

ОТЗЫВ

**на автореферат диссертационной работы
Худойбердиева Хуршеда Атохоновича на тему
«Проектирование и реализация автоматических систем обработки
информации на таджикском языке» представленной на соискание
ученой степени доктора технических наук по специальности
05.13.11 – «Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей»**

Актуальность диссертационной работы обусловлена широким распространением задач компьютерной лингвистики, систем автоматической обработки больших объемов данных, широким развитием технологий формирования цифрового портрета текста и обеспечения правильного контента в глобальной сети интернет. В этих условиях необходимо создавать высокоуровневые, легко используемые и доступные системы обработки информации на естественном языке.

В первой главе диссертационного исследования изложена перспектива развития компьютерной лингвистики в Республики Таджикистан, формирования и реализации математического и программного обеспечения сбора, хранения и обработки информации на таджикском языке. Обзор научных достижений достаточно полон и обеспечивает корректную постановку задачи диссертационной работы.

Во второй главе изложены результаты исследования по реализации математического подхода к построению систем автоматической обработки информации на естественном языке путем их представления в виде математических моделей, представлены алгоритмы автоматической обработки и методология их реализации.

В третьей главе описывается способ проектирования информационных систем на базе объектно-ориентированного подхода. Можно отметить, что само представление системы в виде диаграмм не является новым. Научная новизна заключается в методе его реализации – моделировании поведения системы, структурном моделировании элементов системы и их взаимодействия, а также разработке физической модели информационных систем.

Хочется отдельно отметить последовательное изложение этапов построения информационных систем, проверка правописания, машинный перевод и синтез речи на основе предложенного текста. Автор подробно поясняет различные методы и алгоритмы, их достоинства и недостатки, утверждения, приводящие к итоговой стратегии разработки программных модулей обработки информации.

К основному результату диссертационного исследования следует отнести предложенную методологию обработки текста на естественном языке, которая, в конечном счете, обеспечивает обработку больших объемов информации посредством проверки правописания, машинного перевода и в конечном этапе автоматического преобразования текста в речь.

В трех последующих главах диссертации на основе разработанных моделей представлены материалы о создании систем: проверки правописания; машинного перевода; синтеза речи. Автор предлагает много небольших, практических

оптимизаций, приводит описание эффективного взаимодействия систем посредством анализа элементов текста, таких как буквы, слоги и слова. В содержании каждой главы приведены результаты экспериментальных исследований работоспособности предложенных систем.

Работа не лишена недостатков, которыми являются:

1. В пятой главе приводятся результаты сравнения созданной системы с существующими решениями машинного перевода. При этом созданную систему, сравнивают с примерами, доступными в режиме онлайн, пренебрегая системами, работающими в режиме офлайн.
2. В последнем разделе шестой главы автор упоминает, что исследование проблемы распознавания речи на таджикском языке нашли в свое решение на базе результатов, полученных при проектировании системы синтеза речи. Можно считать, что это представляется некоторым недостатком, так как результаты автора, полученные при разработке системы синтезе речи, базируются на конкатенации слогов и будут преобразованы в случае необходимости в различные фонемы звуков.

Упомянутые недостатки, однако, не снижают ценности полученных результатов и не свидетельствуют о недостаточном качестве диссертационной работы.

По итогам работы были опубликованы научные статьи в журналах, рецензируемых РИНЦ, охватывающие как теоретические так и практические аспекты исследовательской работы: математическое моделирование процессов обработки информации, оптимизированные системы сбора, обработки текстовых данных; разработка алгоритмов проверки правописания и машинного перевода текста; методологию исполнения синтеза речи в голосовом управлении. Покрытие работы научными статьями представляется полным.

По моему мнению, диссертация Худойбердиева Хуршеда Атохоновича является законченным научно-квалификационным исследованием, достаточно полно апробированном на конференциях и семинарах в Республики Таджикистан и Российской Федерации.

Диссертационная работа заслуживает положительную оценку, рекомендую присвоить Худойбердиеву Хуршеду Атохоновичу ученую степень доктора технических наук по специальности 05.13.11 - Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Акмал Рустамович Ахатов,
доктор технических наук, профессор,
Проректор по международному сотрудничеству
Самаркандского государственного университета
имени Шарофа Рашидова.
тел.: +998902716418
email: a-rustamovich@samdu.uz

А. Р. Ахатов
tasdiqlayman
Sharof Rashidov nomidagi
SamDU xodimlar bo'limi boshlig'i

Ахатов



ОТЗЫВ

на автореферат диссертационной работы

Худойбердиева Хуршеда Атохоновича на тему «Проектирование и реализация автоматических систем обработки информации на таджикском языке» на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Диссертационная работа Худойбердиева Х.А. посвящена проблеме разработки и реализации автоматических систем обработки информации на таджикском языке. Особенностью современного этапа развития информационно-телекоммуникационных технологий является рост количества пользователей интернета и потребляемого ими контента на родном языке, в частности на таджикском языке. Исходя из этого, возникают разнообразные задачи и сервисы, связанные с обработкой информации на конкретном естественном языке. Быстрое развитие современных компьютерных технологий является проблемой успешной работы любой организации, так как ежедневно на рынке появляются новые программные обеспечения с поддержки той или иного естественного языка. Сказанное определяет актуальность темы исследования, так как использование таджикского языка в компьютерных технологиях также является современной задачей.

Информационные системы объединяют две области знаний – компьютерные технологии для реализации задач компьютерной лингвистики и системную вычислительную архитектуру правил языка и обработки большого объема информации. В диссертационной работе предпринимается попытка исключить системную архитектуру из существующих методов для создания информационных систем, путём предоставления определенных цифровых портретов элементов текста, которые можно использовать для реализации ряд задач обработки информации.

Диссертационная работа состоит из шести глав. В первой главе даётся обзор существующих решений в области компьютерной лингвистики и постановка задачи. Во второй главе описываются предлагаемые математические модели, которые по ходу решения задач исследования становятся основой реализации в методы и алгоритмы. В третьей главе описана реализация объектно-ориентированного программирования информационных систем. В четвёртой главе приводятся результаты работы построенной системы автоматической проверки правописания, как результаты созданных собственных автором методов и алгоритмов. В пятой главе приведены результаты разработки машинного перевода текста с таджикского языка на русский и реализация в виде веб-приложений. В шестой главе рассматриваются попытка проектирования и реализации системы синтеза речи на таджикском языке.

Научную новизну работы определяют:

- предложены новые научно-технические положения, математические модели, методы и структуры данных, которые в целом составляют теоретическую основу системного анализа и исследования текстовой информации;
- впервые разработаны методы и алгоритмы практического, структурного и объектно-ориентированного проектирования систем автоматической обработки данных;
- предложены новые методы создания программных средств автоматического синтеза речи на таджикском языке, система автоматической проверки орфографии TajSpell в программном пакете Microsoft Office;
- программные модули автоматического перевода текста с таджикского языка на русский и английские языки в виде интернет-приложения, доступного по адресу

tarjumon.tajlingvo.tj.

- на основе раз разработанных методов, моделей и структур данных предложены новые алгоритмы машинного перевода, сформированы компьютерные параллельные корпуса Tajik-Russian-Parallel Corpus и TajikEnglish-Parallel Corpus в виде веб-приложений, а также программные модули автоматического перевода текста с таджикского языка на русский и английский языки;

- разработаны новые модели, методы синтеза речи и компьютерные программы Computer Tajik Text Narrator, Tajik Text-to-Speech, повышающие эффективность практического использования ИКТ для решения актуальных лингвистических задач и речевых технологий в таджикском языке.

Корректность и достоверность результатов диссертационного исследования подтверждается внедрением в несколько организаций и хорошим покрытием приведённых утверждений научными статьями, выступлениями в конференциях и научных семинарах, а также полученных свидетельств государственной регистрации информационных ресурсов и интеллектуальных продуктов.

Можно высказать следующие пожелания, которые соискатель может учесть в будущем:

Так как язык UML имеет концептуальность проектирования информационных систем, сборщик логической и физической структуры разрабатываемого проекта для компьютерных продуктов может ненадолго поддержать обрабатывающую функциональность системы. В диссертации отмечено, что это не приведёт к изменению пакетов, таких как проверки правописания, машинного перевода и синтеза речи на таджикском языке. Однако будет оригинально если все эти пакеты спроектировать в отдельном формате, в конечном счёте непонятно, что автор подразумевает под задачами, объединения всех перечисленных пакетов в одно целое. Например, является ли, такой задачей воссоединения задачи проверки правописания с задачей машинного перевода и синтеза речи? Но, в трех главах с четвертого по шестой приводятся результаты в конкретной реализации отдельно.

Данные замечание не являются принципиальными и не умаляют ценности проведённого исследования.

Диссертационное исследование представляет собой целостную, корректно структурированную, законченную научно-квалификационную работу с практическим результатом, которая с различных сторон освещает исследуемую тему.

Диссертационная работа «Проектирование и реализация автоматических систем обработки информации на таджикском языке» удовлетворяет предъявляемым требованиям, и должна быть положительно оценена. Автор исследования, Худойбердиев Хуршед Атохонович, заслуживает присвоения искомой учёной степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Профессор кафедры информационных технологий и систем управления,
Чебоксарского института (филиала)
Московского политехнического университета,
доктор технических наук,

Иванов Андрей Геннадьевич



Отзыв
на автореферат диссертации
Худойбердиева Хуршеда Атохоновича

«Проектирование и реализация автоматических систем обработки информации на таджикском языке», представленной на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

В современных системах обработки видеоданных одним из способов автоматической обработки текстовой информации на естественном языке является внедрение математических и программных обеспечений на базе методов и алгоритмов определения элементов текста. При этом одной из важнейших задач является реализация цифрового портрета, который должен быть сформирован на базе конкретного элемента текста, например, буквы, слога, слова или словосочетания. Кроме того, метод анализа текстовой информации должен как можно точнее выявить статистические показатели. Актуальность темы диссертационного исследования обоснована необходимостью решения комплекса задач обработки информации с идентификационным образом базовых элементов текстовой информации, которые предоставляют цифровой портрет текста.

Автор диссертационного исследования предлагает программный комплекс TajLINGVO в котором доступны автоматические системы проверки правописания, машинного перевода и синтеза речи на таджикском языке.

В диссертации представлены следующие основные положения:

- концепция автоматической обработки текстовой информации на таджикском языке как объект научного исследования и их программные средства для систематического анализа;
- научно-практический подход к разработке электронных словарей и компьютерных тезаурусов, в рамках которого сформированы примеры решения поисковых задач;
- метод автоматического синтеза речи на таджикском языке, основанный на использовании метода конкатенации слогов;
- методы извлечения, представления и обработки данных, составляющих отдельные элементы текста, а также предложен новый способ решения проблемы автоматического правописания текста на таджикском языке;
- метод автоматического перевода текста с таджикского языка на русский, разработаны модели, методы и алгоритмы, которые дают возможность эффективно решать практические задачи;
- методология совместного использования методов системного анализа, структуры данных, объектно-ориентированного программирования и компьютерной лингвистики для разработки систем автоматической обработки текстовой информации на таджикском языке;
- комплекс компьютерных программ TajLINGVO, для реализации всех представленных методов и систем.

Теоретическая значимость исследования состоит в разработке средств математического и программного обеспечения обработки информации и текстовых

элементов, содержащими статистическими показателями, обеспечивающего повышение устойчивости показателей элементов к преобразованиям цифровые портреты в процессе их обработки.

Практическая значимость диссертационного исследования заключается в разработке комплексов программных продуктов обработки информации, содержащих динамические структуры данных с возможностями автоматической проверки правописания, машинного перевода и синтеза речи.

По результатам диссертационного исследования опубликовано 68 печатных работ, в том числе 25 статьи в изданиях, рекомендованных ВАК, 3 научных монографий, а также получено 18 свидетельств о государственной регистрации программы для ЭВМ и информационных ресурсов.

Исходя из текста автореферата, можно сделать следующие замечания:

1. В третьей главе недостаточно четко определено понятие объектной модели элементов текстовой информации. Из описания приведенных методов и подходов проектирования понятно, что имеется в виду изменение поведения системы в виде диаграммы вариантов использования, однако более удобным для читателя было бы сразу дать определения функциональных возможностей.

2. В реализации задачи синтеза речи отмечен метод конкатенации слогов таджикского языка на основе базы «слог-звук». Есть вероятность появления новых слогов. Более интересно было бы исследовать голосовую базу «буква-слог» для формирования другого подхода синтеза речи.

Отмеченные замечания ни в коем случае не влияют на общую положительную оценку диссертации. Диссертация Худойбердиева Х.А. является законченной научно-исследовательской работой, содержащей новое решение важной научно-технической проблемы в компьютерной лингвистике таджикского языка.

По актуальности избранной темы, глубине проработки всего комплекса частных задач, научной ценности и практической значимости полученных результатов, обоснованности выводов и рекомендаций можно заключить, что диссертационная работа соответствует требованиям ВАК при Президенте Республики Таджикистан, предъявляемым к докторским диссертациям по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», а ее автор, Худойбердиев Хуршед Атохонович, заслуживает присуждения ученой степени доктора технических наук.

Профессор кафедры электроники
государственного образовательного учреждения
«Худжандский государственный университет
имени академика Бободжона Гафурова»
доктор физико-математических наук,



Кодиров Абдулахат Лакимович

Подпись профессора Кодирова Абдулахата Лакимовича заверяю.

Начальник ОК и СД ГОУ ХГУ им. ак. Б. Гафурова

Наврузов Э.

735700, г. Худжанд, проезд Мавлонбекова 1. тел.: (992-3422) 6-52-73, факс: (992-3422) 6-75-18,

www.hgu.tj, e-mail: rector@hgu.tj



UPPSALA
UNIVERSITET

2024-06-14 Dnr: 01

Postal address:

Box 33, 751 05 Uppsala

Visiting address:

Hus 10, Lägerhyddsvägen 1

Murtazo Nazarov

Associate Professor, Docent
Scientific Computing

Telefon: +46 722 81 48 08

E-post: murtazo.nazarov@uu.se

Webbadress: <https://murtazo.github.io>

Organisationsnummer

202100-2932

*на автореферат диссертационной работы
Худойбердиева Хуршеда Атохоновича
на тему «Проектирование и реализация
автоматических систем обработки
информации на таджикском языке» на
соискание ученой степени доктора технических
наук по специальности
05.13.11 – «Математическое и программное
обеспечение вычислительных машин,
комплексов и компьютерных сетей»*

ОТЗЫВ

Диссертационная работа Худойбердиева Х.А. посвящена исследованию задач компьютерной лингвистики на базе моделирования как математических, так и компьютерных решений автоматической обработки информации на таджикском языке. На сегодняшний день исследования в сфере компьютерной лингвистики представляет собой актуальную тему, который связан с общенаучными и междисциплинарными программами. Методы и алгоритмы, основанные на математических моделях, обеспечивают развития и использования современных компьютерных технологий для реализации автоматической обработки информации на естественном языке.

Целью автора диссертации является разработка компьютерного комплекса для исследования и обработки информации на таджикском языке, а также для автоматического управления текстом и речью. Важность развития прикладных аспектов компьютерной лингвистики и управления большими данными подчеркивается в контексте эффективного изучения и анализа таджикского языка. Реализованный комплекс компьютерных программ, доступный онлайн по адресу www.tajlingvoj.tj, представляет собой важный инструмент для исследования и реализации задач компьютерной лингвистики в указанной области.

Диссертация Худойбердиева Х.А. представляет собой важный вклад в область компьютерной лингвистики, отличается своей научной новизной и практической применимостью, обеспечивая новые пути развития данной области исследований.

Предложенные выводы и рекомендации, сделанные в автореферате, обоснованы благодаря ее основательной научной подготовке и заслуживают большого признания за их значимость развития компьютерной лингвистики.

В качестве недостатков можно отметить, что:



UPPSALA
UNIVERSITET

ОТЗЫВ

2024-06-14 Dnr: 01

1. Проведен недостаточно полный теоретический анализ точности и вычислительной сложности методов обработки информации на таджикском языке по сравнению с русским языком.
2. Для апробации разработанных моделей рассмотрена только модельных задач для таджикского языка для реализации задач автоматической проверки правописания, в прочем для задачи машинного перевода и синтеза речи обусловлены взаимные модели с русским языком.

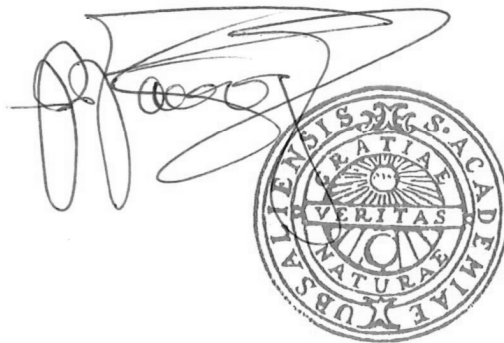
Обозначенные замечания не снижают общую ценность и уровень исследовательской работы. Исследования Худойбердыева Х.А. представляет собой завершенную научную работу с огромной квалификационной базой и достаточным практическим ценностью. Следовательно, считаю, что диссертация удовлетворяет требованиям, предъявляемым к докторским диссертациям по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Автор Худойбердиев Хуршед Атохонович, заслуживает присуждения ученой степени доктора технических наук.

Ассоциированный профессор вычислительных наук

Доцент численного анализа

Доктор технических наук в области прикладной и вычислительной математики

Муртазо Назаров



UPPSALA UNIVERSITET
Institutionen för informationsteknologi
Department of Information Technology



UPPSALA
UNIVERSITET

2024-06-14 Dnr: 01

Postal address:

Box 33, 751 05 Uppsala

Visiting address:

Hus 10, Lägerhyddsvägen 1

Murtazo Nazarov

Associate Professor, Docent
Scientific Computing

Telefon: +46 722 81 48 08

E-post: murtazo.nazarov@uu.se

Webbadress: <https://murtazo.github.io>

Organisationsnummer

202100-2932

*for the introduction of the dissertation work
Khudoiberdiev Khurshed Atokhonovich
on the topic "Design and implementation of
automatic information processing systems in the
Tajik language" for the degree of Doctor of
Technical Sciences in the specialty
05.13.11 – "Mathematical and software for
computers, complexes, and computer networks"*

Review

The dissertation work of Khudoiberdiev K.A. is devoted to the study of computer linguistics problems based on modeling both mathematical and computer solutions for automatic information processing in the Tajik language. In today's research, the field of computational linguistics is a current topic that is associated with general scientific and interdisciplinary programs. Methods and algorithms based on mathematical models ensure the development and use of modern computer technologies to implement automatic processing of information in natural language.

The goal of the author of the dissertation is to develop a computer complex for researching and processing information in the Tajik language, as well as for automatic control of text and speech. The importance of developing applied aspects of computational linguistics and big data management is emphasized in the context of effective study and analysis of the Tajik language. The implemented set of computer programs, available online at www.tajlingvoj.tj, is an important tool for the research and implementation of computational linguistics problems in this area.

The dissertation by Khudoiberdiev K.A. represents an important contribution to the field of computational linguistics, distinguished by its scientific novelty and practical applicability, providing new ways for the development of this field of research.

Her thorough scientific background justifies the proposed conclusions and recommendations made in the abstract and deserves great recognition for their significance in the development of computational linguistics.

The disadvantages include:

1. An insufficiently complete theoretical analysis of the accuracy and computational complexity of information processing methods in the Tajik language in comparison with the Russian language has been carried out.



UPPSALA
UNIVERSITET

Review

2024-06-14 Dnr: 01

2. To test the developed models, only model tasks for the Tajik language were considered for the implementation of automatic spell-checking tasks; in addition, for the task of machine translation and speech synthesis, mutual models with the Russian language were determined.

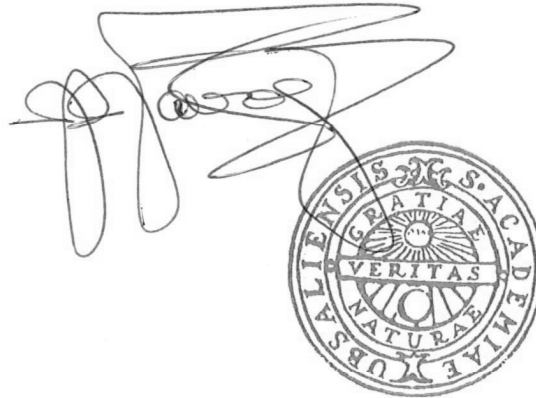
The indicated comments do not reduce the overall value and level of research work. Research by Khudoiberdyev K.A. represents a completed scientific work with a huge qualification base and sufficient practical value. Consequently, I believe the dissertation satisfies the requirements for doctoral dissertations in the specialty 05.13.11 - "Mathematical and software of computers, complexes, and computer networks." The author Khudoiberdiev Khurshed Atokhonovich deserves to be awarded the academic degree of Doctor of Technical Sciences.

Associate Professor in Scientific Computing

Docent in Numerical Analysis

PhD (Teknologie doktor) in Applied and Computational Mathematics

Murtazo Nazarov



UPPSALA UNIVERSITET
Institutionen för informationsteknologi
Department of Information Technology

ОТЗЫВ

на автореферат диссертационной работы
Худойбердиева Хуршеда Атохоновича
на тему «Проектирование и реализация автоматических систем
обработки информации на таджикском языке» представленную на
соискание ученой степени доктора технических наук по
специальности 05.13.11 – «Математическое и программное
обеспечение вычислительных машин, комплексов и
компьютерных сетей»

В диссертационной работе Худойбердиева Х.А. рассмотрена задача проектирования и реализации автоматических систем обработки информации на естественном языке, а именно на таджикском языке. Поставленная задача актуальна в связи с распространением большого объема контента в сети интернет на разных языках, которые позволяют обеспечить к правильным содержаниям электронных документов. Контроль качества контента, которые во многом зависят от действий пользователей, таких как правописание, это первоначальный этап, позволяющий добиться возможности правильного машинного перевода текстовой информации, это второй этап и в последующем третьем этапе синтезировать в речь. В зависимости от требований пользователей перечисленные этапы обработки информации также можно реализовать как отдельные функции одной целой информационной системы.

В работе сделан акцент на возможности функционирования методов и алгоритмов предобработки репрезентативной выборки текстовой информации, основных элементов текста и получение их цифрового портрета. Это обусловило выбор определенных алгоритмов вычисления коэффициентов для заданного элемента текста, таких как буквы, слоги, словосочетания, а также некоторые из них в цифровом формате звуков. Полученные данные определяют новизну и методы решения поставленных задач исследовательской работы для вынесения окончательных результатов.

Значительная часть работы посвящена выбору и обоснованию оптимальных методов для реализации в собственных алгоритмах обработки информации. В качестве классификаторов рассмотрены различные способы представления цифрового портрета текста на таджикском языке. Используемый объектно-ориентированный подход проектирования информационных систем показывает достаточный уровень качественного и по сравнению с аналогичными системами эффективного результата.

В качестве проверки адекватности выбранных признаков рассмотрен также дополнительная задача – проблема реализации распознавания речи на таджикском языке на основе синтезирования по некоторым схожим параметрам.

К сожалению, в тексте диссертации недостаточно сравнительных данных о вычислительной сложности и времени работы предложенного подхода. Именно эти данные вызволили бы оценить о его преимуществе по сравнению с более трудоемкими разработками систем на базе нейросети и машинного обучения. Данное замечания не снижает ценность диссертационной работы.

В целом работа «Проектирование и реализация автоматических систем обработки информации на таджикском языке» выполнена на высоком теоретическом уровне и представляет практическую ценность. Исследовательская работа удовлетворяет требованиям Положения ВАК при Президенте Республики Таджикистан о присуждении ученых степеней. Автор диссертации Худойбердиев Хуршед Атохонович, заслуживает присуждения ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Автор отзыва:

Назимов Акбар Багадулович,
Доктор физико-математических наук,
Сочинский государственный университет,
Профессор кафедры информационных технологий и математики,
n.akbar54@mail.ru

Подпись профессора Назимова Акбара Багадуловича заверяю.



ОТЗЫВ

на автореферат диссертационной работы Худойбердиева Хуршеда Атохоновича на тему «Проектирование и реализация автоматических систем обработки информации на таджикском языке» представленной на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

АВТОР ОТЗЫВА

ФИО: Ускенбаева Раиса Кабиевна

Ученая степень: доктор технических наук

Место работы: КазНИТУ им. К.И. Сатпаева

Должность: проректор по академическим вопросам

Контактная информация: r.k.uskenbayeva@satbayev.university

Широкое использование компьютерных технологий в различных задачах лингвистики вызвало необходимость проведения большого количества исследований посвященных автоматической обработке информации на естественном языке. За последние десять лет в этих работах широко используются математические методы реализации алгоритмов. Практическая ценность подобного исследования зависит от того, рассматриваются ли в нем реальные системы обработки данных. Данная диссертационная работа посвящена анализу цифрового портрета документов, получаемых с помощью компьютерных устройств с целью дальнейшей обработки.

В первой главе сделан обзор ключевых подходов к рассмотрению задач, связанных с определением главных особенностей проектирования и проблем реализации информационных систем обработки информации на таджикском языке.

Во второй главе рассмотрен основной метод компьютерного анализа и синтеза естественного языка, приводится описание основных методов обработки информации на естественном языке, а также методы решения задач математического моделирования, методов анализа и синтеза информации на таджикском языке. Обосновано практическое использование методов анализа текста и созданы классификаторы информации в формате текстовых данных.

В третьей главе описывается полный набор объектно-ориентированного моделирования систем обработки текста на естественном языке. На базе предложенных диаграмм UML разработан типовой проект информационной системы обработки текстовых данных на таджикском языке.

В четвертой главе рассмотрены вопросы проектирования, обработки и реализации системы автоматической проверки орфографии текста на таджикском языке. Представлены результаты исследования и реализации электронных словарей, компьютерного тезауруса таджикского языка, системы автоматической конвертации шрифтов в тексте на стандартные шрифты таджикского языка.

В пятой главе излагаются этапы проектирования, разработки и внедрения автоматического переводчика таджикского языка. Рассмотрены результаты исследования проблем машинного перевода текстов с разных языков на таджикский и наоборот, а также выявлены проблемы художественного перевода и его зависимость от машинного перевода на базе технологий Google.

В шестой главе рассматривается постановка задачи математического моделирования и компьютерной реализации синтеза таджикской речи на основе предложенного текста. Раскрыты проблемы анализа текстовых данных на основе слогов различной структуры, словоупотреблений, слоговых структур слов, встречающихся в таджикском языке.

В данной работе автор предложил новые алгоритмы оценки качества контента документа с точки зрения возможности автоматической проверки его правописания, машинного перевода и синтеза. Проведенные эксперименты показали, что среди методов, пригодных для обработки текстовой информации, предложенное решение показывает наилучшие результаты, немного уступая при этом нейросетевым решениям.

Разработанные методы и алгоритмы полностью удовлетворяют требованиям, сформулированным в постановке задачи диссертации, а также являются гибкими и быстрыми, работают при различных нестандартных шрифтах и с различными форматами документов.

Недостатком метода является его нечувствительность к изменениям качества документов с уровнем обработки выше 95%. Однако, для многих практических применений эта зона не является критически значимой. Другое замечание касается дополнительных примеров использования метода, где набор его признаков необходимо было бы дополнить нетекстовыми признаками, такими как знаки препинания, различные символы, математические формулы.

Вышеуказанные недостатки несколько не снижают уровень полученных результатов исследований и практическую значимость исследовательской работы. Диссертационная работа соответствует требованиям положения о присуждении ученых степеней, а её автор Худойбердиев Хуршед Атохонович заслуживает присуждения ученой степени доктора технических наук по

специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Д.т.н., профессор
проректор по
академическим вопросам
КазНУ им.К.И. Сатпаева

Р.К. Ускенбаева

Подпись профессора Ускенбаевой Раисы Кабиевна заверяю

Директор HR Службы



А.К. Беисова

В диссертационный совет 6D.KOA-049 при Таджикском техническом университете имени академика М.С. Осими по адресу: 734042, г. Душанбе, проспект академиков Раджабовых, 10А

ОТЗЫВ

на автореферат диссертационной работы Худойбердиева Хуршеда Атохоновича на тему «Проектирование и реализация автоматических систем обработки информации на таджикском языке» на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Диссертационная работа Худойбердиева Х.А. посвящена исследованию задач компьютерной лингвистики на базе моделирования как математических, так и компьютерных решений автоматической обработки информации на таджикском языке.

На сегодняшний день исследования в сфере компьютерной лингвистики представляет собой актуальную тему, которая связана с общенаучными и междисциплинарными программами. Методы и алгоритмы, основанные на математических моделях, обеспечивают развитие и использование современных компьютерных технологий для реализации автоматической обработки информации на естественном языке. Повышение эффективности обработки нарастающего объема речевых данных в современном информационном обществе является важнейшей задачей. Работы в области интеллектуальных информационных систем обслуживания и управления, сетей, искусственного интеллекта, научных и образовательных программ весьма востребованы, и актуальность их лишь возрастает со временем. Выполняемые по данной тематике работы объединяют как теоретические исследования, так и практические разработки.

Целью диссертации является разработка компьютерного комплекса для исследования и обработки информации на таджикском языке, а также для автоматического управления текстом и речью. Важность развития прикладных аспектов компьютерной лингвистики и управления большими данными подчеркивается в контексте эффективного изучения и анализа таджикского языка. Реализованный комплекс компьютерных программ, доступный онлайн по адресу www.tajlingvo.tj, представляет собой важный инструмент для исследования и реализации задач компьютерной лингвистики в указанной области.

Диссертация Худойбердиева Х.А. представляет собой важный вклад в область компьютерной лингвистики, отличается своей научной новизной и практической применимостью, обеспечивая новые пути развития данной области исследований.

Теоретические результаты диссертации имеют значение для развития кросс-культурных и национальных лингвистических интеллектуальных автоматизированных систем обработки гетерогенной информации, реализации образовательных программ по изучению таджикского языка и развитию моделей искусственного интеллекта в сфере обработки естественного языка и речи.

Научные и прикладные результаты диссертационной работы могут быть использованы:

- при проведении дальнейших исследований в области фундаментальных проблем применения искусственного интеллекта и обработки больших данных в технических, промышленных, природных и социальных системах;

- разработке и анализе интеллектуальных информационно-сервисных систем управления;

- для мониторинга и анализа информационного пространства, в том числе открытых источников информации;

- для работы с большими массивами неоднородных данных на различных предприятиях.

- для решения задач в рамках разработки новых типов интерфейсов человека-машинного взаимодействия на естественном языке, систем поиска и обработки информации.

Предложенные выводы и рекомендации, сделанные в автореферате, обоснованы благодаря ее основательной научной подготовке и заслуживают высокой оценки за их значимость для развития компьютерной лингвистики.

В качестве *недостатков* можно отметить, что:

1. Проведен недостаточно полный теоретический анализ точности и вычислительной сложности выбранных методов обработки разнородной информации с целью автоматизации процесса обработки данных на таджикском языке по сравнению с другими методами, используемыми для решения подобного класса задач, в том числе применимыми к другим языкам (русский, английский, персидский).

2. Разработанные методы и модели имеют ценность для автоматической обработки информации на таджикском языке (автоматическая проверка правописания, задачи машинного перевода и синтеза речи), но тестирование разработанной системы и оценка эффективности проведены недостаточно.

3. В автореферате приводятся некоторые неточные термины, например: «.....были использованы методы *систематического анализа*», а правильное – «..были использованы методы *системного анализа*».

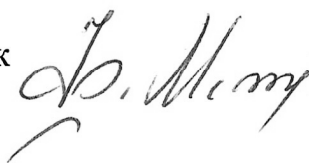
Обозначенные замечания не снижают общую ценность и уровень исследовательской работы.

Исследования Худойбердыева Х.А. представляют собой законченную научную работу с глубокой технической, технологической, математической базой и достаточной практической ценностью. Следовательно, считаю, что диссертация удовлетворяет требованиям, предъявляемым к докторским диссертациям по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Автор Худойбердиев Хуршед Атохонович, заслуживает присуждения ученой степени доктора технических наук.

На включение персональных данных, содержащихся в отзыве, в документы, связанные с работой Диссертационного совета, и их дальнейшую обработку согласен.

Главный научный сотрудник,
заведующий лабораторией №17 «Эргатические системы»
Института проблем управления им. В.А. Трапезникова РАН

доктор технических наук



Фархадов Маис Паша оглы

Фархадов Маис Паша Оглы, главный научный сотрудник, заведующий лабораторией №17 «Эргатических систем», доктор технических наук (специальность 05.13.15 - Вычислительные машины, комплексы и компьютерные сети), научное звание: старший научный сотрудник; email: mais@ipu.ru; телефон: +7 495 198-17-20, доб. 1486.

Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В. А. Трапезникова Российской академии наук; адрес: 117997 Москва, Профсоюзная 65; телефон: +7 495 334-89-10; факс: +7 495 334-93-40; email: dan@ipu.ru; сайт: <https://www.ipu.ru/>.

Подпись

Фархадов

ЗАВЕРЯЮ

ВЕД. ИНЖЕНЕР

ГОРДЕЕВА Ю.Ю.

Гордеева



ТАҚРИЗ

ба автореферати рисолаи илмии Худойбердиев Хуршед Атохонович дар мавзӯи: «Балоихагирӣ ва амалигардонии низомҳои худкори коркарди маълумот бо забони тоҷикӣ» барои гирифтани унвони илмии доктори илмҳои техникӣ аз рӯи ихтисоси 05.13.11 – «Таъминоти математикӣ ва барномавии мошинҳои ҳисоббарор, муҷтамаъҳо ва шабакаҳои компютерӣ»

Тадқиқ ва омӯзиши забоншиносии компютерӣ, ҳалли масъалаҳои амсиласозии математикӣ, тартиб додани усулҳо ва воситаҳои ҳисоббарор аҳамияти хеле муҳими назариявӣ амалӣ доранд. Дар ҷаҳони муосир соҳаи таҳқиқот ва амсиласозии математикӣ бо мақсади коркарди маълумоти электронӣ ба забони тоҷикӣ ва дар барномаҳои компютерӣ татбиқ намудани он вазифаи ниҳоят муҳим мебошад. Ташаккули низомҳои коркарди худкори матн бо забони табиӣ ба маҷмӯи барномаҳо асос ёфтааст. Низомҳои худкори коркарди маълумот ба забони тоҷикӣ имкон медиҳад, ки як қатор вазифаҳои ҷорӣ, аз қабилӣ тафтиши имлои матн, синтези нутқ, тарҷумаи компютериӣ матн аз забони тоҷикӣ ба русӣ ва англисӣ амалӣ карда шаванд.

Рисолаи доктории Худойбердиев Хуршед Атохонович маҳз ба ҳамин самти илмӣ-тадқиқотӣ равона карда шудааст, ки ҷараёни дақиқи математикиро ба амсиласозии равандҳои коркарди иттилоот ба забони тоҷикӣ, коркарди муҷтамаъи барномаҳои компютериӣ амалӣ оид ба автоматикунонии технологияҳо забони тоҷикӣ имконият медиҳанд.

Муаллиф барои тадқиқ ва коркарди ахборот бо забони тоҷикӣ, инчунин барои коркарди матн ва нутқ маҷмӯи барномаҳои компютериӣ TajLINGVO коркард ва пешниҳод намудааст, ки аз нуқтаи назари рушди забоншиносии ҳисоббарор ва самаранок кардани раванди ҳуҷҷатгузорӣ дар ташкилоту корхонаҳо хеле муҳим арзёбӣ карда мешавад. Лоихаҳои коркардшуда дар шабакаи интернет дар суроғаи www.tajlingvo.tj дастрас мебошанд.

Аз ҷиҳати назариявӣ натиҷаҳои асосии рисола дар се монографияи илмӣ, шумораи зиёди мақолаҳои илмӣ дар маҷаллаҳои бонуфуз ва 18 шаходатномаи бақайдгирии маҳсулоти барномавӣ ва захираҳои иттилоотӣ инъикос гардидаанд. Натиҷаҳои амалии рисолаи илмӣ дар якҷанд муассисаҳои илмӣ ва таълимӣ мавриди истифода қарор ёфтаанд, аз қабилӣ Маркази илмии АМИТ дар шаҳри Хучанд, Муассисаи давлатии таълимии Донишгоҳи давлатии Хучанд ба номи академик Б.Ғафуров, Донишгоҳи давлатии ҳуқуқ, бизнес ва сиёсати Тоҷикистон, Донишкадаи политехникии Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ дар шаҳри Хучанд.

Мувофиқи маълумоти дар автореферат овардашуда, рисолаи илмӣ дар 328 саҳифаи матни компютерӣ пешниҳод гардида, аз муқаддима, шаш боби мундариҷаи асосӣ, натиҷа ва пешниҳоди амалӣ оид ба ҳар як боб, хулоса, рӯйхати адабиёт иборат аз 322 номгӯ ва 2 замима иборат аст.

Дар охири автореферат натиҷаҳои асосии тадқиқот пешниҳод карда мешаванд. Барои дар амал татбиқ намудани натиҷаҳо пешниҳодҳои назариявӣ ва амалӣ дода мешаванд.

Бояд гуфт, ки арзиши илмии кори рисолаи Худойбердиев Х.А. аз маҷмӯи амсилаҳои математикӣ ва барномавӣ барои ҳалли масъалаҳои забоншиносии компютерӣ, истифодаи моҳиронаи онҳо дар коркарди комплекси барномаҳои компютерӣ барои тафтиши худкори имло дар матн бо забони тоҷикӣ, тарҷумаи мошинӣ ва синтези нутқ арзиши махсус дорад.

Мо бовар дорем, ки дар доираи иҷрои рисолаи илмӣ дар мавзӯи «Балоихагирӣ ва амалигардонии низомҳои худкори коркарди маълумот бо забони тоҷикӣ» муаллиф якҷанд масъалаҳои илмию амалӣ дар соҳаи забоншиносии компютерӣ, низомҳои иттилоотӣ ва коркарди маълумот бо забони тоҷикӣ ҳалли худро бомуваффақият ёфтаанд. Дар умум чунин мешуморем, ки рисолаи илмӣ ба талаботи рисолаи докторӣ комилаан ҷавобгӯ буда, Худойбердиев Х.А. сазовори гирифтани унвони илмии доктори илмҳои техникӣ аз рӯи ихтисоси 05.13.11 – «Таъминоти математикӣ ва барномавии мошинҳои ҳисоббарор, муҷтамаъҳо ва шабакаҳои компютерӣ» мебошад.

Профессори кафедраи таҳлили математикӣ
ва муодилаҳои дифференсиалии
Донишгоҳи давлатии Бохтар,
доктори илмҳои физикаю математика

Шамсудинов Ф.М.

Мудирӣ кафедраи технологияи
иттилоотӣ ва методикаи таълими
информатика
Донишгоҳи давлатии Бохтар,
номзади илмҳои физика ва математика

Махмадалиев Х.С.

Имзои Ф.М. Шамсудинов ва Х.С. Махмадалиев-ро
тасдиқ мекунам.

Сардори шӯъбаи кадрҳо ва корҳои махсуси ДДБ
ба номи Носири Хусрав



Шукурзод Ҷ.А.