

МОДЕЛИ ИНТЕГРАЦИИ ИНФОРМАЦИОННЫХ СИСТЕМ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ

Models of Integration of Information Systems in Higher Education Institutions

Marat Rakhmatullaev

Uktam Karimov

Tashkent University of Information Technologies named
after Muhammad al Khwarizmi, Uzbekistan

Abstract. *At present a lot of automated systems are developing and implementing to support the educational and research processes in the universities. Often these systems duplicate some functions, databases, and also there are problems of compatibility of these systems. The most common educational systems are systems for creating electronic libraries, access to scientific and educational information, a program for detecting plagiarism, testing knowledge, etc. In this article, models and solutions for the integration of such educational automated systems as the information library system (ILS) and the anti-plagiarism system are examined. Integration of systems is based on the compatibility of databases, if more precisely in the metadata of different information models. At the same time, Cloud technologies are used - data processing technology, in which computer resources are provided to the user of the integrated system as an online service. ILS creates e-library of graduation papers and dissertations on the main server. During the creation of the electronic catalog, the communication format MARC21 is used. The database development is distributed for each department. The subsystem of anti-plagiarism analyzes the full-text database for the similarity of texts (dissertations, diploma works and others). Also it identifies the percentage of coincidence, creates the table of statistical information on the coincidence of tests for each author and division, indicating similar fields. The integrated system was developed and tested at the Tashkent University of Information Technologies to work in the corporate mode of various departments (faculties, departments, TUIT branches).*

Keywords: *integrated system, e-library, plagiarism, database, information library system, cloud technologies, communicative format.*

Введение

Introduction

В последнее время во всем мире осуществляются ряд международных проектов по совершенствованию высшего образования путем внедрения инновационных идей, обогащения материально технической базы, внедрения новейших информационно коммуникационных технологий в

процесс образования. Осваиваются передовые опыты по подготовке кадров, сопоставляются учебные планы и учебные программы с ведущими учебными заведениями мира.

Методы, системы и программные средства оценки знаний и контроля образовательного процесса играют большую роль, при подготовке конкурентоспособных кадров. Одним из таких систем является информационная система обнаружения заимствования текстов, которая нужна не только для обнаружения плагиата, но и для поиска типовых текстов, тем, аналогов. В настоящее время существует достаточно большое количество сервисов и программ, позволяющих каким-либо образом выявить заимствования. В числе таковых можно назвать: систему АПЛИС, Антиплагиат (Dygilev et al., 2011), Advego Plagiatus (Johnston, 2008), miratools.ru (Aushra, 2006), istio.com (Benno& Moshe, 2011), Praide Unique Content Analyser II (Spiteri & Grima 2011), Plagiatinform (Monostori et al., 2000), Copyscape (Bouville, 2008)

Особую популярность представляют системы антиплагиата в учебных заведениях, университетах. Базовые принципы систем обнаружения заимствования текстов были предложены давно, но технологии совершенствуются и развиваются непрерывно. Процесс проверки текста на уникальность основан на математических моделях и алгоритмах, все программы антиплагиата взаимодействуют с поисковыми машинами для ранжирования web страниц или определения заимствований в курсовых и дипломных работах и диссертациях. Сегодня активно развивается шингловый и корреляционный метод проверки уникальности. Актуальные версии программ умеют распознавать стоп-знаки. Поэтому, меняя пунктуацию, расставляя союзы, предлоги и вводные слова, обмануть антиплагиат не получится. Алгоритм безошибочно определяет “вес” слова и сравнивает хэш-функции.

Система *Антиплагиат* разработана компанией Форексис. (Dygilev et al., 2011) Система осуществляет онлайн поиск по большому количеству документов, хранящихся в собственной базе. Обладает рядом недостатков, таких как: отсутствие осуществления поиска по документам, доступным в сети Интернет; ограничение проверяемого текста 3000 символами и 5000 символами после регистрации и ограничение возможностью проверки только по базе имеющихся работ.

Программа *Advego Plagiatus* осуществляет онлайн проверку с использованием поисковых систем (Bouville, 2008). В отличие от аналогичных систем, Advego Plagiatus не использует Яндекс.XML (бесплатный сервис, предоставляющий возможность производить автоматические поисковые запросы к “Яндексу” и публиковать его выдачу у себя на ресурсе). Программа выдаёт процент совпадения текста и выводит

найденные источники. Программа не преобразует буквы, то есть нет преобразования регистра, нет обработки и изменения латинских букв в русских словах на аналогичные буквы русского алфавита для текстов на русском языке. Также отсутствует поддержка поиска по собственной базе; из-за особенностей работы возникают ситуации, когда результаты проверки отличаются от раза к разу.

Сервис *www.miratools.ru* позволяет осуществлять онлайн проверку текста на плагиат (Aushra, 2006). Система использует результаты выдачи поисковых систем. К достоинствам можно отнести возможность замены английских букв на русские. Также имеется возможность изменения длины и шага шингла, используемых для проверки. По результатам проверки выдаётся процент совпадений и найденные источники. Недостатком является то, что система не работает с собственной базой, существует ограничение на длину текста в 3000 символов и ограничение на число проверок в течение суток.

Сервис *www.istio.com* осуществляет проверку текста на наличие заимствованного контента с использованием поисковых систем Яндекс.XML и Yahoo.com (Benno & Moshe, 2011). Возможности сервиса несколько слабее по сравнению с Miratools. По результатам проверки выдаётся сообщение о том, является ли текст уникальным или нет, и выдаётся список подобных страниц сайтов. Сервис предоставляет дополнительные средства для анализа текстов, например, проверку орфографии, анализ наиболее частотных слов и т. д. У системы отсутствует преобразование букв и поиск по собственной базе.

Программа *Praide Unique Content Analyser II* имеет широкие возможности по проверке текстов с использованием поисковых систем (Spiteri & Grima 2011). Имеется возможность выбора используемых поисковых систем, содержит средства добавления новых поисковых систем. Проверка осуществляется шинглами, длину которых можно изменять. Можно задавать количества слов перекрытия шинглов. Выводится подробный отчёт по проверке в каждой поисковой системе. В программе отсутствует замены букв, обработка стоп-слов и нет поддержки работы с собственной базой.

Система *Plagiainform* проверяет документы на наличие заимствований как в локальной базе, так и в сети Интернет (Monostori et al., 2000). Система умеет находить плагиат в виде документов, скомпонованных из “перемешанных” кусков текста нескольких источников. Проверка может осуществляться с использованием быстрого или углублённого поиска. Результаты проверки выдаются в виде наглядного отчёта. Недостатком является отсутствие преобразования букв. Также авторы не предоставляют

возможности свободного использования или тестирования системы, и оценить качество её работы невозможно.

Сервис *Copyscape* позволяет осуществлять поиск копий веб-страниц в Интернете (Bouville, 2008). Система возвращает список веб-страниц, у которых есть похожий по содержанию текст. Сервис осуществляет проверку на наличие заимствованного контента с использованием поисковых систем Google и Yahoo!. Проверяется только контент веб-страницы, то есть, для выяснения уникальности текста необходимо опубликовать текст на сайте и ввести в системе адрес страницы. Без регистрации существует ограничение на количество проверок в месяц и на количество отображаемых результатов – 10 сайтов. Для зарегистрированных пользователей нет ограничений на количество проверок и выводимых результатов, но каждый запрос стоит 5 центов.

По результатам сравнительного анализа вышеуказанных систем можно указать на следующие общие проблемы данных систем:

- Отсутствует поддержка поиска по собственной базе;
- Отсутствия преобразования регистра, нет обработки и изменения латинских букв в русских словах на аналогичные буквы русского алфавита для текстов на русском языке;
- Существует ограничение на длину проверяемого текста в 3000-5000 символов;
- Нет возможности интеграции баз данных с электронными библиотеками.

Целью данного исследования является на основе сравнительного анализа существующих информационных систем определения заимствования текстов, разработка оптимальной информационной модели проверки дипломных работ и диссертаций путем интеграции антиплагиатных систем и электронных библиотек в процесс образования.

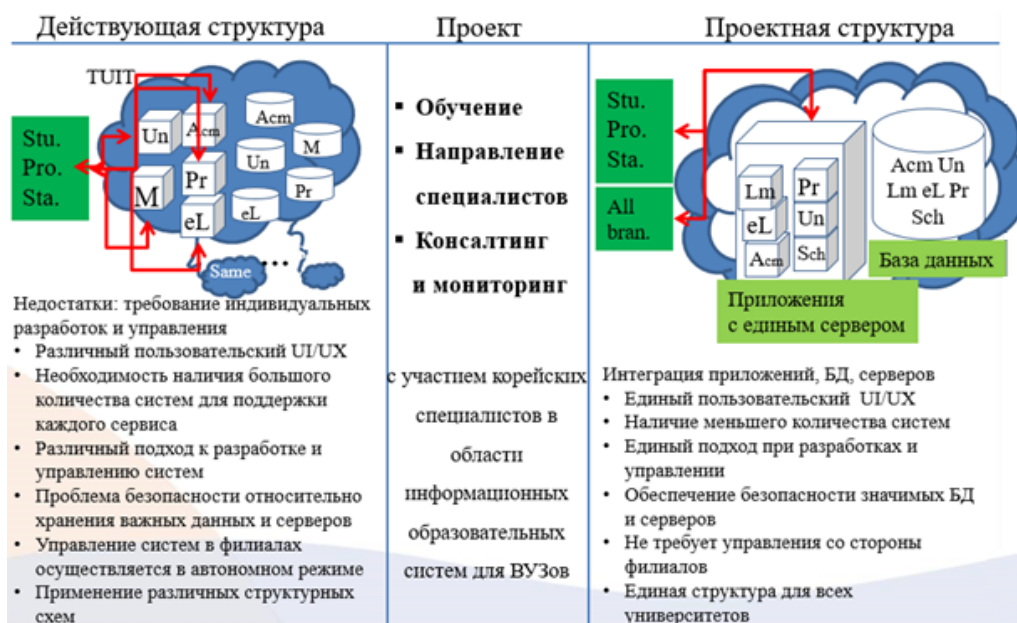
Вклад *Contributions*

С 2013 года в Ташкентском университете Информационных технологий имени Мухаммада аль-Хоразмий (далее – ТУИТ) и его региональных филиалах запущена система определения заимствования текстов в дипломных работах. В базу данных системы введено защищённые дипломные работы и магистерские диссертации с 2013 года по настоящее время. С 2014-2015 учебного года проверяется каждая работа студентов и магистрантов ТУИТ. Для проверки работ студентов и магистров была разработана несколько экспериментальных моделей проверки диссертаций

и дипломных работ на наличие плагиата (Fig-2, 3). На основе двух предыдущих экспериментальных моделей была разработана третья модель проверки диссертаций и дипломных работ на наличие плагиата.

Данная статья написана на основе выработанного практического опыта, результатов экспериментов, анализа документов реализованных проектов по созданию корпоративных сетей библиотек и литератур по созданию электронных библиотек.

В периоде 2018-2021 годах в Узбекистане будет реализован проект под названием “Интегрированная система управления образовательным процессом ТУИТ” с корейскими партнёрами. В рамках которого будет разработана интегрированная система управления учебным процессом на основе современных программно-технических средств (рис. 1).



1.Рис. Информационная модель проекта “Интегрированная система управления образовательным процессом ТУИТ”

Figure 1. Information model of the project “Integrated learning management system of Tashkent university of information technologies”

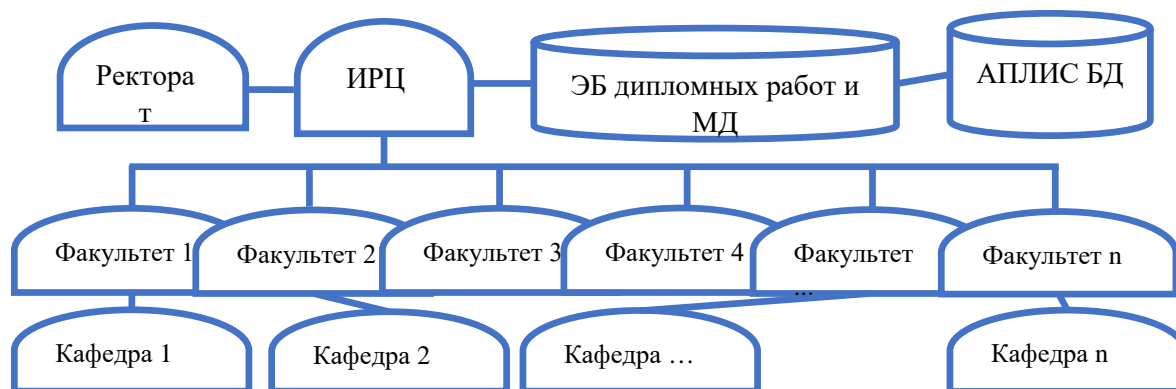
ARMAT++ – система, предназначенная для автоматизации основных библиотечных процессов и формирования базы данных электронной библиотеки. Данной системой организована корпоративная сеть всех информационно-ресурсных центров (библиотек) при вузах Узбекистана. В базе данных данной системы существует более 1 млн. библиографических записей в MARC21 формате и более 50 тыс. полнотекстовых научно образовательных ресурсов.

Система АПЛИС (система определения похожих текстов) разработан преподавателями Ташкентского университета информационных технологий. Данная система отличается от других, тем что она интегрирована с базой данных электронной библиотеки ARMAT++.

Ниже рассмотрены экспериментальные подходы и результат модели интеграции систем для решения задач определения заимствования текстов, предназначенная для использования в вузах и для работы в корпоративных информационных сетях. Подход предполагает формирование электронных библиотек дипломных работ, диссертаций, научных статей и др., в режиме корпоративного взаимодействия кафедр, факультетов и даже филиалов одного вуза. А программный комплекс использует эту базу для определения заимствования текстов, их авторов, формирования статистической информации.

Эксперименты *Experiments*

В данных экспериментах использована организационная структура Ташкентского университета Информационных технологий и его региональных филиалов. Первая модель проверки магистерских диссертаций (далее – МД) и дипломных работ (далее – ДР) студентов разработана в 2013 году (рис. 2).

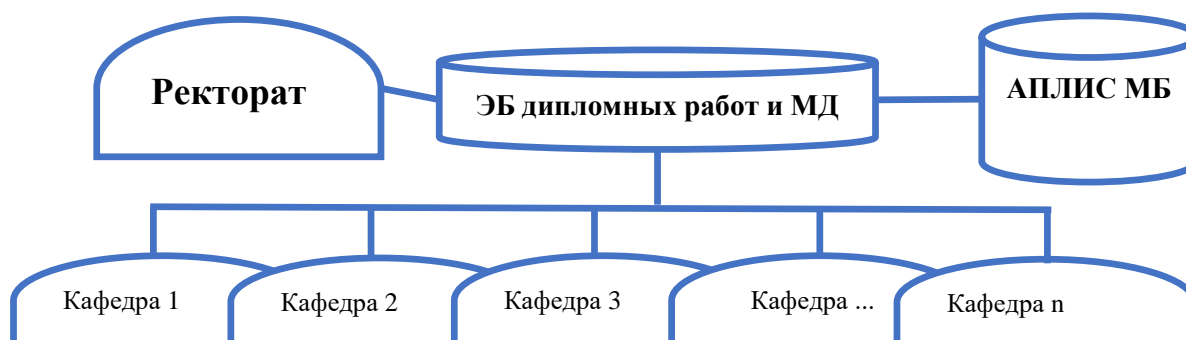


2.рис. **Первая модель проверки магистерских диссертаций и дипломных работ**
Figure 2. The first information model of checking for plagiarism of diploma works and master's dissertations

После предварительной защиты все работы собираются в информационном ресурсном центре (библиотека вуза), вводятся в базу данных электронной библиотеки и проверяются на наличия плагиата в системе АПЛИС. Требуется много времени для подготовки базы данных.

Так как промежуток между предварительной защитой и предоставления заключения государственной аттестационной комиссии (ГАК) очень короткий. Проверка и каталогизация 1500 работ за короткое время не легкая задача. В некоторых случаях из-за поспешного заключения, пропускаются “похожи” работы и отозвания таких работ создают много проблем.

Вторая информационная модель проверки МД и ДР на наличия плагиата разработана в 2014 году (рис. 3).



3.рис. **Вторая модель проверки магистерских диссертаций и дипломных работ**
Figure 3. The second information model of checking for plagiarism of diploma works and master's dissertations

Подготовленные работы рассматриваются научными руководителями, рецензентами и рекомендуются на предварительную защиту. После предварительной защиты, электронные версии работ вводятся в базу данных электронной библиотеки находящегося на сервере ИРЦ. Работы проверяются на наличии плагиата в самой библиотеки с помощью системы АПЛИС. Заключение допуска на защиту выдаётся кафедральной комиссией.

Результат *Results*

В качестве результата двух предыдущих моделей была разработана третья информационная модель проверки МД и ДР на наличия плагиата в 2016 году (рис. 4). Организовывается отдел для сбора, хранения, систематизации, каталогизации и проверки на повторения тем, дипломных работ и магистерских диссертаций. Отдел собирает первичные данные о работах с деканатов.



4.рис. Третья модель проверки магистерских диссертаций и дипломных работ
Figure 4. The third information model of checking for plagiarism of diploma works and master's dissertations

После утверждения тем, сотрудники отдела вводят первичные данные (научные руководитель, автор и другие библиографические элементы) в электронный каталог системы. Процесс формирования электронного каталога работ не продлится позже чем ноябрь месяц, что предотвращает суету перед защитой. Создаётся группа экспертов по направлениям обучения для проверки на наличия плагиата. Отдел предоставляется группе экспертов для выполнения работы. Определённые результаты предоставляются группе экспертов для окончательного заключения. Качество выполненных работ определяет группа экспертов. Председателем группы предлагается проректор учебного заведения по учебным делам. Новый организованный отдел ведёт мониторинг выполнения работ и проверяет на наличия рекомендации производства для данной темы.

Анализ моделей *Analysis of the models*

Если организовать работу по первой модели, то, все работы университету собираются в одном месте, для каталогизации и дальнейшей проверки. В этом случае 4 сотрудника ИРЦ освобождаются от повседневных обязанностей для каталогизации и проверки всех работ.

Организация работ на примере второй модели проявляется огромное доверия кафедрам выпускающих специалистов. В таких случаях, наблюдается ошибки в процессе каталогизации и не попадания работ в саму базу данных. В некоторых случаях студенты сдавали работы с опозданием.

Организация работ на примере третьей модели, процесс каталогизации дипломных работ и магистерских диссертаций выполняется вовремя и качественно, так как оно начинается на начале учебного года и появляется возможность проверки тем на наличия повторения, относительно работам выполненным в прошлых годах. База данных электронной библиотеки МД и дипломных работ формируется вовремя в результате чего, появляется возможность своевременной выдачи заключения экспертной группы на наличия плагиата ГАК.

Summary

Organization of the work by the first model does not yield significant results, since in a short time it is necessary to collect a huge amount of information, catalog and check all the work on the presence of plagiarism. The certification commission does not receive the results of the examination of master's dissertations and diploma theses on time. As a result, plagiarism may arise after the award of diplomas to students.

Organization of work on the example of the second model shows great confidence in the departments of graduates. In such cases, there are many mistakes in the process of cataloging and not getting the work into the database itself. In some cases, students were late in their work. This all leads to the routine work of the departments in which even without this many responsibilities. Checking the work on the presence of plagiarism does not facilitate, but increases the scope of duties of the departments. In this model, too, there is not enough time, as a result of which information is not provided in time by the state attestation commission.

Organization of works on the example of the third model, the process of cataloging the theses and master's theses is carried out on time and qualitatively, since it begins at the beginning of the academic year and it becomes possible to check the topics for the existence of a repetition, with respect to the work done in previous years. The database of the electronic library of master's dissertations and diploma works is formed on time, as a result of which, it becomes possible to issue the expert group's opinion on the availability of plagiarism of the attestation commission in time. After the completion of the cataloging of this year's work, the department's employees begin to enter the work (annotation, titles, content and output data) in the chronological order of the last years thereby enriching the database of the e-library of diploma works and master's theses. After the creation of the e-library of diploma works and master's theses, it is possible to integrate the database of e-books related to the study areas for determining the scope of the study and the degree of analysis of the teaching aids for the work. It will be possible for the supervisor of a fair evaluation for the recommendation of the student's work.

Литература References

- Aushra, A. (2006). Scientific e-library as instrument for detecting plagiarism. *Educational Technology & Society* 9 (3).
- Benno, S., & Moshe, K. (2011). *Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection PAN'07* (pp. 68-71). ACM SIGIR Forum. http://sigir.org/files/forum/2007D/2007d_sigirforum_stein.pdf

- Bouville, M. (2008). *Plagiarism: Words and ideas. Science and Engineering Ethics*. Science and Engineering Ethics. Downloaded from <https://arxiv.org/pdf/0803.1526.pdf>
- Doreen, S., & Grace, G. (2011). *ACT BSSS What is plagiarism?* (pp. 12-19) South Australia: ACT Board of Senior Secondary Studies. Downloaded from https://www.um.edu.mt/__data/assets/pdf_file/0018/261324/avdplagiarism.pdf
- Dyagilev, V., Say, A., & Butakov, S. (2011). *The architecture of service of detecting plagiarism*. Moscow: Vestnik NGU.
- Johnston, B. (2008). *The concept of plagiarism*. UK: Learning, Teaching unit. Downloaded from <http://www.celt.mmu.ac.uk/ltia/issue4/johnston.shtml>
- Leong, A., Lau, H., & Rynson, W. H. (1997). *Check: A Document Plagiarism Detection System* (pp. 70-77). Hong Kong: The Hong Kong Polytechnic University. <http://www.cs.cityu.edu.hk/~rynson/papers/sac97.pdf>
- Monostori, K., Zaslavsky, A., & Schmidt, H. (2000) *Document Overlap Detection System for Distributed Digital Libraries*. South Korea: Chung-Ang university. School of computer science & Engineering. Downloaded from https://link.springer.com/chapter/10.1007%2F11846406_83
- Sharapov, R. V., & Sharapova, E. V. (2011) *Checking system for detecting of loan from other sources*. Russian scientific conference. E-libraries: Perspective methods and technologies of the e-collections.
- Sorokina, D., Davis, J., & Garcia-Molina, H. (2014) *Plagiarism detection in arXiv*. USA: Cornell university
- Ushakin, S. (2001) *Plagiarism? About scientific ethics*. Science and contemporaneity.