

Моделирование системы автоматической обработки текста на таджикском языке

Х.А. Худойбердиев

Аннотация—Рассматриваются вопросы моделирования процессов в системе автоматической обработки текстовой информации на таджикском языке. Разработана методология формирования процессов обработки текстовой информации в системе TajLINGVO. Предложена логическая структура системы с подробным описанием каждого подпроцесса. Определена структура информационной модели системы, которая состоит из множества связанных между собой элементов текста на естественном языке. На основе языка моделирования UML предложена функциональная модель системы TAJLINGVO с подробным описанием диаграмм вариантов использования, активности и классов. Эффективность применения предлагаемых моделей подтвердилась в решении прикладных задач, таких как разработка компьютерного тезауруса, автоматическая проверка орфографии, синтеза речи и машинного перевода для таджикского языка. Полученные результаты и разработанные информационные системы доступны в сети Интернет по адресу www.tajlingvo.tj.

Ключевые слова — информационная система, обработка текста, автоматический процесс, комплекс программ, компьютерное моделирование, таджикский язык.

I. ВВЕДЕНИЕ

Системы автоматической обработки текста (АОТ) на естественном языке поддерживаются множеством пакетов программ и компьютерных приложений, работа которых основаны на математических моделях. Разработка автоматической системы проверки орфографии и её корректировка на базе существующих правил определенного языка, пакетов автоматического синтеза и распознавания устной речи, модулей голосового управления конечными автоматами, а также автоматических систем машинного перевода являются наиболее актуальными задачами в области компьютерной лингвистики.

В России наиболее заметные достижения в области компьютерной лингвистики с применением математического и компьютерного моделирования связаны с работами ученых, таких как Сбоев А.Г.,

Сулейманов Д.Ш., Кузнецов Л.А., Михайлов Д.В., Мурзин Ф.А., Хорошилов А.А., Воронина И.Е., Загоруйко Н.Г., Кривнова О.Ф., Жигулевцев Ю. Н., Евсеев А.И., Потапова Р.К., Прохоров Ю.Н., Галунов В.И. [1-5], Белоруссии – Цирульник Л.И., Лобанов Б.М., Карневская Е.Б., Людовик Т.В., Сажока Н.Н. и Украины – Анисимов А.В., Марченко А.А [6-7].

Наиболее оптимальные решения проблем, связанных с автоматической обработкой текстовой информации, нашли свое широкое применение в исследовательских работах ученых Большакова Е.И., Клышинского Э.С., Ландэ Д.В., Носкова А.А., Песковой О.В., Ягуновой Е.В., Белоногова Г.Г., Палагина А.В. В работах ученых предложены принципиально новые возможности для построения перспективных систем АОТ [8-11].

Разработка методологий, методов и базовых моделей создания автоматических систем обработки текстовой информации имеет многолетнюю историю. В работах ученых Сулейманова Д.Ш., Невзоровой О.А., Соснина П.И., Беляевой Л.Н., Лукашевича Н.В., Татевосова С.Г. предложены способы разработки базовых принципов, архитектуры, технологии построения прагматически-ориентированных лингвистических моделей, которые в последствии реализованы в информационных системах обработки текстов на естественном языке [12].

Наиболее совершенные системы обработки информации на естественном языке и программные продукты, например, такие как Языковой портал bab.la (Шретер А., Юкер П.), On-line словари Издательского Дома "ЭТС" (ETS Publishing House), Словари Ожегова и Зализняка (Старостин С.А.), Lexical FreeNet (Datamuse Corporation), WordNet (Cognitive Science Laboratory, Princeton University), Babylon.com (Babylon.com Ltd.), АBBYU Lingvo-11 (АBBYU Software House), Электронные словари ПРОМТ (ЗАО "ПРОект МТ"), Яндекс.Словари (Яндекс), Lexical FreeNet (Datamuse Corporation) предоставляют как в режиме офлайн, так и в режиме онлайн возможности мультязычного словаря отражающего все возможные толкования слов определенного языка с определением взаимосвязи между ними – в формате компьютерного тезауруса [13]. Функциональные возможности и информационное обеспечение перечисленных систем предоставляют пользователю эффективный процесс изучения основ естественного языка. Исходя из этого, возникает задача

Статья получена *** января 2023.

Худойбердиев Хуршед Атохонович, канд. физ.-мат. наук. Худжандский политехнический институт Таджикского технического университета имени академика М.С.Осими в г. Худжанде, доцент кафедры программирование и информационных систем, (tajlingvo@gmail.com)

моделирования и разработки систем автоматической обработки текстовой информации на таджикском языке.

Широкое использование информационно – коммуникационных технологий в Таджикистане предоставило возможность таджикским ученым – информатикам и лингвистам перейти в принципиально новую сферу науки – компьютерной лингвистики таджикского языка под руководством Академика НАН Таджикистан Усманова З.Д. Исходя из этого, появились актуальные задачи, связанные с моделированием простого нераспространенного предложения, созданием собственных драйверов таджикской графики и решением проблемы стандартизации печатной продукции, автоматическим синтезом таджикского текста, конверсией графических систем письма, автоматический морфологический анализ и распознаванием автора таджикских текстов [14-15].

За последние несколько лет в рамках проводимых научных исследований в области компьютерной лингвистики и продвижения Государственной стратегии развития информационных технологий в Республике Таджикистан достигнуты значительные результаты. Спроектированы и реализованы автоматические системы обработки информации на таджикском языке с целью формирования и получения электронных словарей, синтеза речи, автоматической проверки орфографии и компьютерного перевода текста.

II. МЕТОДОЛОГИЯ

Методология проектирования и архитектуры комплекса программ. Проектируемая модель системы TajLINGVO, состоит из совокупности взаимосвязанных информационных технологий, процессов, алгоритмов, множества текстовых элементов, интерфейсов и совокупности результатов необходимых для формирования цифрового портрета. Их можно представить следующим образом:

$$TajLINGVO = \{T, P, A, TE, I, R\} \quad (1)$$

где T – совокупность информационных технологий; P – комплекс процессов в TajLINGVO, $P_i, i=1..n$; A – комплекс алгоритмов, $A_j, j=1..m$ для реализации процессов $\{P_i\}$; TE – множество элементов текстовой информации, которые передаются для обработки с использованием алгоритмов $\{A_j\}$ в процессах $\{P_i\}$; I – интерфейсы пользователя для реализации ввода, обработки и вывода данных; R – результаты для передачи в обработку в процессах $\{P_i\}$.

III. ЛОГИЧЕСКАЯ СТРУКТУРА СИСТЕМЫ TAJLINGVO

Логическая структура информационных систем разрабатывается в соответствии с какой-либо методологией разработки программного обеспечения. Существует ряд современных методов и инструментов, которые предоставляют разработчикам возможность моделировать системы начиная от этапа анализа до

проектирования. Одним из таких инструментов является SADT (Structured Analysis and Design Technique) – это методология инженерии для разработки и описания систем в виде иерархии подсистем.

Структура системы TajLINGVO, предложенная в соответствии методологией SADT (рис. 1), состоит из четырех подсистем и представляет собой совокупность информационных ресурсов, алгоритмов и программных средств, управляющими процессами АОТ и интерфейсами пользователей. Подсистемы совместно реализуют совокупность алгоритмов автоматической обработки предложенных входных данных. Результаты обработки формируют совокупность текстовых элементов в соответствии с семантическими структурами, которые записываются в базы данных и интегрируются в интерфейс пользователя.

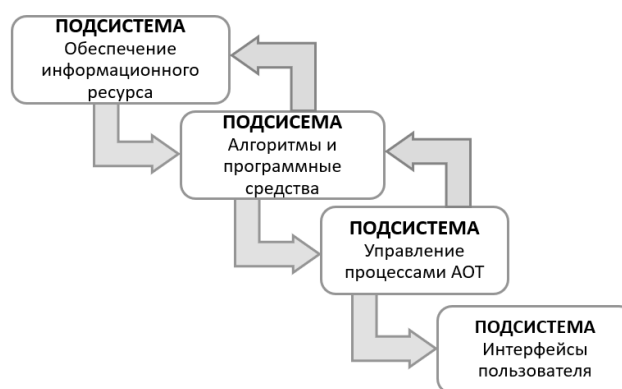


Рис. 1. Блок-схема логической структуры TajLINGVO

Подсистема «*Обеспечение информационного ресурса*» на основе репрезентативной выборки обеспечивает формирования лингвистического корпуса текстов, баз данных языковых и текстовых структур. Подсистема состоит из следующих компонентов: источники текстовой информации, различные базы данных, например, электронные словари, заранее определенные структуры элементов текста, которые является результатом реализации некоторого процесса АОТ.

Подсистема «*Алгоритмы и программные средства*» состоит из совокупности алгоритмов, реализованных в виде программных модулей, функций и процедур для обработки структур элементов текста. Программные средства предоставляют возможность пользователю осуществлять управление процессом АОТ.

Подсистема «*Управление процессами АОТ*» предоставляет предварительную подготовку результатов обработки входных данных. А также, реализуются процедуры контроля и проверки результатов с целью принятия решения пользователя. В случаях получения неоднородных результатов предоставляется возможность повторной обработки данных.

Подсистема «*Интерфейсы пользователя*» осуществляет возможность функционирования поиска,

фильтрации данных и записи результатов в базе данных. Также для наиболее наглядного просмотра результатов пользователю предоставляется возможность получить графические варианты отчетов в виде точечных графиков, диаграмм и гистограмм.

На основе полученной логической структуры для компьютерного моделирования системы TajLINGVO следует разработать функциональную модель системы, информационную модель процессов P и программных средств, реализующих комплекс алгоритмов A. В зависимости от эффективности результатов полученные данные передаются для автоматизации процесса обработки элементов текста, в частности, разработки компьютерного тезауруса, проверки орфографии, синтеза речи и машинного перевода.

IV. ИНФОРМАЦИОННАЯ МОДЕЛЬ СИСТЕМЫ TAJLINGVO

Архитектура современных информационных систем, связанных с естественным языком, состоит из множества элементов текста и формирует концептуальную модель базы знаний. Для получения архитектуры необходимо основываться как на традиционную модель естественного языка, так и на современные методы формальных моделей текста. Далее приводится формальная модель информационной архитектуры:

$$FM = \{LC, SW, SS, DS, GS, CS\} \quad (2)$$

где LC (linguistic structure) – источник текстовой информации для формирования лингвистического корпуса; WS (word structure) – множество словоформ, полученные из LT; SS (semantic structure) – множество семантических структур, которые описывают SW; DS (data structure) – множество лингвистических структур, полученные на основе множества SS в SW; GS (grammar structure) – множество грамматических структур, которые основаны на грамматических правилах естественного языка; CS (code structure) – множество кодовых структур для представления DS в соответствии GS.

Последовательность преобразования текстовой информации $WS \rightarrow CS$ осуществляется процессами поиска, обработки, анализа и понимания элементов текста. Для более удобного описания информационной модели системы TajLINGVO предлагается следующая блок-схема, в которой программными средствами осуществляются процессы обработки текстовой информации (рис. 2).

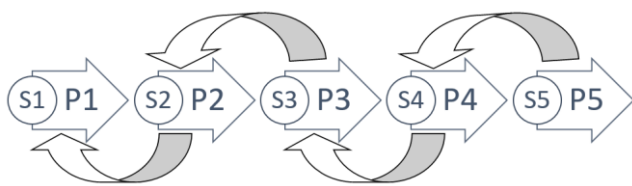


Рис. 2. Схема информационной модели TajLINGVO

Рассмотрим следующие обозначения принятые в информационной модели системы TAJLINGVO:

P1 – формирование репрезентативной выборки на базе текстовых документов (классические и современные произведения);

P2 – предварительная обработка текстовых документов для автоматического лингвистического анализа; в результате выявляется проблема омонимии предлагается возврат к процессу P1;

P3 – процесс извлечения из текстовой информации множества элементов текста с выявлением их структур и запись баз данных; в результате выявления неоднозначности семантической структуры элементов текста возможен возврат к процессу P2;

P4 – процесс формирования структур элементов текста на основе орфографических правил языка; в результате определения несоответствий выявленных структур с правилами естественного языка возможен возврат к процессу P3;

P5 – процесс обработки и управления данными; в результате определения неявного цифрового портрета текста возможен возврат к процессу P4;

S1 – источники текстовых документов;

S2 – корпус текстов;

S3 – семантические структуры элементов текста, которые соответствуют грамматическим правилам естественного языка;

S4 – множество информационных структур после обработки текста;

S5 – база данных и цифровой портрет элементов текстовой информации для формирования базы знаний.

V. ФУНКЦИОНАЛЬНАЯ МОДЕЛЬ СИСТЕМЫ TAJLINGVO

На сегодняшний день для моделирования функциональных моделей программного обеспечения используются стандартные методологии и языки функционального моделирования, такие как IDEF, DFD, UML. Унифицированный язык моделирования (UML) – используется для графического описания и объектного моделирования информационных систем. В рамках визуального моделирования язык UML широко использует стандарты объектно-ориентированную методологию. В UML определяют четыре основных вида моделей:

- статическая модель (static model);
- динамическая модель (dynamic model);
- модель взаимодействия объектов (interaction model)
- физическая модель (physical model).

Функциональную модель информационной системы можно представить совокупностью трёх видов диаграмм: диаграмма вариантов использования, диаграмма активности и диаграмма классов.

Диаграммы вариантов использования определяет общие границы и контекст моделируемой предметной области на начальных этапах

проектирования системы. Диаграмма формулирует общие требования к функциональному поведению системы и раскрывает её концептуальную модель в

форме логических и физических объектов.

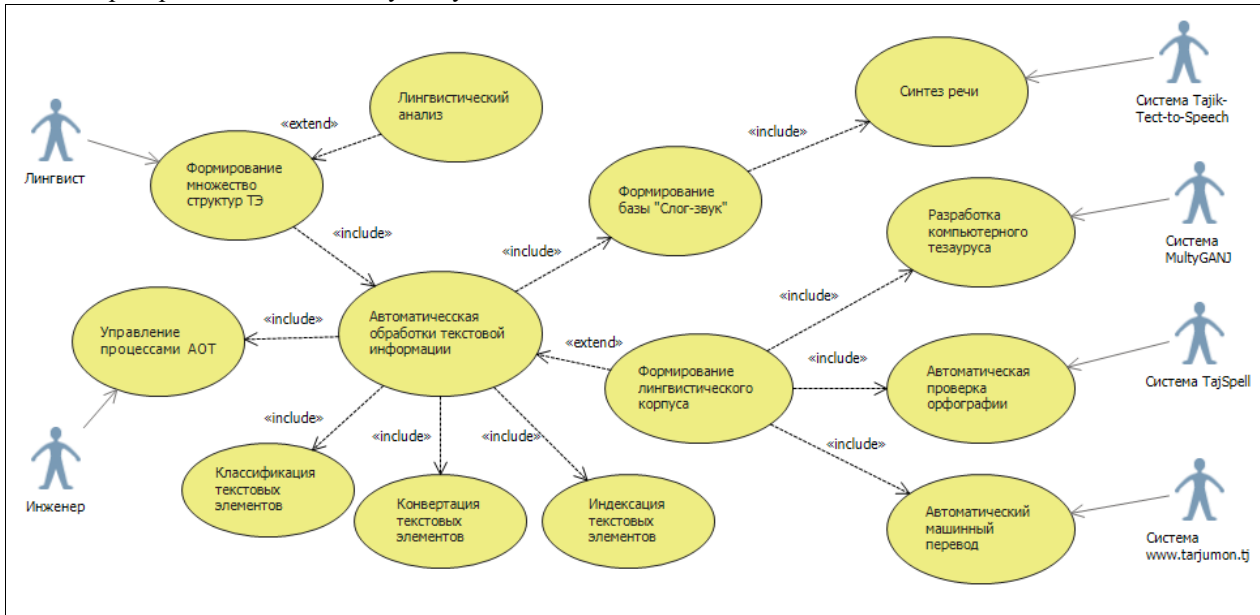


Рис. 3. Диаграмма вариантов использования TajLINGVO

С помощью диаграммы активности (рис. 4) анализируется поведение системы на базе потока данных и процессов управления. Диаграмма отображает абстрактный алгоритм жизненного цикла объекта системы, но отличается от блок-схемы в котором описывается основные шаги алгоритма.

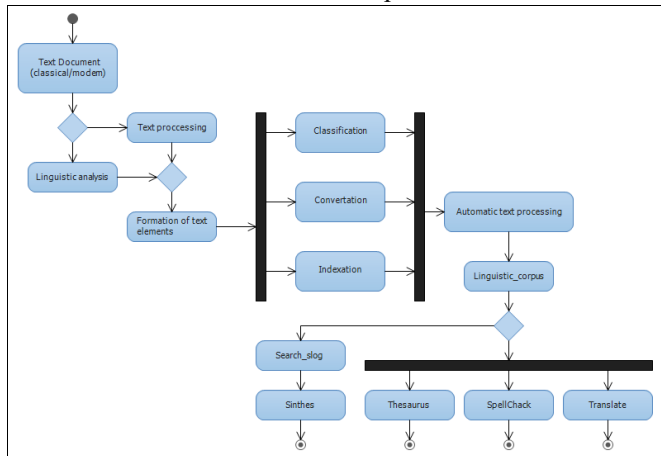


Рис. 4. Диаграмма активности системы TajLINGVO

Для описания структуру объектов системы, взаимоотношения объектов, атрибутов, функций и процедур используется диаграмма классов (рис. 5).

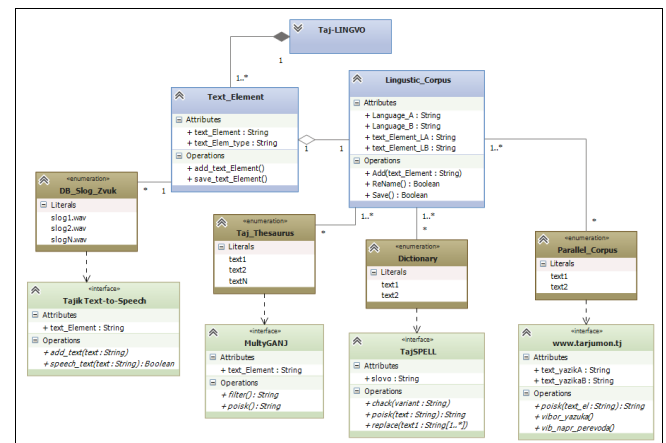


Рис. 5. Диаграмма классов системы TajLINGVO

Структурная модель системы состоит из совокупности классов и их взаимодействий:

- класс-сущность Text_Element реализует технологию композиции обработки текстовой информации;
- класс-сущность Text_Element реализует технологию обработки и управления элементами текстовой информации следующих типов: буквы, биграммы и триграммы букв, слоги, слова, словосочетания, предложения;
- класс-сущность Linguistic_Corpus реализует технологию представления структуры текстовых элементов в виде параллельного корпуса;
- класс-интерфейс Tajik Text-to-Speech предоставляет доступ к внешней библиотеке TTextSpeechLib, реализующей функции синтеза речи на основе базы слог-звук в виде класс-списка DB_Slog_Zvuk;
- класс-интерфейс MultyGANJ предоставляет доступ к внешней библиотеке TThesaurusLib, реализующей функции управления лингвистического тезауруса на

основе базы данных в класса-списка Taj_Thesaurus;

– класс-интерфейс TajSPELL предоставляет доступ к внешней библиотеке TSpellCheckLib, реализующей функции автоматической проверки орфографии на основе базы словаря в виде класса-списка Spell_Dictionary;

– класс-интерфейс www.tarjumon.tj предоставляет доступ к внешней библиотеке TTranslateLib, реализующей функции автоматического машинного перевода текста на основе базы параллельного корпуса текста на таджикском, русском и английском языках в виде класса-списка tj_ru_en_Parallel_Corpus.

Программно-аппаратные средства системы TajLINGVO состоят из графических интерфейсов, в которые включены программные модули управления процессами, реализующими синтез речи, компьютерного тезауруса, автоматической проверки орфографии и машинного перевода [16]. Информационное обеспечение системы TajLINGVO реализованы соответственно в базе данных «слог-звук», «таджикский компьютерный тезаурус», XML-файлы словоформ, базы данных «параллельный корпус текста». Управление базами данных сопровождается языком запросов SQL в СУБД MySQL, а управляющие процессы реализованы средствами программирования пакета MS Visual Studio .Net и веб-программирования PHP, MySQL.

VI. ЗАКЛЮЧЕНИЕ

В работе предложена методология проектирования автоматических систем обработки текстовой информации на примере таджикского языка - TajLINGVO. На основе методологии SADT предложена логическая структура работы системы. Для отображения возможности информационного обеспечения системы и взаимодействия с процессами обработки информации предлагается блок-схема информационной модели системы. На основе возможностей языка UML предложены диаграммы вариантов использования, активности и классов, которые описывают функциональную модель системы TajLINGVO. Результаты исследования стали основой проведения научных исследований в сфере компьютерной лингвистики, а именно анализ основных процессов автоматической обработки текстовой информации. Эффективность применения модели использованы для решения конкретных прикладных задач, таких как разработка компьютерного тезауруса, автоматическая проверка орфографии, синтеза речи и машинный перевод. Полученные результаты и разработанные информационные системы доступны в сети интернет по адресу www.tajlingo.tj [17].

БИБЛИОГРАФИЯ

- [1] Сбоев А.Г., Рыбка Р.Б., Иванов И.И. Численное моделирование процедуры синтаксического разбора с использованием нейронных сетей // Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. – 2015. – № 3. – С. 28-33.
- [2] Кузнецов Л.А., Капнин А.В. Технология автоматического формирования тезауруса русского языка // Информационные системы и технологии. – 2012. – № 4 (72). – С. 14-19.
- [3] Михайлов Д.В. Теоретические основы построения открытых вопросно-ответных систем. Семантическая эквивалентность текстов и модели их распознавания: монография / Д.В. Михайлов, Г.М. Емельянов; НовГУ им. Ярослава Мудрого. – Новгород, 2010. – 286 с.
- [4] Воронина И.Е. Компьютерное моделирование лингвистических объектов: монография. – Воронеж: Издательско- полиграфический центр ВГУ, 2007. – 177 с.
- [5] Загоруйко Н. Г. Прикладные методы анализа данных и знаний. – Новосибирск: ИМ СО РАН, 1999. – 270 с.
- [6] Цирульник Л.И. Алгоритмы синтеза просодических характеристик речи по тексту в системе «Мультифон» / Л.И. Цирульник, Д.В. Жадинец, Б.М. Лобанов, О.Г. Сизонов // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции Диалог'2007, Бекасово. – М.: Издательский центр РГГУ, 2007. – С. 550-558.
- [7] Анисимов А.В., Марченко А.А. Система обработки текстов на естественном языке // Искусственный интеллект. – 2002. – № 4. – С. 157-163.
- [8] Клышинский Э.С. Начальные этапы анализа текста // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. – С. 106-140.
- [9] Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Е.И. Большакова, Э.С. Клышинский, Д.В. Ландэ, А.А. Носков, О.В. Пескова, Е.В. Ягунова. – М.: МИЭМ, 2011. – 272 с.
- [10] Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004. – 248 с.
- [11] Палагин А.В. Онтологические методы и средства обработки предметных знаний: монография. – Луганск: Изд-во ВНУ им. В. Даля, 2012. – 323 с.
- [12] Формальные модели и системы в вычислительной лингвистике. Д.Ш. Сулейманов, О.А. Невзорова, П.И. Соснин, Л.Н. Беляева, Н.В. Лукашевич, С.Г. Татевосов: Научное издание / Под редакцией П. И. Соснина, О. А. Невзоровой // Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: 2016. – 187 с.

- [13] Селезнев К., Владимиров А. Лингвистика и обработка текстов // Открытые системы. – 2013. – № 4. – С.46-49. <https://www.osp.ru/os/2013/04/1303556> (дата обращения 20.10.2022)
- [14] Усманов З.Д. Об одном цифровом портрете текста и его приложении // Политехнический вестник. Серия: Интеллект. Инновации. Инвестиции. – 2019. – № 3(47). – С. 35-38.
- [15] Усманов З.Д. Автоматический поиск и статистические закономерности множества анаграмм / З.Д. Усманов. – Душанбе: "Дониш", 2020. – 75 с.
- [16] Усманов З.Д., Худойбердиев Х.А. Низомҳои худкори коркарди маълумот бо забони тоҷикӣ. Монография. – Хучанд. «Ирфон», 2022. – 186 с.
- [17] Худойбердиев Х.А. Web-приложение «Автоматические системы обработки информации на таджикском языке» www.tajlingvo.tj. – Свидетельство о государственной регистрации информационного ресурса, Республика Таджикистан. №4202200496 от 28/04/2022.

Modeling a system for automatic processing of text in the Tajik language

Kh.A. Khudoyberdiev

Abstract – The article deals with the modeling of processes in the system of automatic processing of text information in the Tajik language. It is proposed a methodology for the formation of text information processing processes in the TajLINGVO system. Also, it is proposed a logical structure of the system with a detailed description of each sub-process. The structure of the information model of the system is deciphered, which consists of a set of interconnected text elements in natural language. Functional model of the TajLINGVO system is proposed with a detailed description of the diagrams of use cases, activities and classes based on the UML modeling language. The utilization effectiveness of the proposed models has acquired their evidence and implementation in solving specific applied problems, such as the development of a computer thesaurus, automatic spelling, speech synthesis and machine translation for the Tajik language. The results obtained and the developed information systems are available on the Internet at www.tajlingvo.tj.

Keywords – information system, text processing, automatic process, software package, computer modeling, Tajik language.

REFERENCES

- [1] A.G.Sboev, R.B.Rybka, I.I.Ivanov. Chislennoe modelirovanie protsedury sintaksicheskogo razbora s ispol'zovaniem neuronnykh setei. Vestnik VGU. Seriya: Lingvistika i mezhdunarodnaya kommunikatsiya. 2015. № 3. – pp. 28-33.
- [2] L.A. Kuznetsov, A.V. Kapnin. Tekhnologiya avtomaticheskogo formirovaniya tezaurusov russkogo yazyka. Informatsionnye sistemy i tekhnologii. 2012. № 4 (72). – pp. 14-19.
- [3] Mikhailov D.V. Teoreticheskie osnovy postroeniya otkrytykh voprosno-otvetnykh sistem. Semanticheskaya ehkvivalentnost' tekstov i modeli ikh raspoznavaniya: monografiya / D.V. Mikhailov, G.M. Emel'yanov; NoVGU im. Yaroslava Mudrogo. Novgorod, 2010. – 286p.
- [4] Voronina I.E. Komp'yuternoe modelirovanie lingvisticheskikh ob'ektov: monografiya / I. E. Voronina. –Voronezh: Izdatel'sko-poligraficheskii tsentr VGU, 2007. – 177p.
- [5] Zagoruiko N. G. Prikladnye metody analiza dannykh i znaniy. Novosibirsk: IM SO RAN, 1999. – 270p.
- [6] Tsirul'nik, L.I. Algoritmy sinteza prosodicheskikh kharakteristik rechi po tekstu v sisteme «Mul'tifoN» / L.I. Tsirul'nik, D.V. Zhadinets, B.M. Lobanov, O.G. Sizonov // Komp'yuternaya lingvistika i intellektual'nye tekhnologii: trudy mezhdunarodnoi konferentsii Dialog'2007, Bekasovo, 30 maya – 3

- iyunya 2007 g. – M.: Izdatel'skii tsentr RGGU, 2007. – pp. 550-558.
- [7] Anisimov A.V., Marchenko A.A. Sistema obrabotki tekstov na estestvennom yazyke // *Iskusstvennyi intellekt*. – 2002. – № 4. – pp. 157-163.
- [8] Klyshinskiy E.H.S. Nachal'nye ehtapy analiza teksta / E.H.S. Klyshinskii // *Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i komp'yuternaya lingvistika: ucheb. posobie*. – M.: MIEHM, 2011. – pp. 106-140.
- [9] *Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i komp'yuternaya lingvistika: ucheb. posobie / E.I. Bol'shakova, E.H.S. Klyshinskiy, D.V. Landeh, A.A. Noskov, O.V. Peskova, E.V. Yagunova*. – M.: MIEHM, 2011. – 272p.
- [10] Belonogov, G.G. *Komp'yuternaya lingvistika i perspektivnye informatsionnye tekhnologii / G.G. Belonogov*. – M.: Russkii mir, 2004. – 248p.
- [11] Palagin A.V. *Ontologicheskie metody i sredstva obrabotki predmetnykh znaniy*. – [monografiya] – Lugansk: izd-vo VNU im. V. Dal'ya, 2012. – 323p.
- [12] *Formal'nye modeli i sistemy v vychislitel'noi lingvistike*. D.S.H. Suleimanov, O.A. Nevzorova, P.I. Sosnin, L.N. Belyaeva, N.V. Lukashevich, S.G. Tatevosov: Nauchnoe izdanie / Pod redaktsiei P. I. Sosnina, O. A. Nevzorovoi – Akademiya nauk RT, Institut prikladnoi semiotiki AN RT. – Kazan': 2016. – 187p.
- [13] Seleznev K., Vladimirov A. *Lingvistika i obrabotka tekstov // Otkrytye sistemy*. – 2013. №04. – pp.46-49. <https://www.osp.ru/os/2013/04/1303556> (data obrashcheniya 20.10.2022)
- [14] Usmanov Z.D. *Ob odnom tsifrovom portrete teksta i ego prilozhenii / Z. D. Usmanov // Politekhnicheskii vestnik. Seriya: Intellekt. Innovatsii. Investitsii*. – 2019. – № 3(47). – pp. 35-38.
- [15] Usmanov Z.D. *Avtomaticheskii poisk i statisticheskie zakonomernosti mnozhestva anagramm / Z.D. Usmanov*. – Dushanbe: "Donish", 2020. – 75p.
- [16] Usmanov Z.D., Khudoyberdiev K.H.A. *Nizomhoi khudkori korkardi ma"lumot bo zaboni tojiki. Monografiya*. Khujand. «Irfon», 2022, – 186p.
- [17] Khudoyberdiev K.H.A. *Web-prilozhenie «Avtomaticheskie sistemy obrabotki informatsii na tadjikskom yazyke»* www.tajlingvo.tj. – Svidetel'stvo o gosudarstvennoi registratsii informatsionnogo resursa, Respublika Tadjikistan. №4202200496, 28/04/2022.