ИНСТИТУТ ИНФОРМАЦИОННЫХ И ВЫЧИСЛИТЕЛЬНЫХ ТЕХНОЛОГИЙ КН МОН РК

ӘЛ-ФАРАБИ атындағы ҚАЗАҚ ҰЛТТЫҚ УНИВЕРСИТЕТІ

TURAN UNIVERSITY

МАТЕРИАЛЫ

II МЕЖДУНАРОДНОЙ НАУЧНОЙ КОНФЕРЕНЦИИ

«ИНФОРМАТИКА И ПРИКЛАДНАЯ МАТЕМАТИКА»

«ИНФОРМАТИКА ЖӘНЕ ҚОЛДАНБАЛЫ МАТЕМАТИКА»

«COMPUTER SCIENCE AND APPLIED MATHEMATICS»

(ЧАСТЬ II)

Алматы

27-30 сентября 2017 года

www.conf2017.ipic.kz

УДК 004:51
ББК 32.973:22.1
И 74

И 74 **Информатика и прикладная математика:** Мат. Межд. науч. конф. (27-30 сентября 2017 г). Ч.2. – Алматы, 2017, – 424 с.

ISBN 978-601-04-3101-0

В сборнике опубликованы доклады, представленные по 5 секциям от Республики Казахстан, Российской Федерации, США, Латвии, Польши, Республики Беларусь, Украины, Азербайджана, Японии, Кореи, Португалии, Испании, Англии, Грешии, Кыргызской Республики и других.

Рассмотрены актуальные вопросы в области математики, информатики и управления: математического моделирования сложных систем и бизнес-процессов, исследования и разработки защищенных и интеллектуальных информационных и телекоммуникационных технологий, математической теории управления, технологий искусственного интеллекта.

Материалы сборника предназначены для научных работников, докторантов и магистрантов, а также студентов старших курсов.

УДК 004:51
ББК 32.973:22.1

ISBN 978-601-04-3101-0

Секция 3. Технологии искусственного интеллекта

21 *Goldschmidt, E., Jacobsen, N.* Genetic and environmental effects on myopia development and progression // The Scientific journal of the Royal College of Ophthalmologists. - № 28 (2). – P. 126-133.

22 Psychological Testing / ed. Sam Goldstein, Jack A. Naglieri // Encyclopedia of Child Behavior and Development, Springer: US, 2011. – P. 1189.

23 http://helpiks.org/3-15841.html. 16.05.2015.

24 *Vaidhehi V.* A Framework to Design a Web based Neuro Fuzzy System for Course Advisor // International Journal of Innovative Research in Advanced Engineering. – 2014. - № 1(1). – P. 186-190.

25 *Garyaev A.V., Garyaeva T.P.* Psycho-fiziologicheskie osobennosti vizualnogo perception informatsii i ih uchet prisozdani i uchebnyh prezentatsy // Vestnik PSPU, 2008. - № 4. – P. 106-113.

26 *Elliot A.J., Fairchild M.D., Franklin A.* Color symbolism and association // Handbook of Color Psychology. – 2015. – P. 295–316.

27 http://www.tikkurila.ru /dlya_professionalov /tsveta/vvedenie_v_teoriyu_tsveta /vospriyatie_tsveta_mozgom 13.01.2017.

28 *Chrysafiadi K., Virvou M.* Fuzzy Logic in Student Modeling // Advances in Personalized Web-Based Education, Springer, 2014. - № 78. – P. 25-60.

29 *Zsolt Csaba Johanyák* Survey on Five Fuzzy Inference-Based Student Evaluation Methods // Computational Intelligence in Engineering, Springer, 2012. - № 313. – P. 219-228.

30 *Samigulina G. A., Shayakhmetova A.S., Suleimen O.*, Certificate of state registration of rights to the object of copyright Ministry of Justice of the Republic of Kazakhstan 0090, 15 January 2016.

TECHNOLOGY OF THE STRUCTURAL MACHINE TRANSLATION RULES GENERATION, BASED ON THE COMPLETE SET OF KAZAKH ENDINGS

Ualsher Tukeyev, Aida Sundetova, Balzhan Abduali, Karibayeva Aidana, Amirova Dina

ualsher.tukeyev@gmail.com, *sun27aida@gmail.com*, *balzhanabdualy@gmail.com*, *a.s.karibayeva@gmail.com*, *amirovatdina@gmail.com*

Al-Farabi Kazakh National University, Information Systems department

Abstract. *In this paper we propose technology of generating a set of machine translation rules, based on the system of the complete set of Kazakh endings. The*

proposed technology is based on the combination of using a complete set of endings types of the rich morphology source language, such as the Kazakh language, free/open-source Apertium platform and rule generation application. Structural transfer rules extraction is shown for English-Kazakh and Kazakh-Russian machine translations systems.

Keywords: *inferring, Apertium platform, machine translation, chunk, transfer rules, endings.*

1 Introduction

Kazakh language, as one of Turkic languages, belongs to an agglutinative language, and it uses vowel harmony. Which means that translating text in Kazakh language into other languages with simpler morphology, such as English or Russian languages, with, for instance, statistical machine translation (SMT), will cause some quality loss because of morphological segmentation.

In previous work [1] we already built application to extract rules from complete set of endings for Kazakh language, where application used only prepared, by hand, template with morphological analysis. In this paper we propose full technology of generating rules from the original phrases, based on the system of complete set of endings. Phrases will be processed by Apertium platform's morphological analyzer and rules generation application. All examples will be shown for English-Kazakh and Kazakh-Russian machine translation systems.

1 Problem of creating rules for Kazakh language

The problem of creating rules is hidden in the methods of their building. In rule-based machine translation (RBMT) systems usually rules are created by humans, which could cause some questions, such as: is set of rules, created by humans, complete or full? In case of SMT, "rules" to translate are applies by considering probabilities of the translations and in fact, it does not guarantee that in particular case phrase with low probability will be translated correctly [2].

In this paper we propose our solution of the problem: usage of complete set of Kazakh endings, which gurantees that rules, are built from this complete set, will be also complete and technology, where human's work will be less than in RBMT, because of automatical rule generalisation. In the next section 3 we will show how complete set of endings was created, in section 4 improved process of rule extraction will be described, in section 5 some experiments and program realisation will be shown.

2 Related works

In the previous work we considered how to wrap all the endings of one language and by automatic system fill in all the rules that you can write for all of these endings. In the previous work, the chunk transfer rule logical templates for Kazakh-English and Kazakh-Russian words with nominal base were constructed.

The question of automatic inferring of the structural rules of machine translation from one language to another are rather actual for machine translation systems based on grammatical rules (RBMT). This is due to the time-consuming process of drawing up the rules for RBMT.

The scheme of inference transfer rules based on the following: the source is a complete set of endings types of the Kazakh language; for each types of endings (template of morphological structure types of endings) of the Kazakh language is constructed a equivalent grammatical structure logical template (pattern) in the target language (for example, Russian and English languages); on the basis of grammatical structure logical template built template of program structure for transfer of the morphological structure of word's endings into the equivalent grammatical structure of the target language [1].

In this paper proposed the technology of generating rules from complete set of Kazakh endings and example of how can work this system. This approach allows generating a complete system of morphological chunk transfer rules.

3 Technology of generating rules from complete set of Kazakh endings

As was described in the previous section, application for rules generation, based on the complete set of Kazakh endings, used templates in the following format:

<n><pl><nom>|<n><m><nn><pl><nom>

Where left side is source language and on the right side is target language phrase's morphological analysis.

This template has some disadvantages:

- To get morphological analysis user has to use additional application;
- Template should be build by hand;
- Only one phrase could be used in one time.

To fix these issues new technology of generating rules was proposed: to do morphological analysis will be used a free/open-source machine translation platform Apertium. There are several reasons:

- Apertium is open source, which means that code of this project could be used or modified to insert to previous extraction application.
- This project has morphological analyzer for Kazakh, Russian and English languages, monolingual resources, such like English, Kazakh, Russian dictionaries would be used in process to get template.

New workflow of generating rules, based on the next schema (Fig. 1) and includes several steps:

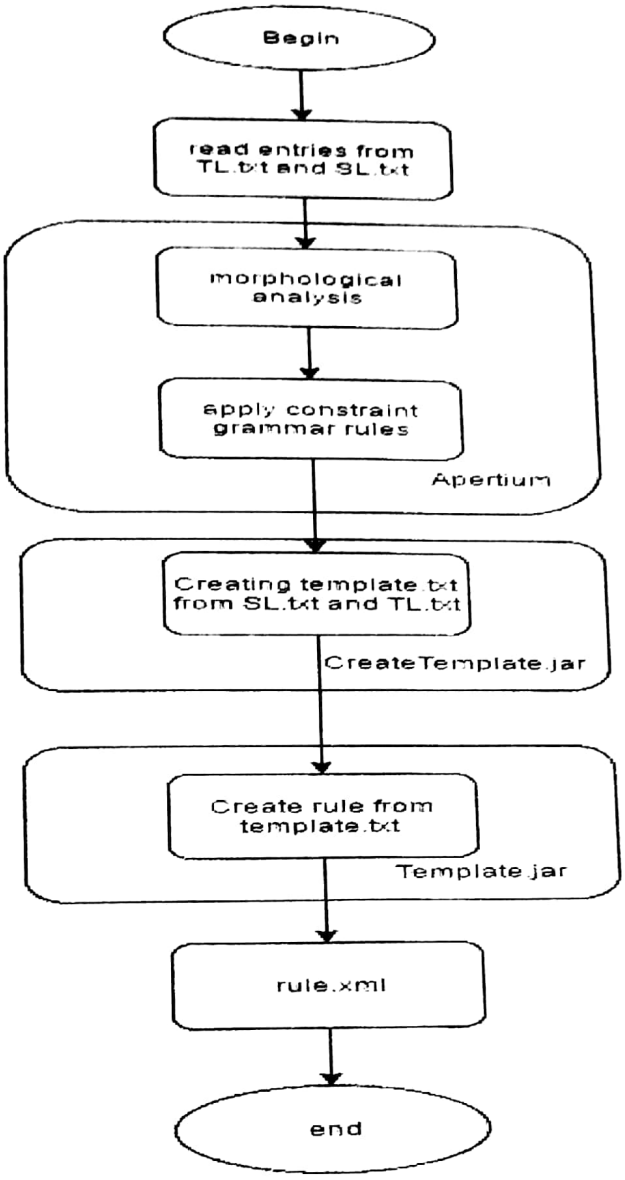

Fig. 1. New workflow of generating rules

- Two files are created: first file, which contains part of phrase in source language and in translation is of phrase is put to second file.
- Apertium platform is used to get morphological analysis of phrases from created files.
- Morphological ambiguity is also solved with Apertium instruments, such like constraint grammar module [3]. If phrase is translated from English, additionally part-of-speech tagger, based on the hidden Markov model will be used [4].
- To create template from both analysis, simple script (CreateTemplate.jar) was built. This script reads two analyses from source and target language phrase translations and creates another file with name of rule to be created and on the second line template itself, for instance:

S-K <n><pl><nom>|<n><m><nn><pl><nom>

• Finally, application(Template.jar), which will generate the rule, reads name of chunk and template from "template" file.

• All process ends with generating one file in .xml format containing rule, which could be used in transfer stage of Apertium platform.

In the section 5 will be shown few results of this technology realization.

4 Example of program realization

To develop workflow, described above, was used Apertium platform with installed apertium-kaz [5], apertium-rus [6], apertium-eng-kaz [7] machine translation systems, application to create templates and application to extract rules from template itself. For instance, let's consider that we are going to build rule for plural form of nouns, which chunk is named as "S-K". Firstly, as was mentioned above, user has to prepare two files: file, containing phrase in source language(SL.txt) and file with translation of this phrase in target language (TL.txt). Then, content of SL.txt will be:

Table 1. *Example of two files for build chunk of "S-K"*

SL.txt	*TL.txt*
in garden	бақшада
in school	мектепте

Afterwards, user will be start script to extract rules, which starts all the process described above. In the result, rule.xml file will be created. Below is shown piece of this file:

```
  <section-rules>
<rule comment="RULE:S-K">
<pattern>
 <pattern-item n="cat_pr"/>
 <pattern-item n="cat_n"/>
</pattern>
  <action>
 <out>
       <chunk name="__S-K__" case="caseFirstWord">
      <tags><tag><lit-tag v="NP"/></tag>       </tags>
  <lu>
  <clip pos="2" side="tl" part="lem"/>
  <clip pos="2" side="tl" part="_attr_n_"/>
<lit-tag v="loc"/>
</lu>
<b/>  </chunk>
  </out>
 </action>
 </rule>
 </section-rules>
</transfer>
```

Fig. 2. The result, rule.xml file

This rule could be easily integrated into first stage of structural transfer of Apertium – chunker level [8], because in the generating script additionally user should add name of pharse, for instance, noun phrase – NP. Below we show the how rule translated the structure "in garden – бақшада (baqshada):

Fig. 3. Translating the phrase "in garden – бақшада (baqshada)

We did experiment for Kazakh-Russian language pair with structure "S-K-T-C", where source language is Russian. Below, in figure 3 could be seen translation with original Apertium Kaz-Rus [9]:

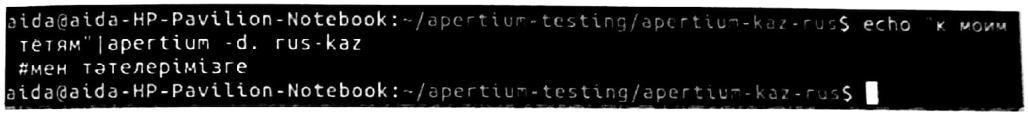

Fig. 4. Experiment for Kazakh-Russian language pair with structure "S-K-T-C"

By using complete set of the Kazakh settings, we were able to identify, that rule for structure "S-K-T-C" does not exist in Apertium-kaz-rus machine translation system. In files named "rus.txt" and "kaz.txt" were put structures to generate rule:

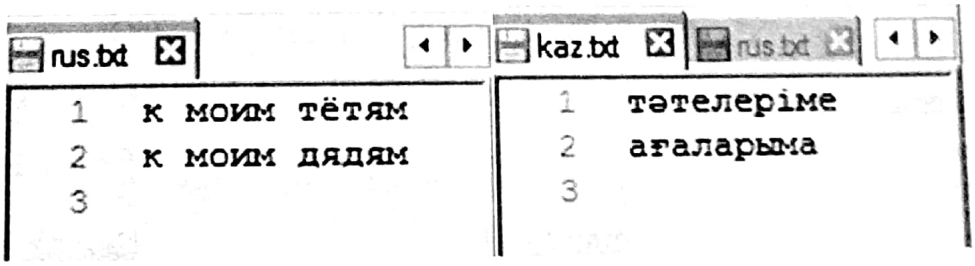

Fig. 5. The structure "S-K-T-C" to generate rule

After running generation script we got the rule:

```
<rule comment="RULE:S-K-T-C">
<pattern>
 <pattern-item n="cat_pr"/>
 <pattern-item n="cat_prn"/>
 <pattern-item n="cat_n"/>

</pattern>
  <action>
 <out>
       <chunk name="__S-K-T-C__" case="caseFirstWord">
      <tags><tag><lit-tag v="NP"/></tag>        </tags>
  <lu>
  <clip pos="3" side="tl" part="lem"/>
<clip pos="3" side="tl" part="_attr_n_"/>
<lit-tag v="px1sg"/>
<lit-tag v="dat"/>
</lu>
<b/>  </chunk>
  </out>
 </action>
 </rule>
 </section-rules>
</transfer>
```

Fig. 6. The rule for structure "S-K-T-C" that were taken from generating

The translation could be seen on the following picture:

```
...$ echo "к моим
...$ echo "к моим
...$ }$^default<default>{^.<sent>$}$
```

Fig. 7. Translation the structure "S-K-T-C"

As could be seen from the screenshot, generated rule translates without any error, as "#" sign, it means that rule is generated correctly and could work with other Apertium platform transfer's stage.

To present how we are going to use complete system of Kazakh endings, firstly we will check current state of rules in the language pairs, which we considered above, we used example structures from the previous work (Table 1 and Table 2 in [1]). During the translation of this structures, it was detected, that the rules for following phrase structures missed in some language pairs, for instance, for Kazakh-Russian pair:

Table 2. *Structure translations with Apertium platform*

Examples in Kazakh	Translation with Apertium	Chunk name
тәтелер	тёти	pr_n2
тәтем	моя тётя	pr_n2
тәтемін	#я тётя	n_pers

тәтеге	К тёте	pr_n1
тәтелерім	Тёти	pr_n2
тәтелерміз	тёти	pr_n2
Тәтелерге	К тётям	pr_n1
Тәтемсіз	Моя тётя	pr_n2
Тәтемге	тәтеме	pr_n1
Тәтесіңдер	Тётя	pr_n2
Тәтелерімсіңдер	тёти	pr_n2
Тәтеліре	тәтелеріме	pr_n1
Тәтелергемін	*тәтелергемін	
Тәтеңненбіз	от твоей тёти	pr_n2
Тәтелерінеміз	*тәтелерінеміз	
барғандар	#идти	r-KL
барғаным	#мой #идти	r-Koptik-bartr
барғанмын	#идти	v-neg-past
барғанға	к #идти	r-KL-SG
барғандарым	#идти	r-KL
барғандармыз	*барғандармыз	
барғандарға	к #идти	r-KL
барғандамыз	*барғандамыз	
барғандарғамын	*барғандарғамын	

On the right column existed chunk names are shown. Empty cell means that there is no rule for structure, for instance, "тәтелергемін".

Translations with symbols "#" and "*" mean that there no correct rule for this structure. Also some structures do not have right endings and word order. To get right translation we can use workflow for generate missed rules.

So next step will be creation of files on source and target languages. Let's create "rus.txt" and "kaz.txt" files for "S-R-K" structure:

Table 3. *Missed sctructures for English-Kazakh language pair*

Kazakh types of endings	Examples in Kazakh	Examples in English	Apertium-eng-kaz
S-K-C-J	Тәте-лер-ге-мін/ Тәте-лер-ге е-ді-м (tate-ler-ge-min/ tate-ler-ge e-di-m)	I am to aunt-s	Мен тәтелерге# е
S-K-T-C-J	Тәте-лер-ің-е-міз (tate-ler-ing-ge-miz)	we are to your aunt-s	Біз сіздің тәтеліріне #е
S-R-K-J	бар-ған-дар-мыз (bar-gan-dar-myn)	We are gones	Біз #бар

S-R-T-C	бар-ған-ым-ның (bar-gan-ym-nyng)	of my gone	менің барған
S-R-C-J	бар-ған-да-мыз (bar-gan-da-myz)	in our gone	біздің барған
S-R-K-T-C	бар-ған-дар-ым-ның (bar-gan-dar-ym-nyng)	of my gone	менің барған
S-R-K-C-J	бар-ған-дар-ға-мын (bar-gan-dar-ga-myn)	I to gone	Барғанға мен

Table 4. Example of two files for build chunk of "S-R-K"

SL.txt – rus.txt	TL.txt – kaz.txt
сходившие	барғанда

Script to generate that structure have to be updated:

```
12 DEFINE_string 'source_language' 'rus' 'source language' 's'
13 DEFINE_string 'target_language' 'kaz' 'target language' 't'
14 DEFINE_string 'data_dir' '/home/aida/apertium-testing/' 'Directory where the source and compiled Apertium dictionaries can be found
   (default: /usr/local/share/apertium/apertium-$s-$t/)' 'd'
15 DEFINE_string 'apertium_prefix' '/usr/local' 'Prefix where Apertium was installed (default: /usr/local/)' 'u'
16 DEFINE_string 'phrase' 'S-R-K' 'name of chunk'
17 DEFINE_string 'chunk' 'NP' 'name of chunk'
18
19 FLAGS "$@" || exit $?
20 eval set -- "${FLAGS_ARGV}"
21
22 SL=${FLAGS_source_language}
23 TL=${FLAGS_target_language}
24 PAIR=${SL}-${TL}
25 APERTIUMPATH="$PAIR"
26 CORPUS=${FLAGS_corpus}
27 PHRASE=${FLAGS_phrase}
28 CHUNK=${FLAGS_chunk}
```

Fig. 8. Updated script for structure "S-R-K"

Generated rules was inserted to current apertium-kaz-rus system:

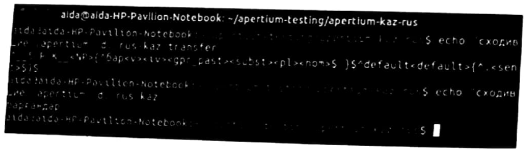

Fig. 9. Generated rule's translation of structure "S-R-K"

As could be seen from the Fig.9, generated rule uses chunk name "__S-R-K__" and phrase name "NP", these variables were defined in the script. Output of the translation appeared without any error.

By using the same sequences of steps, let's add missed rule, according to Table 3 for structure "S-K-C-J" to apertium-eng-kaz MT system.

***Table 5.** Example of two files for build chunk of "S-K-C-J"*

SL.txt – eng.txt	TL.txt – kaz.txt
I am to aunts	Тәтелерге едім

Table 5 shows the input files for "S-K-C-J" generation, where English is source language and Kazakh is target. Translation of generated rule could be seen at Fig.10

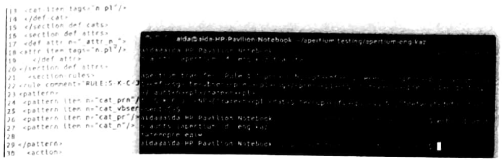

Fig. 10. Generated rule's translation of structure "S-K-C-J"

The workflow of rule generation, which was performed above, shows that application of rule generation provides automatic way to get rules for structures from the complete system of Kazakh endings. Usage of Apertium system helps to reduce amount of hand-work and human participation in whole process of generation. In the next Table 6 was performed number of endings which are already covered by generated and hand-written rules in Apertium English–Kazakh and Kazakh–Russian systems:

***Table 6.** Number of endings covered by rules in Apertium*

Endings	Total amount	English-Kazakh	Kazakh–Russian
Verbal	432	376	208
Participle	560	320	301
Nominal	768	723	577

5 Conclusion and future work

In this paper we performed improved technology of construction the chunk transfers rules, based on the complete set of Kazakh endings, for Kazakh-Russian and Kazakh-English language pairs. Technology, beside using previously developed application to generate rules, is used morphological analyzer of Apertium platform, which helps to do process of rule generation more easily.

In the future work is planned to improve rule generation program to use for verb phrases and structures. Now generation application could be applied for noun phrases only.

References

1. *Tukeyev U., Sundetova A., Abduali B., Akhmadiyeva Zh., Zhanbussunov N.* Inferring of the morphological chunk transfer rules on the base of complete set of Kazakh endings // Proceedinds of 8th International conference ICCCI 2016. – Halkidiki, Greece, 2016. – P. 563-574.

2. *Philipp Koehn* «Statistical Machine Translation». Cambridge University Press The Edinburgh Building, Cambridge CB2 8RU, UK. Published in the United States of America by Cambridge University Press, New York. ISBN-13 978-0-521-87415-1 Hardback.
3. *Karlsson F., Voutilainen A., Heikkilä J., Anttila, A.* Constraint Grammar: A language independent system for parsing unrestricted text. //Mouton de Gruyter. – 1995.
4. *Doug Cutting, Julian Kupiec, Jan Pedersen, and Penelope Sibun.* A Practical Part-ofspeech Tagger. Proceedmgs of ANLP-92. Trento, Italy, 1992.
5. https://svn.code.sf.net/p/apertium/svn/languages/apertium-kaz
6. https://svn.code.sf.net/p/apertium/svn/languages/apertium-rus
7. https://svn.code.sf.net/p/apertium/svn/staging/apertium-eng-kaz
8. *Sundetova, A., Forcada, M. L., Shormakova, A., and Aitkulova, A.* (2013). Structural transfer rules for english-to-kazakh machine translation in the free/open-source platform apertium. In Proceedings of the International Conference on Computer processing of Turkic Languages, pages 317–326.
9. https://svn.code.sf.net/p/apertium/svn/staging/apertium-kaz-rus/

МОБИЛЬДЫ РОБОТТЫҢ ЖЫЛДАМДЫҚ БЕРУ МЕН БҰРЫЛУ МЕХАНИЗМДЕРІНІҢ ӨЗАРА ҚАТЫНАСТАРЫ НЕГІЗІНДЕ КЕҢІСТІКТЕГІ ҚОЗҒАЛЫСЫ ТУРАЛЫ

Джолдасбаев С.К., Елеусинов А.И., Исламгожаев Т.У., Мажитов Ш.С.
serykjoldasbaev@mail.ru

Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан

Андатпа: *Жұмыста барлау жүргізу мақсаттарындағы құрастырылған арнайы жеңіл көліктік роботтың қозғалу принциптері мен орынауыстыруының сипаттамасы және сол сипаттамалар бойынша, оның қозғалыс бағытын, кеңістіктегі орналасу орны мен жағдайын анықтап отыру және бағыттау мәселелірнің қозғалу механизімдеріне тәуелділігі қарастырылады.*

Жеңіл көлік машиналарын құрастыру мен модернизациясында қозғалыстың орта жылдамдықпен бағаланатын жылдам жүруін болжау негізгі мәселелерінің бірі болып табылады. Жұмыс мақсаты әдістік-техникалық есептер шешімдерімен байланысты шектеулерді қарастырмай,

МАТЕРИАЛЫ

Международной научной конференции «Информатика и прикладная математика», 27-30 сентября 2017 года, Алматы, Казахстан

(Часть II)

(Сборник статей)

Подписано в печать 14.09.2017 г. Формат А4.
Печать ризография. Бумага офсетная. Усл. печ. л. 17,28.
Тираж 300 экз. Заказ № 01177.
Отпечатано в типографии НЦ НТИ.
Алматы, ул. Богенбай батыра, 221.