

## ОТЗЫВ

официального оппонента доктора физико-математических наук, академика Национальной академии наук Республики Таджикистан Илолова Мамадшо Илоловича о диссертационной работе 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».

*Актуальность темы исследования и ее связь с общенаучными и общегосударственными программами.* Исследование компьютерной лингвистики, решение актуальных задач на основе математических моделей, разработка компьютерных моделей и средств имеет очень важное как теоретическое, так и практическое значение. В современном мире сфера исследования и математического моделирования в целях автоматической обработки информации на таджикском языке и ее реализация в компьютерные прикладные программы является чрезвычайно важной задачей. Разработка систем автоматической обработки текста на естественном языке основана на комплексе компьютерных программ и приложений, которые базируются на математических моделях. Системы автоматической обработки информации на таджикском языке позволяют реализовать ряд актуальных задач, таких как автоматическая проверка орфографии текста, синтез речи, голосовое управление устройств, а также автоматический машинный перевод текста с таджикского языка на русский и английские языки.

Следует отметить, что компьютерная лингвистика таджикского языка представляет собой весьма новую сферу науки, поэтому не удивительно, что задачам математического и компьютерного моделирования в сфере компьютерной лингвистики посвящено значительное число научных работ. Математические модели помогают разработать оптимальную стратегию обработки информации на основе компьютерного моделирования и позволяют решать важные задачи для автоматизации таджикского языка, в частности проверка правописания, машинный перевод и синтез речи.

Научная основа процессов в компьютерной лингвистике таджикского языка была заложена в работах З.Д. Усманова, в которых впервые в виде математических моделей были объединены элементы текста, так и процессы их обработки. Крупным вкладом в развитие автоматизации таджикского языка и моделирования процессов обработки информации являются работы М.А.Исмаилова, С.А.Зарипова, О.М.Солиева Л.А.Гращенко, Г.М.Довудова, А.А.Косимова, К.С. Бахтеева, выполненные в отделе математического моделирования Института математики имени академика А. Джураева НАНТ.

Именно к этой, весьма актуальной, области исследований относится диссертационная работа Х.А. Худойбердиева, объединяющая строгие

математические подходы к моделированию процессов обработки информации на таджикском языке, создание пакетов прикладных программ, позволяющие реализовать актуальные задачи автоматизации таджикского языка.

Автор диссертации Х.А. Худойбердиева поставил перед собой актуальную научную цель - создание компьютерного комплекса для исследования и обработки информации на таджикском языке, а также для автоматического управления текстом и речью. Указанные проблемы, несомненно, весьма актуальны с точки зрения развития прикладных аспектов компьютерной лингвистики, а также управления и обработки больших массивов данных, так называемой BigData. Существенным, как полагаем мы, является развитие прикладных аспектов данной проблемы в приложении к изучению таджикского языка для формирования общего представления об элементах текста и речи с целью формирования их цифровых портретов. В работе реализован комплекс компьютерных программ для целей исследования и реализации задач компьютерной лингвистики в таджикском языке, который доступен в сети Интернет по адресу [www.tajlingvo.tj](http://www.tajlingvo.tj).

**Объект диссертационной работы** – компьютерное моделирование вычислительных процессов и проектирования программных обеспечений для системы автоматической обработки информации на таджикском языке.

**Предмет диссертационной работы** – методы, модели и алгоритмы обработки информации на таджикском языке для проектирования и реализации электронных словарей, синтеза речи, автоматической проверки орфографии и компьютерного перевода.

**Оценка обоснованности и достоверности.** Результаты, полученные в диссертационной работе, обоснованы предложенными математическими моделями элементов текстовой информации с целью последующей обработки. В свою очередь достоверность спроектированных автоматических систем и компьютерных модулей подтверждена корректным выбором исходных данных и репрезентативной выборки текстовой информации в формировании постановки задач разработки математических и компьютерных средств и их реализации в информационных системах для автоматической обработки информации на таджикском языке. Необходимо отметить, что полученные научные результаты, сформулированные положения, выводы и рекомендации диссертационной работы обеспечиваются обоснованностью методологии системного подхода исследования с привлечением фактических материалов и его статистическим анализом, научной концепцией компьютерного проектирования, опорой на теоретические и практические методы исследования, анализом объектов исследования, результатами тестовых экспериментов, практическим утверждением, применением множества методов, адекватных задачам

исследования лингвистики, а также актов применения результатов исследования на практике.

**Научная новизна исследования.** В результате научно-исследовательской работы и разработки автоматических систем автором предложен ряд новых методических подходов к исследованию, анализу и автоматической обработке текстовой информации на таджикском языке. В частности:

- 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](http://tarjumon.tajlingvo.tj);

- на основе разработанных методов, моделей и структур данных предложены новые алгоритмы машинного перевода, сформированы компьютерные параллельные корпуса Tajik-Russian-Parallel Corpus и Tajik-English-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.

Все полученные результаты реализованы в программном комплексе TajLINGVO, который позволяет:

- существенно сократить время изучения таджикского языка как для пользователей в Республике Таджикистан, так и за рубежом;

- повысить уровень обоснованности принимаемых решений по компьютерной лингвистике и задачам таджикского языка;

- обеспечить формирование и использование корректного контента на таджикском языке в сети Интернет.

**Содержание диссертации.** Представленная на защиту диссертационная работа изложена на 328 страницах компьютерного текста и состоит из введения, шести глав основного содержания, выводов по каждой главе, заключения, списка использованной литературы, включающей 322 наименования и двух приложений.

Во введении дается обоснование актуальности темы диссертационной работы, описана степень научной разработанности проблемы и методологическая основа исследования. Определены цели и задачи исследования, указаны научная новизна, теоретическая и практическая значимость полученных результатов, а также сформулированы основные положения диссертации, выносимые на защиту.

В первой главе рассмотрены вопросы, связанные с определением существенных особенностей проектирования и проблем реализации систем автоматической обработки информации на таджикском языке.

Во второй главе дается описание методов автоматической обработки информации на естественном языке, а также их решения с использованием математического моделирования, методов анализа и синтеза информации на таджикском языке.

В третьей главе исследуется компьютерное моделирование разработки систем автоматической обработки информации на естественном языке с учетом объектно-ориентированного подхода с применением языка UML. В этом разделе представлен разработанный автором проект информационной системы обработки текстовых данных на таджикском языке с учетом модели действия, взаимодействия, структуры и реальности.

Четвертая глава посвящена вопросам проектирования, обработки и реализации системы автоматической проверки орфографии текста на таджикском языке.

В пятой главе дается описание проектирования, разработки и внедрения таджикского автоматического переводчика. Здесь также рассмотрены задачи реализации машинного перевода текста с таджикского языка на русский.

Шестая глава посвящена задачам математического моделирования и компьютерной реализации синтеза таджикской речи на основе предложенного текста. В данной главе детально рассмотрены методы конкатенации фонем на примере слогов таджикского языка, предложены подходы и методы последующего проектирования систем распознавания речи на таджикском языке.

В заключении диссертации сформулированы основные результаты исследования, даны предложения по практическому внедрению полученных в диссертационной работе результатов. В приложении приведены свидетельство о регистрации программного продукта, выданного Национальным патентным центром Министерства экономического развития и торговли Республики Таджикистан, акты внедрения результатов диссертационной работы.

#### ***Степень достоверности и апробация результатов исследования.***

Основные результаты диссертации опубликованы в трех монографиях, 50 статьях, 25 из которых в рецензируемых журналах, рекомендованных ВАК при Президенте Республики Таджикистан и ВАК Российской Федерации для публикации основных материалов докторских диссертаций.

Получено 18 свидетельств регистрации программного продукта. Полученные в диссертационной работе результаты внедрены в отдельные организации, такие как Худжандский научный центр НАНТ, Управление по инвестициям и управлению государственным имуществом Согдийской области, ГОУ Худжандский государственный университет имени академика Б. Гафурова, кафедра таджикского языка Таджикского государственного университета права, бизнеса и политики, Политехнический институт Таджикского технического университет имени академика М.С. Осими в городе Худжанде, ЗАО «Душанбе Сити Банк».

*Степень обоснованности и достоверности результатов, заключений и рекомендаций, сформулированных в диссертации*

Научная ценность диссертационной работы Худойбердиева Х. А. обусловлена комплексным моделированием задач компьютерной лингвистики к изучению элементов текстовой информации на таджикском языке, умелом использовании математического и компьютерного обеспечения, что позволило разработать эффективный компьютерный комплекс для автоматизации задач обработки информации на таджикском языке. Особую ценность работе, как представляется нам, придает возможность реализовать полученные данные в задачах автоматической проверки правописания, машинного перевода и синтеза речи по тексту.

*Недостатки по содержанию и оформлению диссертации.*

Проведенный нами анализ диссертационной работы, наряду с достоинствами, которые являются бесспорными, выделил и некоторые недостатки, к числу которых можно отнести следующие:

1. В диссертации предложены методы и алгоритмы практического, структурного и объектно-ориентированного проектирования систем автоматической обработки данных. Какое именно развитие получили теории программирования, создания и сопровождения программных средств, из приведенной формулировки неясно. Дальнейшая отсылка к полученным результатам, скорее, относится к практической, а не теоретической значимости.

2. Предложены новые алгоритмы машинного перевода. В чем конкретно состоит новизна алгоритмов автоматического перевода по правилам, перевод на основе машинного обучения и статистический машинный перевод, не сформулированы. Видимо, предлагается считать новую программную реализацию алгоритмов в информационных системах.

3. Избыточную детализацию выводов по главам, в то время как общие выводы, представленные в заключении диссертации, были даны конспективно. Для улучшения структуры и понимания работы рекомендуется более подробное изложение общих выводов в конце каждой главы и обоснованные основные положения, выносимые на защиту, во введении. Улучшение структуры и содержания выводов позволит более

четко выделить суть и значимость проделанной работы, сделав ее более полной и информативной для читателя.

4. В шестой главе предлагаются ряд методов и алгоритмов синтеза речи на таджикском языке. По полученным результатам разработаны программные средства для автоматического озвучивания введенного текста в компьютер. А также, в разделе 6.5 исследованы проблемы распознавания речи на таджикском языке. в диссертации нет четких формулировок применения методов синтеза и распознавания речи для решения задачи автоматического голосового диалога человека с компьютером на таджикском языке.

5. По тексту диссертации имеются замечания редакционного характера.

Однако указанные замечания не снижают научной ценности полученных в работе результатов, а носят характер пожелания и вопросы уточняющий характер.

**Заключение.** Считаю, что в диссертационной работе Худойбердиева Хуршеда Атохоновича на тему «Проектирование и реализация автоматических систем обработки информации на таджикском языке» решены научная и практическая задача по весьма актуальной проблеме компьютерной лингвистики, автоматических систем и обработки информации на таджикском языке. Диссертация по актуальности, научной новизне и практической значимости, содержанию и объему, достоверности полученных результатов, обоснованности научных положений и заключений полностью отвечает требованиям положениям «Типовое положение о Диссертационных советах», «О присуждении ученой степени кандидата наук и доктора наук на диссертационном совете» ВАК при Президенте Республики Таджикистан, предъявляемым к докторским диссертациям, а сам автор заслуживает присуждения учёной степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Отзыв составлен доктором физико-математических наук, академиком Национальной академии наук Республики Таджикистан, заведующим отделом математического моделирования динамических процессов центра инновационного развития науки и новых технологий при НАНТ.

Официальный оппонент  
Илолов Мамадшо Илолович  
доктор физико-математических наук,  
академик Национальной Академии наук Таджикистан,

Подпись академика Илолова М.И. **Удостоверено**



## ОТЗЫВ

официального оппонента, доктора технических наук  
Пруцкова Александра Викторовича на диссертационную работу  
Худойбердиева Хуршеда Атохоновича на тему «Проектирование  
и реализация автоматических систем обработки информации  
на таджикском языке», представленную к защите на соискание  
ученой степени доктора технических наук по специальности  
05.13.11 – «Математическое и программное обеспечение вычислительных  
машин, комплексов и компьютерных сетей»

**1. Актуальность темы.** Информатизация является одним из видов автоматизации. Одним из проявлений информатизации является внедрение информационных систем в различные сферы нашей социально-экономической жизни. Информационные системы используют в качестве обмена данными в том числе текст, поэтому в этих системах требуется реализация функций по обработке текстов на естественном языке, синтеза речи.

Для таджикского языка эти функции не были разработаны так, чтобы давать удовлетворительное решение. Считаю тему диссертационного исследования, посвященную разработке новых моделей и методов обработки текстов на таджикском языке, синтеза речи, актуальной.

**2. Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверность и новизна.** Научные положения, выводы и рекомендации, сформулированные в диссертации обосновываются практической реализацией предложенных моделей, методов и алгоритмов. Достоверность полученных в диссертации результатов подтверждается корректным применением методов компьютерной лингвистики, разработанным программным обеспечением, апробацией и обсуждением результатов на научных конференциях, рецензированием и экспертизой научных статей, опубликованных в научных журналах.

При решении поставленных задач диссертационного исследования автором были заявлены результаты, претендующие на научную новизну.

"- предложены новые научно-технические положения, математические модели, методы и структуры данных, которые в целом составляют теоретическую основу системного анализа и исследования текстовой информации;"

Это положение относится ко всей диссертации. В диссертации разработаны методы и алгоритмы объектно-ориентированного проектирования систем автоматической обработки данных, предложены новые алгоритмы машинного перевода, разработаны модели, методы синтеза речи и компьютерные программы, предложено использовать бинарное дерево для хранения словаря, хэш-таблица для поиска в словаре.

"- впервые разработаны методы и алгоритмы практического, структурного и объектно-ориентированного проектирования систем автоматической обработки данных;"

В главе 3 проектируется информационная система автоматической обработки текста с использованием языка UML. Методы и алгоритмы в явном виде не представлены.

"- 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](http://tarjumon.tajlingvo.tj);"

В главах 4-6 подробно описано проектирование перечисленных в положении программных средств. Рассмотрены особенности предметной области, обоснован выбор структур данных, спроектирована структура программных систем, приведены некоторые алгоритмы работы.

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

В главе 5 спроектирована системы машинного перевода для пар языков «русский–таджикский», «таджикский–английский». Система разработана на основе статистического подхода к машинному переводу и подхода на основе правил.

"- 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."

В главе 6 исследован слоговый строй слов таджикского языка, проанализированы правила фонетики. Разработаны алгоритмы синтеза речи. Результаты исследования реализованы программно.

Все перечисленные положения обладают научной новизной и имеют отличия от известных теоретических положений.

**3. Значимость для науки и практики результатов диссертации и конкретные пути их использования.** Теоретическая значимость диссертационного исследования заключается в разработке методов и алгоритмов для проверки орфографии, машинного перевода и синтеза речи на таджикском языке.

Практическая значимость работы состоит в разработке программ для проверки орфографии, машинного перевода и синтеза речи на таджикском языке. Предложенный подход может применяться на практике информационных системах машинного перевода, проверки орфографии и синтеза речи текстов на таджикском языке во всех сферах национальной экономики.



#### **4. Оценка содержания диссертации, её завершенность.**

Диссертационная работа состоит из введения, общего описания исследования, шести глав, обсуждения полученных результатов, заключения, выводов, рекомендаций по практическому использованию результатов исследования, списка литературы из 322 наименований и двух приложений. Диссертация изложена на 328 страницах, содержит 15 рисунков и 19 таблиц.

Диссертация имеет завершенный характер, качество оформления соответствует предъявляемым требованиям.

Во введении рассматривается актуальность темы исследования, степень научной разработанности изучаемой проблемы, связь работы с научными программами (проектами), темами.

В первой главе перечислены известные подходы к проверке орфографии, машинному переводу и синтезу речи, проанализированы работы по автоматической обработке текстов на таджикском языке, описана программная система TajLINGVO.

Во второй главе проанализированы этапы обработки текстов на естественных языках, перечислены математические модели обработки текстов (в том числе работы З.Д. Усманова), выполнен обзор методов проверки правописания текста, машинного перевода, синтеза речи.

В третьей главе рассмотрено моделирование информационных процессов, моделирование информационной системы обработки текста, спроектированы концептуальная модель и структура классов этой информационной системы.

Четвертая глава посвящена разработке методов и программного обеспечения проверки орфографии текстов на таджикском языке. Здесь же рассмотрены электронный словарь и тезаурус таджикского языка. Выявлены особенности автоматической системы проверки орфографии на таджикском языке. Предложен алгоритм проверки орфографии с описанием используемых структур данных. Спроектирована программная система проверки орфографии для таджикского языка.

В пятой главе рассмотрены современные системы машинного перевода. Предложены метод и программная система транслитерации таджикского, русского и английского языков. Спроектирована информационная система для машинного перевода русского и таджикского языков с использованием модели нейронного машинного перевода.

Шестая глава посвящена проектированию алгоритмов и программного обеспечения синтеза речи по тексту на таджикском языке.

**5. Подтверждение опубликования основных результатов диссертации в научных изданиях.** Автор по теме диссертации имеет 68 научных работ, из них 25 статей опубликованы в изданиях, рекомендованных ВАК РТ. В этих работах содержатся теоретические и экспериментальные результаты диссертационного исследования.

**6. Недостатки по содержанию и оформлению работы.** Выявлены следующие недостатки диссертационной работы:

1. Первая и вторая главы содержат общеизвестные сведения. Общий объем этих глав составляет почти 30% от суммарного объема всех глав диссертации, что очень много.

2. На мой взгляд, в докторской диссертации должны быть представлены обобщения моделей, методов и алгоритмов, которые могут быть применены при решении различных однотипных задач. В рассматриваемой диссертации представлены частные модели, методы и алгоритмы для решения одной задачи и какие-то обобщения отсутствуют. Этот недостаток снижает научную ценность диссертации.

Указанные замечания не снижают практическую значимость диссертационной работы.

**7. Соответствие автореферата основному содержанию диссертации.** Автореферат диссертации достаточно полно отражает её содержание.

**8. Заключение о соответствии диссертации критериям, установленным «Положением о присуждении учёных степеней».** На основании вышесказанного считаю, что диссертационное исследование Худойбердиева Х.А. соответствует всем требованиям ВАК при Президенте Республики Таджикистан, предъявляемым к диссертациям, представленным к защите на соискание учёной степени доктора технических наук, а её автор Худойбердиев Х.А. заслуживает присуждения ему учёной степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент

Профессор кафедры вычислительной и прикладной математики Федерального государственного бюджетного образовательного учреждения высшего образования «Рязанский государственный радиотехнический университет имени В. Ф. Уткина» (РГРТУ), доктор технических наук (специальность 05.13.11), доцент



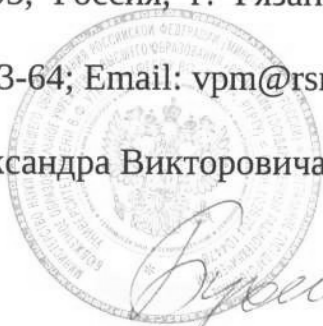
Пруцков Александр Викторович

Почтовый адрес: 390005, Россия, г. Рязань, ул. Гагарина, д. 59/1, РГРТУ, кафедра ВПМ

Телефон: +7 (4912) 72-03-64; Email: vpm@rsreu.ru

Подпись Пруцкова Александра Викторовича удостоверяю:

Ученый секретарь  
ученого совета РГРТУ



К.В. Бухенский

## ОТЗЫВ

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

**1. Актуальность темы.** Диссертация Худойбердиева Хуршеда Атохоновича посвящена проектированию и реализации автоматических систем обработки информации на таджикском языке. С ростом объема информации, электронных документов, системам их обработки требуются все новые методы повышения эффективности использования компьютерных средств и программных обеспечений в процессе решения задач компьютерной лингвистики. Наиболее из оптимальных способов повышения производительности и обработки являются применения математических моделей и информационных систем с вычислительным аппаратом различной архитектуры согласно цифровому портрету текстовой информации. В последние годы было создано множество специализированных инструментов обработки информации на естественном языке, которые обеспечивают механизм реализации синтаксического, морфологического и семантического анализа в соответствии правил конкретного языка. Изучение и использование естественного языка в процессе формирования электронного документа и цифрового контента становятся более доступны с помощью таких инструментов.

Объект и предмет диссертационного исследования, в частности таджикский язык является одним из важных языков национального состава, особенно в регионе Центральной Азии. Развитие автоматических систем обработки информации на этом языке имеет большое значение для поддержки информационной инфраструктуры и коммуникации в странах, где он используется. С развитием информационных технологий все больше возрастает потребность в адаптации этих технологий к различным языкам, включая редкие и малораспространенные. Проектирование автоматических

систем обработки информации на таджикском языке позволит расширить доступность информационных ресурсов для носителей этого языка.

Многие страны, где используется таджикский язык, являются мультязычными, и обеспечение поддержки этого языка в автоматических системах обработки информации поможет улучшить коммуникацию и взаимодействие между различными языковыми группами.

В сравнении с более распространенными языками, ресурсы и инструменты для обработки текстов на таджикском языке часто оказываются ограниченными. Разработка автоматических систем обработки информации поможет заполнить этот пробел и сделать информацию на таджикском языке более доступной и удобной для использования.

Проведенный анализ существующих методов использования информационных систем с целью последующей обработки информации в диссертационной работе Худойбердиева Х.А. показывает актуальность задач автоматизации проверки правописания текста, машинного перевода и синтеза речи для повышения качественного использования таджикского языка.

Диссертационная работа на тему "Проектирование и реализация автоматических систем обработки информации на таджикском языке" представляет собой актуальное исследование, которое вносит значительный вклад в развитие технологий обработки естественного языка. Автор исследования провел обширный анализ существующих методов и подходов к автоматической обработке информации на таджикском языке, выявил их преимущества и недостатки, и предложил собственное решение задачи.

В работе четко структурированы методы улучшения качества обработки информации на таджикском языке, разработаны и реализованы соответствующие алгоритмы и программное обеспечение. Проведенные эксперименты и оценка результатов показывают, что предложенное решение эффективно и значительно улучшает процесс обработки информации на таджикском языке.

Диссертационная работа является важным вкладом в область автоматической обработки информации на малоизученных языках и может быть использована как основа для дальнейших исследований в этой области. Автор заслуживает похвалы за проделанную работу и достигнутые результаты.

Таким образом, тема и цель исследования, а также поставленные задачи в диссертационной работе Худойбердиева Х.А. являются актуальными и отвечают реальным потребностям области компьютерной лингвистики на сегодняшний день.

**2. Степень обоснованности результатов, сформулированных в диссертации, достоверность и новизна.** В диссертационной работе, все полученные результаты обоснованы предложенными математическими моделями, которые реализованы для обработки элементов текстовой информации. Достоверность автоматических систем и компьютерных модулей можно подтвердить корректным выбором исходных данных. В результате исследовательской работы, проектирования и реализации автоматических систем предложен ряд методических подходов к исследованию, анализу и автоматической обработке текстовой информации на таджикском языке. Все полученные результаты реализованы в программном комплексе TajLINGVO, который позволяет сократить время изучения таджикского языка как для пользователей в Республике Таджикистан, так и за рубежом, увеличить степень обоснованности принимаемых решений по компьютерной лингвистике и задачам таджикского языка, а также обеспечить качественный и корректный контент в сети Интернет на таджикском языке.

**3. Теоретическая и практическая значимость и внедрение результатов диссертации.** В диссертационной работе представлены совокупность математических моделей, методов и алгоритмов для реализации обработки элементов текста и звукового сигнала на естественном языке, способствующие изучению природы таджикского языка. В результате исследования апробированы и усовершенствованы системы программном комплексе TajLINGVO. Практическое значение и значимость основных положений исследования подтверждает опыт создания программных средств для реализации электронных словарей, электронного тезауруса, автоматического синтеза речи, проверки орфографии, автоматического перевода. Полученные теоретические результаты нашли свои применения в научных центрах и учебных заведениях, а также программные комплексы практически используются в процессе делопроизводства организаций.

**4. Оценка содержания диссертации, её завершенность.**

Диссертационная работа Худойбердиева Х.А. состоит из введения, шести глав, заключения и приложения.

Во введении сформулированы цели и задачи диссертации, обоснована их актуальность, показаны новизна, теоретическая и практическая ценность результатов, приведены основные положения, выносимые на защиту.

В первой главе диссертации определены понятия задачи реализации компьютерной лингвистики в таджикском языке. Описаны основные механизмы для оптимизации и эффективного выполнения результатов исследований компьютерной лингвистики в таджикском языке. Приведен подробный обзор научных результатов как в Республики Таджикистан, так и за рубежом на тему диссертации, что существующие исследования имеют ряд значительных недостатков. Они порождают ряд важных и актуальных задач, решённых в диссертационной работе. Дается формальная постановка задачи реализации автоматических систем обработки информации на таджикском языке. Анализированы существующие методы с точки зрения их применения в информационных системах для повышения эффективности использования компьютерных средств и обработки информации на естественном языке.

Во второй главе диссертации представлены методология и алгоритмы необходимы для решения задачи обработки информации на естественном языке. Соискателем разработаны методы к построению математических моделей реализации задач автоматической проверки правописания, машинного перевода текста и синтеза речи. Для управления процессами обработки информации разработаны математический аппарат и новая модель на основе элементов текста и их цифровых портретов. Элементы текста представляют собой реализации структур данных и алгоритмов и позволяют формировать цифровые портреты текстовой информации. Предложены алгоритмы автоматической проверки правописания, машинного перевода и синтеза речи текстовой информации на естественном языке.

В третьей главе представлена реализация информационной системы на основе современных технологий проектирования и объектно-ориентированного программирования. Описана модель поведения, общая архитектура и физическая модель информационной системы TajLINGVO. Продемонстрированы возможности языка UML и CASE-инструментов, таких

как MS Visual Studio .Net, IBM Rational Rose, Enterprise Architect для компьютерного моделирования программного обеспечения. Эксперименты показывают, что предложенные алгоритмы и методология позволяют ускорять процесс разработки информационных систем.

В четвертой главе рассмотрены результаты разработки и реализации компьютерного тезауруса таджикского языка. На основе интерактивного электронного словаря MultiGANJ создано описание лингвистического тезауруса таджикского языка, насчитывающего более 150 000 словарных единиц. С учетом того, что на практике часто используются нестандартные шрифты, предложен алгоритм преобразования символов текста в государственный стандарт раскладки букв таджикского алфавита. Предлагается собственный алгоритм проверки орфографии на таджикском языке адаптирован к кодировке символов таджикского алфавита. На основе полученных результатов разработан модуль TajSpell с возможностью исправления таджикского текста, проверкой орфографии, расстановки переноса слов и тезаурусом таджикского языка. Таким образом можно проверять таджикские тексты, написанные в стандартной кодировке UNICODE.

В пятой главе предложены результаты исследования в рамках проектирования и реализации задачи машинного перевода текста с таджикского языка на другие языки. На основе научных исследований и анализа мировых проектов анализированы методы и алгоритмы машинного перевода для перевода текста, выявлены математические модели и методы, используемые в популярной системе машинного перевода онлайн-переводчика Google. Изучены природа возникновения ошибок в процессе машинного перевода, как орфографические, так и стилистические. Следовательно, предлагаются ряд методов и алгоритмы исправления этих ошибок в тексте на таджикском языке. С целью информационного обеспечения системы машинного перевода для были разработана параллельные корпуса Taj-Rus-Corpus и Taj-Eng-Corpus. В результате полученных результатов, для обеспечения машинного перевода текста на таджикский язык предлагается информационная система в виде Web-приложения. Проект доступен в Интернете по адресу

[www.tarjumon.tajlingvo.tj](http://www.tarjumon.tajlingvo.tj) для онлайн-перевода текстовой информации с таджикского языка на русский и английский языки и наоборот.

В шестой главе выявлены статистические закономерности текстовых данных в таджикском языке, выявлено всего 274 различных структуры слов (элемент, гласная - 1, согласная - 0) в объеме 1724472 слов. Установлено, что 8 единиц («01», «010», «01010», «01001», «10», «0101», «010101», «100101») покрывают 50%, а 23 элементов покрывают 75% таджикских текстов. На основе распределения структур слов выявлено всего 9 различных сочетаний слогов в таджикском языке, 6 из которых соответствуют правилам таджикского языка: «1», «10», «01», «010», «100», «0100». Согласно полученным структурам предложен алгоритм разбиения слов на слоги. На основе разработанного алгоритма проведен статистический анализ многообразия слогов таджикского языка. На этапе обработки цифровой структуры слогов разработан базы данных слог-звук, состоящий из  $2 \times 3259$  таджикских слогов, который реализован в автоматической системе синтеза речи на таджикском языке Tajik text-to-speech и синтезатори OS Windows в качестве диктора «Tajik Text Narrator».

В заключении резюмируются основные результаты и выводы диссертационной работы, а также рекомендации для использования результатов исследований.

#### **5. Подтверждение опубликования основных результатов диссертации в научных изданиях.**

Диссертация написана хорошим языком, логично структурирована и соответствует паспорту специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Полнота изложенного материала в диссертационной работе позволяет понять основные идеи соискателя и опубликованы 68 научных работ в научных журналах, международных конференциях и семинарах, а также получены свидетельства о регистрации интеллектуальных продуктов и информационных ресурсов.

#### **6. Недостатки по содержанию и оформлению работы.**

По содержанию работы имеются следующие замечания:

1. Необходимо обратить внимание на корректность использования определенных аспектов/методов обработки информации на таджикском



языке, которые не в полной мере охватывают полный спектр проблем и возможностей в этой области; на глубокий анализ полученных результатов и сравнение с существующими методами и технологиями в области обработки информации на таджикском языке;

2. Рекомендуется обратить внимание на то, что обзор литературы должен быть более глубоким и широким, включая как классические, так и современные исследования в области автоматической обработки естественных языков, а также специфические работы по обработке таджикского языка и провести сравнительный анализ существующих решений в области обработки информации на таджикском языке и выявить преимущества и недостатки разработанной системы;
3. Важно также уделить внимание адекватности и эффективности выбранных методов и алгоритмов обработки информации на таджикском языке, чтобы гарантировать высокое качество и точность результатов;
4. Следует также описать технические характеристики и требования к реализации автоматических систем обработки информации на таджикском языке для обеспечения их успешной работы и использования;
5. Рекомендуется уделить внимание анализу и оценке эффективности и результативности автоматических систем обработки информации на таджикском языке; предоставить более подробные результаты экспериментальных исследований, включая описание используемых наборов данных, методов оценки и критериев успеха. Следует, обратить внимание на необходимость проведения адекватной оценки качества и производительности разработанной системы;
6. В заключительной части работы стоит обратить внимание на возможные направления дальнейших исследований и улучшений в области проектирования и реализации автоматических систем обработки информации на таджикском языке.

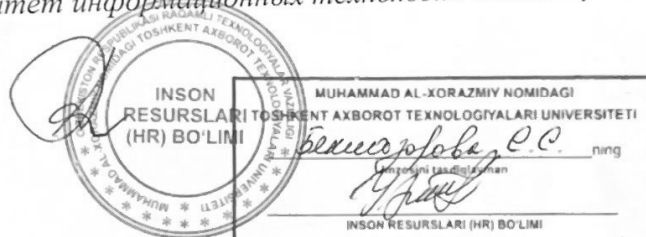
Следует отметить, что указанные недостатки, в целом, не снижают уровня диссертации. Эти замечания не являются критичными и не уменьшают значимость и важность диссертационной работы и полученных результатов.

**7. Соответствие автореферата основному содержанию диссертации.**

Текст автореферата на диссертацию полностью отражает содержание диссертации и написан в соответствии существующих требований.

**8. Заключение.** Диссертация Худойбердиева Хуршеда Атохоновича является законченной научно-квалификационной работой, выполненной автором самостоятельно и на высоком уровне. Полученные автором результаты являются новыми, обоснованными и достоверными. Работа отвечает требованиям «Положения о присуждении ученых степеней кандидата наук, доктора наук ВАК при Президенте Республики Таджикистан», предъявляемым к диссертациям на соискание ученой степени доктора технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

*Бекназарова Саида Сафибуллаевна, доктор технических наук, профессор кафедры «Телевизионные и медиа технологии», Ташкентский университет информационных технологий имени Мухаммада ал-Хоразмий.*



Почтовый адрес: 100200, Узбекистан, г. Ташкент, ул. А. Темура, 108, кафедра  
Телевизионные и медиа технологии  
Телефон: +9 (9890) 3276666  
Email: [saida.beknazarova@gmail.com](mailto:saida.beknazarova@gmail.com)

Подпись Бекназарова Саида Сафибуллаевна  
подтверждаю:

Должность утверждающего лица:

Ташкентский университет информационных технологий имени Мухаммада  
ал-Хоразмий

Ф.И.О.: Абулмусаввитов Р