

DIGITAL TECHNOLOGIES FOR ANALYZING AND EVALUATING THE LEVEL OF EDUCATIONAL RESULTS OF STUDENTS

S. V. Tarkhov ^{1a}, L. M. Tarkhova ^{2b}

¹ Ufa State Aviation Technical University (UGATU)

² Bashkir State Agrarian University (BSAU)

^a tarkhov@inbox.ru, ^b tarkhova@inbox.ru

Submitted 2022, August 12

Abstract. A set-theoretic mathematical model of content analysis of students' educational works is presented. A method of information support for the decision-making process on the quality of students' educational work has been developed, based on the use of digital technologies for analyzing and evaluating the level of students' educational results. A software implementation of a multifunctional text analyzer has been developed. The algorithms of the program and its description are given. Content analysis is carried out on the basis of highlighting the key blocks of the student's analyzed academic work. The basic objects are words and word forms made up of the symbols of the alphabets of the natural language and the alphabet artificially created for the purposes of analysis.

Keywords: computer content analysis; semantic analysis; automated text analysis; electronic document; contextual analysis of documents; Porter stemming; digital technologies; information technology; decision support; congruence of texts; multifunctional text analyzer; text analysis algorithm.

ЦИФРОВЫЕ ТЕХНОЛОГИИ АНАЛИЗА И ОЦЕНИВАНИЯ УРОВНЯ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ ОБУЧАЮЩИХСЯ

С. В. Тархов ^{1а}, Л. М. Тархова ^{2б}

¹ ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

² ФГБОУ ВО «Башкирский государственный аграрный университет» (БГАУ)

^a tarkhov@inbox.ru, ^b tarkhova@inbox.ru

Поступила в редакцию 12.08.2022

Аннотация. Приведена теоретико-множественная математическая модель контент-анализа учебных работ обучающихся. Разработан метод контент-анализа выполненных обучающимися «учебных работ», обеспечивающий информационную поддержку процесса анализа и оценивания уровня достижения обучающимися образовательных результатов и принятия решений о качестве «учебных работ», базирующийся на применении цифровых технологий. Разработана программная реализация мультифункционального анализатора текстов. Приведены алгоритмы работы программы и ее описание. Контент-анализ проводится на основе вы-

деления ключевых блоков анализируемой учебной работы обучающегося. Базовыми объектами являются слова и словоформы, составленные из символов алфавитов естественного языка и искусственно созданного для целей анализа алфавита.

Ключевые слова: компьютерный контент-анализ; семантический анализ; автоматизированный анализ текстов; электронный документ; контекстный анализ документов; стемминг Портера; цифровые технологии; информационные технологии; поддержка принятия решений; конгруэнтность текстов; мультифункциональный анализатор текстов; алгоритм анализа текста.

ВВЕДЕНИЕ

Современный уровень развития цифровых информационных технологий закономерно привел к тому, что большинство работ, выполняемых обучающимися, представляется в электронном виде [1, 2]. Работа преподавателей в образовательных учреждениях различного уровня неразрывно связана с обработкой учебных и научно-технических работ обучающихся с целью определения их качества и требует существенных затрат времени. Процесс обработки заключается в детальном и всестороннем анализе содержательной части таких работ с выставлением оценки, отражающей степень качества выполненных обучающимися работ [3]. Учебные и научно-технические работы обучающихся представляются в цифровом формате данных в виде: рефератов, пояснительных записок к курсовым работам и проектам, выпускных квалификационных работ, статей, докладов, диссертаций и др. (далее в статье «учебные работы»). Традиционные подходы к проведению анализа содержательной части «учебных работ» трудоемки и во многом субъективны и требуют существенной формализации с целью обеспечения эффективной информационной поддержки при принятии решений об оценке их качества [4, 5]. Повысить качество анализа «учебных работ», значительно снизить непродуктивные затраты времени на их формальную проверку, а также уменьшить число ошибок, неизбежно допускаемых в процессе их проверки, можно посредством применения компьютерных технологий [6]. Современные компьютерные технологии позволяют обеспечить всестороннюю информационную поддержку процессов формализованного автоматизированного анализа текстов «учебных работ» и поддержки принятия решений при оценке их качества.

В настоящее время проблема автоматизированного семантического анализа текстовых документов является предметом научных исследований как зарубежных, так и российских ученых. Существенное внимание уделяется проблемам кластеризации и категоризации текстовых документов, позволяющей реорганизовать большие коллекции документов в меньшее количество управляемых кластеров [7]. Специфика представления текстовой информации достаточно сильно зависит от сферы ее применения, а также семантического содержания конкретного документа [8]. В этой связи задача проведения анализа произвольной категории текстовых документов представляется плохо формализуемой и достаточно сложной [9]. В то же время, задача анализа и оценки научно-технической документации несколько упрощается, поскольку ее структура, состав и оформление регламентированы определенными требованиями или стандартами. При этом, в большинстве случаев для оценки могут быть применены как количественные, так и качественные критерии. Зачастую объем информации, представленной в различных документах, настолько велик, что сравнение их семантики даже по принципу сопоставления отдельных фрагментов, частоты их упоминания, а также других показателей, поддающихся четко формализованному анализу, является достаточно трудоемкой задачей, требующей большого количества времени [10]. Использование современных информационных технологий позволяет в достаточной мере упростить эту задачу, применив методы многофункционального автоматизированного анализа текста документа [11].

Для анализа и оценки содержательной части текстовых документов в настоящее время используют два основных подхода: традиционный (качественный) и формализованный (количественный). Традиционный (качественный) подход базируется на определенной последовательности логических действий эксперта, целью которых в конечном итоге является раскрытие содержания документа. При этом результаты анализа и преследующей оценки будут во многом зависеть от точки зрения и опыта эксперта, а также целей и условий проводимого исследования. Таким образом, ключевым недостатком традиционного (качественного) подхода является субъективность результатов оценки документа экспертом. Формализованный (количественный) подход основан на возможности квалитметрии, т.е. измерении характеристик содержащейся в «учебной работе» информации. При этом могут использоваться различные параметры и индикаторы, такие, как, частота использования тех или иных терминов, процент цитирования источников, приведенных в списке литературы, количества совпадений фрагментов текста в двух и более анализируемых документах (проверка на плагиат), конгруэнтность разделов документа и др.

В настоящее время для автоматизации процесса анализа текстов документов разработаны и используются различные программные инструментальные средства и on-line сервисы. Они различаются по целям и задачам. Так, например, известны программы и интернет сервисы, предназначенные для семантического анализа и проверки текста документа на уникальность, среди которых следует отметить Advego Plagiatus [12]. Широко используются также on-line сервисы и программы подсчета статистических характеристик текста, например, TextAnalyze [13]. Эффективным инструментальным средством, предназначенным для поиска и определения степени совпадения текстов документов является программа Fast Duplicate File Finder [14]. Одним из методов формальной количественной оценки содержащейся в документе информации выступает контент-анализ, позволяющий оценивать смысловое подобие текстов документов [15]. Существенный интерес в рамках данной статьи представляет работа «Инструментальное средство оценки качества научно-технических документов» [16]. В ней для оценки качества научно-технических документов авторы используют комбинированный подход, который позволяет учитывать различные категории автоматически рассчитываемых характеристик качества документов: библиометрические и наукометрические характеристики; характеристики, полученные в результате семантического анализа текстов научно-технических документов с применением эвристических правил; характеристики, полученные на основе методов оценки наличия прямых текстовых заимствований (плагиата). Предложенная авторами реализация экспериментальной системы направлена на повышение качества оценки научно-технических документов.

В настоящей статье рассматриваются новая версия программной реализации мультифункционального анализатора текстов Multifunctional Text Analyzer [17], а также положенные в его основу модели и алгоритмы, информационной поддержки процессов анализа и оценки «учебных работ». В отличие от публицистических и художественных текстов, «учебные работы» содержат четко сформулированные определения и однозначный понятийный аппарат, а также достаточно строгую структуру. В них, как правило, не допускаются синонимы для базовых понятий, что несколько упрощает процесс семантического анализа текста «учебных работ».

Цель исследования. Повышение эффективности процесса анализа и объективности принятия решений о выставлении итоговой оценки, отражающей степень качества выполненных обучающимися учебных и научно-технических работ («учебных работ») на основе применения программного продукта автоматизированного анализа текстов.

Задачи исследования:

1. Разработать метод контент-анализа выполненных обучающимися «учебных работ», обеспечивающий информационную поддержку процесса оценивания уровня достижения обучающимися образовательных результатов и принятия решений о качестве «учебных работ».

2. Разработать алгоритмы и специальное программное обеспечение, реализующие цифровые технологии анализа и оценивания уровня образовательных результатов обучающихся на основе контент-анализа выполненных обучающимися «учебных работ».

3. Выполнить исследование эффективности специального программного обеспечения, разработанного на основе предложенных методов и алгоритмов на примере оценивания качества «учебных работ» студентов, обучающихся по техническим специальностям и направлениям подготовки.

МЕТОД ИССЛЕДОВАНИЯ

Участники эксперимента

Исследование проводилось на базе:

– кафедры информатики федерального государственного образовательного учреждения высшего образования «Уфимский государственный авиационный технический университет» (УГАТУ);

– кафедры прикладной механики и компьютерного инжиниринга федерального государственного образовательного учреждения высшего образования «Башкирский государственный аграрный университет» (БГАУ).

Экспериментальное исследование посвящено проверке эффективности специального программного обеспечения – мультифункционального анализатора текстов, разработанного на основе предложенных методов и алгоритмов. С использованием мультифункционального анализатора текстов оценивалось качество «учебных работ», выполненных студентами вторых-пятых курсов очной и заочной форм обучения, обучающимися по техническим специальностям и направлениям подготовки в УГАТУ и БГАУ, а также выпускных квалификационных работ, выполненных студентами Бирского филиала БашГУ. Исследование проводилось с 2019–2020 по 2021–2022 учебный год.

Инструментарий исследования

Оценка качества «учебных работ» проводилась как традиционным методом анализа содержательной части представленных студентами электронных документов, так и с использованием разработанного мультифункционального анализатора текстов Multifunctional Text Analyzer. Для сбора данных использовались контрольные листы, в которых фиксировались результаты оценки «учебных работ». Все собранные данные результатов анализа «учебных работ», выполненного традиционным методом, вводились в рабочую книгу табличного процессора MS Excel. Данные, полученные в программе Multifunctional Text Analyzer также сохранялись в формате MS Excel. Систематизация, структурирование и статистическая обработка экспериментальных данных, собранных с использованием контрольных листов и полученных в программе Multifunctional Text Analyzer выполнялась в табличном процессоре MS Excel.

Метод контент-анализа и оценки качества «учебных работ»

Информационная поддержка процесса комплексного анализа и оценки качества «учебных работ» включает два ключевых этапа:

– семантический анализ текста «учебной работы»;

– количественную оценку параметров (показателей и индикаторов), определяющих качество «учебной работы».

Под качеством «учебной работы» в соответствии со стандартом ISO 9000 будем понимать степень соответствия совокупности присущих ей характеристик установленным требованиям.

Математическую модель, положенную в основу метода анализа и оценки качества «учебных работ», построим с использованием теоретико-множественного подхода. Для этого выполним структуризацию «учебной работы», представив ее содержательную часть на первом уровне детализации как конечное упорядоченное множество D . В свою очередь, конечное упорядоченное множество D будет состоять из подмножеств, представляющих собой логически и семантически завершенные базовые фрагменты F . В «учебной работе», с учетом ее специфики, базовые фрагменты F имеют вполне определенный и при этом достаточно жестко регламентированный порядок следования. Как и в большинстве научно-технических документов, «учебная работа» включает: N – название; A – аннотацию; K – ключевые слова; I – введение; P – основную часть, представленную в виде нескольких глав и параграфов (разделы, подразделы); S – выводы по ним; C – заключение; L – список использованной литературы (библиографию); App – приложения.

$$D \in (\{ F_1, F_2, \dots, F_n \} \vee \{ N, A, K, I, P, S, C, L, App \}). \quad (1)$$

После того, как «учебная работа» формально представлена, как совокупность n базовых фрагментов F , (по вполне понятным причинам, это достаточно условно, поскольку она, как уже было сказано ранее, по умолчанию должна иметь заранее определенную структуру и необходимо лишь проанализировать их вид, а также количество) следует определить совокупность критериев, по которым она будет анализироваться и оцениваться. При этом можно утверждать, что для каждого подмножества F_i существует множество критериев Y , таких, что Y функционально зависит от F_i . При этом один и тот же критерий Y_i может применяться к нескольким фрагментам F_i «учебной работы». В то же время, одни и те же фрагменты F могут быть проанализированы по множеству критериев Y . Поскольку базовые фрагменты F «учебной работы» включают в свой состав множества текстовых фрагментов T_i и объектов O_k (рисунки, схемы, графики, диаграммы, формулы и т.д.), запишем

$$D = T_i \subset O_k. \quad (2)$$

Проводя более детальную структуризацию, выделим в подмножествах F_i текстовые фрагменты T_p , состоящие из множества символов (знаков) S , приведенных в табл. 1.

Таблица 1. Символы, используемые в текстовых электронных документах

Обозначение	Характеристика	Назначение	Пояснение
AL	Символы искусственного алфавита	Формирование слов	Цифры, знаки математических операций др.
ALN	Символы естественного алфавита языка	Формирование слов	Латинский алфавит, алфавит кириллицы и др.
PM	Знаки препинания	Формирование семантических конструкций	Выделения предложений, слов и словосочетаний; акцентирования на значимости и эмоциональной окраски предложений; указания на логические и грамматические отношения между словами; указания на законченность и коммуникативный тип предложения и др.;
SC	Специальные символы	Разделители и служебные символы	Звездочка (*), маркеры оформления списков (буллит) (•) или тире (–), коммерческое at (@), знак решетки (окто-торп #), знак номера (№), знак параграфа (§), амперсанд (&), знаки валют (₽, €, ¥ и др.), знаки интеллектуальной собственности (©, ®, ™), знаки разметки электронного документа (¶, ↵, →).

Ключевым объектом анализа содержательной части «учебной работы», представленной в виде D в каждой группе ее базовых фрагментов F , включающих текст T_i выступают слова W . При этом, элементы множества PM (знаки препинания) выполняют семантические функции разделителей слов. Разделителями слов могут быть также некоторые специальные символы множества SC , такие как, например, табулятор, знак абзаца, символ перехода на новую строку и др. Следует отметить, что в большинстве случаев в процессе анализа нет необходимости учитывать различие прописных и строчных букв алфавита естественного языка ALN .

В процессе подготовки к проведению семантического анализа, выполним преобразование текстовых фрагментов T_i в специальные текстовые фрагменты T_i^W , для чего удалим из T_i знаки препинания PM , оставив в них пробелы $\{ \}$, как символы разделителей слов, при этом сохранив в T_i^W только слова, составленные из символов алфавита ALN ,

$$T_i^W = (T_i \setminus (PM \setminus \{ \})) \cap (ALN \cup \{ \}) . \quad (3)$$

Множество специальных текстовых фрагментов T_i^W представим, как вектор \bar{W} , количество элементов которого равно мощности множества T_i^W .

В соответствии с выбранным ранее множеством критериев Y , в соответствии с которыми предполагается провести анализ содержательной части совокупности базовых фрагментов F «учебной работы» D , на основе выделения множества слов W , составленных из символов алфавита естественного языка ALN и используя сформированный вектор \bar{W} , можно выполнить:

- анализ библиографических ссылок в списке использованной литературы «учебной работы» D на предмет их оформления в соответствии с требованиями стандарта;
- анализ актуальности библиографических ссылок с учетом года опубликования работ, включенных в список использованной литературы «учебной работы» D ;
- поиск ссылок на литературу (в квадратных скобках) в тексте «учебной работы» D , с учетом их оформления в соответствии с требованиями стандарта;
- анализ полноты использования в тексте «учебной работы» D ссылок на литературу, приведенную в списке использованной литературы;
- поиск абзацев в соответствии с заданным поисковым образом во фрагментах текста T_i , или в тексте «учебной работы» целом D с учетом/без учета регистра символов;
- поиск слов T_i^W (на выбранном языке с учетом ограничения их длины) во фрагментах текста T_i (тексте в целом), из выбранных базовых фрагментов F «учебной работы» D ;
- анализ конгруэнтности фрагментов F или выбранных фрагментов текста T_i , «учебной работы» D , с использованием алгоритмов поиска основ слов, встречающихся в тексте.

Метод комплексного анализа «учебных работ» предусматривает выделение в них базовых фрагментов F , которые фактически представляют собой разделы «учебной работы», или совокупность разделов, объединенных по их назначению (семантическому содержанию). Базовыми объектами в процессе анализа «учебных работ» D являются слова (словоформы) W , составленные из алфавитов естественного языка ALN , а в отдельных случаях и искусственно созданного для целей анализа алфавита. Процесс автоматизированного анализа и последующей оценки текста «учебных работ» D состоит из пяти основных этапов:

1. Анализ «учебной работы» в целом с выделением ее структурных компонентов, которые будут служить основой для сопоставления и проведения контент-анализа по отдельным фрагментам, таким как, введение и заключение, главы и выводы по ним, аннотация и т.п., перечень использованной литературы.

2. Выбор и обоснование критериев, по которым в дальнейшем будет выполняться поиск контента в «учебной работе», проводится анализ и формируется итоговая оценка качества «учебной работы», как степень ее соответствия установленным требованиям.

3. Формирование списка поисковых образов (шаблонов) в соответствии с множеством сформированных критериев для анализа и оценки «учебной работы».

4. Поиск по «учебной работе» в целом, а также ее отдельным разделам или совокупности разделов на основе сформированных поисковых образов (шаблонов).

5. Определение оценок по отдельным критериям и группам критериев, а также формирование интегральной (итоговой) оценки «учебной работы» на основе полученных численных результатов поиска по тексту.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Структура и алгоритмы программы мультифункционального анализатора текстов

Поддержка принятия решений при определении уровня качества «учебных работ», выполненных обучающимися, основана на комплексном и всестороннем анализе их текстов с использованием разработанной авторами программы МТА (Multifunctional Text Analyzer). Программа МТА позволяет автоматизировать все основные этапы проверки (анализ и оценку) «учебных работ». Результатом проверки являются рекомендации лицу, принимающему решение о соответствии «учебной работы» установленным требованиям, а также информация о степени ее качества (количественная оценка значимых параметров). Состав модулей, включенных в программный продукт МТА, показан на рис. 1.



Рис. 1. Модули программы Multifunctional Text Analyzer

Рассмотрим назначение, базовые возможности, и основные результаты работы алгоритмов модулей программы МТА, предоставляющих экспертам информацию для принятия решений об уровне качества «учебных работ».

Модуль поиска слов на определенном (в рамках локализации программы) языке в тексте «учебной работы» позволяет искать:

- ключевые слова, а также слова и словосочетания, встречающиеся в названии и аннотации;
- произвольные слова и словосочетания, заданные экспертом.

На рис. 2 показаны основные процедуры модуля поиска слов на выбранном языке (например, русских слов с учетом ограничения их длины) во фрагментах текста или в тексте в целом, из выбранных разделов F «учебной работы». Модуль позволяет реализовать перечисленные ниже функции этапа анализа «учебной работы», связанные с поиском русских слов:

- выполнить анализ использования ключевых слов, словосочетаний, словоформ (включая числа), с учетом установленных ограничений на количество символов в словах;
- определить количество найденных искомых слов, словосочетаний, словоформ;
- сформировать список искомых слов, словосочетаний, словоформ с указанием количества.

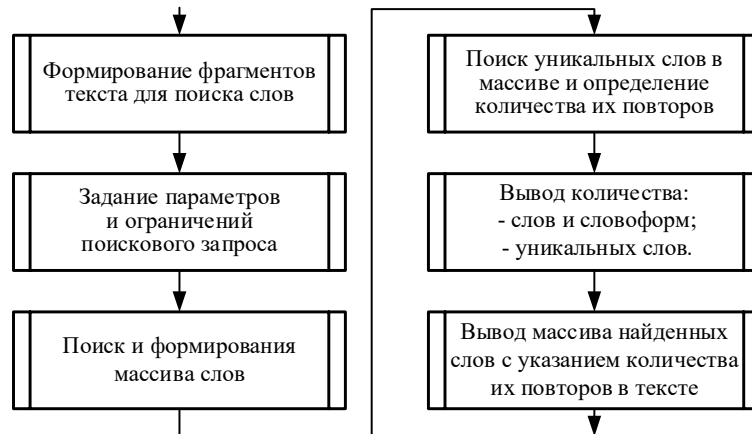


Рис. 2. Алгоритм работы модуля поиска слов, словосочетаний, словоформ

Извлечение тематически зависимого контента позволяет анализировать содержимое, которое имеет отношение к образцам тематических аспектов, выделенных пользователями, и формировать рейтинг релевантности предложения в соответствии с конкретным тематическим аспектом. На рис. 3 показаны процедуры модуля, позволяющего на основе заданного поискового образа искать абзацы текста T_p , выбранных фрагментах F «учебной работы».

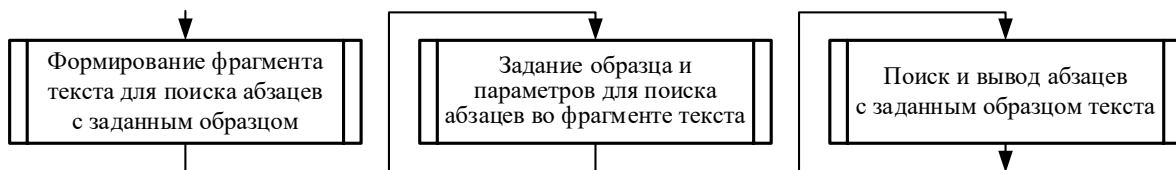


Рис. 3. Алгоритм работы модуля поиска абзацев, содержащих заданный поисковый образ

Поиск может осуществляться как с учетом, так и без учета регистра символов естественного алфавита ALN . Итогом работы алгоритма будет текст, сформированный из найденных абзацев. Этот текст при необходимости можно скопировать в буфер обмена, а также сохранить в файл. Для удобства дальнейшей работы со сформированным текстом заданный ранее поисковый образ будет отмечен специальным символом или же группой символов в зависимости от предпочтений пользователя. Это позволяет визуально находить интересующий эксперта фрагмент и выполнять дальнейшую неформальную оценку содержания анализируемой «учебной работы».

Модуль поиска абзацев с заданным образцом текста программы МГА перед выполнением поиска определяет количество абзацев в анализируемом фрагменте «учебной работы». Модуль позволяет реализовать перечисленные ниже функции анализа наличия ссылок на библиографические записи из списка использованной литературы:

- рассчитать полноту использования в «учебной работе» ссылок (в процентном отношении) на библиографические записи из списка использованной литературы;
- определить общее количество ссылок на библиографические записи из списка использованной литературы; количество уникальных ссылок;
- сформировать массив записей полного списка библиографических ссылок, имеющих в тексте «учебной работы» и оформленных по требованиям стандарта (указанных в квадратных скобках);
- определить количество ссылок на каждую библиографическую запись, приведенную в списке использованной литературы в тексте «учебной работы»;

– сформировать перечень номеров и указать библиографические записи из списка использованной литературы, на которые в тексте «учебной работы» нет ссылок.

Корректность и полнота цитирования является одним из ключевых показателей «учебной работы». На рис. 4 показаны процедуры модуля поиска ссылок на литературу в тексте «учебной работы», приведенных в квадратных скобках в соответствии с требованиями стандарта (ГОСТ 7.0.5–2008).

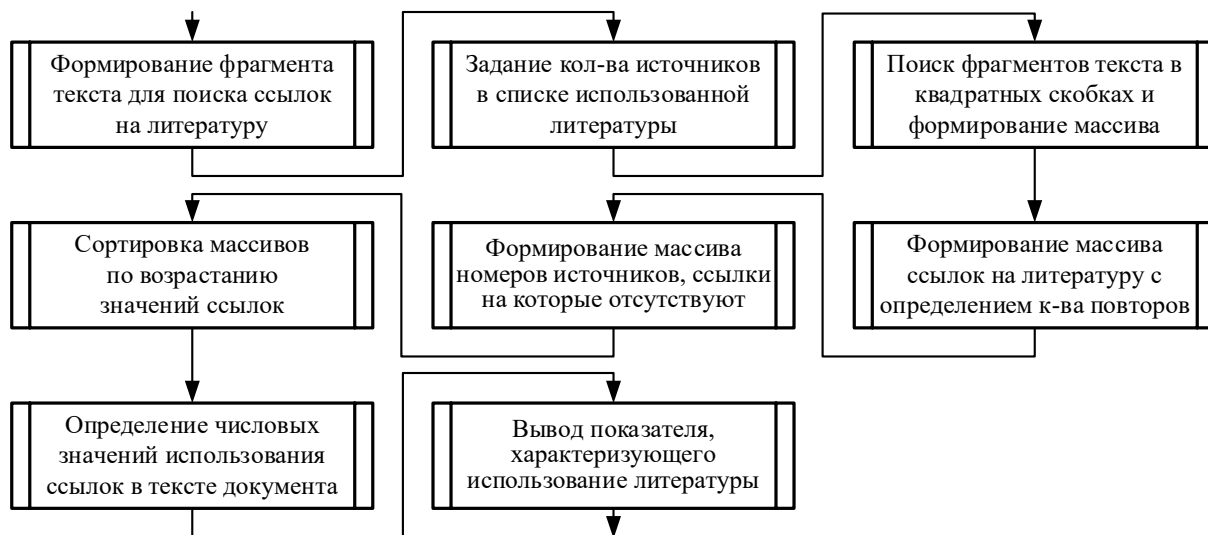


Рис. 4. Алгоритм работы модуля поиска ссылок на библиографические записи из списка использованной литературы

В программе МТА предусмотрена возможность сохранения в структурированном текстовом файле результатов анализа наличия ссылок на библиографические записи из списка использованной литературы. Это позволяет в дальнейшем использовать табличный процессор для более детальной и всесторонней обработки полученных данных, в том числе с использованием средств иллюстративной графики.

Контекстный анализ научно-технических документов с учетом количественных показателей качества (информационная емкость, значимость и независимость содержания), а также традиционных библиометрических и наукометрических показателей (индекса цитирования документа и импакт-фактора журнала) обеспечивает объективную оценку качества документа. На рис. 5 показаны основные этапы работы алгоритма модуля анализа списка использованной литературы, приведенного в «учебной работе».



Рис. 5. Алгоритм работы модуля анализа списка использованной литературы

Список использованной литературы должен быть представлен в «учебной работе» в форме затекстовых библиографических ссылок, оформленных в соответствии со стандартом.

Модуль позволяет реализовать перечисленные ниже функции анализа библиографических записей, содержащихся в списке использованной литературы «учебной работы»:

- по заданному поисковому образу, представленному в формате текстовой строки, выполнить поиск интересующих эксперта библиографических записей;
- выделить в списке использованной литературы библиографические записи, соответствующие указанному экспертом месту опубликования работы (например, журналу);
- определить количество библиографических записей, соответствующих публикации результатов исследований в материалах конференций;
- сформировать массив данных, содержащий информацию о библиографических записях, содержащихся в списке использованной литературы «учебной работы» по каждому году опубликования работы.

Поисковые запросы анализа списка использованной литературы могут выполняться в определяемом пользователем диапазоне дат. По умолчанию установлена нижняя граница поиска 1900 год и верхняя граница поиска – текущий год.

Алгоритм модуля анализа списка использованной литературы «учебной работы» позволяет сформировать выборку из библиографических записей. Они могут быть удобны для последующего анализа экспертом заинтересовавших его библиографических записей. Такая выборка может быть создана на основе поискового образа, содержащего перечень номеров и/или их диапазонов, например: «2, 13–15, 34, 41». В этом случае будут показаны библиографические записи с номерами 2, 13, 14, 15, 34, 41.

Следует также отметить определенную сложность и неоднозначность алгоритма формального поиска библиографических записей в списке использованной литературы. Она обусловлена тем, что в библиографической записи могут содержаться сведения о более чем одной дате с указанием года, причем как одинаковые, так и различные по значению. В этом случае для получения адекватного результата анализа следует выполнить дополнительную проверку списка библиографических записей, содержащих две и более дат в своей записи. Модуль анализа списка использованной литературы программы МТА позволяет находить такие библиографические записи и предупреждать об их наличии. При этом может быть сформирован отдельный список таких библиографических записей. В качестве примера приведем библиографическую ссылку «Tarkhov S. V., Tarkhova L. M., Shagiyeva Yu. R. Information support of decision-taking in staff recruitment management process based on complex diagnostics // Proceedings of the 15th Workshop Computer Science and Information Technologies (CSIT) 2013 (Ufa –Vienna – Budapest – Bratislava). 2013. Vol. 2. Pp. 138-142». Она содержит две одинаковые даты и автоматический их учет при анализе по годам опубликования приведет к неверному результату.

В программе МТА предусмотрена возможность сохранения результатов анализа списка использованной литературы в структурированном текстовом файле. Специальные разделители позволяют открыть этот файл в табличном процессоре для более детальной и всесторонней обработки полученных данных, в том числе с использованием средств иллюстративной графики.

Алгоритм базируется на поиске основ слов (части слова, представляющей собой его неизменяемую часть, выражающую его лексическое значение), встречающихся в тексте. Для реализации процесса поиска основы слова в заданном исходном слове (словоформе) применяется стемминг Портера [18]. Модуль анализа конгруэнтности текстовых фрагментов программы МТА позволяет одновременно сравнивать до четырех заранее выбранных фрагментов текста.

Это могут быть любые фрагменты «учебной работы», в том числе ее название и ключевые слова. Модуль анализа конгруэнтности позволяет определить логическое и семантическое соответствие текстовых фрагментов «учебной работы»:

- общее количество слов и словоформ в анализируемых фрагментах с учетом заранее заданного ограничения на длину слов;
- количество повторно встретившихся в каждом анализируемом фрагменте найденной основы слова или словоформы;
- количество уникальных основ слов и словоформ в анализируемом фрагменте;
- отношение количества одинаковых слов, основ слов и словоформ в двух сравниваемых между собой фрагментах, к их общему количеству в одном из них – базовом фрагменте, т.е. рассчитать конгруэнтность фрагментов текста, выраженную в процентах;
- комплексный (итоговый) показатель конгруэнтности сравниваемых фрагментов текста с учетом повтора слов, выраженный в процентах;
- комплексный (итоговый) показатель конгруэнтности сравниваемых фрагментов текста без учета повтора слов, выраженный в процентах.

На рис. 6 показаны основные этапы работы алгоритма модуля анализа конгруэнтности фрагментов текста, скопированных из «учебной работы» в рабочее окно программы МТА.

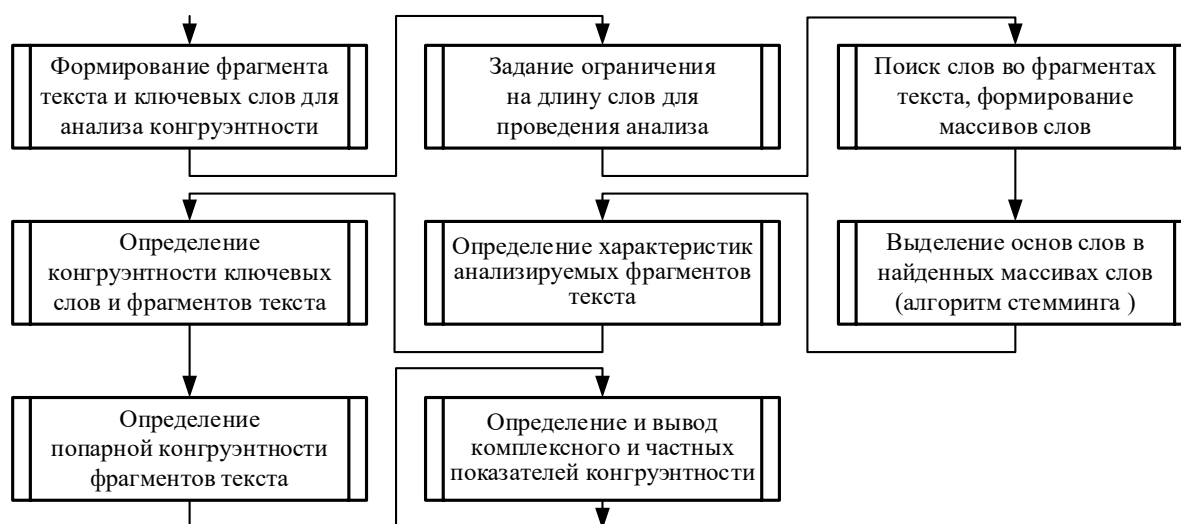


Рис. 6. Алгоритм работы модуля анализа на конгруэнтность текстовых фрагментов

Результаты анализа на конгруэнтность текстовых фрагментов могут быть сохранены в структурированном текстовом файле и в рабочей книге электронной таблицы.

На базе оригинального метода автоматизированного цифрового анализа и оценки текстов создан программный продукт МТА, позволяющий реализовать поддержку принятия решений экспертами (преподавателями или сотрудниками) об уровне качества «учебных работ». При том следует отметить, что программный продукт МТА лишь упрощает за счет использования средств автоматизации контент-анализ «учебных работ», а не полностью его автоматизирует. Не стоит также полагаться исключительно на результаты анализа «учебных работ» в программе МТА. Оптимального результата при оценке качества «учебных работ» можно достичь исключительно на основе комплексного подхода, предусматривающего применение как субъективных (экспертных) методов оценки, так объективных (качественных) методов оценки (методов квалиметрии) с использованием компьютерных инструментальных средств.

Поддержка принятия решений об уровне достижения обучающимися образовательных результатов

Математическую модель поддержки принятия решений об уровне достижения обучающимися образовательных результатов с использованием экспертных оценок построим на основе

применения методов нечеткой логики. Для этого зададим функцию принадлежности и определим интервальные значения (уровни) ранговой шкалы. В такой шкале фактическим значениям соответствующих показателей и индикаторов придается конкретный смысл, связанный с выполняемой оценкой работы.

Функция принадлежности с девятиуровневой ранговой шкалой Харрингтона (с основными и промежуточными (дополнительными) уровнями), позволяет экспертам выполнять формализованную оценку «учебных работ».

$$P = e^{-e^{-R}}, \quad (4)$$

где P – значения шкалы предпочтений; R – значения лингвистической шкалы $L = [-4; 4]$, разбитой на несколько уровней [1; 9].

Ранговая шкала придает конкретный смысл фактически измеренным значениям показателей и индикаторов, характеризующих качество оцениваемой «учебной работы». Значение функции принадлежности $P = 0$ характеризует уровень качества «учебной работы» как неприемлемый. Значение функции принадлежности $P = 1$ для девятиуровневой лингвистической шкалы $L = [-4; 4]$ и функции принадлежности (4) ($P = 0,981851073$), в полной мере соответствует абсолютному уровню качества «учебной работы». Значения показателей для проведения экспертных оценок «учебных работ» на основе представленной ранговой шкалы приведены в табл. 2.

Таблица 2. Характеристики ранговой шкалы для оценки «учебных работ»

Уровни шкалы	Значения функции принадлежности	Значения показателя
Основные уровни ранговой шкалы		
9	0,80–1,00	Высокий
7	0,64–0,80	Хороший
5	0,37–0,64	Средний
3	0,20–0,37	Низкий
1	0,00–0,20	Минимальный
Дополнительные уровни ранговой шкалы		
2, 4 6, 8	Промежуточные значения	«Выше чем» «Не достаточно»

Множество показателей и индикаторов S , выраженных в числовой (количественной) S_q и в формальной качественной (текстовой) S_c форме позволяет определить общую (интегральную) оценку качества «учебной работы». Для практической реализации процесса поддержки принятия решений об уровне качества «учебных работ» введем функцию принадлежности $\vartheta(x)$, где x – значение параметра или индикатора, характеризующего конкретный анализируемый показатель качества. Выбор вида функции $\vartheta(x)$ в значительной мере субъективен и зависит от разработчика системы поддержки принятия решений. Желательно, чтобы в информационной системе поддержки принятия решений, предназначенной для оценки качества «учебных работ» такая возможность была предоставлена лицу, принимающему решение (эксперту).

Для вычисления итогового показателя по всему множеству параметров и индикаторов, характеризующих «учебную работу» по показателям, выраженным в качественной форме S_c , вычислим сумму.

$$\theta_c = \sum_{i=1}^{S_c} \vartheta_i^c(x). \quad (5)$$

При оценке показателей и индикаторов, выраженных в количественной форме S_q , значение функции $\vartheta(x)$ по выбранной заранее шкале измерений будет тождественно фактическому измеренному значению параметра. В этом случае значение агрегирующей функции принадлежности для множества показателей и индикаторов, характеризующих «учебную работу», вычислим как сумму значений функций принадлежности $\vartheta(x)$ по совокупности отдельных значений параметров и индикаторов

$$\theta_c = \sum_{i=1}^{S^c} \vartheta_i^c(x). \quad (6)$$

Интегральный показатель, характеризующий уровень качества «учебной работы» (общая оценка «учебной работы»), определим по формуле

$$\psi = \frac{\sum_{i=1}^{S^q} \vartheta_i^q(x) + \sum_{i=1}^{S^c} \vartheta_i^c(x)}{\sum_{i=1}^{S^q} \vartheta_i^q(x^{ref}) + \sum_{i=1}^{S^c} \vartheta_i^c(x^{ref})}, \quad (7)$$

где $\vartheta_i^q(x^{ref})$ и $\vartheta_i^c(x^{ref})$ – эталонные значения количественных и качественных показателей и индикаторов.

Метод экспертных оценок и алгоритмы его реализации планируется включить в следующую версию программного продукта МТА. В настоящее время описанная выше модель реализована средствами анализа данных в рабочей книге табличного процессора. Результаты контент-анализа «учебных работ» также выгружаются из программного продукта МТА в рабочую книгу табличного процессора, что в итоге позволяет обеспечить поддержку принятия решений при совместном использовании квалиметрического и экспертного подхода к оценке качества «учебных работ».

Разработанный метод контент-анализа «учебных работ» обеспечивает информационную поддержку процесса формализованного оценивания уровня достижения обучающимися образовательных результатов и принятия решений о качестве «учебных работ». В основу метода положен принцип разбиения текста «учебной работы» на логически и семантически завершенные базовые фрагменты с учетом ее специфики. Глубина деления текста на фрагменты – вплоть до слов и словоформ, составленных из символов алфавитов естественного языка и искусственно созданного для целей анализа алфавита, что позволило применить как традиционные алгоритмы поиска и семантического анализа текстов (наличие фрагментов, соответствующих поисковым образам; их количество; местонахождение в тексте), так и модифицированный алгоритм стемминга Портера для анализа конгруэнтности текстовых фрагментов.

Ключевое внимание в данной статье уделено совершенствованию механизмов цифрового анализа и оценки качества «учебных работ», проводимых экспертами с использованием программного продукта МТА. Отличительной особенностью использования программного продукта МТА является возможность всестороннего многофакторного анализа «учебных работ». Ключевая идея метода анализа и оценки текстов, положенного в основу программного продукта МТА базируется на поиске и выделении слов и словоформ в естественных и формальных алфавитах. Для поиска основ слов (например, в алфавите кириллицы) применен стемминг, являющийся модифицированным алгоритмом стеммера Портера [19]. В процессе анализа библиографических записей, приведенных в списках литературы «учебных работ», выделяются слова, сформированные на основе искусственного алфавита, с учетом правил, установленных соответствующими стандартами. Разработанные в процессе создания программного продукта МТА метод и реализованные на его основе алгоритмы позволяют экспертам быстро и эффективно проводить всесторонний многофакторный анализ и оценку качества «учебных работ».

Авторами с использованием программного продукта МТА было проанализировано значительное количество «учебных работ» (пояснительных записок к выпускным квалификационным работам, конкурсных научно-исследовательских работ, диссертаций). Анализ снижения трудозатрат преподавателей, оценивающих «учебные работы» показал, что время, затрачиваемое на формализованную проверку снизилось в среднем на 47 %. В качестве формальных критериев были приняты (рис. 7): время анализа списка использованной литературы с точки зрения ее новизны (снижение на 49 %), время анализа наличия ссылок на литературу в тексте учебной работы и полноты использования библиографических ссылок (снижение на 42 %); время поиска текстовых фрагментов в анализируемой «учебной работе» (снижение на 39 %); время анализа смыслового соответствия введения, цели и задач «учебной работы», а также ее результатов и выводов (снижение на 57 %) – в разработанном методе контент анализа учебных работ это анализ на конгруэнтность текстовых фрагментов из разных разделов работы.

При этом обеспечивался высокий уровень соответствия заключения об общем уровне качества «учебной работы» – идентичная оценка (отклонение не более 15 %) одних и тех же работ при их независимой проверке выставлялась двумя и более преподавателями-экспертами.

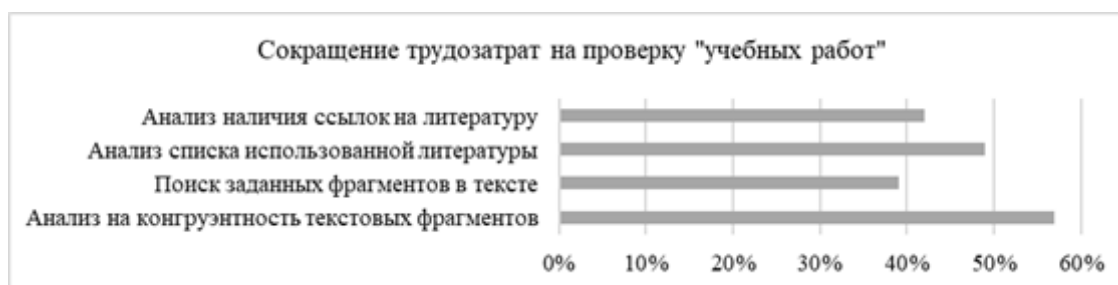


Рис. 7. Результаты оценки сокращения трудозатрат

Практическое использование программного продукта МТА при анализе качества «учебных работ» показало его высокую эффективность. Так, при оценке научно-технических работ у эксперта-рецензента появилась реальная возможность детально проанализировать использование автором работы современных достижений науки и техники на основе приведенных в тексте работы списка литературы и ссылок на библиографические записи из списка литературы. Текст анализируемых «учебных работ» может быть оценен как качественный, если комплексный показатель конгруэнтности фрагментов текста работы (название, цель, задачи, новизна, заключение (выводы)) превышает 70 %.

ЗАКЛЮЧЕНИЕ

Разработан метод контент-анализа выполненных обучающимися «учебных работ», обеспечивающий информационную поддержку процесса оценивания уровня достижения обучающимися образовательных результатов и принятия решений о качестве «учебных работ», в основу которого положен принцип разбиения текста «учебной работы» на логически и семантически завершенные базовые фрагменты с учетом ее специфики и глубиной деления вплоть до слов и словоформ, что позволило применить как традиционные алгоритмы поиска и семантического анализа текстов, так и модифицированный алгоритм стемминга Портера для анализа конгруэнтности текстовых фрагментов.

Рассмотрена практическая реализация метода многофункционального анализа и оценки текста («учебных работ») с использованием цифровых технологий. Информационная поддержка процесса анализа и оценки преподавателями-экспертами учебных научно-технических

работ обучающихся обеспечивается использованием разработанного авторами программного продукта Multifunctional Text Analyzer (МТА), который, как показало его практическое применение, позволяет:

– в значительной мере снизить влияние субъективных факторов на результаты оценки работы: идентичная оценка (отклонение не более 15 %) одних и тех же работ при их независимой проверке выставлялась двумя и более преподавателями-экспертами;

– существенно сократить непродуктивную работу преподавателей и сотрудников (экспертов), в чьи функциональные обязанности входит проверка и оценка учебных научно-технических работ: трудозатраты (в часах) уменьшились на 40–60 %;

– определять значения показателей, оценка которых без использования компьютерных технологий нецелесообразна по причине высоких трудозатрат (например, использование в тексте статьи ссылок на литературу), а в ряде случаев практически неосуществима – анализ конгруэнтности текстов, вместо которого обычно поводится анализ на смысловое соответствие.

Установлено, что анализируемая «учебная работа» может быть оценена в целом как качественная, если комплексный показатель конгруэнтности фрагментов текста работы (название, цель, задачи, новизна, заключение) превышает 70 %.

СПИСОК ЛИТЕРАТУРЫ

1. **Ashlee Humphreys, Jen-Hui Wang Rebecca.** Automated Text Analysis for Consumer Research // *Journal of Consumer Research*. 2018. Vol. 44, No. 6. Pp. 1274-1306.
2. **Infographics** and their application in the educational process / L. Tarkhova, et al. // *International Journal of Emerging Technologies in Learning*. 2020. Vol. 15, No. 13. Pp. 63-80.
3. **Kiyavitskaya N., Zeni N., Mich L.** Text mining through semi automatic semantic annotation //6-th International Conference on Practical Aspects of Knowledge Management (November 30 – December 1, 2006, Vienna, Austria). 2006. Vol. 4333. Pp. 143-151.
4. **Chakraborty B., Bhattacharjee T.** A review on textual analysis of corporate disclosure according to the evolution of different automated methods // *Journal of Financial Reporting & Accounting*. 2020. Vol. 18, No. 4. Pp. 757-777.
5. **Tools** for assessing the quality of scientific-technical documents / S. V. Gerasimov, et al. // *Proceedings of the Institute for System Programming of the Russian Academy of Sciences*. 2013. V. 24. Pp. 359-378.
6. **Kunz N.** An automated quantitative content analysis process for humanitarian logistics research // *Journal of Humanitarian Logistics and Supply Chain Management*. 2019. Vol. 9, No. 3. Pp. 475-491.
7. **Sisodia D. S., Choudhary T. M., Rai V. R.** Agglomerative Similarity Measure Based Automated Clustering of Scholarly Articles // *Machine Intelligence and Signal Analysis*. 2018. Vol. 784. Pp. 533-544.
8. **Минасова Н. С., Тархов С. В., Тархова Л. М.** Модели формирования и практическая реализация скомпилированных учебных модулей в системе электронного обучения // *Открытое образование*. 2006. № 5. С. 21–29. [N. S. Minasova, S. V. Tarkhov, L. M. Tarkhova, "Formation models and practical implementation of compiled training modules in the e-learning system", (in Russian), in *Otkrytoe obrazovanie*, no. 5, pp. 21-29, 2006.]
9. **Storage** and processing technologies of cognitive content for e-learning systems / Tarkhov S. V., et al. // *Proceedings of the 3rd Russian-Pacific Conference on Computer Technology and Applications* (August 18-25, Vladivostok, Russia). 2018. Article number 8482185.
10. **Yadav C., Sharan A.** A New LSA and Entropy-Based Approach for Automatic Text Document Summarization // *International Journal on Semantic Web and Information Systems*. 2018. Vol. 14, No. 4. Pp. 1-32.
11. **Tarkhov S. V., Minasova N. S.** Support for decision making in the evaluation of educational scientific-technical works on the basis of multifunctional automated analysis of text document // *Proceedings of the 17th Workshop Computer Science and Information Technologies (CSIT'2015)*, (Rome, Italy, September 22-26). 2015. Vol. 1. Pp. 186-190.
12. **Программа** для проверки уникальности текста Advego Plagiatus. [Электронный ресурс]. URL: <https://advego.com/plagiatus/> (дата обращения 07.07.2022). [The program for checking the uniqueness of the text Advego Plagiatus (2022, Jul. 7). [Online]. Available: <https://advego.com/plagiatus/>]
13. **Анализатор** текста. [Электронный ресурс]. URL: <https://www.textanalyzer.ru/> (дата обращения 24.01.2021). [Text analyzer (2021, Jan. 24). [Online]. Available: <https://www.textanalyzer.ru/>]
14. **Fast Duplicate File Finder.** [Electronic resource]. URL: <https://www.mindgems.com/products/Fast-Duplicate-File-Finder/Fast-Duplicate-File-Finder-About.htm> (accessed 24.01.2021).
15. **Information** content measures of semantic similarity between documents based on Hadoop system / Birjali Marouane, et al. // *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*. 2016. Pp. 187-192.

16. **Симанков В. С., Толкачев Д. М.** Автоматическая оценка семантического сходства текстов // Материалы XXXVII Международной научно-практической конференции. № 8 (33). Новосибирск: «СибАК», 2014. [Электронный ресурс]. URL: <http://sibac.info/15679> (дата обращения 24.05.2022). [V. S. Simankov, D. M. Tolkachev (2022, May 24), "Automatic evaluation semantic similarity of texts [Online], (in Russian), in *Materials of the XXXVII International Scientific and Practical Conference*, 2014, No. 8 (33). Available: <http://sibac.info/15679>]
17. **Tarkhov S. V., Minasova N. S., Kalimullina G. R.** Multifunctional text analyzer (MTA): Certificate of state registration of computer programs No. 2015612998. Moscow: Rospatent, 2015.
18. **Popovič M., Willett P.** The Effectiveness of Stemming for Natural-Language Access to Slovene Textual Data // Journal of American Society for Information Science. 1992. Vol. 43, No. 5. Pp. 384-390.
19. **Porter M. F.** An algorithm for suffix stripping // Program. 1980. Vol. 14, No. 3. Pp. 130-137.

ОБ АВТОРАХ

ТАРХОВ Сергей Владимирович, проф. каф. информатики УГАТУ. Дипл. инженер-механик (УАИ, 1980). Д-р техн. наук по упр. в соц. и экон. системах. Иссл. в обл. управления в социальных системах.

ТАРХОВА Ляйля Мукаддасовна, доц. каф. прикладной механики и компьютерного инжиниринга БГАУ. Дипл. инженер-механик (УАИ, 1980). Канд. техн. наук по упр. в соц. и экон. системах. Иссл. в обл. управления обучающими системами.

TARKHOV, Sergey Vladimirovich, Prof., Dept. of Computer Science. Dipl. Mechanical Engineer (USATU, 1980). Dr. of Tech. Sci. (USATU, 2010). Research in the area of Management in Social Systems.

TARKHOVA Lyaylya Mukaddasovna, Assoc. Prof. Head of Dept. of Mechanics and Engineering Graphics. Dipl. Mechanical Engineer (USATU, 1980). Cand. of Tech. Sci. (USATU, 2001). Research in the area of Management in Educational Systems.

Language: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 26, no. 4 (98), pp. 72-87, 2022. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).