11.2018).
- [30] *Официальный интернет ресурс SocialBakers* [Электрон. ресурс]. – URL: <https://socialbakers.com/> (дата обращения: 24.11.2018).
- [31] Тихонова Е. Что такое Engagement Rate и для чего это нужно? [Электрон. ресурс]. – 2018. – URL: <https://blog.sociate.ru/chto-takoe-engagement-rate-i-dlya-chego-eto-nuzhno> (дата обращения: 29.11.2018)
- [32] *Engagement Rate: A Metric You Can Count On* [Электрон. ресурс]. – 2018. – URL: <https://www.socialbakers.com/blog/1427-engagement-rate-a-metric-you-can-count-on> (дата обращения: 29.11.2018)
- [33] Румянцева Д. ER - методы расчета и значение показателей [Электрон. ресурс]. – 2015. – URL: <https://www.cossa.ru/155/101492/> (дата обращения: 29.11.2018)
- [34] *Официальный интернет ресурс Комитета по статистике Министерства национальной экономики РК* [Электрон. ресурс]. – URL: <https://stat.gov.kz> (дата обращения: 29.11.2018).

References

- [1] Benedetto F., Tedeschi A., "Big Data Sentiment Analysis for Brand Monitoring in Social Media Streams by Cloud Computing", (*Sentiment Analysis and Ontology Engineering*, 639 (2016), accessed on November 14, 2018, https://doi.org/10.1007/978-3-319-30319-2_14
- [2] Gu J., Xiao B., Gu Z., "Software Design for Network Monitoring System", (*International Conference on Information Computing and Applications*, 2012) : 41-46.
- [3] Laine M.O.J., Frühwirth C., "Monitoring Social Media: Tools, Characteristics and Implications", (*International Conference of Software Business*, 2010) : 193-198.
- [4] Liu Y., "Social Media Monitoring", (*Social Media in China*, 2016), accessed on November 14, 2018, https://doi.org/10.1007/978-3-658-11231-8_10
- [5] Gesell, S.B., Barkin, S.L. & Valente, T.W., "Social network diagnostics: a tool for monitoring group interventions", (*Implementation Science*, 2013), accessed on November 14, 2018, <https://doi.org/10.1186/1748-5908-8-116>
- [6] Tesconi M., Gazzè D., Duca A.L., "SocialTrends: A Web Application for Monitoring and Visualizing Users in Social Media", (*International Conference on Social Informatics*, 2012) : 535-538.
- [7] Averchenkov V., Budylskii D., Podvesovskii A., Averchenkov A., Rytov M., Yakimov A., "Hierarchical Deep Learning: A Promising Technique for Opinion Monitoring and Sentiment Analysis in Russian-Language Social Networks" in *Creativity in Intelligent Technologies and Data Science*, ed Kravets A., Shcherbakov M., Kultsova M., Shabalina O. (Springer, Cham, 2015) : 583
- [8] Psallidas F., Ntoulas A., Delis A., "Soc Web: Efficient Monitoring of Social Network Activities", (*International Conference on Web Information Systems Engineering*, 2013) : 118-136.

- [9] Kirsch B., Giesselbach S., Knodt D., Rping S., "Robust End-User-Driven Social Media Monitoring for Law Enforcement and Emergency Monitoring" in *Community-Oriented Policing and Technological Innovations*, ed Leventakis G., Haberfeld M. (SpringerBriefs in Criminology. Springer, Cham, 2018) : 29.
- [10] Xiaoyan W., kai X., Ying S., Jian-long T., Li G., "Research of New Words Identification in Social Network for Monitoring Public Opinion", (International Conference on Trustworthy Computing and Services, 2013) : 598-603
- [11] Guo Z., Wang Z., Zhang R., "MMCRD: An Effective Algorithm for Deploying Monitoring Point on Social Network", (China Conference on Knowledge Graph and Semantic Computing, 2018) : 40-51.
- [12] Schinas M., Papadopoulos S., Apostolidis L., Kompatsiaris Y., Mitkas P.A., "Open-Source Monitoring, Search and Analytics Over Social Media", (International Conference on Internet Science, 2017) : 361-369
- [13] Grülbauer J., Haric P., "Social Media Monitoring Tools as Instruments of Strategic Issues Management" in *Handbook of Social Media Management. Media Business and Innovation*, ed Friedrichsen M., Mühl-Benninghaus W. (Springer, Berlin, Heidelberg, 2013) : 671
- [14] Gyarmati L., Trinh T.A., "Measurement Methods of User Behavior in Online Social Networks", (Computational Social Networks, 2012), accessed on November 14, 2018, https://doi.org/10.1007/978-1-4471-4051-1_7
- [15] Stiehm S., Welter F., Richert A., Jeschke S., "Shaping the Future Through Cybernetic Approaches of Social Media Monitoring", in Automation, Communication and Cybernetics in Science and Engineering 2016, ed Jeschke S., Isenhardt I., Hees F., Henning K. (Springer, Cham) : 179
- [16] Souza R.C.S.N.P., de Brito D.E.F., Cardoso R.L., de Oliveira D.M., Meira W., Pappa G.L., "An Evolutionary Methodology for Handling Data Scarcity and Noise in Monitoring Real Events from Social Media Data", (Ibero-American Conference on Artificial Intelligence, 2014) : 295-306
- [17] Sparks R., "Social Network Monitoring: Aiming to Identify Periods of Unusually Increased Communications Between Parties of Interest", in Frontiers in Statistical Quality Control 11, ed Knoth S., Schmid W. (Frontiers in Statistical Quality Control. Springer, Cham, 2015) : 3
- [18] Official internet resource Hootsuite, accessed November 21, 2018, <http://socialmention.com/about/>
- [19] Official internet resource Socialmention, accessed November 21, 2018, <http://socialmention.com/about/>
- [20] Official internet resource BuzzLook, accessed November 21, 2018, <https://br-analytics.ru/about/>
- [21] Official internet resource IQBuzz, accessed November 21, 2018, <http://iqbuzz.pro/about.php>
- [22] Official internet resource Socialbakers, accessed November 21, 2018, <https://www.socialbakers.com/company/>
- [23] Official internet resource Keyhole, accessed November 23, 2018, <https://keyhole.co/about-us>
- [24] Official internet resource Brandwatch, accessed November 23, 2018, <https://www.brandwatch.com/about/>
- [25] Official internet resource YouScan, accessed November 23, 2018, <https://youscan.io/product/>
- [26] Official internet resource iMAS, accessed November 24, 2018, <https://imas.kz>
- [27] Official internet resource Alem Media Monitoring, accessed November 24, 2018, <https://alem.kz>
- [28] Official internet resource Brand Analytics, accessed November 24, 2018, <https://br-analytics.ru>
- [29] Official internet resource Microsoft engagement, accessed November 24, 2018, <https://dynamics.microsoft.com/ru-ru/microsoft-social-engagement>
- [30] Official internet resource SocialBakers, accessed November 24, 2018, <https://socialbakers.com/>
- [31] Tihonova E., "Что такое Engagement Rate и для чего это нужно?", [What is the Engagement Rate and what is it for?], accessed November 29, 2018, <https://blog.sociate.ru/chto-takoe-engagement-rate-i-dlya-chego-eto-nuzhno/>
- [32] "Engagement Rate: A Metric You Can Count On", accessed November 29, 2018, <https://www.socialbakers.com/blog/1427-engagement-rate-a-metric-you-can-count-on>
- [33] Rumyantseva D., "ER - методы расчета и значение показателей", [ER - calculation methods and indicators], accessed November 29, 2018, <https://www.cossa.ru/155/101492/>
- [34] Official internet resource Committee on Statistics Ministry of National Economy of the Republic of Kazakhstan, accessed November 29, 2018, <https://stat.gov.kz>

МРНТИ 20.23.25; 16.31.21; 28.23.39

Разработка единого метаязыка морфологии тюркских языков

Шарипбай А.А., НИИ «Искусственный интеллект»,
Евразийский национальный университет имени Л.Н. Гумилева,
г. Астана, Республика Казахстан, E-mail:sharalt@mail.ru

Гатиатуллин А.Р., Институт прикладной семиотики Академии наук Республики Татарстан, г. Казань, Татарстан, Россия, E-mail: ayrat.gatiatullin@gmail.com

Ергеш Б.Ж., НИИ «Искусственный интеллект»,
Евразийский национальный университет имени Л.Н. Гумилева,
г. Астана, Республика Казахстан, E-mail:b.yergesh@gmail.com

Қажымұхан Дағынұлы, НИИ «Искусственный интеллект»,
Евразийский национальный университет имени Л.Н. Гумилева,
г. Астана, Республика Казахстан, E-mail:askarovna0105@mail.ru

В настоящее время в связи с резким увеличением объема информации на естественных языках в интернете и социальных сетях исследование и разработки в области вычислительной лингвистики становятся чрезвычайно актуальными. Как известно, вычислительная лингвистика является новым научным направлением и входит в состав вычислительной науки - информатики. Вычислительная лингвистика включает в себя компьютерную обработку естественных языков (ОЕЯ) – Natural Language Processing (NLP). Создание унифицированной системы разметки для тюркских языков (UniTurk) является актуальной задачей для обработки тюркских языков. Единая система разметок позволит унифицировать разметки, облегчить их понимание и использовать общее программное обеспечение, а также проводить различные исследования по лингво-статистическому сравнительному анализу среди тюркских языков. В статье представлены некоторые результаты, полученные в ходе выполнения запланированных работ над международным проектом по созданию многоязычной онтологии и единого метаязыка морфологии тюркских языков. С помощью онтологических моделей formalизованы морфологические правила тюркских (казахского, киргизского, татарского, турецкого и узбекского) языков. Результат этих работ могут быть использованы для решения задач NLP, например, для разметки корпусов, в системах извлечения знаний, системах информационного поиска, машинного перевода и т.д.

Ключевые слова: морфология, тюркские языки, метаязык, тезаурус.

Түркі тілдері морфологиясының бірегей метатілін әзірлеу
Шәріпбай А.Ә., «Жасанды интеллект» ғылыми-зерттеу институты,

Л. Н. Гумилев атындағы Еуразия ұлттық университеті,
Астана қ., Қазақстан Республикасы, E-mail:sharalt@mail.ru

Гатиатуллин А.Р., Қолданбалы семиотика институты, Татарстан Республикасының ғылым академиясы, Казань қ., Татарстан, Ресей, E-mail: ayrat.gatiatullin@gmail.com

Ергеш Б.Ж., «Жасанды интеллект» ғылыми-зерттеу институты,
Л. Н. Гумилев атындағы Еуразия ұлттық университеті,

Астана қ., Қазақстан Республикасы, E-mail: b.yergesh@gmail.com

Қажымұхан Дағынұлы, «Жасанды интеллект» ғылыми-зерттеу институты,
Л. Н. Гумилев атындағы Еуразия ұлттық университеті,

Астана қ., Қазақстан Республикасы, E-mail: askarovna0105@mail.ru

Соңғы кездегі Интернет пен әлеуметтік желілердегі табиғи тілдердегі ақпараттың көлемінің күрт артуы компьютерлік лингвистика саласындағы зерттеулер өте өзекті мәселе. Компьютерлік лингвистика информатика сияқты есептеу ғылымының жаңа бағыты екені белгілі. Компьютерлік лингвистика табиғи тілдерді өңдеу (Natural Language Processing-NLP) есебін шешумен айналысады. Түркі тілдерінің бірегей белгілеу жүйесін әзірлеу түркі тілдерін компьютерлік өңдеу саласының өзекті есебінің бірі. Бірегей белгілеу жүйесі белгілеудерді бірегейлендіріп, оларды түсінуді жеңілдетеді және түркі тілдеріне ортақ қосымшаларды пайдалануга, түрлі лингвистикалық-статистикалық зерттеулер жүргізуге және салыстырмалы талдау жасауга мүмкіндік береді. Бұл мақалада түркі тілдерінің көптілді онтологиясын және бірегей ортақ метатіл құру бойынша орындалып жатқан халықаралық жобада жоспарланған жұмыстардың кейбір нәтижесі көрсетілген. Онтологиялық моделдер арқылы түркі (қазақ, қыргыз, татар, түрк, өзбек) тілдерінің морфологиялық ережелері формалданған. Бұл жұмыстың нәтижелерін NLP есептерін шешу үшін қолдануға болады. Мысалы, корпустарды белгілеуде, білімдерді шыгару жүйелері, ақпараттық іздеу жүйелері, машиналық аудары және басқалары.

Түйін сөздер: морфология, түкі тілдері, метатіл, тезаурус.

Development of an unified meta language of the turkic languages morphology

Sharipbay A., Scientific-Research Institute "Artificial intelligence",

L. N. Gumilyov Eurasian National University,

Astana, Republic of Kazakhstan, E-mail: sharalt@mail.ru

Gatiatullin A., Institute of Applied Semiotics of the Academy of Sciences Republic of Tatarstan,

Kazan, Tatarstan, Russia, E-mail: ayrat.gatiatullin@gmail.com

Yergesh B., Scientific-Research Institute "Artificial intelligence",

L. N. Gumilyov Eurasian National University,

Astana, Republic of Kazakhstan, E-mail: b.yergesh@gmail.com

Kazhymukhan D., Scientific-Research Institute "Artificial intelligence",

L. N. Gumilyov Eurasian National University,

Astana, Republic of Kazakhstan, E-mail: askarovna0105@mail.ru

Currently, due to the sharp increase of information amount in natural languages on the Internet and social networks, research and development in the field of computational linguistics is becoming extremely relevant. As is known, computational linguistics is a new scientific field and part of computer science. Computational linguistics includes the Natural Language Processing (NLP). Creating a unified metalanguage for Turkic languages (UniTurk) is an important task for processing Turkic languages. An unified metalanguage system will allow to unify tags, facilitate their understanding and use common software, as well as conduct various studies on linguistic-statistical comparative analysis among the Turkic languages. The article presents some of the results obtained on an international project to create a multilingual ontology and unified meta-language of the Turkic languages morphology. Using ontological models, the morphological rules of the Turkic (Kazakh, Kyrgyz, Tatar, Turkish, and Uzbek) languages are formalized. The result of these works can be used in the NLP applications, for example, for corpus tagging, in knowledge extraction systems, information retrieval systems, machine translation, etc.

Key words: morphology, turkic languages, metalanguage, thesaurus.

1 Введение

Для компьютерной обработкой любых естественных языков требуются формализация их грамматических (морфологических и синтаксических) правил, разработка алгоритмов анализа и синтеза слов и предложений по этим правилам, программная реализация всех этих алгоритмов, создание тезаурусов по предметным областям аналогично WordNet [1], построение текстовых корпусов (база данных размеченных текстов) и аудио корпусов (база данных размеченных аудиозаписей) и других программ для анализа и обработки текстов.

Начиная с 2013 года усилиями ученых по компьютерной лингвистике из Казахстана, Татарстана, Кыргызстана и Турции проводятся ежегодные международные научно-практические конференции по компьютерной обработке тюркских языков «TurkLang», в которых обсуждаются проблемы компьютерной обработки тюркских языков. В рамках этой конференции организовывается традиционный круглый стол UniTurk, где в 2014 году была принята резолюция о разработке унифицированной системы грамматической разметки для электронных текстов на тюркских языках. Тогда же был представлен рабочий вариант унифицированной морфологической разметки, который ориентируется на морфемную структуру тюркской словоформы и должен максимально полно отразить морфологические особенности всех тюркских языков [2].

2 Обзор литературы

Метаязык – язык, предназначенный для описания другого языка, называемого объектным языком. Метаязык – язык лингвистики. метаязык некоторый искусственный язык описывающий естественный язык. Систематическая формализация научного языка [3]. Метаязык обладает следующими свойствами: с помощью его языковых средств можно выразить все, что выразимо средствами объектного языка, и обозначить все знаки, выражения объектного языка, для которых имеются имена; на метаязыке можно говорить о свойствах выражения объектного языка и отношениях между ними; на нем можно сформулировать определения, обозначения, правила образования и преобразования для выражений объектного языка [4]. Имеются известные системы разметки, такие как Penn Treebank [5], система разметки CLAWS [7], который применяется для разметки Британского корпуса [6], система разметки Брауновского корпуса, в американском национальном корпусе [8, 9] используются несколько систем разметки, про системы разметки более подробно описаны в [10, 11]. Все эти системы в основном применяются для разметки английского языка.

В настоящее время существует больше десяти электронных корпусов для тюркских языков: корпус казахского языка [12, 13, 14]; корпус татарского языка «Туган тел» [15, 16], корпус турецкого языка [17], корпус крымскотатарского языка [18], чувашского языка [19] и др. Одним из основных компонентов этих корпусов является система морфологической, синтаксической и семантической разметки, базовой среди которых является система морфологической разметки, которая разняется в некоторых корпусах. Поэтому необходимость создания единой системы разметки не вызывает сомнения. Единая система разметок позволит унифицировать разметки, облегчить их понимание и использовать общее программное обеспечение, а также проводить различные исследования по лингво-статистическому сравнительному анализу среди тюркских языков [20, 21, 22]. Создание унифицированной системы разметки (метаязыка) для тюркских языков (UniTurk) является актуальной задачей.

3 Материал и методы

Для создания такой системы разметок необходим общий ресурс с которым могли бы работать все разработчики тюркских электронных корпусов. Такой ресурс мог бы выполнять роль справочной системы как для разработчиков, так и для

пользователей тюркских электронных корпусов. Наиболее подходящими компонентами такого ресурса, которые соответствуют требуемым условиям являются онтологические модели грамматики тюркских языков. Положительной чертой онтологической модели является возможность поэтапного уточнения ресурса, начиная с описания категорий и концептов, добавляя к ним систему формализованных свойств функций и отношений между ними, а также добавление к ним тезаурусов. В настоящее время существует множество тезаурусов, наиболее известным среди которых является Принстонский WordNet [23]. В этом тезаурусе присутствуют большие фрагменты с описанием лингвистических единиц и лингвистических категорий (Рисунок 1), однако эти фрагменты тезауруса невозможно использовать для работы напрямую с тюркскими языками, так как они не отражают в полной мере все структурные и семантические особенности тюркских языков.

The screenshot shows the 'WordNet Search - 3.1' interface. At the top, there is a navigation bar with links to the 'WordNet home page', 'Glossary', and 'Help'. Below the navigation bar, there is a search bar labeled 'Word to search for: content word' and a 'Search WordNet' button. Underneath the search bar, there are 'Display Options' and a dropdown menu set to '(Select option to change)'. There is also a 'Change' button. A key definition is provided: 'Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations'. Below the key, it says 'Display options for sense: (gloss) "an example sentence"'.

Noun

- S: (n) **content word**, open-class word (a word to which an independent meaning can be assigned)
 - direct hyponym / full hyponym
 - S: (n) **headword**, **head word** (a content word that can be qualified by a modifier)
 - S: (n) **noun** (a content word that can be used to refer to a person, place, thing, quality, or action)
 - direct hyponym / full hyponym
 - S: (n) **collective noun** (a noun that is singular in form but refers to a group of people or things)
 - S: (n) **mass noun** (a noun that does not form plurals)
 - S: (n) **count noun** (a noun that forms plurals)
 - S: (n) **generic noun** (a noun that does not specify either masculine or feminine gender)
 - S: (n) **proper noun, proper name** (a noun that denotes a particular thing; usually capitalized)
 - S: (n) **common noun** (a noun that denotes any or all members of a class)

Рисунок 1: Фрагмент Принстонского WordNet

Особенность нашей работы заключается в том, что параллельно ведется работа по пяти тюркским языкам: казахский, татарский, киргизский, узбекский и турецкий. Такая параллельность в процессе построения онтологии грамматики позволяет производить сравнительный анализ лингвистических концептов этих языков и создавать единые концепты для описания их грамматик. Ожидаемым результатом этой работы должен стать единый, многоязычный тезаурус грамматики тюркских языков, в основе которого лежат их лингвистические концепты с системами отношений между ними. Текстовыми входами тезауруса являются лексемы по грамматике тюркских языков.

В таблице 1 представлен фрагмент единого многоязычного тезауруса. В левой части таблицы представлено обозначение концептов в виде названия на английском языке, а также обозначения тэга. Обозначение тэга используется для разметки электронных

корпусов тюркских языков. В центральной части таблицы представлены название тэгов на тюркских языках.

Таблица 1: Фрагмент единого многоязычного тезауруса

Tag	English	Kazakh	Tatar	Kyrgyz	Uzbek	Turkish	Russian
N	Noun	Зат есім	Исем	Зат атооч	От сўз туркуми	Isim (Ad)	Имя существительное
SIMP	Simple Noun	Дара зат есім	Гади исем	Жөнөкөй	Оддий		Простые существительные
CMPL	Complex Noun	Күрделі зат есім	ясалма исем	Татаал	Мураккаб		Сложные существительные
FUSW	Fused words	Біріккен зат есім	Күшма исем	Бириккен	Кўшма сўз		слитные (сложные) существительные
PAIR	Pair Noun	Қосарланған зат есім	Парлы исем	Кош	Жуфт сўз		парные существительные
CMPN	Compound Noun	Тіркесті (құрама) зат есім	Тезмә исем	Кошмок	Мураккаб сўз		составные существительные
ABBR	Abbreviations	Қыскартған зат есім	қыскарттылма исем	Қыскарттылган	Қисқартмалар		сокращенные существительные
UNDR	underivatives Noun	Негізгі зат есім		Тубаса негиз	Туб сўз	Basit isim	непроизводные существительные
DRV	derivatives Noun	Түйнды зат есім	ясалма исем	Туунду негиз	Ясама сўз	Türemis ad	производные существительные
COMP	compound					Birlesik	
ANIM	animate Noun	Жаңды зат есім	жанлы исем	Адамзаттык	Жонли	Canlı	одушевленные существительные
INAM	inanimate Noun	Жансыз зат есім	Жансыз исем	Адамзаттык эмес	Жонсиз	Cansız	неодушевленные существительные

Текстовая часть на тюркских языках состоит не только из названия концепта, а

также из описания, вопросов и примеров на этих языках. В таблице 2 представлен фрагмент описания концептов на казахском языке. Аналогичное описание концептов создано для каждого из тюркских языков, представленного в проекте.

Таблица 2: Фрагмент описания концептов на казахском языке

Part of speech	properties	types	types	definition	Questions	example
Зат есім				заттың, құбылыстың, оқиғаның атын білдіреді	кім? не?	Адам, бала, кітап
Кұрамы	жалаң	Дара зат есім		Бір сөзден жасалған зат есім		бала, ағаш, буын, орақ, жауын, жасау
	курделі	Курделі зат есім		кеміндегі екі я одан да көп түбір морфемадан құралған зат есімдер		ата-ана, ақтас, теміржол
	біріккен		Біріккен зат есім	кеміндегі екі түбір морфемадан бірге арқылы жасалған зат есімдер.		Жаңаарқа, Алтынтөбе, Ақтөбе, теміржол
	Қосарланған		Қосарланған зат есім	екі түбір морфемадан қосарланып жасалған зат есімдер		Бала-шаға, ата-ана
	тіркесті		Тіркесті (құрама) зат есім	кеміндегі екі я онан астам түбір сөздерден әр қылыш жолмен тіркесу арқылы жасалған зат есімдер.		ауылдың маңы, үйдің жаны

Компьютерная реализация онтологии грамматики выполнена в системе Pro-Tege (Рисунок 2), одном из наиболее известных инструментов для создания и редактирования онтологий общего назначения [24, 25]. Данный продукт является разработкой Центра биомедицинской информатики Стенфордского университета (Stanford Center for Biomedical Informatics Research) и распространяется под свободной

лицензией. Программа имеет две реализации: в виде кроссплатформенного десктопного приложения и в виде web-приложения WebProtege.

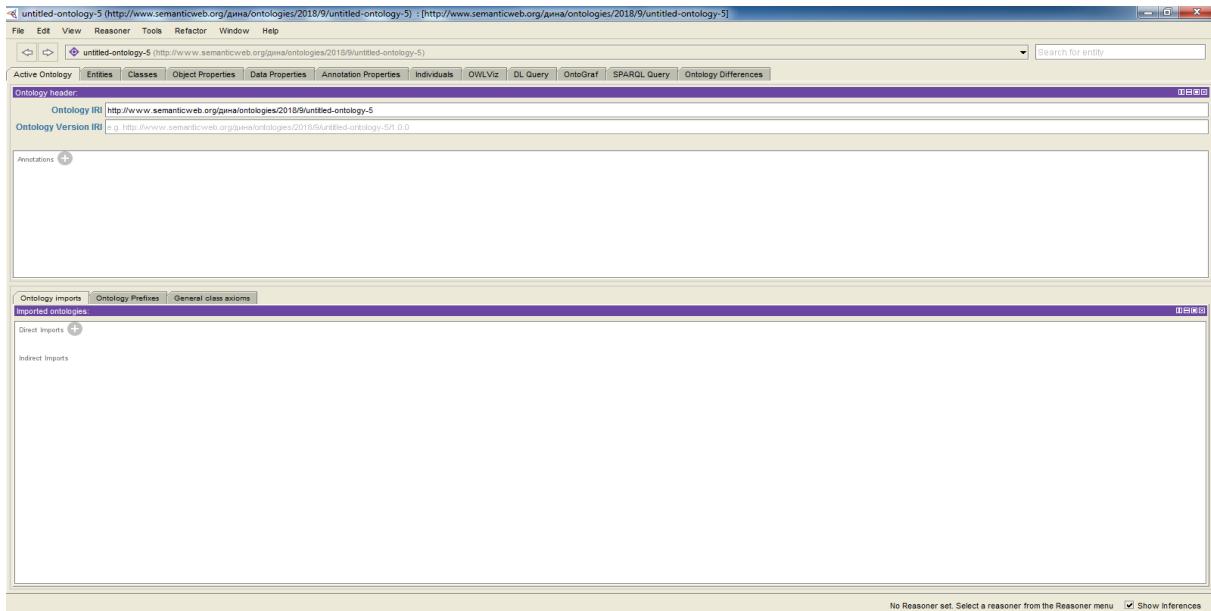


Рисунок 2: Фрагмент интерфейса системы Protege

4 Результаты и обсуждение

В результате исследования получена система единой разметки морфологии тюркских языков(казахский, татарский, киргизский, узбекский и турецкий). В создании метаязыка принимали участие ученые, молодые исследователи из Кыргызстана, Татарстана, Турции и Узбекистана. Также с помощью единого метаязыка разработаны онтологические модели морфологии этих языков. В будущем планируется создание единого метаязыка синтаксиса и их онтологических моделей, а также разработка тезауруса тюркских языков.

5 Заключение

Работа над созданием многоязычного тезауруса грамматик тюркских языков показывает насколько тюркские народы отдаляются друг от друга в создании научной терминологии. Создаваемый многоязычный информационный ресурс в Интернете должен, с одной стороны, способствовать взаимопониманию терминологии между тюркскими языками, а с другой стороны, стать многоязычной базой данных, которая будет использоваться в системах многоязычного поиска, машинного перевода между тюркскими языками, автореферирование тюркских текстов, а также в информационно-справочных и обучающих системах. Машинный перевод между тюркскими языками, многоязычный семантический поиск и извлечение знаний приведет к усилению экономических, образовательных, социальных и культурных

связей между тюркскими странами, так как будет уменьшаться языковой барьер. Тесное научное сотрудничество между коллективами ученых тюркских стран приведет к еще большему расширению научных и академических связей. Однако, также понятно, что создаваемые онтологические модели грамматики тюркских языков являются лишь небольшой частью тех лингвистических ресурсов, которые необходимы для полноценной реализации, перечисленных программных продуктов и технологий. Созданный информационный ресурс будет открытым и возможно его расширение путем добавление концептуальных описаний грамматик других тюркских языков.

6 Благодарности

Работа выполнена при поддержке грантового финансирования научно-технических программ и проектов Министерством науки и образования Республики Казахстан (грант № АР05132249, 2018-2020 годы)

Список литературы

- [1] *Wordnet. A large lexical database of English* [Электр.ресурс]. – 2018. – URL: <https://wordnet.princeton.edu> (дата обращения: 10. 09. 2018)
- [2] *Резолюция научно-практического семинара «Унификация систем грамматической разметки в корпусах тюркских языков (семинар UniTurk)* [Электр.ресурс]. – 2014. – URL: <http://www.turklang.tatar/ru/резолюция/> (дата обращения: 10.10.2018)
- [3] *Мельчук И.А.* Курс общей морфологии. Том 1. – Москва - Вена: «Языки русской культуры», Венский славистический альманах, Издательская группа «Прогресс», 1997. – 416 с.
- [4] *Залевская А.А.* Введение в психолингвистику: учебник . – М.: Российск. гос. гуманит. ун-т, 2000. – 382 с.
- [5] *Marcus, Mitchell P. Beatrice Santorini, and Mary Ann Marcinkiewicz.* Building a large annotated corpus of English: the Penn Treebank. // Computational Linguistics. – 1993. – 19(2). С. 313–330.
- [6] *British National Corpus* [Электрон.ресурс]. – 2018. – URL: <http://www.natcorp.ox.ac.uk> (дата обращения: 10.09. 2018)
- [7] *Garside, R.* The CLAWS Word-tagging System // In: R. Garside, G. Leech and G. Sampson (eds), The Computational Analysis of English: A Corpus-based Approach. – London: Longman, 1987.
- [8] *The Open American National Corpus* [Электр.ресурс]. – 2018. – URL: <http://www.anc.org> (дата обращения: 10.10.2018)
- [9] *Ide, N.* The American National Corpus: Then, Now, and Tomorrow // Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages, Cascadilla Proceedings Project, Sommerville, MA, - 2008.
- [10] *Jurafsky, Daniel and James H. Martin.* : Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. – 2nd Edition. – Prentice-Hall, 2009. – 988 p.
- [11] *Nitin Indurkhya and Fred J. Damerau.* Handbook of Natural Language Processing. – 2nd ed. – Chapman and Hall, 2010. – 704 p.
- [12] *Kazakh Language Corpus* [Электрон.ресурс]. – 2018. – URL: <http://kazcorpus.kz/> (дата обращения: 10.09.2018)
- [13] *Makhambetov O., Makazhanov A., Yessenbayev Zh., Matkarimov B., Sabyrgaliyev I., and Sharafudinov A.* 2013. Assembling the Kazakh Language Corpus // In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013. – р. 1022–1031.
- [14] *Мадиева Г.Б., Уматова Ж.М.* Об Алматинском корпусе казахского языка // Вестник КазНУ. Серия «Филология». – Алматы, 2015. – №5 (157). – С. 99-103.

- [15] Татарский национальный корпус «Туган тел» [Электр.ресурс]. – 2018. – URL: <http://tugantel.tatar> (дата обращения: 10.10.2018)
- [16] Galieva A., Khakimov B., Gatiatullin A. On the Way to the Relevant Grammatical Tagset for the Tatar National Corpus // EPiC Series in Language and Linguistics . Volume 1, CILC2016. 8th International Conference on Corpus Linguistics, 2016. – P.121–129.
- [17] Turkish National Corpus (TNC) [Электрон.ресурс]. – 2018. – URL: <http://www.tnc.org.tr/> (дата обращения: 10.10.2018)
- [18] Kubedinova L., Gatiatullin A. Morphological tagging of crimean tatar electronic corpus //Proceedings of the international conference «Turkic languages processing» TurkLang-2015. – Kazan, Tatarstan, 2015. – P.331-337
- [19] Zheltov P. Morphological annotation system for the national corpus of the chuvash language //Proceedings of the international conference «Turkic languages processing» TurkLang-2015. – Kazan, Tatarstan, 2015. – P. 328-331.
- [20] Sharipbay A., Mukanova A., Yergesh B., Zhetkenbay L., Zulkhazhav A., Yelibayeva G. Ontology modeling of morphological rules of the kazakh and turkish languages // Abstract of the VI international conference «modern problems of applied mathematics and information technology - al-Khorezmiy 2018». – Tashkent, Uzbekistan, 2018. – P. 51-52.
- [21] Zhetkenbay L., Sharipbay A., Bekmanova G., Kamanur U. «Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages // Journal of Theoretical and Applied Information Technology, - 2016. – Vol. 91. No.2. – P. 257- 263.
- [22] Arripov M., Sharipbay A., Abdurakhmonova N., Razakhova B. Ontology of grammar rules as example of noun of Uzbek and Kazakh languages // Abstract of the VI international conference «modern problems of applied mathematics and information technology - al-Khorezmiy 2018». – Tashkent, Uzbekistan, 2018. – p.37-38.
- [23] Fellbaum, Christiane. WordNet and wordnets // In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, - Oxford: Elsevier, 2006. – p. 665-670.
- [24] Protege [Электрон.ресурс]. – 2018. – URL: <http://protege.stanford.edu> (дата обращения: 10.10.2018)
- [25] Musen, M.A. The Protege project: A look back and a look forward // AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence. - 2015. – 1(4). – P.4-12. DOI: 10.1145/2557001.25757003.

References

- [1] "WordNet. A large lexical database of English accessed September 10, 2018, <https://wordnet.princeton.edu>.
- [2] "Resolution of the scientific-practical seminar "Unification of the systems of grammatical annotation in the Turkic languages corpora (UniTurk seminar) accessed October 10, 2018, <http://www.turklang.tatar/ru/резолюция/>.
- [3] Melchuk I.A. *Kurs obschey morfologii /General morphology course/*. Moskva - Vena: «Progress», Vol 1, 1997.
- [4] Zalevskaya A.A. *Vvedenie v psiholingvistiku /Introduction to Psycholinguistics/*. M.: Rossiysk. gos. gumanit. un-t, 2000.
- [5] Marcus, Mitchell P. Beatrice Santorini, and Mary Ann Marcinkiewicz. "Building a large annotated corpus of English: the Penn Treebank." *Computational Linguistics*.19(2)(1993):313–330.
- [6] "British National Corpus", accessed September 10, 2018, <http://www.natcorp.ox.ac.uk>.
- [7] Garside, R. "The CLAWS Word-tagging System." In: R. Garside, G. Leech and G. Sampson (eds), The Computational Analysis of English: A Corpus-based Approach. London: Longman, 1987.
- [8] "The Open American National Corpus", accessed October 10, 2018, <http://www.anc.org>.
- [9] Ide, N. "The American National Corpus: Then, Now, and Tomorrow." Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages. Sommerville, MA, 2008.
- [10] Jurafsky, Daniel and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd Edition. Prentice-Hall, 2009.
- [11] Nitin Indurkhya and Fred J. Damerau. *Handbook of Natural Language Processing*. 2nd Edition. Chapman and Hall/CRC, 2010.

- [12] "Kazakh Language Corpus accessed September 10, 2018, <http://kazcorpus.kz/>.
- [13] Makhambetov O., Makazhanov A., Yessenbayev Zh., Matkarimov B., Sabyrgaliyev I., and Sharafudinov A. "Assembling the Kazakh Language Corpus" In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*(2013): 1022–1031.
- [14] Madieva G.B., Umatova Zh.M. "Ob Almatinskem korpuze kazahskogo yazyika[About Almaty Corpus of the Kazakh language]"*KazNU Bulletin. Philology series. no.5(157)* (2015): 99-103.
- [15] "Tugan Tel" Tatar National Corpus accessed October 10, 2018, <http://tugantel.tatar>.
- [16] Galieva A., Khakimov B., Gatiatullin A. "On the Way to the Relevant Grammatical Tagset for the Tatar National Corpus"*EPiC Series in Language and Linguistics .CILC2016. 8th International Conference on Corpus Linguistics, Volume 1* (2016): 121–129.
- [17] "Turkish National Corpus (TNC)", accessed October 10, 2018, <http://www.tnc.org.tr/>.
- [18] Kubedinova L., Gatiatullin A. "Morphological tagging of crimean tatar electronic corpus." Proceedings of the international conference "Turkic languages processing"TurkLang-2015(2015):331-337.
- [19] Zheltov P. "Morphological annotation system for the national corpus of the chuvash language." *Proceedings of the international conference "Turkic languages processing"TurkLang-2015* (2015):328-331.
- [20] Sharipbay A., Mukanova A., Yergesh B., Zhetkenbay L., Zulkhazhav A., Yelibayeva G. "Ontology modeling of morphological rules of the kazakh and turkish languages." Abstract of the VI international conference "Modern problems of applied mathematics and information technology - al-Khorezmiy 2018"(2018): 51-52.
- [21] Zhetkenbay L., Sharipbay A., Bekmanova G., Kamanur U. "Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages." *Journal of Theoretical and Applied Information Technology , Vol. 91. No.2* (2016): 257- 263.
- [22] Aripov M., Sharipbay A., Abdurakhmonova N., Razakhova B. "Ontology of grammar rules as example of noun of Uzbek and Kazakh languages." Abstract of the VI international conference "Modern problems of applied mathematics and information technology - al-Khorezmiy 2018"(2018): 37-38.
- [23] Fellbaum, Christiane. "WordNet and wordnets." *Encyclopedia of Language and Linguistics*, ed. Brown, Keith et al., 2nd Edition, Oxford: Elsevier (2007): 665-670.
- [24] "Protege accessed October 10, 2018, <http://protege.stanford.edu>.
- [25] Musen, M.A. "The Protege project: A look back and a look forward." AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4) (2015): 4-12. doi: 10.1145/2557001.25757003.

3-бөлім**Қолданылмаған
математика**

МРНТИ 27.35.17

Раздел 3**Прикладная
математика****Section 3****Applied
Mathematics****Численный алгоритм для решения задачи моделирования динамики крупномасштабного термика**

Жакебаев Д.Б., Казахский национальный университет имени аль-Фараби,
г. Алматы, Республика Казахстан, E-mail: dauren.zhakebayev@gmail.com

Моисеева Е.С., Казахский национальный университет имени аль-Фараби,
г. Алматы, Республика Казахстан, E-mail: k.moisseyeva@gmail.com

Хребтов М.Ю., Новосибирский государственный университет,
г. Новосибирск, Российская Федерация, E-mail: weexov@yandex.ru

Цой Н.В., Казахский национальный университет имени аль-Фараби,
г. Алматы, Республика Казахстан, E-mail: nktatsoy@gmail.com

В данной статье рассматривается динамика крупномасштабного термика под действием силы плавучести, с учетом турбулентного перемешивания и адиабатического расширения. Математическая модель строится на основе уравнений Навье – Стокса, уравнения неразрывности и уравнения полной энергии. Численное моделирование осуществляется на основе решения решеточных уравнений Больцмана в трехмерной постановке с применением D3Q27 модели. Первое приближение решения уравнения Больцмана приводит к гидродинамическому уравнению Навье – Стокса. Приведены результаты проверки численного алгоритма на примере тестовой задачи течения Пуазейля, в рамках которой были посчитаны ошибки нормы для различных размеров расчетной сетки. Был произведен ряд численных экспериментов, при различных начальных условиях для температуры и плотности внутри и вне крупномасштабного термика. Получена зависимость высоты подъема облака от начальной температуры. В качестве результатов приведена динамика распространения поля температуры для начального значения 1800° К в момент времени 5 с, 15 с и 35 с.

Ключевые слова: крупномасштабный термик, решеточный метод Больцмана, D3Q27.

Кең ауқымды термалдардың динамикасың моделдеудің сандық әдісі

Жакебаев Д.Б., PhD, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ.,
Қазақстан Республикасы, E-mail: dauren.zhakebayev@gmail.com

Моисеева Е.С., PhD, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ.,
Қазақстан Республикасы, E-mail: k.moisseyeva@gmail.com

Хребтов М.Ю., ф.-м.ғ.к. , Новосибирск Мемлекеттік Университеті, Новосибирск қ.,
Ресей Федерациясы, E-mail: weexov@yandex.ru

Цой Н.В., докторант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ.,
Қазақстан Республикасы, E-mail: nktatsoy@gmail.com

Берілген мақалада, турбулентті алмасу мен адиабаталық кеңеюді ескеретін жүзгіштік күштің әсерінен қозғалысқа келетін улкен масштабты термиктің динамикасы қарастырылды. Математикалық модель Навье - Стокс теңдеуі, энергияның сакталу заңы және үзіліссіздік теңдеулерінің негізінде құрылды. Сандық пішіндеуді үшілшемді, торлық Больцман теңдеуінің шешімі негізінде, D3Q27 пішінін қолдану арқылы құрамыз. Больцман теңдеуінің бірінші жуықтау шешімі гидродинамикалық Навье - Стокс теңдеуіне келтіреді. Пуазейл ағымының тестілік есебінің мысалы негізінде, әртүрлі өлшемдерінде есептеу торының қателіктерін есептейтін, сандық алгоритмді тексеру нәтижелері көрсетілді. Ирі масштабты

термиктің ішінде жататын және жатпайтын, әртүрлі бастапқы шарттармен берілген сандық эксперименттер жасалынды. Бұлттың көтерілу биіктігінің бастапқы температурага тәуелділікі шамасы алынды. Нәтиже ретінде 5 сек, 15 сек және 35 сек уақыт мезетіне сәйкес келетін, бастапқы мәні 1800° К болатын, температуралың тарапу аумағының динамикасы алынды.

Түйін сөздер: кең ауқымды термик, торлық Больцман әдісі, D3Q27.

Numerical algorithm for solving the problem of modeling the dynamics of a large-scale thermal

Zhakebayev D.B., PhD, al-Farabi Kazakh National University, Almaty, Republic of Kazakhstan,
E-mail: dauren.zhakebayev@gmail.com

Moisseyeva Ye., PhD, al-Farabi Kazakh National University, Almaty, Republic of Kazakhstan,
E-mail: k.moisseyeva@gmail.com

Hrebtov M.Y., c.ph.-m.s., Novosibirsk State University, Novosibirsk, Russian Federation,
E-mail: weexov@yandex.ru

Tsoy N.V., PhD student, al-Farabi Kazakh National University, Almaty, Republic of Kazakhstan,
E-mail: nktatsoy@gmail.com

The article presents research results on dynamics of large-scale thermals influenced by buoyancy force, turbulent mixing and adiabatic expansion. Mathematical model is based on the Navier – Stokes equations, the continuity equation, and the total energy equation. Numerical simulation realized by finding solution of three dimension Lattice Boltzmann Equations applying D3Q27 model. First approximation of the solution of Boltzmann equation leads to the hydrodynamic Navier – Stokes equation. Numerical verification of the algorithm was carried out using benchmark Poiseuille flow problem. A number of numerical experiments were carried out, with different initial conditions for temperature and density inside and outside the large-scale thermal. The dependence of the height of cloud rise on the initial temperature was obtained. Dynamics of the temperature field propagation is given for an initial value of 1800 ° K at the time instant of 5 s, 15 s, and 35 s.

Key words: large-scale thermal, Lattice Boltzmann Method, D3Q27.

1 Введение

Чрезвычайные ситуации (ЧС) природного и антропогенного происхождения, такие как лесные пожары, пожары на промышленных предприятиях, аварийные падения ракет-носителей, часто сопровождаются наземным взрывом, при котором образуется крупномасштабный высокотемпературный термический поток – термик. Высокая температура и давление, а также сложный химический состав газопылевого облака приводят к образованию экологически опасных компонентов, распространяющихся на многие километры как в горизонтальном, так и вертикальном направлении, что представляет реальную угрозу окружающей среде, живым организмам и изменению климата.

Крупномасштабный высокотемпературный термик, – явление не до конца изученное. Согласно известным на сегодняшний день законам физики, холодный воздух вытесняет теплый, поэтому перемещения тепловых потоков наблюдается повсеместно. Термиком принято считать восходящий поток теплого воздуха [1]. Образование термика происходит, при нагревании поверхности земли на солнце или при чрезвычайных ситуациях природного или антропогенного происхождения. При этом очевидно, что чем контрастнее погодные условия, тем сильнее и выраженнее термические потоки. Следует отметить, что в штилевую погоду термики располагаются практически вертикально. При наличии ветра термический поток наклоняется по направлению ветра. При достаточном уровне влажности воздуха вода из воздушного потока конденсируется, и

над термиком образуются кучевые облака. В сухую погоду термические потоки часто бывают невидимыми (так называемые «голубые термики») [1]. В случае образования термиков по причине техногенной катастрофы, термики могут переносить различные примеси и визуализироваться дымом, пылью (пылевые столбы), пухом растений и прочие.

Одним из лучших методов для изучения образования и перемещения термических потоков является математическое моделирование, поскольку оно предполагает разработку математического описания, реализацию модели посредством применения наиболее точных современных численных методов, а также составление программного кода для высокопроизводительной вычислительной машины.

Также нужно отметить тот факт, что внутри термических потоков имеет место турбулентное перемещение воздушных масс, изучение которого, при его математическом описании, также затрудняется бесчисленным количеством степеней свободы рассматриваемых элементов системы потока.

В данной статье проводится изучение крупномасштабных термиков, посредством разработки и реализации динамической модели движения высокотемпературных потоков. Представлено математическое описание в виде системы уравнений в частных производных с поставленными краевыми и начальными условиями. Проведена проверка адекватности разработанной математической модели движения крупномасштабных термических потоков, на основе решения тестовых задач. Осуществлен подбор численного метода для реализации разработанной математической модели, удовлетворяющий критериям точности. Разработан специальный алгоритм применения методов для решения системы дифференциальных уравнений в частных производных при определенных граничных и начальных условиях. Приведено описание результатов моделирования движения крупномасштабных термиков в атмосферном приземном слое. Визуализированы вычислительная область и трехмерное поле температуры, показана времененная эволюция температуры и высоты подъема крупномасштабного термика.

2 Обзор литературы

Основные подходы к моделированию турбулентных течений в настоящее время – это DNS (Direct Numerical Simulation, прямое численное моделирование) и LES (Large Eddy Simulation, моделирование крупных вихрей). Поскольку в методах DNS рассчитываются величины всех масштабов турбулентности (от Колмогоровского до интегрального), для их использования требуются огромные вычислительные ресурсы, а прямое численное моделирование сжимаемых турбулентных течений на данный момент представляет собой практически неразрешимую задачу вследствие крайне высокого объема вычислений [1, 2, 3, 4, 5]. Методы моделирования крупных вихрей (LES) основываются на гипотезе о статистической независимости крупных и мелких вихрей. LES-методы разрешают крупные вихри, а мелкомасштабные структуры исключаются из уравнений Навье–Стокса при помощи пространственной фильтрации и моделируются с использованием подсеточных моделей (SGS, Sub-Grid Scale). К таким моделям относятся, например, модель Смагоринского и ее модификации [6], динамические подсеточные модели [7, 8] неявное моделирование подсеточной турбулентности Implicit-

it LES (ILES) [9, 10]. Таким образом, этот подход требует меньших вычислительных ресурсов в сравнении с методами DNS, поскольку шаги сетки и временные интервалы могут быть намного больше колмогоровских, обеспечивая при этом высокую точность расчета[2]. Кроме того, наряду с DNS и LES на практике часто применяются методы RANS (Reynolds averaged Navier-Stokes, осредненные по Рейнольдсу уравнения Навье-Стокса) для описания средних характеристик турбулентности, поскольку не требуют больших вычислительных затрат [11]. Численное изучение эволюции термика представлено в работах [12, 13]. Так, в работе [14] на основе осредненных по Фавру уравнений Навье-Стокса, замкнутых k-epsilon моделью турбулентности с учетом сжимаемости, построена математическая модель процесса поднятия термика в устойчивой атмосфере, записанная в цилиндрической системе координат. Для инициирования термика в начальный момент в вычислительную область вводится сферическое облако с повышенной температурой: значение температуры на нижней границе области $T_0 = 288.1K$, вертикальный градиент температуры $grad T = -6.5K \cdot \text{км}^{-1}$, начальный радиус термика $r_0 = 500\text{ м}$, его центр находится на высоте $h = 1250\text{ м}$ относительная избыточная температура равна $(T_{in} - T_0)/T_0 = 0.2$. Численное решение осуществлялось с применением схемы переменных направлений, дополненной процедурой монотонизации. k-epsilon модель решается отдельно неявной схемой. Получено, что распространение турбулентных параметров достаточно неоднородно и не стационарно: на начальной стадии турбулентное смешение в основном сосредоточено в верхней части термика, тогда как на финальной стадии эта область сдвигается к периферии термика и в области вихревого следа, также наблюдается сильное подавление турбулентности в вихревом ядре термика.

В работе [15] отмечено подобие термиков, возникающих в атмосфере, и очень плавучих пузырей с большой избыточной удельной энтропией относительно окружающей среды, изучавшийся в [16]. Используя это, в работе [17] численно изучается эволюция термика на основе динамики трехмерного плавучего изолированного пузыря в стратифицированном слое. Вследствие предположения, что стратификация температуры имеет адиабатический характер, равновесие незначительно устойчиво при развитии конвекции. Таким образом, при вводе горячего плавучего пузыря в данную среду он будет подниматься без влияния эффектов стратификации за исключением эффектов, вызванных ростом и расширением поднимающегося пузыря. В начальный момент времени в вычислительную область помещается горячая область с центром в начале координат и задается профиль энтропии, определяющий форму пузыря. Также задается начальное равенство давлений. В модельных уравнениях принято $\nu = 6 \cdot 10^{-3}$, т. е. задается постоянный коэффициент кинематической вязкости. Задача решается численно с применением компактной схемы Леле шестого порядка точности по пространству и схемы Хаймана третьего порядка точности по времени. В результате получено, что динамика плавучего пузыря состоит из двух фаз: в первой фазе силы плавучести равны скорости изменения момента, во второй — появляется эффект торможения под влиянием граничных условий (искусственный эффект) и торможения плавучести (физический эффект). Выявлено, что верхняя часть плавучего пузыря поднимается, достигая поверхности.

3 Материалы и методы

3.1 Математическая постановка задачи

Пусть в начальный момент времени у нижней границы расчетной области находится крупномасштабный термик, обозначенный – G , с радиусом – r , начальной температурой термика – T_0 , начальной температурой вокруг области G – T_1 и начальной плотностью – ρ (Рис. 1).

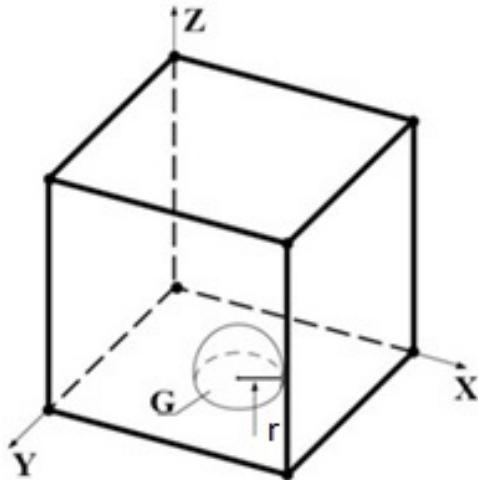


Рисунок 1: Схематичная иллюстрация постановки задачи.

Для решения рассматриваемой задачи используются модель Чепмена-Энскога [18, 19]:

Уравнение неразрывности:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho u) = 0 \quad (1)$$

Уравнение движения:

$$\frac{\partial(\rho u)}{\partial t} + \nabla(\rho uu) = -\nabla p + \nabla \cdot \tau + \rho g \quad (2)$$

Уравнение полной энергии:

$$\frac{\partial(\rho E)}{\partial t} + \nabla \cdot [(\rho E + p)u_m] = \nabla(\lambda \nabla T) + \nabla \cdot (\tau \cdot u) + \rho u \cdot g \quad (3)$$

где давление – $p = \rho RT$, тензор вязкого напряжения – $\tau = \mu [S - (2/D)(\nabla \cdot u)I]$, вязкость – $\mu = \tau_{fp}$, теплопроводность – $\lambda = \frac{(D+2)R}{2}\tau_h p = c_p \tau_h p$, здесь $c_p = \frac{(D+2)R}{2}$ – удельная теплоемкость при постоянном давлении, $E = c_v T + |u|^2/2$ – полная энергия.

Уравнение температуры:

$$c_v \left[\frac{\partial(\rho T)}{\partial t} + \nabla \cdot (\rho u T) \right] = \nabla \cdot (\lambda \nabla T) - p \nabla \cdot u + \tau \cdot \nabla u \quad (4)$$

где $c_v = DR/2$ удельная теплоемкость.

Данные уравнения решаются с заданными начальными значениями температуры и плотности внутри и вне области G . На нижней границе используется условие Дирихле, а на всех остальных границах условие Неймана.

3.2 Численный метод

Для моделирования динамики термиков применяется модель D3Q27 [20] решения уравнения Больцмана. Решеточные уравнения Больцмана (LBE) является одной из перспективных технологий последнего десятилетия для моделирования физических и атмосферных процессов. Многие сложные проблемы в традиционном CFD могут быть решены в LBM, например, турбулентные течения [21, 22, 23, 24], многофазные течения [25, 26], теплообмен [27, 28], микрофлюиды [29], течения через пористую среду [30] и фрактальную геометрию [31]. Функции распределения f вычисляются путем решения решеточного уравнения Больцмана, которое является особой дискретизацией кинетического уравнения Больцмана. После введения приближения Бхатнагара-Гросса-Крука (BGK) можно сформулировать уравнение Больцмана для потока жидкости и энергии в следующем виде:

$$f_i(x + e_i \Delta t, t + \Delta t) - f_i(x, t) = -\frac{\Delta t}{\tau} [f_i(x, t) - f_i^{eq}(x, t)] \quad (5)$$

$$g_i(x + e_i \Delta t, t + \Delta t) - g_i(x, t) = -\frac{\Delta t}{\tau_\theta} [g_i(x, t) - g_i^{eq}(x, t)] \quad (6)$$

где x – радиус-вектор узла решетки, Δt – обозначает решеточный временной шаг, τ – решеточное время релаксации, τ_θ – тепловое время релаксации, e_i – дискретная решеточная скорость в направлении i , f_i – функция распределения плотности в узле i , f_i^{eq} – равновесная функция распределения, зависящая от макроскопических параметров в узле i , g_i – функция распределения энергии в узле i , g_i^{eq} – равновесная функция распределения, зависящая от макроскопических параметров в узле i .

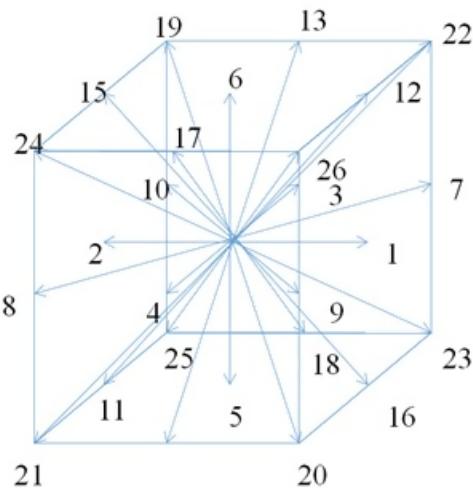


Рисунок 2: Схематичная иллюстрация решетки D3Q27.

Векторы скорости e_0, \dots, e_{26} модели D3Q27 показаны на рисунке 2. Для каждого вектора скорости хранится значение функции распределения f_i .

$$e_i = \begin{cases} (0, 0, 0)c & i = 0 \\ (\pm 1, 0, 0)c, (0, \pm 1, 0)c, (0, 0, \pm 1)c & i = 1 - 6 \\ (\pm 1, \pm 1, 0)c, (0, \pm 1, \pm 1)c, (\pm 1, 0, \pm 1)c & i = 7 - 18 \\ (\pm 1, \pm 1, \pm 1)c & i = 19 - 26 \end{cases}$$

где $c = \Delta x / \Delta t$ – решеточная скорость и i – направление решеточной скорости.

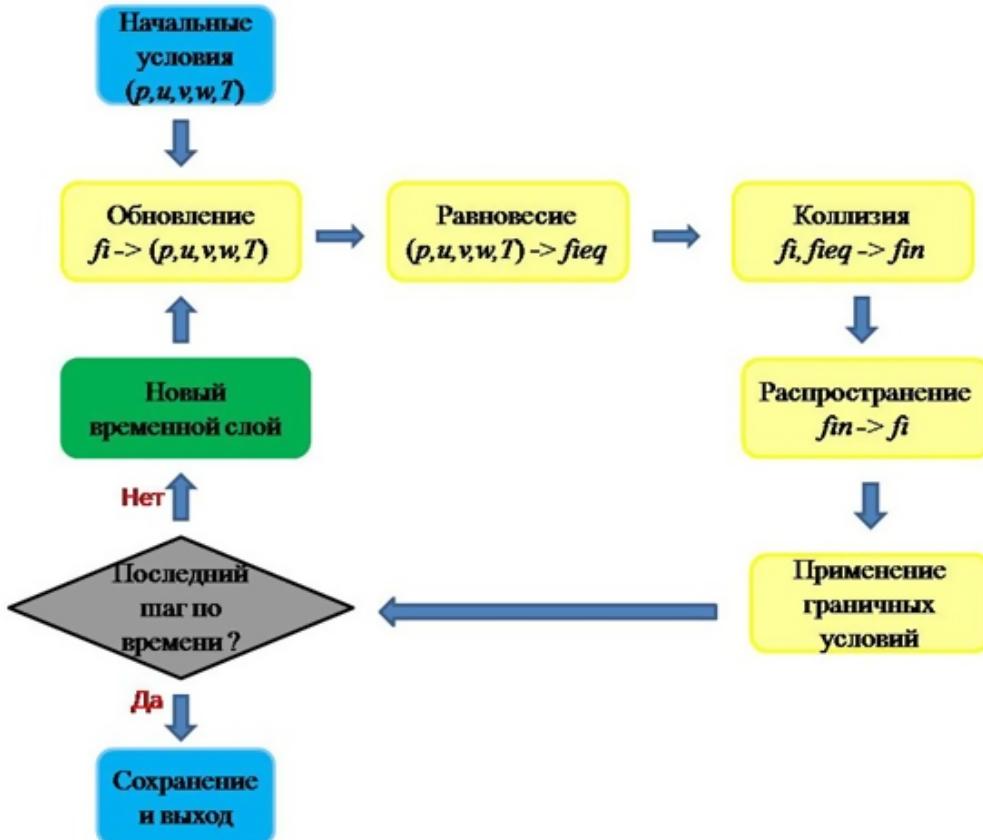


Рисунок 3: Алгоритм реализации решеточного метода Больцмана.

Решеточное время релаксации τ определяется по формуле:

$$\tau = 3\nu - \frac{1}{2}$$

Равновесные функции распределения плотности и энергии Максвелла вычисляются по формуле:

$$f^{eq}(T) = \omega_i \rho \left[1 + \frac{e_i \cdot u}{RT} + \frac{1}{2} \left(\frac{e_i \cdot u}{RT} \right)^2 - \frac{u^2}{2RT} \right]$$

$$g^{eq}(T) = \omega_i \rho \left[\frac{e_i \cdot u}{RT} + \left(\frac{e_i \cdot u}{RT} \right)^2 - \frac{u^2}{2RT} + \frac{1}{2} \left(\frac{e_i^2}{RT} - D \right) \right] + E \cdot f^{eq}$$

где ρ – плотность среды, T – температура среды, $c_s = 1/\sqrt{3} \cdot c$ – решеточная скорость звука и значения весовых коэффициентов ω_i таковы:

$$\omega_i = \begin{cases} 8/27 & i = 0 \\ 2/27 & i = 1 - 6 \\ 1/54 & i = 7 - 18 \\ 1/216 & i = 19 - 26 \end{cases}$$

Уравнение (5) можно разделить на два этапа, столкновение и распространение:

1. $\tilde{f}_i(x, t) = f_i(x, t) - \frac{\Delta t}{\tau} [f_i(x, t) - f_i^{eq}(x, t)],$
2. $f_i(x + e_i \Delta t, t + \Delta t) = \tilde{f}_i(x, t).$

Здесь $\tilde{f}_i(x, t)$ обозначает массу частиц, пришедших в узел по направлению i , столкнувшихся с остальными частицами. На втором этапе столкнувшиеся частицы переносятся в соседние узлы по направлениям i . Данная процедура является аналогичным для функции распределения энергии g_i .

После второго этапа необходимо пересчитать макроскопические переменные плотности, скорости и температуры жидкости как первые два момента функций распределения для каждого узла [18]:

$$\rho = \sum_{i=0}^{26} f_i, \quad \rho u = \sum_{i=0}^{26} e_i f_i, \quad \rho E = \sum_{i=0}^{26} g_i, \quad T = \left[E - \frac{|u|^2}{2} \right] \frac{1}{c_v}$$

На рисунке 3 представлен алгоритм применения решеточного метода Больцмана.

3.3 Численная валидация

Рассмотрим тестовую задачу течение Пуазейля для валидации численного алгоритма. Рассматривается ламинарное течение в трехмерной области с неподвижными параллельными стенками. В неподвижных стенах применена схема mid-link bounce back (BB) в качестве граничных условий. Согласно схеме граница стенки расположена на пол шага от границ узлов жидкости (boundary fluid nodes). Функции распределения направленные внутрь от границы определяются по формуле:

$$f_i(x_B, t + \Delta t) = \tilde{f}_i(x_B, t)$$

где x_B – координаты точек в границе.

В таблице 2 заданы параметры расчетов тестовой задачи. Высота канала H равна $32h$. Моделирование произведено при различных размерах расчетной области: $N_x \cdot N_y \cdot N_z = 32^3, 64^3, 96^3$ и 128^3 . Максимальная скорость в канале и скорость звука c_s равны 0.1 и 0.5773, соответственно. Кинематическая вязкость $\nu = 9.021 \cdot 10^{-3}$. Число Рейнольдса $Re = V_{max} \cdot H / \nu = 30$. Параметры релаксации $\tau = 1.2$.

Вдоль оси X поддерживается постоянная разность давлений:

$$\frac{\Delta p}{x_{out} - x_{in}} = \frac{8\eta V_{max}}{(y_{top} - y_{bot})^2}$$

где Δp – разница в давлениях, $\Delta p = p_{out} - p_{in}$, где p_{out} и p_{in} – давление на выходе и входе канала, соответственно, η – динамическая вязкость, V_{max} – максимальная скорость течения в канале, x_{out} и x_{in} – границы выхода и входа соответственно, y_{top} и y_{bot} – границы верхней и нижней стенки соответственно.

Аналитическое решение задачи вычисляются из следующей формулы:

$$u(y) = -\frac{1}{2\eta} \frac{\Delta p}{x_{out} - x_{in}} (y - y_{bot})(y - y_{top}), \quad y_{bot} \leq y \leq y_{top}$$

На рисунке 4 представлено сравнение численного решения с аналитическим решением в сечении $x/H = 0.5$. Течение устанавливается при $t \cdot \nu/H^2 = 0.4875$. Как видно из рисунка численные результаты хорошо согласуются с аналитическим решением.

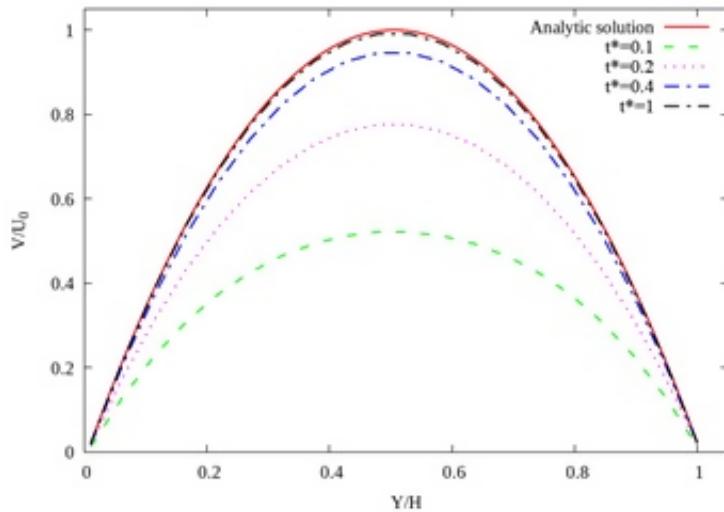


Рисунок 4: Сравнение численного результата с аналитическим решением при различных моментах времени.

На рисунках 5а и 5б представлены профили нормы ошибки L_1 и L_2 при различных размерах сетки. Ошибки нормы L_1 и L_2 были вычислены по следующим формулам:

$$\varepsilon_{L1}(t) = \frac{\sum_{x,y} |q_n(x, y, z, t) - q_b(x, y, z, t)|}{\sum_{x,y} |q_b(x, y, z, t)|}$$

$$\varepsilon_{L2}(t) = \frac{\sqrt{\sum_{x,y} |q_n(x, y, z, t) - q_b(x, y, z, t)|^2}}{\sum_{x,y} |q_b(x, y, z, t)|^2}$$

Как видно из рисунков увеличение разрешения сетки способствует к уменьшению норм ошибок. Если предположить, что известна норма ошибок при различных размерах сетки и что их соотношение размеров каждой сетки к начальному равен m , то можно определить порядок точности по следующей формуле:

$$n(t) = \log_m \left(\frac{\varepsilon_0(t)}{\varepsilon_m(t)} \right).$$

В тестовой задаче ошибки нормы были вычислены в момент времени $t \cdot \nu/H^3 = 1$. Порядки точности численного алгоритма в зависимости размера сетки представлены на таблице 3.

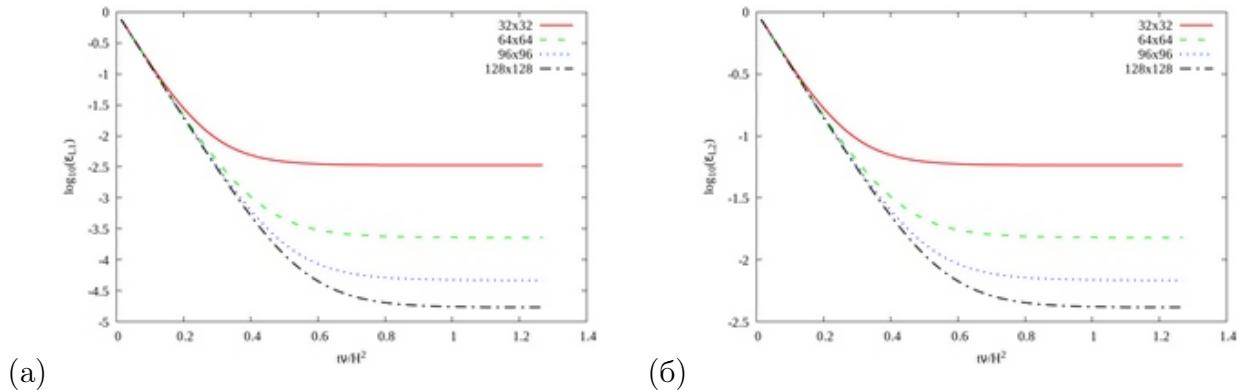


Рисунок 5: Профили ошибки нормы (а) L_1 и (б) L_2 при различных моментах времени.

Таблица 2 – Параметры моделирования.

Параметры	
Фактор масштабирования, $scale$	$scale = 1 \div 3$
Количество точек по оси, x , N_x	$N_x = 32 \cdot scale$
Количество точек по оси, y , N_y	$N_y = 32 \cdot scale$
Количество точек по оси, z , N_z	$N_z = 32 \cdot scale$
Параметр релаксации, τ	$\tau = \sqrt{3/16} + 0.5$
Максимальная скорость в канале, V_{\max}	$V_{\max} = 0.1/scale$
Кинематическая вязкость, ν	$\nu = (2\tau - 1)/6$
Число Рейнольдса, Re	$Re = V_{\max}H/\nu$
Давление на выходе канала, p_{out}	$p_{out} = 1$

Таблица 3 – Порядок точности L_1 и L_2 в зависимости от размера сетки при $t \cdot \nu/H^3 = 1$.

Размер сетки	$u(L_1)$	Порядок (n)	$u(L_1)$	Порядок (n)
$32 \cdot 32 \cdot 32$	3.3738×10^{-4}	–	5.8084×10^{-2}	–
$64 \cdot 64 \cdot 64$	2.3047×10^{-4}	3.8716	1.5181×10^{-2}	1.9358
$96 \cdot 96 \cdot 69$	4.6600×10^{-4}	3.8978	6.8264×10^{-2}	1.9489
$128 \cdot 128 \cdot 128$	1.6972×10^{-4}	3.8175	4.1198×10^{-3}	1.9087

4 Результаты и обсуждение

Представлены результаты численного моделирования эволюции крупномасштабного термика на основе метода LBM. Определена динамика изменения высоты подъема термика зависимости от начальной температуры. Полученные результаты моделирования, определяющий высоту поднятия облака, соответствует результатам работы Онуфриева [12].

Численное моделирование этапа формирования и распространения термика проведено в трехмерной области с физическим размером 300 м высоты, 120 м длины и 120 м ширины и расчетной сеткой $256 \times 128 \times 128$.

Численный эксперимент был проведен при начальной температуре облака $T_0 = 1800 K$ и метеорологических параметрах приведенных в рисунке 6.

На рисунках 6-7 представлены графики изменения температуры и высоты термика с течением времени, а также динамика распространения поля температуры при заданных начальных условиях.

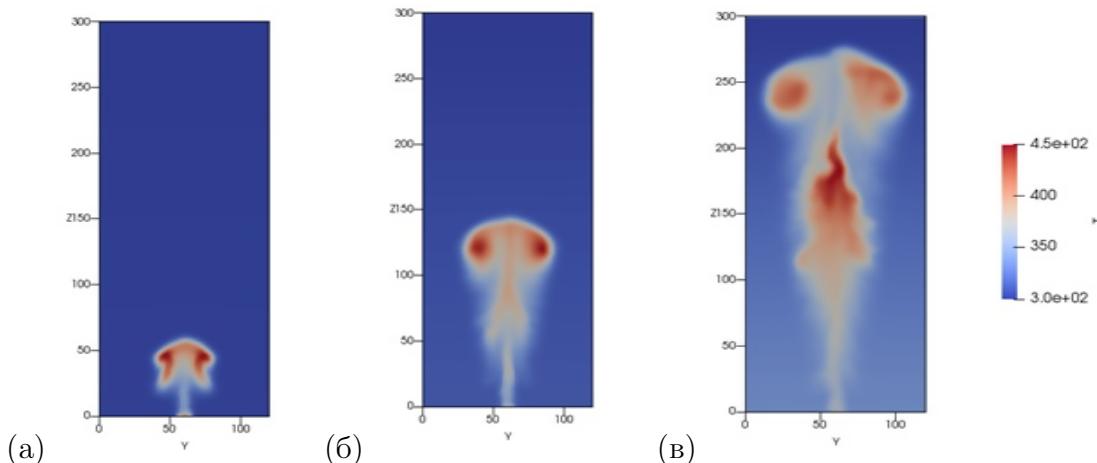


Рисунок 6: Динамика распространения поля температуры при начальном значении $T_0 = 1800 K$
 (а) $t = 5s$ и (б) $t = 15s$ (в) $t = 35s$ при различных моментах времени.

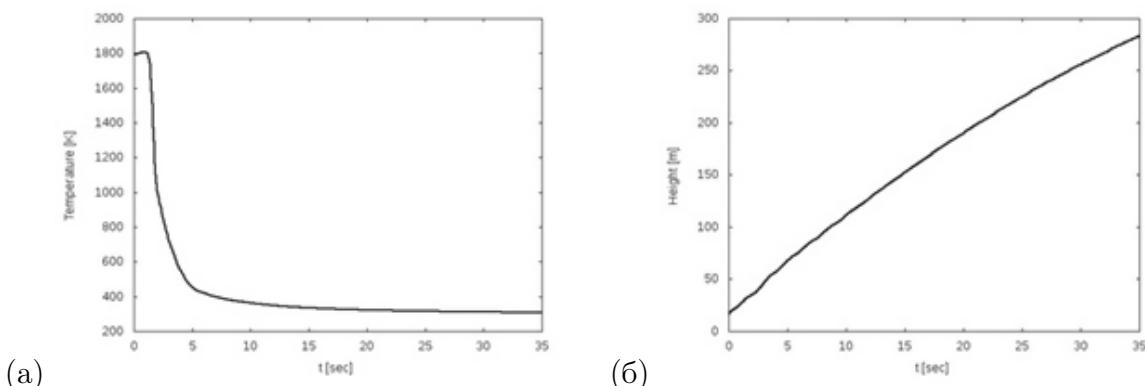


Рисунок 7: Изменение (а) температуры и (б) высоты крупномасштабного термика при начальной температуре $1800 K$.

5 Заключение

В данной работе была рассмотрена динамика крупномасштабных термиков под действием силы плавучести, с учетом эффектов турбулентного перемешивания и адиабатического расширения.

Математическая модель крупномасштабного термика построена на основе системы дифференциальных уравнений в частных производных. Были учтены уравнение неразрывности, количества движения, температуры, давления и энергии.

Была поставлена задача о поднятии крупномасштабного термика с заданными начальными температурой и плотностью. К расчетной области применены граничные условия Неймана и Дирихле.

Численный алгоритм основан на решении решеточных уравнений Больцмана в сжимаемой постановке. Использованный численный алгоритм имеет второй порядок точности по времени и по пространству. Порядок точности численного алгоритма оценен с помощью тестовой задачи течение Пуазейля.

Получены результаты численного моделирования динамики крупномасштабных термиков. Определенная зависимость между начальной температурой и высотой поднятия облака соответствуют результатам работы Онуфриева [12].

Список литературы

- [1] Speziale C.G., So R.M.C. Turbulence modeling and simulation. -Florida: Press LLC, 2016. -73
- [2] Гарбарук А.В., Стрелец М.Х., Травин А.К., Шур М.Л. Современные подходы к моделированию турбулентности: учеб. пособие. – СПб.: Изд-во Политехн. ун-та, 2016. – 234 с.
- [3] Kizbayev A., Zhakebayev D., Abdibekov U., Khikmetov A. Mathematical modeling of electron irradiation of oil// Engineering Computations, Vol. 35 Issue: 5, -PP.1998-2009.
- [4] Abdibekov S., Zhakebayev D., Karzhaubayev K., and Abdibekov U. Large eddy simulation the evolution of the cloud explosion of a launch vehicle// Совместный выпуск Вестник Восточно-Казахстанского государственного технического университета им. Д. Серикбаева и научного журнала Института вычислительных технологий Сибирского отделения РАН ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ. -2018, Т. 1, N 3. -С. 7-17.
- [5] Жумагулов Б.Т., Жакебаев Д.Б., Абдигеков У.С., Математическое моделирование вырождения энергии турбулентности на основе гибридного метода// «Вестник НИА РК» № 3 – 2018 г., -С. 9-15.
- [6] Ben-Nasra O., Hadjadj A., Chaudhurib A., Shadloo M.S. Assessment of subgrid-scale modeling for large-eddy simulation of a spatially-evolving compressible turbulent boundary layer // Computers and Fluids. - 2017. - Vol. 151. - P. 144-158.
- [7] Meneveau C., Katz J. Scale-invariance and turbulence models for large eddy simulation // Annu. Rev. Fluid Mech. – 2000. – Vol. 32. – P.1-32.
- [8] You D., Moin P. A dynamic global-coefficient subgrid-scale eddy-viscosity model for large-eddy simulation in complex geometries // Physics of Fluids. – 2007. – Vol. 19, No. 6. –065110.
- [9] Oran E.S., Boris J.P. Numerical Simulation of reactive flow. – Cambridge University Press, 2001. – 550 p.
- [10] Grinstein F.F., Margolin L.G., Rider W.J. Implicit Large Eddy Simulation. – Cambridge University Press, 2007. – 577 p.
- [11] Spalart P.R. Strategies for turbulence modeling and simulations // Int. J. of Heat and Fluid Flow. – 2000. – Vol. 21. – P. 252-263.
- [12] Онуфриев А.Т. Теория движения вихревого кольца под действием силы тяжести. Подъем облака атомного взрыва// Прикладная механика и техническая физика 2 -1967. -С. 3-15.
- [13] Кудряшов Н.А. Влияние вязкости и теплопроводности на всплытие термика под действием сил плавучести, - Москва: Наука, 1985. -135 с.

- [14] Dovgalyuk Y.A., Zatevakhin M.A., Stankova E.N. Numerical Simulation of a Buoyant Thermal Using the Turbulence Model // Journal of applied Meteorology. -1994. -P. 1118-1126.
- [15] Sinjab, I. M., Robertson, J. A., Connon Smith, R. The dissipation factor in contact binaries revisited // Monthly Notices of the Royal Astronomical Society. -1990. -P. 619.
- [16] Hazlehurst, J. The dissipation factor in contact binaries // Astronomy and Astrophysics. -1985. -P. 481-488.
- [17] Brandenburg A., Hazlehurst J. Evolution of highly buoyant thermals in a stratified layer // Astronomy and Astrophysics. -2001. -P. 1092-1102.
- [18] Zhaoli G., Chaguang Z., Baochang S. Thermal lattice Boltzmann equation for low Mach number flow // Physical Review E. -2007. -P. 367.
- [19] Woods L.C. An Introduction to the Kinetic Theory of Gases and Magnetoplasmas. -Oxford: Oxford University Press, 1993. -372 p.
- [20] Kruger, T., Kusumaatmaja, H., Kuzmin, A., Shardt, O., Silva, G., Viggen, E. M. The Lattice Boltzmann Method. - Springer, 2017. -35 p.
- [21] Chang S.C., yang Y.T., Chen C.K., Chen W.L. Application of the lattice Boltzmann method combined with large-eddy simulations to turbulent convective heat transfer // International Journal of Heat and Mass Transfer. -2013. Vol. 66. -P. 338-348.
- [22] Pradhan A., Yadav S. Large Eddy Simulation using Lattice Boltzmann Method based on Sigma Model // Procedia Engineering. 2015. Vol 127. P. 177-184.
- [23] Sagaut P. Toward advanced subgrid models for Lattice-Boltzmann-based Large-eddy simulation: Theoretical formulations // Computers and Mathematics with Applications. 2010. Vol. 59. P. 2194-2199.
- [24] Grunau D., Chen S., Eggert K. A lattice Boltzmann model for multiphase fluid flows// Phys. Fluids, -1993. -Vol. 5. -P. 2557-2562.
- [25] Amirshaghghi H., Rahimian M.H., Safari H., Krafczyk M. Large Eddy Simulation of liquid sheet breakup using a two-phase lattice Boltzmann method // Computers and Fluids. - 2018. -Vol. 160. - P. 93-107.
- [26] Chen C.K., Chang S.C., Sun S.Y. Lattice Boltzmann method simulation of channel flow with square pillars inside by the field synergy principle // CMES-Comput. Model. Eng. Sci. 2007. Vol. P. 203–215.
- [27] Yang Y.T., Chang S.C., Chiou C.S. Lattice Boltzmann method and large-eddy simulation for turbulent impinging jet cooling // International Journal of Heat and Mass Transfer. 2013.- Vol. 61. - P. 543-553.
- [28] Zhang J. Lattice Boltzmann method for microfluidics: models and applications // Microfluid. Nanofluid. - 2011. - Vol. 10. - P. 1–28.
- [29] Guo Z., Zhao T.S. Lattice Boltzmann model for incompressible flows through porous media // Phys.Rev.E. -2002. -Vol. 66. - 036304.
- [30] Kuwata Y., Suga K. Large eddy simulations of pore-scale turbulent flows in porous media by the lattice Boltzmann method // International Journal of Heat and Mass Transfer.- 2015. -Vol. 55. -P. 143-157.
- [31] Chang S.C., Hsu Y.S., Chen C.L. Lattice Boltzmann simulation of fluid flows with fractal geometry: an unknown-index algorithm // J.Chin.Soc.Mech.Eng. -2011. -Vol. 32. - P. 523–531.

References

- [1] Speziale C.G., So R.M.C. *Turbulence Modeling and Simulation Handbook of Fluid Dynamics* (Florida: Press LLC, 2016):1-73
- [2] Garbaruk A.V., Strelec M.H., Travin A.K., Shur M.L. "Sovremennye podhody k modelirovaniyu turbulentnosti: ucheb. posobie" [Modern approaches for modeling of turbulence] (SPb, Izd-voPolitehn. un-ta, 2016),234 s.
- [3] Assylzhan Kizbayev, Dauren Zhakebayev, Ualikhan Abdibekov, Askar Khikmetov, "Mathematical modeling of electron irradiation of oil", Engineering Computations(2018),pp.1998-2009.

- [4] Abdibekov S., Zhakebayev D., Karzhaubayev K., Abdibekov U., "Large eddy simulation the evolution of the cloud explosion of a launch vehicle", VYCHISLITELNYE TEHNOLOGII (2018), 7-17 p.
- [5] Zhumagulov B.T., Zhakebaev D.B., Abdibekov U.S., "Matematicheskoe modelirovanie vyrozhdenija jenergii turbulentnosti na osnove gibriddenogo metoda"/[Mathematical modeling of turbulence energy by gibridden method], Vestnik NIA RK № 3(2018), 9-15 p.
- [6] Ben-Nasra O., Hadjadj A., Chaudhuri A., Shadloo M.S. "Assessment of subgrid-scale modeling for large-eddy simulation of a spatially-evolving compressible turbulent boundary layer", Computers and Fluids (2017), 144-158 p.
- [7] Meneveau C., Katz J. Scale-invariance and turbulence models for large eddy simulation, Annu. Rev. Fluid Mech(2000), 1-32 p.
- [8] You D., Moin P. A dynamic global-coefficient subgrid-scale eddy-viscosity model for large-eddy simulation in complex geometries, // Physics of Fluids. – 2007. – Vol. 19, No. 6. –065110.
- [9] Oran E.S., Boris J.P. Numerical Simulation of reactive flow (Cambridge University Press, 2001), 550 p.
- [10] Grinstein F.F., Margolin L.G., Rider W.J. Implicit Large Eddy Simulation(Cambridge University Press, 2007), 577 p.
- [11] Spalart P.R. "Strategies for turbulence modeling and simulations Int. J. of Heat and Fluid Flow (2000), 252-263 p.
- [12] Onufriev A.T. "Teoriya dvizhenija vikhrevogo kol'ca pod dejstviem sily tjazhesti. Podyem oblaka atomnogo vzryva,"*Prikladnaja mehanika i tehnicheskaja fizika*(1967), 3-15 p.
- [13] Kudrjashov N.A. "Vlijanie vjazkosti i teploprovodnosti na vslyvanie termika pod dejstviem sil plavuchesti"/[Humid viscosity and heat capacity of thermals](Moskva: Nauka, 1985), 135 p.
- [14] Dovgalyuk Y.A., Zatevakhin M.A., Stankova E.N. "Numerical Simulation of a Buoyant Thermal Using the Turbulence Model," *Journal of applied Meteorology* 33 (1994): 1118-1126
- [15] Sinjab, I. M., Robertson, J. A., Connon Smith, R. "The dissipation factor in contact binaries revisited," *Monthly Notices of the Royal Astronomical Society* 244 (1990):619
- [16] Hazlehurst, J. "The dissipation factor in contact binaries," *Astronomy and Astrophysics* 145 (1985): 481-488
- [17] Brandenburg A., Hazlehurst J. "Evolution of highly buoyant thermals in a stratified layer," *Astronomy and Astrophysics* 370 (2001): 1092-1102
- [18] Zhaoli G., Chaguang Z., Baochang S. "Thermal lattice Boltzmann equation for low Mach number flow," *Physical Review E* 75 (2007): 036704
- [19] Woods L.C. "An Introduction to the Kinetic Theory of Gases and Magnetoplasmas," *Oxford University Press* (Oxford: 1993)
- [20] Kruger, T., Kusumaatmaja, H., Kuzmin, A., Shardt, O., Silva, G., Viggen, E. M. "The Lattice Boltzmann Method," *Springer* 215 (2017): 35
- [21] Chang S.C., yang Y.T., Chen C.K., Chen W.L. "Application of the lattice Boltzmann method combined with large-eddy simulations to turbulent convective heat transfer" *International Journal of Heat and Mass Transfer* (2013), 338-348 p.
- [22] Pradhan A., Yadav S. "Large Eddy Simulation using Lattice Boltzmann Method based on Sigma Model Procedia Engineering" (2015), 177-184 p.
- [23] Sagaut P. "Toward advanced subgrid models for Lattice-Boltzmann-based Large-eddy simulation: Theoretical formulations" *Computers and Mathematics with Applications* (2010), 2194-2199 p.
- [24] Grunau D., Chen S., Eggert K. "A lattice Boltzmann model for multiphase fluid flows Phys. Fluids. (1993), 2557-2562 p.
- [25] Amirshaghagh H., Rahimian M.H., Safari H., Krafczyk M. "Large Eddy Simulation of liquid sheet breakup using a two-phase lattice Boltzmann method" *Computers and Fluids*(2018), 93-107 p.
- [26] Chen C.K., Chang S.C., Sun S.Y. "Lattice Boltzmann method simulation of channel flow with square pillars inside by the field synergy principle CMES-Comput. Model. Eng. Sci. (2007), 203-215.
- [27] Yang Y.T., Chang S.C., Chiou C.S. "Lattice Boltzmann method and large-eddy simulation for turbulent impinging jet cooling" *International Journal of Heat and Mass Transfer*(2013), 543-553.

-
- [28] Zhang J. "Lattice Boltzmann method for microfluidics: models and applications" *Microfluid. Nanofluid.* (2011), 1–28.
 - [29] Guo Z., Zhao T.S. "Lattice Boltzmann model for incompressible flows through porous media" *Phys. Rev. E.* (2002), 036304.
 - [30] Kuwata Y., Suga K. "Large eddy simulations of pore-scale turbulent flows in porous media by the lattice Boltzmann method" *International Journal of Heat and Mass Transfer* (2015), 143–157.
 - [31] Chang S.C., Hsu Y.S., Chen C.L. "Lattice Boltzmann simulation of fluid flows with fractal geometry: an unknown-index algorithm" *J. Chin. Soc. Mech. Eng.* (2011), 523–531.

МРНТИ 27.35.63

Численный алгоритм решения задач электрического зондирования среды с рельефом земной поверхности

Миргаликызы Т., Евразийский национальный университет имени Л.Н.Гумилева,
г. Астана, Республика Казахстан, E-mail: m_t85@mail.ru
Муканова Б.Г., Евразийский национальный университет имени Л.Н.Гумилева,
г. Астана, Республика Казахстан, E-mail: mbsha01@gmail.com

На сегодняшний день проблема влияния рельефа земной поверхности на распределение электрического поля является актуальным вопросом в интерпретации электромагнитных полей, изучаемых в электроразведке постоянным током. В работе рассматривается тестирование алгоритма численного решения задачи электрического зондирования среды с рельефом земной поверхности путем моделирования методом интегральных уравнений. Идея метода интегральных уравнений заключается в представлении электрического поля, как суммы первичного поля и поля вторичных зарядов. Контактные границы и поверхность геоэлектрического разреза выступают как вторичные возбудители электрического поля. Задача расчета полей сводится к системе интегральных уравнений на плотности вторичных источников, индуцируемых на поверхностях контакта проводящих сред и на рельефной поверхности среды. Математическое описание этого явления приводит к уравнениям Фредгольма II рода с полярным ядром. Алгоритм расчета протестирован путем сравнения результатов с решениями, приведенными в работах в открытом доступе по учету влияния рельефа, задавая такие же параметры среды. Выполнены сравнения полученными разными подходами конечно-элементных и конечно-разностных методов для учета влияния рельефа в 2D средах. Также выполнены сравнения полученными данными физического моделирования. Можно сказать, что выполненные нами результаты расчетов кривых кажущихся сопротивлений находятся в хорошем согласии с имеющимися исследованиями в этой области.

Ключевые слова: метод интегральных уравнений, численный алгоритм, прямая задача электроразведки, рельеф земной поверхности.

Жер беті бедері бар орта үшін электрлік барлау есебін шешүге арналған сандық алгоритм

Миргаликызы Т., Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ.,
Қазақстан Республикасы, E-mail: m_t85@mail.ru
Муканова Б.Г., Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ.,
Қазақстан Республикасы, E-mail: mbsha01@gmail.com

Электр өрісінің таралуына жер беті бедерінің әсері түрақты тоқпен электр барлауы арқылы зерттелетін электромагнитті өрістерді сипаттауда казыргы уақытта өзекті мәселелердің бірі болып табылады. Бұл жұмыста интегралдық теңдеулер әдісімен модельдеу жолымен жер беті бедері бар органды электрлік зондтау есебін сандық шешудің алгоритмін тестілеу қарастырылады. Интегралды теңдеулер әдісінің идеясы электр өрісін алғашқы өріспен екіншілік зарядтар өрісінің қосындысы ретінде қарастыру болып табылады. Жанасу шекаралары және геоэлектрлік қима беті электр өрісінің екіншілік қоздырғышы болып табылады. өрістерді есептеу есебі өткізгіш орталар беттері мен бедерлі жер беті шекарасында индукцияланатын екіншілік тоқ көздерінің тығыздығы бойынша құрылатын интегралдық теңдеулер жүйесіне негізделеді. Бұл құбылыстың математикалық сипатталуы полярлық ядросы бар II түрдегі Фредгольм теңдеулеріне алып келеді. Есептеу алгоритмі нәтижелерді жер бедерінің әсерін есепке алу бойынша ашық қол жетімді жұмыстарда келтірілген шешімдермен салыстыру жолымен, органдың дәл сондай параметрлерін бере отырып, тестілеу өткізілді. 2D орталардағы жер беті бедерінің әсерін есепке алу үшін ақтық-элементтік және ақтық-айырымдық әдістер арқылы, әр түрлі тәсілмен алынған салыстыру орындалды.

Сонымен қатар физикалық модельдеудің алынған мәліметтерімен салыстыру жүргізілді. Біз орындаған жорамал кедергінің қисықтарын есептеулер нәтижелері осы саладағы зерттеулермен жақын үйлесім табады деп айтуга болады.

Түйін сөздер: интегралдық теңдеулер әдісі, сандық алгоритм, электрлік барлаудың тұра есебі, жер беті бедері.

Numerical algorithm for solving of electrical prospecting problems for medium with a surface relief

Mirgalikyzy T., L.N. Gumilyov Eurasian national university, Astana, Republic of Kazakhstan,
E-mail: m_t85@mail.ru

Mukanova B.G., L.N. Gumilyov Eurasian national university, Astana, Republic of Kazakhstan,
E-mail: mbsha01@gmail.com

Nowadays the problem of the influence of the ground surface relief on the distribution of the electric field is a pressing issue in the interpretation of electromagnetic fields studied in the direct current electrical prospecting. The paper deals with testing the algorithm for the numerical solution of the problem of electrical sensing of a medium with the ground surface relief by means of modeling using integral equations. The idea of the method of integral equations is to represent the electric field as the sum of the primary field and the field of the secondary charges. The contact boundaries and the surface of the geoelectric section act as a secondary creators of the electric field. The problem of calculating the fields is reduced to the systems of integral equations on the density of secondary sources induced on the contact surfaces of conducting media and on the relief surface of the medium. A mathematical description of this phenomenon leads to Fredholm equations of the second kind with a polar core. The calculation algorithm was tested by comparing the results with the solutions given in the open access works on taking into account the influence of the relief, setting the same environmental parameters. Comparisons are made using the finite-element and finite-difference methods obtained by different approaches to take into account the effect of relief in 2D environments. Comparisons are also made with the physical modeling data obtained. We can say that our results of calculations curves of apparent resistivity are in good agreement with the available research in this area.

Key words: method of integral equations, numerical algorithm, direct problem of electrical prospecting, surface relief.

1 Введение

Настоящее время характеризуется бурным развитием геофизических методов для решения разнообразных задач поиска и разведки объектов, расположенных в сложной горизонтально-неоднородной вмещающей среде. В частности в последнее время получил развитие и внедрение такой метод как электротомография. Электротомография - это современная методика геофизических исследований методом сопротивлений [1]. Практическая реализация электротомографических исследований потребовало создания специализированных аппаратных систем, а также развитых программных средств для обработки данных и осуществления инверсии.

В задачах вертикального электрического зондирования важное значение имеет расчет влияния условий эксперимента на кривые зондирования, в частности, рельефа зондируемой поверхности. В настоящее время построение геоэлектрического разреза на основе измерений кажущегося сопротивления проводится с помощью программ 2D и 3D инверсии. Эти программы выполняют решение обратной задачи (например, Res2DInv, M.H.Loke и ZondRes2D, A.E.Каминский) [2],[3]. В большинстве случаев имеющиеся резкие геоэлектрические границы получаются размытыми, а искажения кривых кажущегося сопротивления, связанные с рельефом поверхности, генерируют

ложные аномалии. В этих условиях двумерная инверсия не является идеальным способом перехода от электрического поля к реальному геоэлектрическому разрезу. Другим источником ошибок является то обстоятельство, что до сих пор не существует программ, которые с высокой точностью количественно учитывают влияние искажений, связанных с рельефом земной поверхности. Обработка данных на основе существующих программ двумерной инверсии показывает, что вблизи отклонений поверхности среды от плоскости образуются фиктивные аномалии геоэлектрического разреза [4].

Задача в условиях сложного рельефа земной поверхности в настоящее время полностью не решена. Имеются лишь отдельные результаты, полученные на физических моделях среды и расчетные данные, полученные в приближении однородного внешнего поля и для узко специальных типов рельефа [5],[6].

Данная работа является продолжением исследований авторов на эту тему [7],[8]. Новизна нашего подхода заключается в том, что моделирование проводится на основе численного решения системы интегральных уравнений. Характерной особенностью разработанных нами алгоритмов является, то, что они позволяют выполнять расчеты для более разнообразных форм рельефов, включая формы, определяемые из натурных экспериментов.

2 Обзор литературы

В связи с большой важностью проблемы учета влияния рельефа земной поверхности изучение влияния рельефа на данные методов постоянного тока предпринималось достаточно давно. К ним можно отнести работы Запорожец В.М. 1938 [9], Чантуришивили Л.С. 1959 [10], Вешев А.В. 1959, 1980 [11],[5]. В настоящее время основными математическими методами, используемыми на практике при решении прямых и обратных задач, являются сеточные методы (конечных разностей и конечных элементов) [12]-[17]. В рамках этих методов разработано несколько способов учета влияния рельефа [18]-[25].

В методе конечных разностей широко используется алгоритм геометрической трансформации ячеек сетки адаптированной к рельефу земной поверхности [18]. Второй алгоритм в методе конечных разностей основан на триангуляции расчетной области, где элементы сетки формируются путем дополнения диагоналей в прямоугольных клетках, обеспечивая гибкость дизайна модели [19]-[20]. Также и в методе конечных элементов существует ряд алгоритмов учета влияния рельефа земной поверхности [21]-[24]. Один из алгоритмов схожий с первым из описанных для конечных разностей - адаптированная к рельефу трансформация ячеек сетки. Другой алгоритм основан в построении специальной нерегулярной сети в зависимости от формы рельефа [25],[26] и степени неоднородности исследуемой среды.

Все указанные способы расчета основаны на единой общей идее и физическом принципе. Универсальность, гибкость и ряд вычислительных преимуществ сеточных методов определили их широкое распространение во многих сферах применения численного моделирования, и, в том числе, в методах постоянного тока. Однако при задании модели с рельефом в большинстве из них конфигурация ячеек сетки, так или иначе, повторяет конфигурацию рельефа. Это может вносить ошибки в решение как прямых, так и, главное, обратных задач. Поэтому для контроля работы сеточных

методов целесообразно применять другие методы, основанные на других идеях и также позволяющие решать соответствующие прямые задачи метода сопротивления. В рамках настоящей работы был выбран метод интегральных уравнений, имеющий простой физический смысл и хорошо зарекомендовавший себя при проведении моделирования в методе сопротивлений [27]-[32].

3 Материал и методы

Математическая модель и численный алгоритм

Метод интегральных уравнений основан на теории потенциала для решений уравнения Лапласа [33]. Как показано в [7], задача расчета поля постоянного тока в однородной среде с рельефом дневной поверхности сводится к решению интегрального уравнения Фредгольма второго рода с ядром со слабой особенностью:

$$q(P) = \iint_{\Gamma} q(M) \frac{\cos\psi_{PM}}{|\overrightarrow{PM}|^2} d(M) + F_0(P) \quad (1)$$

Здесь M, P - точки поверхности среды Γ , по которой берется интеграл, $q(P)$ - плотность простого слоя на границе Γ , через которую в дальнейшем рассчитывается потенциал поля, угол ψ_{PM} - угол между направлением нормали к поверхности в точке P и вектором PM , $F_0(P)$ - заданная функция.

Численная реализация метода интегральных уравнений выполнена с использованием метода радиальных базисных функций (RBF) для аппроксимации произвольного рельефа поверхности и адаптивной сетки с учетом геометрии измерительной установки [34].

Далее описывается пошаговый алгоритм численного решения интегральных уравнений с использованием метода RBF.

1 Описание модели среды.

Описывается исследуемая среда: длина измерительной линий, количество и высоты приемных электродов относительно питающего электрода, длина приемной линии, сила тока питающего электрода.

2 Составление и выбор функции RBF аппроксимирующей рельеф дневной поверхности:

- Для заданного набора точек плоскости $Oxy \{x_i\}$ ($i = \overline{1, n}$) и соответствующих значений $\{f_i\}_{i=1}^n$ строится интерполирующая функция $s(x,y)$ двух переменных.

- Определение радиальной базисной функции. Выполняется выбор подходящей радиальной функции, подбирая требуемую гладкость формы поверхности для решения задачи [34].

- Строится матрица по формуле $A_{i,j} = \phi(\|x_i - x_j\|)$, ($i, j = \overline{1, n}$).

- Методом LU-разложения с выбором главного элемента решается система $A\lambda = f$. Коэффициенты λ_i составляют вектор решения. Чтобы вычислить значение интерполируемой функции в точке (x,y) используется формула $s(x,y) = \sum_{i=1}^n \lambda_i \phi(\|(x,y) - x_i\|)$, где ϕ - радиальные функции.

3 Построение адаптивной сетки, с учетом геометрии измерительной установки.

Строится сетка с триангуляцией для разбиения границы на ячейки, аппроксимирующие вторичные источники тока с учетом особенностей нашей сетки – послойное расположение треугольников и логарифмическое расширение сетки при удалении от оси расчетной области. Для расчета каждого сопротивления и потенциала поля в качестве расчетной области используется область в форме овала. Овал образуется двумя полукругами и прямоугольником между ними. Ось, соединяющая центры окружностей, соответствует положению измерительной линии, одна из центров окружностей - положению питающего электрода. Определяются параметры каждой ячейки: номер узла, координаты геометрического центра, площадь:

- Верхняя и нижняя части овала делятся на заданное количество слоев (например на N) по правилу концентрических полуокружностей (полуволов) радиуса $r_i = \exp(i * hs - 1) / \alpha$, i - номер концентрического полуовала, $hs = \ln(1 + a * \alpha) / N$ - шаг сетки по радиусу в логарифмических преобразованных координатах, α - коэффициент неравномерности сетки. Каждый слой разбивается на треугольники, начиная с первого слоя.

- Каждый слой разбивается на треугольники, начиная с первого слоя. Определяется массив $uz(i,j)$, где $i=0, \dots, N$ - номер слоя, на верхней границе которого лежит узел, j - номер узла на этом слое. Массиву узлов покоординатно присваиваются значения. После выполняется триангуляция: заводится динамический массив tr типа «треугольник» для верхней части овала и $tr1$ – для нижней части овала, которые в качестве вершины треугольника содержат номера узлов, образующих эту вершину.

- Вычисление площади ячеек проводились по формулам аналитической геометрии, на основе определителя построенного по декартовым координатам узлов ячеек.

4 Составление матрицы коэффициентов взаимного влияния.

5 Решение системы линейных алгебраических уравнений и определение плотности вторичных источников тока.

6 Расчет потенциалов на поверхности.

7 Расчет функций каждого сопротивления вдоль поверхности.

4 Результаты и обсуждение

Выполнено тестирование численных результатов путем сравнения их с результатами приведенными в работах по учету влияния рельефа, задавая такие же параметры среды.

В трудах Erhan Erdogan и Ismail Demirci (2008, 2012 гг.) для учета влияния рельефа в 2D средах описываются разные подходы использования конечно-элементных (FE) и конечно-разностных (FD) методов [19], [20]. В отличие от других исследователей, которые также используют прямое решение FD, они использовали треугольные элементы, которые формируются путем дополнения диагоналей в прямоугольных клетках, обеспечивая гибкость дизайна модели, который обычно доступны только при использовании метода FE. Авторы добавляют рельеф в прямую задачу, с использованием трех подходов, основанных на методе FE:

1 Присвоили значения с высоким удельным сопротивлением (представляющий часть воздуха) ко всем треугольным элементам, которые представляют собой клетки границы воздух/земля. Этот FE полный воздушный (full-air) подход (рисунок 1а) представляет собой рельефную границу в лестничной форме.

2 Используя преимущество, полученное путем применения треугольных элементов, моделировали рельефную границу там, где она пересекала прямоугольные ячейки и присвоили значения с высоким удельным сопротивлением к треугольным элементам на границе ячеек воздуха/земля. Это половинно воздушный (half-air) подход FE представляет собой более плавный и точный рельефный край (рисунок 1b).

3 Равномерно искривляли FE сетки по отношению к геометрии поверхности, чтобы имитировать эффект рельефа (Holcombe и Jiracek, 1984; Loke, 2000). Такой подход называется FE искривление сетки (FE distorted - mesh) (рисунок 1c).

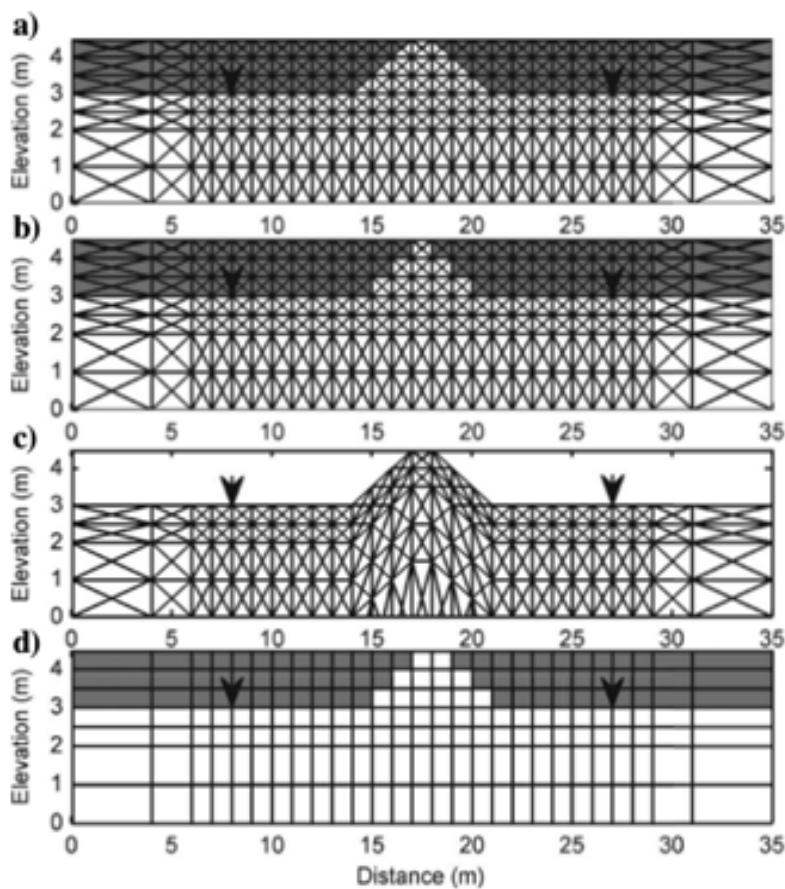


Рисунок 1: Схема 2D моделирования сетки, которые включают рельеф: (a) FE full-air, (b) FE half-air и FD half-air, (c) FE distorted-mesh, и (d) FD full-air

Сетка FD состоит из прямоугольных ячеек и не может быть реорганизован; рельеф можно моделировать только с помощью ячеек поверхностных сопротивлений, которые представляют воздействие воздуха в методе FD. Первый подход, который они использовали для имитации рельефа, присвоили значениям с высоким удельным сопротивлением ко всем прямоугольным ячейкам на границе воздух-земля и назвали его FD полный воздушный (full - air) подход (рисунок 1d). После этого использовали формулировку Вивера (Weaver 1994) и разделили прямоугольные ячейки в треугольные элементы (рисунок 2), используя ту же сетку, что и в методе FE на рисунок 1b и назвали

его FD половинно воздушный (half - air) подход.

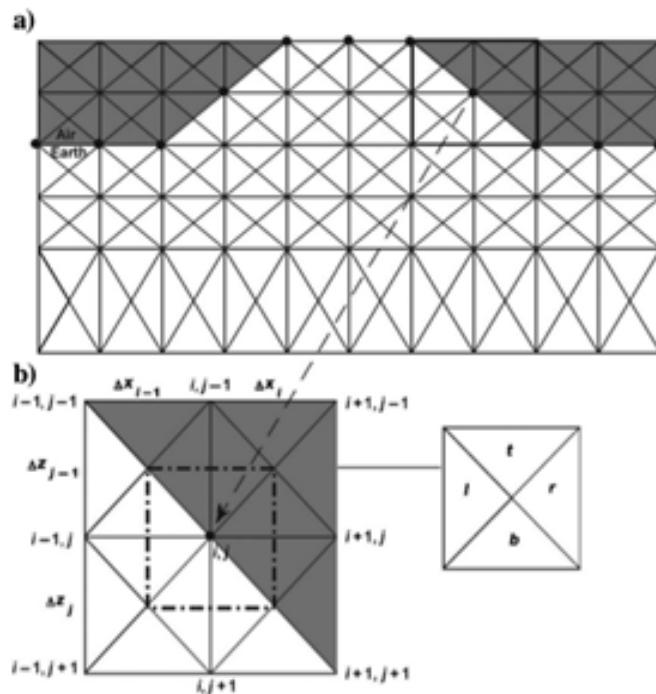


Рисунок 2: (а) FD half-air сетка. Каждая ячейка подразделяется на четыре элемента триангуляции. (б) Дискретизация ячеек FD триангуляции между интерфейсом воздух-земля

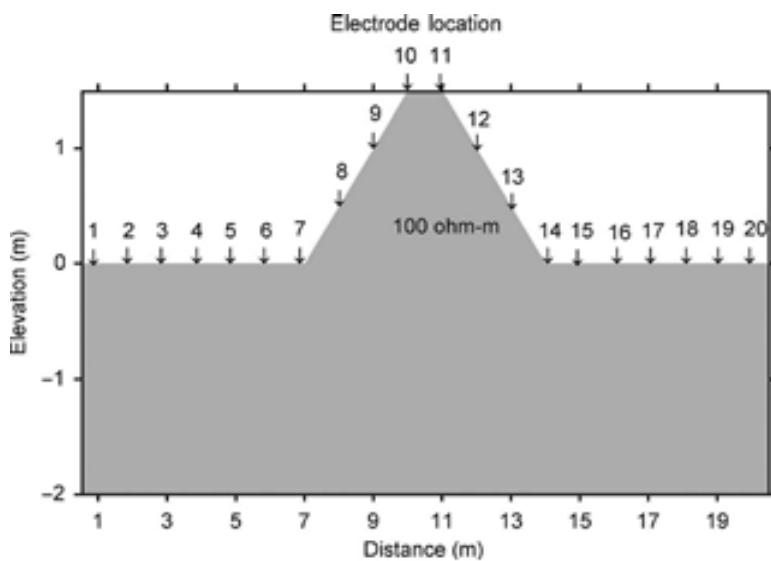


Рисунок 3: Однородная модель холма с углом уклона, равным 30°

В работе приводятся результаты моделирования на рельефной поверхности для

однородной среды с удельным сопротивлением 100 Ом.м. Рельеф поверхности - на 30° отклоненный холм (рисунок 3). Результаты моделирования для установки Веннер-Шлюмберже для всех подходов приведены на рисунке 4. Во всех решениях в центральном районе можно увидеть аномалию низкого кажущегося сопротивления в нижней области холма.

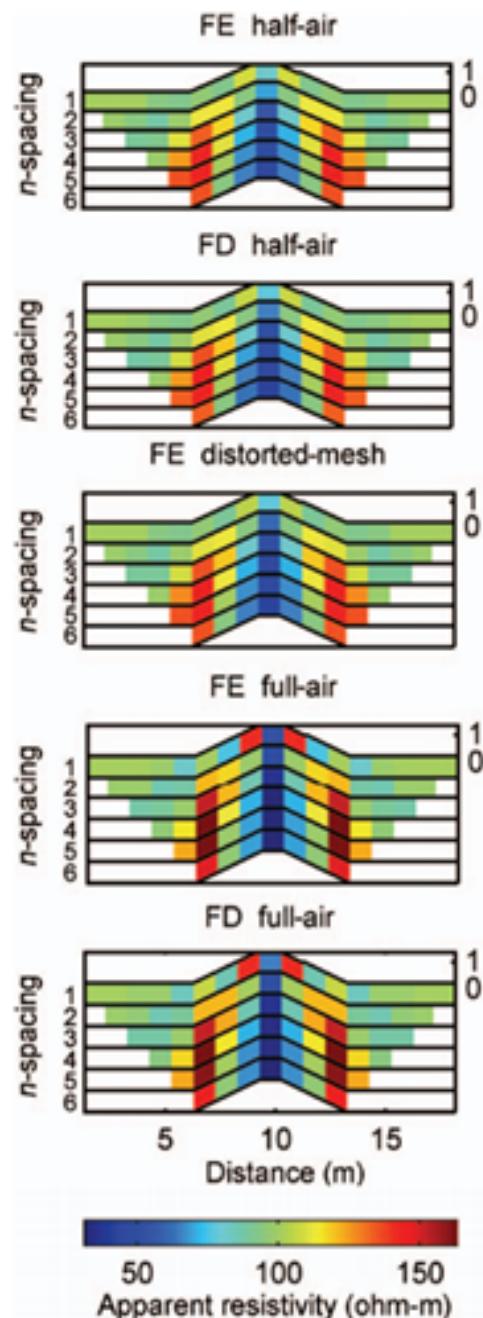


Рисунок 4: Псевдоразрез кажущегося сопротивления однородной модели холма, с углом уклона 30°

Результаты полученные нашим методом для описанной модели представлены на рисунке 5. По мере приближения к угловой точке наблюдается повышение и резкое увеличение значении кажущегося сопротивления, а вершина отмечается минимумом кажущегося сопротивления. Затем в окрестности следующей угловой точки также наблюдается повышение ρ_k . В дальнейшем происходит плавное уменьшение и приближение к удельному сопротивлению среды. Такое поведение кажущегося сопротивления можно увидеть на каждой картинке полученных разными подходами методов FE и FD.

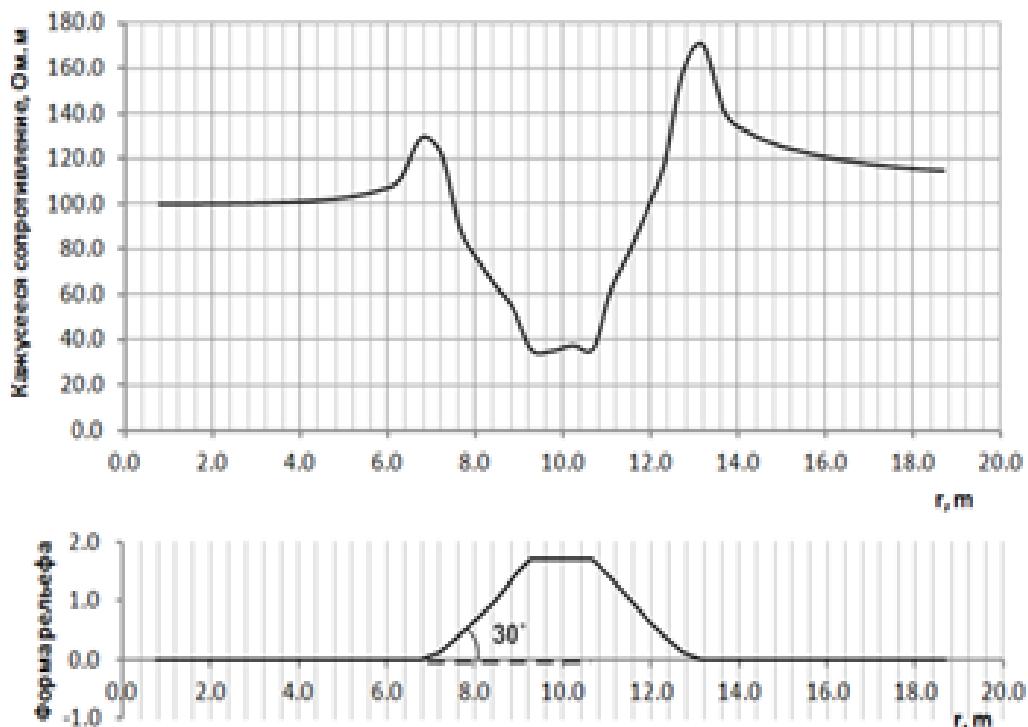


Рисунок 5: Кажущиеся сопротивления, полученные методом интегральных уравнений и форма рельефа для однородной модели холма, угол наклона 30°

Такие результаты встречаются также в работах R.C. Fox, G.W. Hohmann 1978, I.Tsourlos, E.Szymanski и др. 1999, QIANG Jian-Ke и др. 2007, Sedat Yilmaz, Nart Coşkun 2011, Sebastien Penz и др. 2013.

Дополнительное сравнение результатов программы выполнено полученными данными физического моделирования, представленных в работе Вешева А.В. [11]. Например, на рисунке 6, скопированном со статьи Вешева А.В. построены кажущиеся сопротивления для разных параметров установки для однородной модели полусферической выемки. На рисунке 7 построен график кажущегося сопротивления, полученный применяемым нами методом интегральных уравнений для данной модели. Можно сказать, результаты практически совпадают, имеющиеся незначительные отклонения, связаны с разными геометрическими параметрами установки.

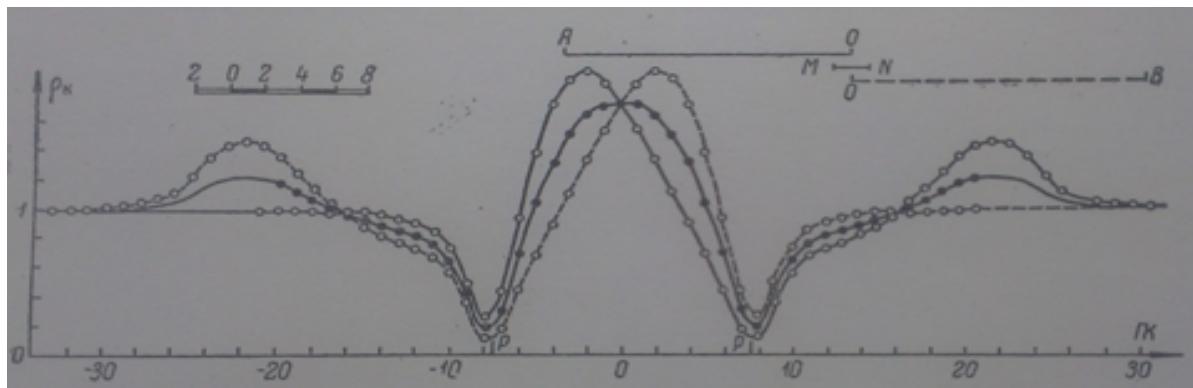


Рисунок 6: Результаты физического моделирования Вешева А.В. для модели полусферической выемки

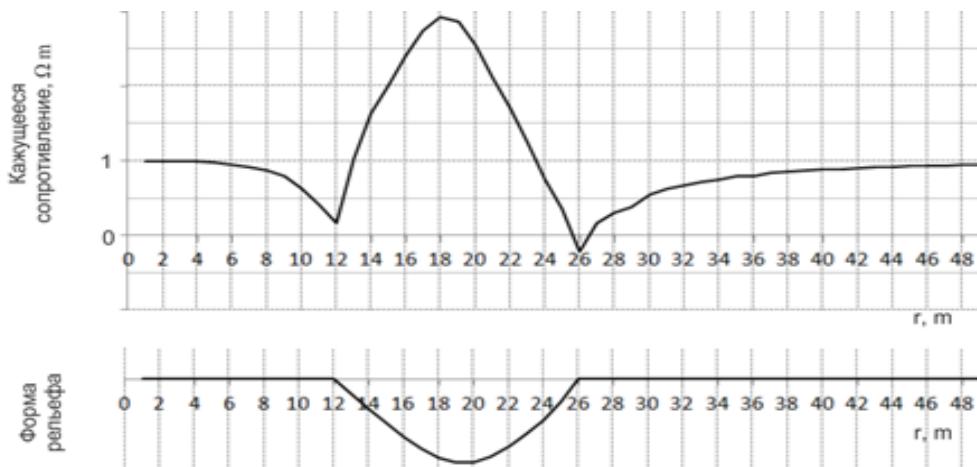


Рисунок 7: Результаты, полученные методом интегральных уравнений для однородной модели полусферической выемки

5 Заключение

Разработан численный алгоритм решения задач электрического зондирования, который позволил построить функцию кажущегося сопротивления, с учетом основных закономерностей формирования электрического поля, связанных с влиянием рельефа. Выполнено тестирование численных результатов, сравнивая с результатами, приведенными в работах в открытом доступе по учету влияния рельефа, задавая такие же параметры среды. Выполнены сравнения полученными разными подходами конечно-элементных и конечно-разностных методов для учета влияния рельефа в 2D средах для заданной модели с рельефной поверхностью. Также выполнены сравнения полученными данными физического моделирования, представленных в работе Вешева А.В. Можно сказать, что результаты совпадают, но есть незначительные отклонения, связанные с разными параметрами геометрической установки. Выполненные нами

результаты расчетов кривых кажущихся сопротивлений находятся в хорошем согласии с имеющимися исследованиями в этой области. Однако, разработанные нами программы позволяют выполнять расчеты для более разнообразных форм рельефов, включая также формы, определяемые из натурных экспериментов. Дальнейшие исследования будут связаны с расчетом эффекта влияния рельефа, взятого из натурных экспериментов и для усложненных моделей сред. Конечная цель данного исследования – это задача исключения из результатов интерпретации ложных аномалий, связанных с рельефом.

Список литературы

- [1] Шевнин В.А. Электроразведка: Пособие по электроразведочной практике для студентов геофизических специальностей. – М.: МГУ, 2013. - Том II. – 124 с.
- [2] Loke M.H., Barker R.D. Practical techniques for 3D resistivity surveys and data inversion // Geophysical Prospecting. - 1996. – Vol. 44. – P. 499-523.
- [3] Gunther T., Rucker C. Boundless Electrical Resistivity Tomography // BERT 2 - the user tutorial. – 2013. – Ver. 2.0.
- [4] Баранчук К.И., Миргаликызы Т., Модин И.Н., Муканова Б.Г. Физическое моделирование электрической томографии на поверхности со сложным рельефом // Инженерные изыскания. - Москва, 2017. - №11. - С.56-65.
- [5] Вешев А.В. Электропрофилирование на постоянном и переменном токе. - Л.: Недра, 1980. - 2-е изд.
- [6] Бобачев А.А. Решение прямых и обратных задач электроразведки методом сопротивлений для сложно-построенных сред: Диссертация на соискание ученой степени физико-математических наук. - Москва, 2003. - 95 с.
- [7] Mirgalikyzy T., Mukanova B., Modin I. Method of Integral Equations for the Problem of Electrical Tomography in a Medium with Ground Surface Relief // Journal of Applied Mathematic. - Volume 2015.
- [8] Balgaisha Mukanova, Tolkyn Mirgalikyzy and Dilyara Rakisheva Modelling the Influence of Ground Surface Relief on Electric Sounding Curves Using the Integral Equations Method // Mathematical Problems in Engineering. - Volume 2017.
- [9] Запорожец В.М. Влияние рельефа на результаты замеров сопротивления (по работам С. Г. Комарова и Л.П. Горбенко) // Элкгр. – 1938. – № 4(12).
- [10] Чантуриашвили Л.С. О количественном учете влияния рельефа для некоторых случаев разведки постоянным током // Труды Института геофизики. – Тбилиси, 1955. – Т. 14. – С. 199-209.
- [11] Вешев А.В. Влияние рельефа на результаты работ комбинированным электропрофилированием // Ученые записки ЛГУ. – 1959. – №278.
- [12] Loke M.H., Barker R.D. Rapid least-squares inversion of apparent resistivity pseudo sections using a quasi-Newton method // Geophysical Prospecting. – 1996. – Vol. 44, no. 1. – P. 131-152.
- [13] Coggon J.H. Electromagnetic and electrical modeling by the finite element method // Geophysics. – 1971. – Vol. 36, no. 1. – P. 132-155.
- [14] Mufti I.R. Finite-difference modeling for arbitrary-shaped two dimensional structures // Geophysics. – 1976. – Vol. 41, no. 1. – P. 62-78.
- [15] Pelton W.H., Rijo L., Swift C.M. Inversion of two dimensional resistivity and Induced Polarization data // Geophysics. – 1978. – Vol. 43, no. 4. – P. 788-803.
- [16] Dey A., Morrison H.F. Resistivity modeling for arbitrary shaped two-dimensional structures // Geophysical Prospecting. – 1979. – Vol. 27, no. 1. – P. 106-136.
- [17] Loke M.H., Barker R.D. Least-squares deconvolution of apparent resistivity pseudosections // Geophysics. – 1995. – Vol. 60, no. 6. – P. 1682-1690.
- [18] Loke M.H. Topographic modelling in resistivity imaging inversion // Extended Abstracts of 62nd EAGE Conference and Technical Exhibition. - Glasgow, Scotland, 2000.

- [19] Erdogan E., Demirci I., Candasayar M.E. Incorporating topography into 2D resistivity modeling using finite-element and finite-difference approaches // Geophysics. – 2008. – vol. 73. – no. 3. – P. 135-142.
- [20] Demirci I., Erdogan E., Candasayar M.E. Two-dimensional inversion of direct current resistivity data incorporating topography by using finite difference techniques with triangle cells: Investigation of Kera fault zone in western Crete // Geophysics. – 2012. – Vol. 77, no. 1. – P. 67-75.
- [21] Penz S., Chauris H., Donno D., Mehl C. Resistivity modeling with topography // Geophys. J. Int. – 2013. – Vol. 194, no. 3. – P. 1486-1497.
- [22] Fox R.C., Hohmann G.W., Killpack T.J., Rijo L. Topographic effects in resistivity and induced-polarization surveys // Geophysics. – 1980. – Vol. 45, no. 1. – P. 75-93.
- [23] Tsourlos P.I., Szymanski J.E., Tsokas G.N. The effect of topography on commonly used resistivity arrays // Geophysics. – 1999. – Vol. 64, no. 5. – P. 1357-1363.
- [24] Plattner A.D. Adaptive wavelet methods for geoelectric modeling and inversion: Dissertation. - Basel, 2011.
- [25] Gunther T., Rucker C., Spitzer K. Three-dimensional modelling and inversion of dc resistivity data incorporating topography - I. Modelling // Geophys. J. Int. – 2006. – Vol. 166. – P. 495-505.
- [26] Gunther T., Rucker C. Boundless Electrical Resistivity Tomography // BERT 2 - the user tutorial. – 2013. – Ver. 2.0.
- [27] Альпин Л.М. Источники поля в теории электрической разведки // Прикладная геофизика. – М., 1947. - Вып. 3. – С. 56-200.
- [28] Dieter K., Paterson N.R and Grant F.S. KP and resistivity type awes for three- dimensional bodies // Geophysics. - 1969. - 34. - P. 615-632.
- [29] Hohmann G.W. Three dimensional induced polarization and electromagnetic modeling //Geophysics. - 1975. - 40. - P. 309-324.
- [30] Eloranta E. A method for calculation mise-a-la-masse anomalies in the case of high conductivity contrast by the integral equation technique // Geoexploration. - 1984. - 22. - P. 77-88.
- [31] Schenkel C.J. The Electrical Resistivity Method in Cased Boreholes: Phodissertation. Published as report LBL-31139 / Lawrence Berkeley National Laboratory. - Berkeley, California, 1991.
- [32] Orunkhanov M., Mukanova B. The integral equations method in problems of electrical sounding // In book: Advances in High Performance Computing and Computational Sciences, Springer-Berlin-Heidelberg. – 2006. – P.15-21.
- [33] Орунханов М.К., Муканова Б.Г., Сарбасова Б.К. Численная реализация метода потенциалов в задаче зондирования над наклонным пластом // Вычислительные технологии, Специальный выпуск: труды Совещания российско-казахстанской рабочей группы по вычислительным и информационным технологиям. – Новосибирск, 2004. – Т. 9. – С.45-48.
- [34] Ракишева Д.С., Миргаликызы Т., Муканова Б.Г. Аппроксимация поверхности рельефа дневной поверхности методом RBF // Вестник Национальной академии наук Республики Казахстан. - 2017. - Vol. 1, Number 365. - С.210-215.

References

- [1] Shevnin V.A. "Elektrorazvedka. Posobie po elektrorazvedochnoj praktike dlya studentov geofizicheskikh spetsial'nostej [Electrical prospecting. Manual on electrical exploration practice for students of geophysical specialties] Moscow, (2013), 124.
- [2] Loke M.H., Barker R.D. "Practical techniques for 3D resistivity surveys and data inversion."Geophysical Prospecting 44 (1996): 499-523.
- [3] Gunther T., Rucker C. Boundless Electrical Resistivity Tomography. BERT 2 - the user tutorial(2013).
- [4] Baranchuk K.I., Mirgalikyzy T., Modin I.N., Mukanova B.G. "Fizicheskoe modelirovaniye elektricheskoy tomografii na poverhnosti so slozhnym rel'eform [Physical modeling of electrical tomography on the surface with a complex relief]."Engineering Surveys 11 (2017): 56-65.

- [5] Veshev A.V. "Elektroprofilirovaniye na postoyannom i peremennom toke [Electrical profiling on direct and alternating current]." 2 (1980).
- [6] Bobachev A.A. "Reshenie pramykh i obratnykh zadach elektrorazvedki metodom soprotivlenij dlya slozhno-postroennykh sred [The solution of direct and inverse problems of electrical prospecting by the method of resistance for complex-built environments]" (Thesis for the degree of physical and mathematical sciences, MSU, 2003).
- [7] Mirkalikyzy T., Mukanova B., Modin I. "Method of Integral Equations for the Problem of Electrical Tomography in a Medium with Ground Surface Relief." Journal of Applied Mathematic(2015).
- [8] Balgaisha Mukanova, Tolkyn Mirkalikyzy and Dilyara Rakisheva "Modelling the Influence of Ground Surface Relief on Electric Sounding Curves Using the Integral Equations Method." Mathematical Problems in Engineering (2017).
- [9] Zaporozhets V.M. "Vliyanie rel'efa na rezul'taty zamerov soprotivleniya (po rabotam S. G. Komarova i L.P. Gorbenko) [The influence of the relief on the results of resistance measurements]." Elkgr 4 (12)(1938).
- [10] Chanturishvili L.S. "O kolichestvennom uchete vliyaniya rel'efa dlya nekotorykh sluchaej razvedki postoyannym tokom [On the quantitative account of the influence of the relief for some cases of direct current intelligence]." Proceedings of the Institute of Geophysics 14(1955): 199-209.
- [11] Veshev A.V. "Vliyanie rel'efa na rezul'taty rabot kombinirovannym ehlektroprofilirovaniem [Influence of relief on the results of work by combined electric profiling]." Scientific notes of Leningrad State University 278 (1959).
- [12] Loke M.H., Barker R.D. "Rapid least-squares inversion of apparent resistivity pseudo sections using a quasi-Newton method." Geophysical Prospecting 44(1)(1996): 131-152.
- [13] Coggon J.H. "Electromagnetic and electrical modeling by the finite element method." Geophysics 36(1) 1971: 132-155.
- [14] Mufti I.R. "Finite-difference modeling for arbitrary-shaped two dimensional structures." Geophysics 41(1)(1976): 62-78.
- [15] Pelton W.H., Rijo L., Swift C.M. "Inversion of two dimensional resistivity and Induced Polarization data." Geophysics 43(4) (1978): 788-803.
- [16] Dey A., Morrison H.F. "Resistivity modeling for arbitrary shaped two-dimensional structures." Geophysical Prospecting 27(1) (1979): 106-136.
- [17] Loke M.H., Barker R.D. "Least-squares deconvolution of apparent resistivity pseudosections." Geophysics 60(6) (1995): 1682-1690.
- [18] Loke M.H. "Topographic modelling in resistivity imaging inversion." 62nd EAGE Conference and Technical Exhibition, Extended Abstracts (2000).
- [19] Erdogan E., Demirci I., Candasayar M.E. "Incorporating topography into 2D resistivity modeling using finite-element and finite-difference approaches." Geophysics 73(3) (2008): 135-142.
- [20] Demirci I., Erdogan E., Candasayar M.E. "Two-dimensional inversion of direct current resistivity data incorporating topography by using finite difference techniques with triangle cells: Investigation of Kera fault zone in western Crete." Geophysics 77(1) (2012): 67-75.
- [21] Penz S., Chauris H., Donno D., Mehl C. "Resistivity modeling with topography." Geophys. J. Int. 194(3) (2013): 1486-1497.
- [22] Fox R.C., Hohmann G.W., Killpack T.J., Rijo L. "Topographic effects in resistivity and induced-polarization surveys." Geophysics 45(1) (1980): 75-93.
- [23] Tsourlos P.I., Szymanski J.E., Tsokas G.N. "The effect of topography on commonly used resistivity arrays." Geophysics 64(5) (1999): 1357-1363.
- [24] Plattner A.D. "Adaptive wavelet methods for geoelectric modeling and inversion" (Dissertation, 2011).
- [25] Gunther T., Rucker C., Spitzer K. "Three-dimensional modelling and inversion of dc resistivity data incorporating topography - I. Modelling." Geophys. J. Int. 166 (2006): 495-505.
- [26] Gunther T., Rucker C. "Boundless Electrical Resistivity Tomography." BERT 2 - the user tutorial (2013).
- [27] Alpin L.M. "Istochniki polya v teorii ehlektricheskoy razvedki [Sources of field in the theory of electrical intelligence]." Applied Geophysics 3 (1947): 56-200.

-
- [28] Dieter K., Paterson N.R and Grant F.S. "KP and resistivity type awes for three-dimensional bodies."Geophysics 34 (1969): 615-632.
 - [29] Hohmann G.W. "Three dimensional induced polarization and electromagnetic modeling."Geophysics 40 (1975): 309-324.
 - [30] Eloranta E. "A method for calculation mise-a-la-masse anomalies in the case of high conductivity contrast by the integral equation technique."Geoexploration 22 (1984): 77-88.
 - [31] Schenkel C.J. "The Electrical Resistivity Method in Cased Boreholes"(Phodissertation, University of California, 1991).
 - [32] Orunkhanov M., Mukanova B. "The integral equations method in problems of electrical sounding."Advances in High Performance Computing and Computational Sciences (2006): P.15-21.
 - [33] Orunkhanov M.K., Mukanova B.G., Sarbasova B.K. "CHislennaya realizatsiya metoda potentsialov v zadache zondirovaniya nad naklonnym plastom [Numerical implementation of the method of potentials in the problem of probing over an inclined stratum]."Special Issue Proceedings of the Meeting of the Russian-Kazakhstan working group on computational and information technologies 9 (2004): 45-48.
 - [34] Rakisheva D. S., Mirkalikyzy T., Mukanova B. G. "Approksimatsiya poverkhnosti rel'efa dnevnoj poverkhnosti metodom RBF [Approximation of the surface relief of the day surface by the RBF method]."Bulletin of the National Academy of Sciences of the Republic of Kazakhstan 1(365) (2017): 210-215.

К СВЕДЕНИЮ АВТОРОВ

1. В журнал «Вестник КазНУ. Серия математика, механика, информатика» (в английской версии «Journal of Mathematics, Mechanics and Computer Science Series») принимаются набранные только в текстовом формате L^AT_EX2ε на казахском, русском или английском языках, ранее не опубликованные проблемные, обзорные, дискуссионные статьи в области естественных наук, где освещаются результаты фундаментальных и прикладных исследований.
2. Материалы следует направлять по адресу: 050040 Алматы, ул. аль-Фараби, 71, корпус 13, Научно-исследовательский институт механики и математики КазНУ им. аль-Фараби, каб. 125, тел. 377-32-23. Электронная почта: Lazat.dairbayeva@gmail.com (ответственный секретарь редколлегии, Даирбаева Л.М.)
3. Статья должна сопровождаться письмом от учреждения, в котором выполнена данная работа, где указываются сведения об авторах: Ф.И.О. полностью, место их работы (название вуза, центра без сокращений), рабочий или моб. телефон, e-mail, домашний адрес и контактный телефон.
4. В редакцию необходимо представить электронную версию статьи: tex-файлы работы и файлы рисунков на одном диске. Для файлов рисунков рекомендуется использовать средства основного пакета L^AT_EX2ε или формат eps [см. п.7]. В редакцию также представляется оттиск работы в двух экземплярах.
5. Объем статьи, включая список литературы, таблицы и рисунки с подрисуточными надписями, аннотации, не должен превышать 17 страниц печатного текста. Минимальный объем статьи - 7 страниц.

Структура статьи.

Первая страница:

- 1) Первая строка - номер МРНТИ (IRSTI) (можно взять здесь: <http://grnti.ru/>), выравнивание - по левому краю, шрифт - полужирный.
- 2) Название статьи (Заголовок) должно отражать суть и содержание статьи и привлекать внимание читателя. Название должно быть кратким, информативным и не содержать жаргонизмов или аббревиатур. Оптимальная длина заголовка - 5-7 слов (в некоторых случаях 10-12 слов). Название статьи должно быть представлено на русском, казахском и английском языках. Название статьи представляется полужирным шрифтом строчными буквами, выравнивание - по центру.
- 3) Автор(ы) статьи - Инициалы и фамилия, место работы (аффилиация), город, страна, email - на русском, казахском и английском языках. Сведения об авторах представляются обычным шрифтом строчными буквами, выравнивание - по центру.
- 4) Аннотация объемом 150-500 слов на русском, казахском и английском языках. Структура аннотации включает в себя следующие ОБЯЗАТЕЛЬНЫЕ пункты: "Вступительное слово о теме исследования. "Цель, основные направления и идеи научного исследования. "Краткое описание научной и практической значимости работы. "Краткое описание методологии исследования. "Основные результаты и анализ, выводы исследовательской работы. "Ценность проведенного исследования (внесенный вклад данной работы в соответствующую область знаний). "Практическое значение итогов работы.
- 5) Ключевые слова/словосочетания - количеством 3-5 на русском, казахском и английском языках.

Последующая страница (новая):

Стандартные разделы статьи: **Введение, Обзор литературы, Материал и методы, Результаты и обсуждение, Заключение, Благодарности (если имеются), Список литературы** (названия разделов не менять)

- 6) **Введение.** Введение состоит из следующих основных элементов: "Обоснование выбора темы; актуальность темы или проблемы. В обосновании выбора темы на основе описания

опыта предшественников сообщается о наличии проблемной ситуации (отсутствие каких-либо исследований, появление нового объекта и т.д.). Актуальность темы определяется общим интересом к изученности данного объекта, но отсутствием исчерпывающих ответов на имеющиеся вопросы, она доказывается теоретической или практической значимостью темы. "Определение объекта, предмета, целей, задач, методов, подходов, гипотезы и значения вашей работы. Цель исследования связана с доказательством тезиса, то есть представлением предмета исследования в избранном автором аспекте.

7) **Обзор литературы.** В разделе обзор литературы должны быть охвачены фундаментальные и новые труды по исследуемой тематике зарубежных авторов на английском языке (не менее 15 трудов), анализ данных трудов с точки зрения их научного вклада, а также пробелы в исследовании, которые Вы дополняете в своей статье. НЕДОПУСТИМО наличие множества ссылок, не имеющих отношения к работе, или неуместные суждения о ваших собственных достижениях, ссылки на Ваши предыдущие работы.

8) **Материал и методы.** Раздел должен состоять из описания материалов и хода работы, а также полного описания использованных методов. Характеристика или описание материала исследования включает его представление в качественном и количественном отношении. Характеристика материала - один из факторов, определяющий достоверность выводов и методов исследования. В этом разделе описывается, как проблема была изучена: подробная информация без повторения ранее опубликованных установленных процедур; используется идентификация оборудования (программного обеспечения) и описание материалов, с обязательным внесением новизны при использовании материалов и методов. Научная методология должна включать в себя: - исследовательский вопрос(-ы); - выдвигаемую гипотезу (тезис); - этапы исследования; - методы исследования; - результаты исследования.

9) **Результаты и обсуждение.** В этом разделе приводятся анализ и обсуждение полученных вами результатов исследования. Приводятся выводы по полученным в ходе исследования результатам, раскрывается основная суть. И это один из самых важных разделов статьи. В нем необходимо провести анализ результатов своей работы и обсуждение соответствующих результатов в сравнении с предыдущими работами, анализами и выводами.

10) **Заключение.** Обобщение и подведение итогов работы на данном этапе; подтверждение истинности выдвигаемого утверждения, высказанного автором, и заключение автора об изменении научного знания с учетом полученных результатов. Выводы не должны быть абстрактными, они должны быть использованы для обобщения результатов исследования в той или иной научной области, с описанием предложений или возможностей дальнейшей работы. Структура заключения должна содержать следующие вопросы: Каковы цели и методы исследования? Какие результаты получены? Каковы выводы? Каковы перспективы и возможности внедрения, применения разработки?

11) **Благодарности** (если имеются). Например: Работа выполнена при поддержке грантового финансирования научно-технических программ и проектов Министерством науки и образования Республики Казахстан (грант «Наименование темы гранта», 2018-2020 годы).

12) **Список литературы/References.** (оба списка, если статья на русском или казахском. Если статья на английском, то только один список по стилю Чикаго). Список используемой литературы, или Библиографический список состоит из не менее 30 наименований литературы, и из них 50% на английском языке. В случае наличия в списке литературы работ, представленных на кириллице, необходимо представить список литературы в двух вариантах: первый - в оригинале, второй - романизированным алфавитом (транслитерация). Романизированный список литературы должен выглядеть в следующем виде: автор(-ы) (транслитерация) -> название статьи в транслитерированном варианте [перевод названия статьи на английский язык в квадратных скобках], название русскоязычного источника (транслитерация, либо английское название - если есть), выходные данные с обозначениями на английском языке (год в круглых скобках) -> страницы. Например: Gokhberg L., Kuznetsova T. Strategiya-2020: novye kontury rossiskoi innovatsionnoi politiki [Strategy 2020: New Outlines of Innovation Policy]. Foresight-Russia, vol. 5, no 4 (2011): 8-30. Список литературы представляется по мере цитирования, и ТОЛЬКО

те работы, которые цитируются в тексте. Ссылки на литературу оформляются в квадратных скобках с указанием номера литературы. Стиль оформления "Список литературы" на русском и казахском языке согласно ГОСТ 7.1-2003 "Библиографическая запись. Библиографическое описание. Общие требования и правила составления"(требование к изданиям, входящих в перечень ККСОН). Стиль оформления "References" романизированного списка литературы (см. выше), а также источников на английском (другом иностранном) языке для естественнонаучных и технических направлений согласно Chicago Style (www.chicagomanualofstyle.org).

В данном разделе необходимо учесть:

- а) Цитируются основные научные публикации, передовые методы исследования, которые применяются в данной области науки и на которых основана работа автора.
 - б) Избегайте чрезмерных самоцитирований.
 - в) Избегайте чрезмерных ссылок на публикации авторов СНГ/СССР, используйте мировой опыт.
 - г) Библиографический список должен содержать фундаментальные и наиболее актуальные труды, опубликованные известными зарубежными авторами и исследователями по теме статьи.
6. Журнал придерживается единого стиля и поэтому предъявляет ряд общих требований к оформлению работ. Исходный (неотранслированный) tex-файл должен целиком помещаться в горизонтальных рамках экрана за возможным исключением матриц и таблиц и транслироваться без протестов L^AT_EX2ε и сообщений о кратных и неопределенных метках, больших переполненных и незаполненных боксах. Не следует определять много новых команд, изобретая собственный сленг. Авторы могут подгружать другие стандартные стилевые пакеты, но только те, которые не входят в противоречие с пакетами amsmath и amssymb. Естественно файл, кроме всего прочего, должен быть проверен на отсутствие грамматических и стилистических ошибок. Статьи, не удовлетворяющие этим требованиям, возвращаются на доработку.
- Эталонный образец работы с демонстрацией графики, с преамбулой устраивающей редакцию, списки типичных ошибок оформления и методы их устранения можно получить в редакции или на сайте КазНУ им. аль-Фараби <http://journal.kaznu.kz>.
7. Графические файлы с рисунками должны быть только качественными черно-белыми в формате .eps , либо выполнеными в латеховском формате. Рисунки в этих форматах делаются, например, с помощью мощных математических пакетов Maple, Mathematica или с помощью пакета Latex-cad. Качественные графические файлы сделанные другими графическими программами должны быть сконвертированы в формат .eps с помощью Adobe Photoshop или конвертера Conversion Artist. Все рисунки должны быть уже импортированными в tex-файл и представляются в редакцию вместе с основным файлом статьи. Графические форматы, отличные от выше указанных, отвергаются.
- Редакция вправе отказаться от включения в работу рисунка, если автор не в состоянии обеспечить его надлежащее качество.

Уважаемые читатели, вы можете подписаться на наш журнал "Вестник КазНУ. Серия математика, механика, информатика", который включен в каталог АО "Казпочта""ГАЗЕТЫ И ЖУРНАЛЫ". Количество номеров в год – 4. Индекс для индивидуальных подписчиков, предприятий и организаций – 75872, подписная цена за год – 1200 тенге; индекс льготной подписки для студентов – 25872, подписная цена за год для студентов – 600 тенге.

МАЗМУНЫ - СОДЕРЖАНИЕ

1-бөлім

Математика

Denis Potapov, Fedor Sukochev

Arens Algebras and Matricial Spaces 3

Suleimenov Zh.

On the existence of a conditionally periodic solution of one quasilinear differential system in the critical case 8

2-бөлім

Информатика

Joldasbayev S.K., Balakayeva G.T., Aidarov K.A., Chris Phillips

Dynamic request distribution for enhanced Quality of Service 18

Shalakhmetov A., Aubakirov S.

Optical character recognition with neural networks 28

Жеткенбай Л., Шарипбай А.А., Бекманова Г.Т., Кажымұхан Д., Каманур Ү.

Сравнение морфологических правил глагола казахского и турецкого языков 42

Искаков К.Т., Хасенова З.Т., Мынбаева Б.Н., Оралбекова Ж.О., Мухтарова Ш.О.

Разработка формата приложений для информационно-моделирующей системы экологического мониторинга 52

Мамыкова Ж.Д., Мутанов Г.М., Сундетова Ж.Т., Торекул С.М.

Подходы к разработке информационной системы мониторинга мнений и оценки социального самочувствия 63

Шарипбай А.А., Гатиатуллин А.Р., Ергеш Б.Ж., Қажымұхан Да.

Разработка единого метаязыка морфологии тюркских языков 78

3-бөлім

Колданылмаған математика

Жакебаев Д.Б., Мусеева Е.С., Хребтов М.Ю., Цой Н.В.

Численный алгоритм для решения задачи моделирования динамики крупномасштабного термика . 88

Миргалиқызы Т., Мұжанова Б.Г.

Численный алгоритм решения задачи электрического зондирования среды с рельефом земной поверхности 103

К сведению авторов 117

Раздел 1

Математика

Раздел 2

Информатика

Раздел 3

Прикладная математика

CONTENS
**Section 1
Mathematics**

<i>Denis Potapov, Fedor Sukochev</i>	
Arens Algebras and Matricial Spaces	3

<i>Suleimenov Zh.</i>	
On the existence of a conditionally periodic solution of one quasilinear differential system in the critical case	8

**Section 2
Computer science**

<i>Joldasbayev S.K., Balakayeva G.T., Aidarov K.A., Chris Phillips</i>	
Dynamic request distribution for enhanced Quality of Service	18

<i>Shalakhmetov A., Aubakirov S.</i>	
Optical character recognition with neural networks	28

<i>Zhetkenbay L., Sharipbay A.A., Bekmanova G.T., Kazhymukhan D., Kamanur U.</i>	
Comparison of the morphological rules of the Kazakh and Turkish languages	42

<i>Iskakov K.T., Khassenova Z.T., Mynbayeva B.N., Oralbekova Zh.O., Mukhtarova Sh.O.</i>	
Development of format applications for information and modeling system environmental monitoring	52

<i>Mamykova Zh.D., Mutanov G.M., Sundetova Zh.T., Torekul S.M.</i>	
Approaches of the development of information monitoring system of social wellness opinions and evaluation	63

<i>Sharipbay A., Gatiatullin A., Yergesh B., Kazhymukhan D.</i>	
Development of an unified meta language of the turkic languages morphology	78

**Section 3
Applied Mathematics**

<i>Zhakebayev D.B., Moisseyeva Ye., Hreblov M.Y., Tsot N.V.</i>	
Numerical algorithm for solving the problem of modeling the dynamics of a large-scale thermal	88

<i>Mirgalikyzy T., Mukanova B.G.</i>	
Numerical algorithm for solving of electrical prospecting problems for medium with a surface relief	103

Note by authors	117
-----------------------	-----