



БАЯНДАМАЛАР ЖИНАҒЫ

Республикалық ғылыми-практикалық конференция

«Математикалық және компьютерлік модельдеудің заманауи мәселелері

Қазақстанның цифрлы индустриясының дамуы жағдайында»

3-5 мамыр 2018 жыл, Астана, Қазақстан

СБОРНИК МАТЕРИАЛОВ

Республиканская научно-практическая конференция

«Современные проблемы математического и компьютерного моделирования

в условиях развития цифровой индустрии Казахстана»

3-5 мая 2018 года, Астана, Казахстан

ӘОЖ 004+519+316

КБЖ 22

М 49

В подготовке Сборника принимали участие:

Адамов А.А., Нугманова Г.Н., Сергибаев Р.А., Байдавлетов А.Т.

Математикалық және компьютерлік моделдеудің заманауи мәселелері Қазақстанның цифрлы индустриясының дамуы жағдайында: Республикалық ғылыми-практикалық конференциясының БАЯНДАМАЛАР ЖИНАҒЫ = Современные проблемы математического и компьютерного моделирования в условиях развития цифровой индустрии Казахстана: СБОРНИК МАТЕРИАЛОВ Республиканской научно-практической конференции. Қазақша, орысша, ағылшынша. – Астана, 2018, 161 б.

ISBN 978-601-337-014-9

Жинаққа студенттердің, магистранттардың, докторанттардың және ғалымдардың механика, математика, математикалық және компьютерлік моделдеу, математиканы оқыту әдістемесінің өзекті мәселелері бойынша баяндамалары енгізілген.

В Сборник вошли доклады студентов, магистрантов, докторантов и ученых по актуальным вопросам механики, математики, математического и компьютерного моделирования и методики преподавания математики.

Тексты докладов представлены в авторской редакции

ISBN 978-601-337-014-9

ӘОЖ 004+519+316

КБЖ 22.1

представление, для того чтобы, подробно ознакомиться с графической информацией по данному вопросу.

Для каждого документа, который хранится в базе знаний, имеются каталоги с графическими файлами документа. Это безусловно увеличивает требования к объему памяти компьютера.

На рисунке 5 представлен еще один пример запроса к базе знаний и ответа системы.

The image shows a user interface for a knowledge base query. On the left, there is a list of grammatical categories with radio buttons: **Существительное** (selected), **Прилагательное**, **Глагол**, **Определение**, and **Гипероним**. Under **Существительное**, the text reads: "холодильная камера". Under **Определение**, it lists technical details: "Охлаждаемая емкость, - наименование детали: - Шайба 20.65Г ГОСТ 6402-70, Гайка М12.5 ГОСТ 5915-70, - Шайба 12.65Г ГОСТ 6402-70, Гайка М16.5 ГОСТ 5915-70, - Болт М16х30.58 ГОСТ 7796-70, Шайба 16.65Г ГОСТ 6402-70". On the right, there is a list of grammatical categories: **Гипоним**, **Холоним**, **Синоним**, **Онтоним**, **Омоним**, and **Мероним**. A magnifying glass icon is positioned over a search results window on the right, which contains a list of related terms: **холодильная камера**, **Стихийные условия**, **Механизация**, **Воздухозаборник**, **Система Водяного охлад**, and **Сторонние механизмы**.

Рисунок 5 – Ответ системы на запрос пользователя «Холодильная камера»

В результате выполненной работы создано специальное программное обеспечение, которое автоматизирует операции включения понятий в онтологию предметной области. Предлагаемое решение уменьшает количество операций, выполняемых вручную экспертом предметной области в процессе создания онтологии, исключает влияние субъективности эксперта и необходимость принятия 9 интуитивных решений при включении новых понятий в онтологию[2]. Указанные свойства повышают потребительские качества и расширяют возможности программных средств, используемых для построения онтологии предметной области, в случае применения в них разработанного метода.

Заключение: Созданные онтологические модели и базы знаний тепловоза, а также программные средства станут основой для построения систем семантически ориентированного доступа к информационным ресурсам, включая навигационные и поисковые системы, электронные обучающие системы (в частности, симуляторы тепловозов), которые используют онтологическое структурирование элементы знаний и учебные объекты, информационные системы научно-исследовательских целей, обеспечивающие построение единой концепции, согласованной между моделью предметной области и базами данных специалистов.

Список использованных литератур:

- 1 Gruber T.R. A translation approach to portable ontology specifications // Knowledge Systems Laboratory, Stanford University, California. – 1972. – 103 p
- 2 Gruber T.R. What is an Ontology? // Knowledge Systems Laboratory, Stanford University, California. – 2001.– 94 p.

ANALYZING THE CONTENT OF INTERNATIONAL EDUCATIONAL MASTER PROGRAMS IN COMPUTATIONAL LINGUISTICS

A.Sharipbay, K.Sagindykov, R.Niyazova, R.Turebayeva

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

E-mail: sharalt@mail.ru, ksagin@mail.ru, rozamgul@List.ru, 58stud@mail.ru

Abstract: Training of national cadres in the field of computer linguistics should be carried out through the development and implementation of quality educational programs based on the study of international experience in the development and implementation of educational programs in com-

puter linguistics, tuning methodology tools, as well as the questioning, diagnosis and classification of training needs.

To develop the educational program Erasmus + we have analyzed the educational programs of foreign universities and institutions. And it was clear that, although the content of their educational programs of some universities relate to computer linguistics, but are called differently.

Based on the questioning of potential employers and the analysis of educational programs on computer linguistics of various universities, a list of competencies and disciplines was defined for the creation of a draft curriculum on computer linguistics.

Presented here is a master's educational program "Computer Linguistics", implemented under the Erasmus + project.

Key words: educational programs, computational linguistics, Mathematical Foundations of Computer Linguistics, Semantic analysis of texts, Methods of speech synthesis

1. Introduction

Modern computer linguistics covers subjects, related to automatic analysis of natural language. At first, These includes machine translation, information retrieval, voiced data entry, emotional analysis of texts and messages, etc. Second, the creation of educational and reference language resources, experimental tasks in the field of language theory. Third, the development of variety of applications that rely on language data.

Computer linguistics is necessary in startups, developing new linguistic technologies - for example, to create robots, understanding speech and responding in natural language, to automatically recognize emotions in the texts of users of social networks; in companies, engaged in processing large amounts of unstructured text data.

Currently, specialists in computer linguistics are very much in demand in the largest companies, involved in development in the field of linguistics. In Central Asia, educational task of training such personnel is quite new, and it became apparent only recently in connection with the rapid development of artificial intelligence. At the same time, there are no educational programs for targeted mastering in the field of computer linguistics, which proves again the obvious relevance of the master's educational program on computer linguistics, implemented under the Erasmus+ project, which includes the following member universities:

1. University of Santiago de Compostela, USC
2. University of a Coruña, UDC
3. Technological Educational Institution of Athens, TEIATH
4. University of Porto, U.PORTO
5. Adam Mickiewicz University in Poznań, AMU
6. Urgan State University, UrSU
7. Samarkand State Institute of Foreign Languages, SamSIFL
8. Tashkent State University of the Uzbek language and literature, TSUUL
9. National University of Uzbekistan, NUUZ
10. Republican State Enterprise operating under the right of economic management A. Baitursynov Kostanay State University of Ministry of Education and Science of the Republic of Kazakhstan, KSU
11. L.N. Gumilyov Eurasian National University, ENU
12. Al-Farabi Kazakh National University, KazNU[1].

In Kazakhstan, the classifier of specialties for higher education does not include specialty "Computer Linguistics", but at L.N. Gumilyov ENU, since 2013 the training path "Computer Linguistics" has been developed and implemented within the framework of educational programs for bachelor's specialties - "5B060200-Informatics" and magistracy - "6M060200-Informatics". For that we conducted a survey of representatives of employers of specialists in the field of computer linguistics, studied educational programs on computer linguistics of foreign universities.

2. Questioning of employers of specialists in the field of computer linguistics

In order to find out the employers' opinion about the prospective competencies required for specialists in the field of computer linguistics, we conducted a questionnaire using the following application:

https://docs.google.com/forms/d/e/1FAIpQLScT6iJA5e-iYKIC4mAIZcF_sMch77CRylhGUwmi_FQ1afiRWA/viewform?c=0&w=1&usp=mail_form_link

The purpose of the questionnaire is to determine the competencies and needs of specialists in the field of computer linguistics based on interviewing potential employers.

The result of the questionnaire: the definition of the list of competencies and the needs of specialists in the field of computer linguistics.

3. Analysis of educational programs on computer linguistics of foreign universitie

The purpose of the analysis of educational programs is to determine the level of education and the list of disciplines that provide these level. The result of the analysis of educational programs is the definition of a list of disciplines on the basis of an analysis of the content of international educational master's programs in the field of computer linguistics.

Training of national cadres in the field of computer linguistics should be carried out through the development and implementation of quality educational programs based on the study of international experience in the development and implementation of educational programs in computer linguistics, tuning methodology tools, as well as the questioning, diagnosis and classification of training needs.

To develop the master's educational program "Computer Linguistics", implemented by the Erasmus+ project, we analyzed the educational programs of the following universities and institutes:

Table 1 - List of disciplines in Computational Linguistics

№	Country, name university, Web-site	Disciplines
1.	RGGU Russian State University for the Humanities, www.rgggu.ru	Introduction to fundamental linguistics Typology, comparativistics, areal linguistics Modern syntactic theories Case and experimental methods in semantics Introduction to Computational Linguistics Computer Sociolinguistics Mathematical Foundations of Linguistics Statistical models in linguistics Methods of artificial intelligence in Computational Linguistics Programming of linguistic tasks Linguistic annotation / markup Specialized linguistic databases Methods for evaluating AOT systems Models and methods of Computational Linguistics Classification methods and machine learning Linguistic basis of machine translation Computer Parsing Analysis of oral speech

2.	SPbSU St. Petersburg University, spbu.ru	<p>Methods and models of ontological engineering Methods of knowledge engineering in humanitarian research Text understanding systems Text analysis models and their software implementation Statistical methods in language engineering Hull methods in language engineering Linguistics of the text and theory of verbal communication Languages and standards for describing information resources Expert systems and methods of inductive generalizations Methods of decision support Methods of software implementation of intelligent information technologies Mathematical modeling in data processing technologies Methodology and technology of designing information systems Information Society and Problems of Applied Informatics Business English Philosophical problems of science and technology</p>
3.	HSE High School of Economics Computational Linguistics	<p>Linguistic data: quantitative analysis and visualization Introduction to Linguistics Mathematics Formal models in linguistics Functional and cognitive models in linguistics Computational Linguistics Programming (Python) Analysis of linguistic data: quantitative methods and visualization (taught in English) Mathematical foundations of Computational Linguistics Machine learning Experimental Linguistics Database Ontologies and semantic technologies Digital Humanitarian Technologies: Resources, Tools, Case Studies Designing of linguistic resources and systems</p>
4.	MPhTI Moscow University of Physics and Technology, mipt.ru	<p>Mathematical Foundations of Linguistics Statistical models in linguistics Introduction to fundamental linguistics Typology, comparativistics, areal linguistics Russian corpus grammar Introduction to Computational Linguistics Computer Sociolinguistics Modern syntactic theories Typology of grammatical categories English for professional communication Models and methods of Computational Linguistics Data structures and basic algorithms The main algorithms of linguistic analysis Analysis of oral speech. Corpus linguistics: building and using enclosures Classification methods and machine learning Computer models of discourse Linguistic basis of machine translation Formal models and resources of world languages Linguistic annotation / markup Methods for evaluating AOT systems Computer parsing. Methods of artificial intelligence in Computational Linguistics Application Packages for Linguistic Studies Specialized linguistic databases Linguistic support of the tasks of document analysis Automatic assessment of the complexity of texts</p>

5.	ITMO (St. Petersburg) St. Petersburg National Research University of Information Technologies, Mechanics and Optics, ifmo.ru	Information Technology: <ul style="list-style-type: none"> • System analysis and modeling of information processes and systems; • Designing information systems; • Organization of design and development of distributed systems software; • Organization of software design and development for embedded systems; • Software testing; • Quality management software development. Speech technologies: <ul style="list-style-type: none"> • Digital signal processing; • Digital processing of speech signals; • Mathematical modeling and decision theory; • Pattern recognition; • Recognition and synthesis of speech; • Recognition of the speaker (speaking by voice); • Multimodal Biometric Systems.
6.	University of Oxford http://www.ox.ac.uk	Analysis of functional and structural data images of the brain. Physiological neuroimaging. Brain disorders. Diffusion of the image. Speech and the brain. Visualization. Neurodegeneration. Cognition. Psychiatry
7.	University of California, Los Angeles (UCLA) http://www.ucla.edu	Phonetics. Phonology. Syntax. Semantics. Psycholinguistics. Matlingvistics. Historical linguistics. African, Indian languages.
8.	Harvard University https://www.harvard.edu	Fundamental studies of the speech apparatus and speech functions. Clinical studies of human voice and speech abnormalities. Mechanics, biophysics, physiology and / or molecular biology of the middle and inner ear. Acquired or congenital abnormalities of the mechanisms of hearing. Neurophysiological or modeling approaches in the study of nerve cells and circuits underlying auditory processing. Neurovisual studies of the mechanisms of tinnitus. Cognitive neurobiology of language signal processing. Design, development and improvement of the hardware and software system for hearing aids, ear implants, vestibular prostheses or algorithms for automatic speech recognition.
9.	Cambridge University https://www.cam.ac.uk	Acoustic modeling (statistical models). Fundamental research in machine learning. Optimize dialogue using reinforcement learning. Recognition on large dictionaries. Pattern recognition. Speech recognition on mobile devices. Dictator independence and noise cancellation. Dialog systems and VoiceXML. Statistical language modeling. Statistical machine translation. Processing and transcription of recognized speech
10.	Carnegie Mellon University https://www.cmu.edu	User Interface Software Cognitive models. Speech recognition. Understanding of natural language.

		Computer graphics. Handwriting recognition. Visualization of data, visual design, multimedia. Computer support for teamwork. Computer music and theatrical skill. Social technologies
11.	Johns Hopkins University https://www.jhu.edu	Language modeling. Natural language processing. Neural treatment. Acoustic processing. The theory of optimization. Language Entry

Hence it is clear that, although the educational programs of some universities are named differently, but in terms of their content, they refer to computer linguistics.

4. Purpose and description of the master's degree program

The master's educational program "Computer linguistics" has 3 directions: "Linguistics", "Speech technologies", "Text technologies" (Text processing).

Bachelors-philologists and bachelors-psychologists enter the direction of "Linguistics".

Graduate students with a philological education are necessary for the training of specialists in the engineering of knowledge in the field of linguistics, without the need to study the mathematical foundations of computer science and programming. They will study the methods of composing the thesauri for certain subject areas, mark-up languages for audio and text records for creating text and audio corpora, language models for speech technologies.

In the direction "Speech technologies" and "Text technologies"

Graduate students who graduated from the bachelor's degree in computer science will be trained in two parallel trajectories: speech technologies and word processing technologies, which will provide mathematical and the basis and methods for creating these technologies.

Preparation of masters in the direction of "Linguistics" requires: a deep study of the fundamental foundations of linguistics with an emphasis on the methods of creating operational formal models of the language system, the adequate complexity of such tasks of natural language processing, as recognition and synthesis of speech and text, semantic analysis and understanding of text and speech.

Preparation of masters in the direction of "Speech Technology", requires in-depth study Methods of transcription of sounds and speech, Methods of Speech Recognition, Methods of speech synthesis, Methods of transcription of sounds and speech, Methods of Speech Recognition , Methods of speech synthesis

Preparation of masters in the direction of "Text Technology" requires Methods and tools for creating text corpora

Methods and tools for creating audio corpora, Syntax analysis of texts, Morphological analysis of texts, Semantic analysis of texts

It should be noted that any educational program consists of mandatory disciplines (Mandatory part) and elective disciplines (Variational part).

1. Mandatory part: History and philosophy of science, Foreign language (professional), Pedagogics, Psychology, Software development technology
2. Variational part:
 Linguistics: Ontology design tools, Tools for processing visual data, Tools for processing audio data, Tools creating thesauri, Semantic Search Tools.
 Speech technologies: Methods of transcription of sounds and speech, Methods of Speech Recognition, Methods of speech synthesis, Methods of transcription of sounds and speech, Methods of Speech Recognition, Methods of speech synthesis.
 Text technology: Methods and tools for creating text corpora, Methods and tools for creating audio corpora, Syntax analysis of texts, Morphological analysis of texts, Semantic analysis of texts.
 The table below presents the master's educational program "Computer linguistics"
 List of disciplines

Table 2 - Semester 1

UE	Objectif	Modules	ECTS	Lectures	TP	TL	W pers.	Total
UE1	History and philosophy of science	Methodology of master's degree students training	3	15	15		60	90
UE2	Foreign language (professional)		3		30		60	90
UE3	Statistical methods in Natural Language Processing	Methods of natural language processing	5	15	30		90	135
UE4	Mathematical Foundations of Computational Linguistics		5	15	30		90	135
	course by choice 1*							
UE5	Tools for creating text corpora	Tools for the creation of language corpora	5	15	30		90	135
UE6	Tools for creating audio corpora		5	15	30		90	135
UE7	1.1 Programming in Python	Languages for symbol processing	5	15	30		90	135
UE8	1.2 Programming in Prolog		5	15	30		90	135
UE9	1.3 Methods of digital processing of speech signals	Digital Signal Processing	5	15	30		90	135
UE10	1.4 Software for processing speech signals		5	15	30		90	135
UE11	Scientific-research work of graduate students (master)		4					120

*- the student has the right to choose two of the 4 courses. Each course is 5ECTS (135 hours)

Conventions:

TP: Practical work

TL: Laboratory work

Pers. work: Independent work (libraries, at home, in practice, etc.)

Table 3. - Semester 2 UE (2 semester = 30 ECTS)

UE	Objectives	Modules	ECTS	Lectures	TP	TL	Pers. work	Total
UE1 2	Pedagogics	Methodology of master's degree students training	3	15	15		60	90
UE1 3	Psychology		3	15	15		60	90
UE1 4	Software development technology		3	15	15		60	90
	Course by choice 2**							
UE1 5	Ontology design tools	Tools for processing natural languages	5	15	30		90	135
UE1 6	Tools for processing visual data		5	15	30		90	135
UE1 7	Tools for processing audio data		3	15	15		60	90
UE1 8	Syntax analysis of texts	Text understanding	5	15	30		90	135
UE1 9	Morphological analysis of texts		5	15	30		90	135
UE2 0	Semantic analysis of texts		3	15	15		60	90
UE2 1	Methods of transcription of sounds and speech	Speech recognition	5	15	30		90	135
UE2 2	Methods of Speech Recognition		5	15	30		90	135
UE2 3	Methods of speech synthesis		3	15	15		60	90
UE2 4	Scientific-research work of graduate students (including scientific internships)		8					240

** - the student has the right to choose three out of 6 courses. Each course - 3 credits (90 hours) or 5 credits (135 hours)

Table 4. - Semester 3 UE (3 semester = 30 ECTS)

UE	Objectives	Modules	ECTS	Lectures	TP	TL	Pers. work	Total
UE25	Specialized linguistic databases	Big Data	5	15	30		90	135

UE26	Data Mining	processing and storing	4	15	30		90	135
	course by choice3***							
UE27	Tools creating thesauri	Search Tools	5	15	30		90	135
UE28	Semantic Search Tools		5	15	30		90	135
UE29	Sentiment analysis of natural language texts	Creating of text corpus	5	15	30		90	135
UE30	Method for processing of text corpora		5	15	30		90	135
UE31	Synthesis of speech analysis of natural language	Creating of audio corpus	5	15	30		90	135
UE32	Methods for processing of audio corpora		5	15	30		90	135
UE33	Scientific-research work of graduate students		8					240
UE34	Teaching practice		3					90

***- the student has the right to choose two of the 4 courses. Each course - 5 credits (135 hours)

Table 5. - Semester4 UE (4semester = 30 ECTS)

UE	Objectives	Modules	ECTS	Lectures	TP	TL	Per s. work	Total
UE35	Scientific-research work of graduate students		8					240
UE36	Research internship		12					360
UE37	Complex exam		3					105
UE38	Writing and defense of Master's degree thesis		7					315

Conclusion

Based on the analysis of questionnaires of employers and educational programs, we defined competencies, a list of disciplines, an ontological model and the educational program on computer linguistics have been built, with three training trajectory.

This educational program can be used not only in the L.N. Gumilyov ENU, but in other universities as well.

References

1. Partners [Электронный ресурс],-2018.-URL:<http://www.comling.enu.kz>.(Дата обращения:18.04.2018).
2. Gruber T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing / Gruber T.R. // International Journal Human-Computer Studies. –43, 907-928 (1995)

3. Yergesh B., Mukanova A., Sharipbay A., Bekmanova G., and Razakhova B. Semantic Hyper- graph Based Representation of Nouns in the Kazakh Language. *Computación y Sistemas* Vol. 18, No. 3, 2014 pp. 627–635. DOI: 10.13053/CyS-18-3-2041.
4. Zhetkenbay L., Sharipbay A., Bekmanova G., Khaby lashimuly M., Kamanur U.. These matical, ontological models and formalization rules Kazakh compound words. *Turklang'14 II International Conference on Computer processing of Turkic Languages*, Istanbul, Turkey, 2014– Istanbul, 2014. – P.107-113.
5. Zhetkenbay L., Sharipbay A.A., Bekmanova G.T., Kamanur U. The ontological model of noun for Kazakh-Turkish machine translation system. *Turklang'15// III International Conference on Computer processing of Turkic Languages*, Russia, Kazan, Tatarstan. – Kazan,2015. –P.15-24.