

Разработка прототипа информационной системы определения степени уникальности выпускных квалификационных работ вуза

Геворг Петросян

Рустам Саакян

Ирина Шпехт

DOI: <https://doi.org/10.58726/27382923-ne2023.1-66>

Ключевые слова: *степень уникальности текста, информационная система, выпускная квалификационная работа, поиск заимствований*

Аннотация: Необходимость определения степени уникальности научных работ остается актуальной, поскольку плагиат и недостаток оригинальности научных работ могут негативно сказаться на качестве образовательного процесса и доверии к высшему учебному заведению. Так как традиционное ручное обнаружение заимствований не точный и трудоемкий процесс, автоматизированные системы выявления плагиата и заимствований привлекают большое внимание. В этом контексте ставится вопрос о необходимости разработки системы по определению заимствований в работах, выполняемых студентами. Такая система позволит достоверно определить, что мысли автора ранее никем не были опубликованы, и являются его интеллектуальной собственностью. Целью данной работы является разработка прототипа информационной системы определения уникальности выпускных квалификационных работ вуза. Такая система позволит сравнить новые работы с имеющимися в локальной базе данных на наличие заимствований и выявить степень уникальности работ.

Введение

Вопрос определения степени уникальности научных работ становится все более актуальным в связи с очень быстрым ростом интернет-технологий. Легкость, с которой можно найти и включить нужную информацию в свою работу без изменений или с минимальными поправками, привела к ухудшению качества работ обучающихся и уменьшению степени их уникальности. Работы с большой степенью заимствования отрицательно влияют на доверие к научному сообществу, искажают результаты исследований и нарушают этические принципы научной деятельности.

Одной из мер по борьбе с данной проблемой является проверка работы на наличие заимствований с помощью специализированных информационных систем. Такие системы используют алгоритмы сравнения текстов и могут определять степень схожести между текстами, выделять фрагменты, заимствованные из других источников, и оценивать степень уникальности текста. На сегодня популярными являются такие системы, как «Антиплагиат» («Анти-Плагиат» ЗАО), Turnitin («Advance Publications» акционерное общество), Plagscan (PlagScan GmbH), Grammarly («Grammarly» акционерное общество) и другие, которые проводят автоматический анализ текстов на наличие заимствований и предоставляют отчеты о степени их уникальности.

Тем не менее, отсутствуют подобные специализированные системы обнаружения заимствований на многих языках мира. В таких случаях приходится прибегать к использованию инструментов, не адаптированных к определенному языку, однако эти решения, как правило, не учитывают грамматические и синтаксические особенности языка и не показывают достаточный уровень качества обработки [1, 4].

Такая ситуация оценки уникальности текстов научно-исследовательских работ характерна для стран, где создание и применение научно-исследовательских работ осуществляется на национальных языках. В этом случае оценка уникальности работ является сложной задачей, так как отсутствует единая система выявления степени уникальности [3, 99].

Заимствованием, кроме дословного копирования, считается сокрытие заимствований путем перефразирования и перевода [1,4]. Проблема также усугубляется в связи с отсутствием в интернет-среде соответствующих научно-исследовательских работ.

Определение плагиата в фрагменте текста является обязанностью эксперта, а не компьютерной программы. Автоматические системы предназначены для того, чтобы помочь эксперту находить потенциальные случаи и при возможности, принимать всесторонние решения [5, 2].

Постановка задачи

Целью данной работы является разработка прототипа информационной системы (ИС) определения уникальности выпускных квалификационных работ (ВКР) вуза. В рамках проведенных работ были достигнуты следующие результаты:

1. Построена структурная схема представления ИС.

2. Спроектирована графическая модель ИС с помощью UML диаграммы.
3. Разработан модуль пополнения базы по двум аспектам – названиям работ и содержанию работ, с возможностью выполнения проверки соответствия содержания названию работы в базе.
4. Разработан модуль поиска уникальности текста по названию работы с возможностью генерации отчета.
5. При реализации поиска заимствований в названии работы был разработан авторский метод первичной обработки слов с использованием массива прямой формы полезных слов с учетом грамматической и синтаксической специфики армянского языка.

Разработка и внедрение ИС в учебный процесс вуза поможет повысить степень уникальности научно исследовательских работ, выполняемых студентами.

Графическая модель ИС с помощью UML диаграммы

Первым этапом является разработка диаграммы взаимодействия UML, она используется для определения необходимых требований к системе. Диаграмма взаимодействия описывает отношения между субъектами (действующими лицами, в том числе и неодушевленными) и прецедентами (возможностями проектируемой системы) в различных условиях их поведения. Действующими лицами в системе являются пользователь и администратор (рис. 1).

Структурная схема отображает основные функции и задачи данной системы, а также связи между ними. На схеме представлены компоненты форм ввода-вывода информации базы данных. Входной информацией являются список ВКР студентов и их содержания. На основе этой информации пополняется база данных. Выходной информацией является отчет результатов сравнения названий новой работы с имеющимися в базе (рис. 2).

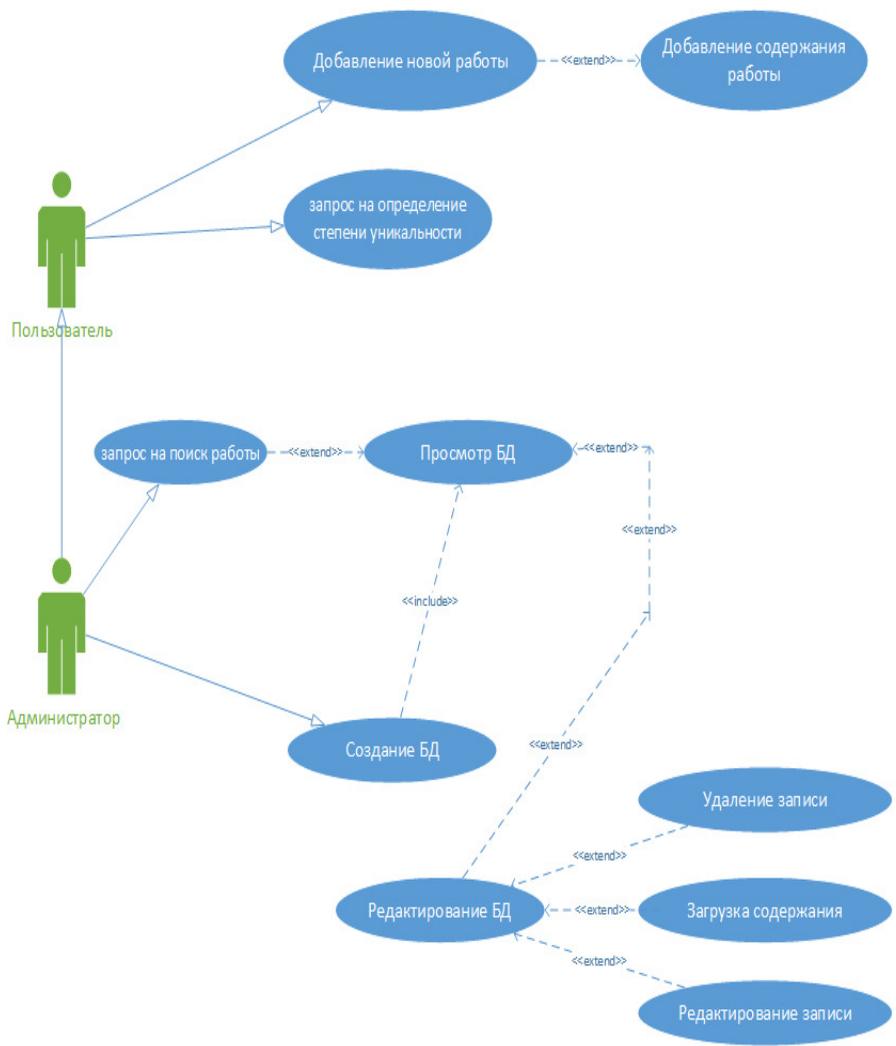


Рисунок 1. Графическая модель системы с помощью UML диаграммы Структурная схема представления ИС

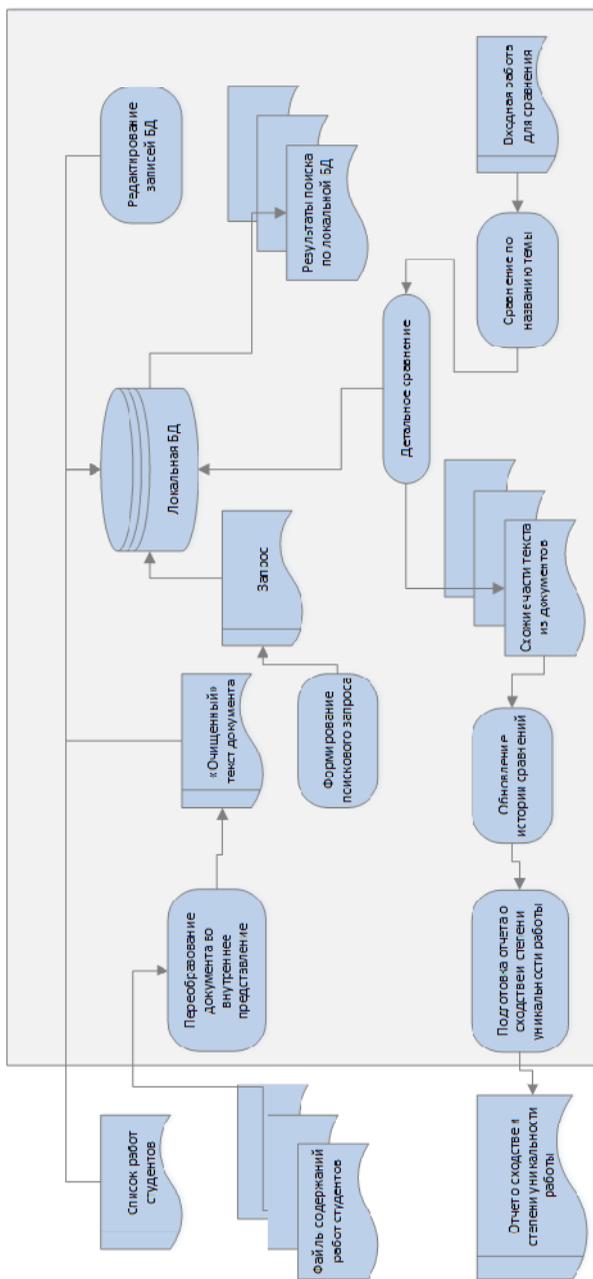


Рисунок 2. Структурная схема проектируемой системы

Описание системы

В ИС реализованы 2 уровня доступа – администратор и пользователь (рис. 3).



Рисунок 3. Окно Входа

База пополняется администратором с помощью заранее заполненного файла в MSExcel. Каждая строка в базе данных хранит следующую информацию о работе – порядковый номер работы, название темы, имя студента, имя профессора, дату работы и информацию о наличии файла содержания работы. Есть возможность загружать в систему несколько различных файлов, они будут располагаться один за другим в базе данных (рис. 4)

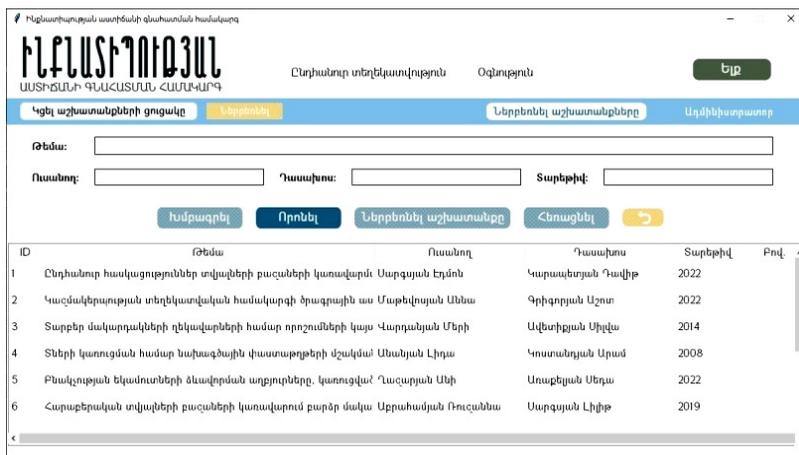


Рисунок 4. Окно страницы администратора после пополнения базы

Содержания работ пополняются администратором в формате MSWord и связываются с работами в базе в соответствии с названием (рис. 5). После загрузки содержаний появляется информация о наличии содержания в соответствующих полях.



Рисунок 5. Окно страницы администратора после загрузки содержаний работ

Пользователь может сравнить новую работу с имеющимися в базе по названию темы. Система показывает те работы из базы, у которых есть минимум 2 совпадающих слов (для сравнения слова преобразуются в прямую форму по авторскому методу [3, с. 80]) с исходным названием, если число слов в исходном названии меньше 6, в противном случае при условии минимум 3 совпадений (рис. 6).

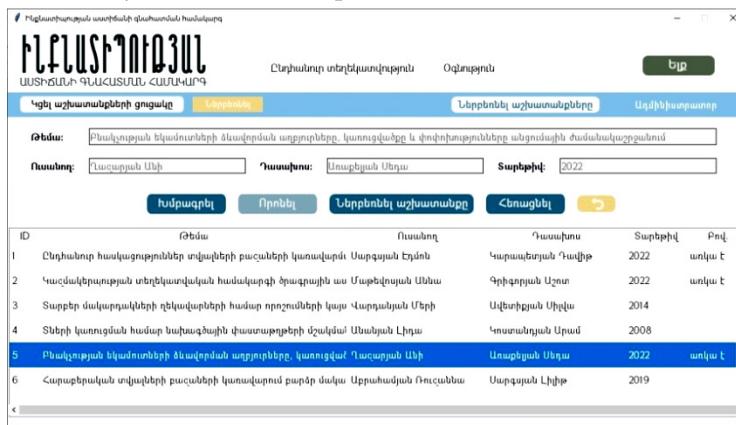


Рисунок 6. Окно страницы пользователя после сравнения новой работы по названию

Также в системе есть возможность автоматически генерировать и сохранять отчет о результатах проверки (рис. 7).

ԻՆՔԼԱՏԻՊՈՒԹՅԱՆ
ԱՄՏԻՃՆԱԻ ԳՆԱՀԱՏՄԱՆ ՀԱՄԱԿԱՐԳ

Հաշվետվություն

Ամսաթիվ: 11/04/2023, 22:44:38

Հաշվետվություն N: 1

Թեմա: **Տվյալների բազաների կառավարման համակարգերի առավելությունները**
և **թերությունները**

Համընկնումների քանակը: 3

1) Ընդհանուր հասկացություններ **տվյալների բազաների կառավարման համակարգերի մասին**,
Ուսանող: Սարգսյան Էդմոն, *Դասախոս:* Կարապետյան Դավիթ, 2022

2) Հարաբերական **տվյալների բազաների կառավարում** բարձր մակարդակի ծրագրավորման
լեզուներով, *Ուսանող:* Աբրահամյան Ռուզաննա, *Դասախոս:* Սարգսյան Լիլիթ, 2019

3) Տեխնիկայի խանութում ապրանքաշրջանառության կազմակերպման և վճարումների
կառավարման **տվյալների բազայի** մշակում, *Ուսանող:* Սկրտչյան Աշոտ, *Դասախոս:* Հովհաննիսյան
Մարիամ, 2019

Рисунок 7. Отчет результатов сравнения по названию

Система сохраняет все результаты сравнений в истории. Пользователь может открыть результат сравнения конкретной работы из истории для просмотра и снова при необходимости сохранить отчет.

Заключение

В рамках работы был разработан прототип ИС определения уникальности ВКР вуза. Одним из результатов исследования стало представление структурной схемы и UML диаграммы ИС. Также были разработаны модули пополнения базы по названию и содержанию работ, и поиска уникальности по названию работы с возможностью генерации отчета. Поиск заимствований в названии текста был реализован с помощью авторского метода. Система, разработанная на основе спроектированного прототипа, будет использоваться для сравнения представленных работ с локальной базой данных, раскрывая потенциальные случаи заимствования и опре-

деляя степень уникальности работы, с целью повышения академической честности студентов и поддержки интегритета научных исследований. Следующим шагом является сравнение работ по содержанию с применением языковых моделей и марковских цепей, по которому уже ведутся работы [4].

DOI: <https://doi.org/10.58726/27382923-ne2023.1-66>

Литература

1. Гукасян Ц. Г. Методы и программные средства для выявления заимствований в текстах на армянском языке // Москва, 2021, 187 с.
2. Петросян Г.А, Саакян Р.Р Подготовка к определению степени уникальности текста на основе формирования массива прямой формы полезных слов// Научные ведомости Ванадзорского государственного университета, Естественные и точные науки, N2, 2022, с. 77-86.
3. Саакян Р.Р, Петросян Г.А Проектирование системы оценки степени оригинальности исследовательских работ //Вестник Инженерной академии Армении, Том 19, N1, 2022, с. 98-103.
4. Саакян Р.Р., Шпехт И.А., Петросян Г.А. Нахождение наличия заимствований в научных работах на основе марковских цепей // Вестник Санкт-Петербургского университета, Прикладная математика. Информатика. Процессы управления. Том 19. Выпуск 1, с. 43-51.
5. Alberto Barron Cedeno L. On the Mono- and Cross-Language Detection of Text Re-Use and Plagiarism //2012.

**Բուհի ավարտական որակավորման աշխատանքների
ինքնատիպության աստիճանի գնահատման տեղեկատվական
համակարգի նախատիպի մշակում**

Գևորգ Պետրոսյան
Ռուստամ Սահակյան
Իրինա Շայխստ

Ամփոփում

Հանգուցային բառեր. տեքստի ինքնատիպության աստիճան, տեղեկատվական համակարգ, ավարտական որակավորման աշխատանք, փոխառությունների որոնում

Այս աշխատանքի նպատակն է՝ մշակել տեղեկատվական համակարգի նախատիպ բուհի ավարտական որակավորման աշխատանքների ինքնատիպությունը որոշելու համար: Կատարված աշխատանքի շրջանակներում ստացվել են հետևյալ արդյունքները. կառուցվել է տեղեկատվական համակարգի կառուցվածքային սխեմա. նախագծվել է տեղեկատվական համակարգի գրաֆիկական մոդել՝ UML դիագրամի օգնությամբ, մշակվել է տվյալների բազայի համալրման մոդուլ երկու եղանակով՝ ըստ աշխատանքների անվանման և ըստ աշխատանքների բովանդակության՝ տվյալների բազայում բովանդակության համապատասխանությունը աշխատանքի անվանման հետ ստուգելու հնարավորությամբ:

Մշակվել է աշխատանքի անվանման ինքնատիպությունը որոշելու մոդուլ՝ հաշվետվություն ստեղծելու հնարավորությամբ: Աշխատանքի անվանման մեջ փոխառությունների որոնումն իրականացնելիս մշակվել է բառերի առաջնային մշակման հեղինակային մեթոդ՝ օգտագործելով օգտակար բառերի ուղիղ ձևի զանգվածը՝ հաշվի առնելով հայերենի քերականական և շարահյուսական առանձնահատկությունները: Տեղեկատվական համակարգի մշակումն ու ներդրումը բուհի ուսումնական գործընթացում կնպաստի ուսանողների կատարած գիտահետազոտական աշխատանքների ինքնատիպության աստիճանի բարձրացմանը:

Development of an Information System Prototype for Determining the Degree of Uniqueness of Final Qualification Works of the University

Gevorg Petrosyan

Rustam Saakyan

Irina Shpekht

Summary

Key words: *degree of uniqueness of the text, information system, final qualification work, search for borrowings*

This work aims to develop a prototype of an information system for determining the uniqueness of the final qualifying works of the university. The following results were achieved as a part of the work: a block diagram of the informational system's representation was built, and a graphical model of the informational system was designed using UML diagrams. A module for replenishing the database has been developed in two aspects: the names of the works and the content of the works, with the ability to check the compliance of the content with the name of the work in the database. A module generating a report and searching for the text's uniqueness based on the work's title has been developed. When implementing the search for borrowings in the title of the work, the author's method of word primary processing was developed using an array of direct forms of valuable words, taking into account the grammatical and syntactic specifics of the Armenian language. The development and implementation of an information system in the university's educational process will contribute to increasing the degree of uniqueness of student research work.

Ներկայացվել է 19.03.2023 թ.

Գրախոսվել է 21.04.2023 թ.

Ընդունվել է տպագրության 25.05.2023 թ.