

ОГЛАВЛЕНИЕ

А. А. Витухновская

СТРАТЕГИЯ ИНФОРМАЦИОННОГО ПОИСКА В ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМАХ И ОЦЕНКА КОМПЕТЕНЦИЙ СТУДЕНТОВ

Д. С. Зуев, М. Ф. Насрутдинов, А. Ф. Хасьянов

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ТЕКСТОВОЙ АНАЛИТИКИ ЮРИДИЧЕСКИХ ДОКУМЕНТОВ

А. О. Федоров

СИСТЕМА АВТОМАТИЗИРОВАННОГО УЧЕТА ОПЛАТЫ ОБУЧЕНИЯ В ВУЗЕ

Р. В. Хабипов

АВТОМАТИЗАЦИЯ ПОСТРОЕНИЯ РАСПИСАНИЯ ЗАНЯТИЙ В ВУЗЕ: МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДЫ РЕАЛИЗАЦИИ

УДК 378:004

СТРАТЕГИЯ ИНФОРМАЦИОННОГО ПОИСКА В ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМАХ И ОЦЕНКА КОМПЕТЕНЦИЙ СТУДЕНТОВ

А. А. Витухновская¹

кандидат педагогических наук, доцент (ветеран Петрозаводского государственного университета)

alla.vit@inbox.ru

Аннотация

Статья посвящена рассмотрению стратегии поиска информации в информационно-поисковой системе (ИПС). Подробно проанализирован подготовительный этап поиска информации, выделены простые и сложные объекты и аспекты поиска, сформулировано правило деления запроса на подзапросы. Обоснована роль логических операций для построения стратегии поиска, предложены правила проектирования и реализации стратегии поиска в ИПС. Проанализированы результаты анкетирования студентов, позволяющие утверждать, что многие из них не готовы грамотно использовать логические операции при поиске информации в информационно-поисковых системах.

Ключевые слова: *технология поиска информации, стратегия поиска информации, информационно-поисковые системы, поисковый образ запроса, язык запросов, логические операции, информационно-поисковые компетенции, студенты, анкетирование.*

ТЕХНОЛОГИЯ ИНФОРМАЦИОННОГО ПОИСКА

Рассматривая проблемы информационного поиска, стоит различать технологию и стратегию поиска. Некоторые авторы, например, И.С. Галеева, выделяют еще и тактику информационного поиска [3].

Под *технологией* информационного поиска мы понимаем *последовательность этапов* (процессов и операций), направленных на получение информации в ответ на информационный запрос пользователя. Ниже представлены технологические процессы, представляющие собой этапы информационного поиска:

1. Формулировка информационного запроса, соответствующего информационной потребности;
2. Подготовительный этап информационного поиска:
 - a. Анализ информационного запроса;
 - b. Деление запроса на подзапросы (при определенных условиях);
3. Стратегия поиска информации в комплексе ИПС:
 - a. Отбор круга информационно-поисковых систем для поиска информации, релевантной информационному запросу;
 - b. Построение общей программы (последовательности) поиска информации по запросу в комплексе отобранных систем;
4. Стратегия поиска информации в конкретной информационно-поисковой системе:
 - a. Составление поискового образа запроса (ПОЗ) на языке ключевых слов с использованием логических операций;
 - b. Составление и ввод поискового образа запроса с использованием языка запросов ИПС;
 - c. Оценка полученных результатов;
 - d. Переформулировка запроса (с использованием синонимичных терминов, ассоциативных, видовых и, нередко, родовых понятий);
 - e. Составление и ввод нового варианта поискового образа запроса.

В табл. 3 представлен перечень технологических процессов, составляющих информационный поиск, проиллюстрированный на конкретном примере. Все перечисленные этапы являются обязательными при осуществлении информационно-поисковой деятельности. Подробно все они описаны в нашей статье [2].

СТРАТЕГИЯ ПОИСКА ИНФОРМАЦИИ

Для характеристики понятия «стратегия поиска» остановимся на родовом по отношению к нему понятии «стратегия». В «Толковом словаре по психологии» стратегия определяется как «... план поведения или действия, сознательно составленный набор операций для решения некоторой проблемы или достижения некоторой цели»¹. В «Экономическом словаре» дано следующее определение

¹Толковый словарь по психологии (https://psychology_dictionary.academic.ru/)

стратегии: «план действий в условиях неопределенности ..., набор правил, согласно которым предпринимаемые действия должны зависеть от обстоятельств, включая естественные события и действия других людей»². В «Системе стандартов по информации, библиотечному и издательскому делу. Поиск и распространение информации. Термины и определения» представлено понятие *стратегия поиска*, трактуемое как «определение последовательности операций, осуществляемых в процессе информационного поиска, с целью повышения его эффективности»³ Это определение, на наш взгляд, скорее относится к понятию «технология поиска». Понятие же «стратегия поиска» стоит определить, опираясь на родовое понятие «стратегия», и рассматривать его как план действий, составленный человеком сознательно, с учетом обстоятельств для решения задачи получения релевантной информации в ответ на информационный запрос. Такими обстоятельствами можно считать характер информационного запроса, требуемые показатели полноты и точности поиска, доступность информационно-поисковых систем, уровень информационно-поисковых компетенций пользователя, время, которым он располагает, и некоторые другие.

На наш взгляд, следует выделить два вида стратегии поиска:

- стратегия поиска информации в комплексе информационно-поисковых систем;
- стратегия поиска информации в конкретной ИПС.

В настоящем исследовании внимание акцентируется на вопросах, касающихся стратегии поиска информации в отдельных информационно-поисковых системах. Как мы отмечали, технология информационного поиска начинается с формулировки и анализа информационного запроса. От правильного анализа информационного запроса зависит выбор правильной стратегии поиска как в целом (в комплексе ИПС), так и в частности (в конкретной информационно-поисковой системе).

²https://dic.academic.ru/dic.nsf/econ_dict/19736/%D0%A1%D0%A2%D0%A0%D0%90%D0%A2%D0%95%D0%93%D0%98%D0%AF

³ ГОСТ 7.73-96 http://allgosts.ru/01/140/gost_7.73-96 (34.13).

ПОДГОТОВИТЕЛЬНЫЙ ЭТАП ИНФОРМАЦИОННОГО ПОИСКА

Результат информационного поиска во многом зависит от умения пользователя анализировать собственный информационный запрос. Перед ним обычно встают два вопроса: *что* нужно найти (*объект поиска*⁴), и *что известно*, т. е. по какому признаку нужно вести поиск нужного объекта (*аспект поиска*⁵). Еще И.Г. Моргенштерн и Б.Т. Уткин в своей «Занимательной библиографии» задали вопрос: «Так с чего начинать поиск?», и сами же ответили: «Сначала надо определить: о чем хотите читать, что искать?» [7, с. 66]. Перечни значений объекта и аспекта поиска представлены в нашей статье [1]. Здесь приведем лишь фрагмент этого списка.

Таблица 1. Фрагмент перечня значений объектов и аспектов поиска информации

Объект поиска: виды	Значения объекта поиска
1. Первичный документ	a) Книга
	b) Периодическое издание
	c) Статья
	d) Компьютерная программа
	e) Видеозапись, аудиозапись
	f) Цифровые образовательные ресурсы (ЦОР) простой структуры
	g) Рекламная информация
	h) Другие виды
2. Вторичный документ	a) Адрес электронного документа
	b) Библиографическое описание (БО) документа
	c) Элементы БО (заглавие, год издания и др.)
	d) Аннотация, реферат
	e) Метаданные

⁴Объект поиска – это то, что пользователь хочет найти в ответ на свой запрос (фактографическую или документальную информацию; конкретные виды документов, элементы библиографической записи; метаданные и др.).

⁵Аспект поиска – те признаки (значения атрибутов) объекта, заданные в запросе, по которым его можно отыскать в ИПС. Аспект поиска может содержать несколько атрибутов (например, тему и дату издания документов).

	f) Другие
3. Определение или значение слова	
4. Фактические сведения	a) Фактографическое описание
	b) Статистические данные
	c) Другие
Аспект поиска	Атрибуты
1. Формальные признаки	a) Полное библиографическое описание
	b) Элементы БО
	c) Адрес электронного документа (веб-страницы)
	d) Гиперссылка
	e) Метаданные
2. Содержательные признаки	a) Тема
	b) Предмет
	c) Персоналия
	d) Другие
3. Слово	

Значения объекта поиска и аспекта поиска (атрибутов) фактически выражают содержание запроса. Мы выделяем простые и сложные объекты и аспекты поиска. Объект поиска будем считать простым, если он содержит одно значение (например, «книги» или «определение слова»). Объект, содержащий более одного значения, будем считать сложным. В некоторых запросах указание на сложность объекта в явном виде не выражено, например, «найти публикации по определенной теме». Здесь использовано широкое понятие «публикации», требующее видového уточнения путем перечисления значений.

Аспект поиска тоже может быть простым и сложным. Простым аспектом мы считаем такой, который ограничен одним атрибутом и одним значением этого атрибута. Со сложными аспектами дело обстоит сложнее. В запросах нередко содержится несколько атрибутов, а для каждого из них может быть задано более одного значения (например, мы можем искать статьи о базах данных и информационных системах). В приведенном запросе один атрибут (тема) содержит два значения. Сложный аспект может содержать несколько атрибутов и/или

значений любого из атрибутов. Примеры простых и сложных объектов и аспектов поиска приведены в таблице 2.

Таблица 2. Примеры простых и сложных объектов и аспектов поиска

Примеры запросов	Объект поиска	Вид объекта поиска	Аспект поиска	Вид аспекта поиска	Число подзапросов
1. Составить список статей и книг по проектированию баз данных	БО книг БО статей	Сложный	Проектирование баз данных (атрибут – тема)	Простой	2
2. Составить список книг по проектированию баз данных	БО книг	Простой	Проектирование баз данных (атрибут – тема)	Простой	1
3. Составить список статей и книг по проектированию баз данных и информационных систем	БО книг БО статей	Сложный	Проектирование баз данных Проектирование информационных систем (атрибут – тема)	Сложный (два значения атрибута «тема»)	4
4. Составить список публикаций по проектированию баз данных	БО книг, БО статей, БО диссертаций, других видов документов (указать)	Сложный (в неявном виде)	Проектирование баз данных (атрибут – тема)	Простой	3 и более

5. Каков состав населения Финляндии и Швеции?	Фактографическое описание	Простой	Финляндия, Швеция (атрибут – название страны)	Сложный	2
6. Найти текст статьи А.В. Соколова о природе информации, опубликованной в 2011 году	Статья (текст)	Простой	А.В. Соколов (атрибут – автор), природа информации (атрибут – тема), 2011 год (атрибут – дата издания)	Сложный	1

Анализ запроса направлен на определение объекта и аспекта поиска и выяснение их состава. Если объект сложный, необходимо однозначно разделить запрос на подзапросы соответственно числу значений объекта (см. запросы 1, 3, 4). Если аспект содержит несколько атрибутов, то деление на подзапросы не требуется (см. запрос 6).

Иначе обстоит дело в случае, если один и тот же атрибут содержит несколько значений. В этом случае требуется деление на подзапросы. Аспект с содержательными признаками (тема или предмет) требует дополнительного анализа. Определять нужно число значений атрибута, а не число слов, которыми он выражается. Так, в запросе «составить список книг по использованию компьютера в обучении математике и физике» содержится сложный аспект, а в запросе «найти книги по использованию электронных образовательных ресурсов в обучении математике» – простой. В первом случае речь идет о двух темах, а во втором – только об одной. Таким образом, если атрибут аспекта имеет несколько значений, то выделяется несколько подзапросов соответственно числу значений атрибута (см. запросы 3, 5).

СТРАТЕГИЯ ПОИСКА ИНФОРМАЦИИ В КОНКРЕТНОЙ ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЕ

Стратегия поиска в отдельной ИПС базируется на правильном определении объекта и аспекта поиска и обоснованном выборе ключевых слов из текста запроса. Построение стратегии поиска включает в себя два этапа:

- составление поисковой формулы (поискового образа запроса) на языке ключевых слов;
- составление поисковой формулы на информационно-поисковом языке с учетом языка запросов данной ИПС.

На первом этапе из текста запроса выбираются ключевые слова, которые фактически выражают значения объекта и аспекта поиска информационного запроса. Между ключевыми словами устанавливаются логические операции (булевские операторы, логические операторы), которые используются для достижения полноты и точности поиска информации. Практически все современные электронные ИПС (каталоги и библиотеки, базы данных) имеют свой язык запросов, который предусматривает использование логических операций конъюнкции, дизъюнкции и отрицания.

Конъюнкция подразумевает обязательное наличие всех лексических единиц запроса, объединенных этой операцией, в искомом документе. Она способствует снижению уровня информационного шума и повышению точности поиска. В разных информационно-поисковых системах конъюнкция обозначается по-разному: чаще всего этой операции соответствует пробел; кроме этого, используются знаки И, &, AND. «В большинстве поисковых систем пробел эквивалентен логическому оператору "и"» [11, с. 113].

Использование *дизъюнкции* предполагает, что результатом поиска будет совокупность документов, каждый из которых содержит хотя бы одну из лексических единиц запроса. Дизъюнкция увеличивает полноту выдачи информации. В разных системах используются разные обозначения: ИЛИ, OR, |. «Эффективнее всего использовать этот оператор при поиске терминов-синонимов, терминов, эквивалентных по значению и альтернативных терминов» [4, с. 34], а также при соединении подзапросов одного запроса в процессе поиска.

Логическая операция отрицания используется для исключения лексических единиц, заданных в поисковом образе запроса после него.

Существует несколько правил, которые необходимо учитывать при проектировании стратегии поиска информации в ИПС:

- ключевые слова (значения объекта и аспекта поиска) одного подзапроса или запроса, который не делится на подзапросы, соединяются операцией конъюнкции;
- в запросах, содержащих сложный объект, значения этого объекта в обязательном порядке соединяются операцией дизъюнкции;
- в запросах, содержащих несколько атрибутов сложного аспекта поиска, значения разных атрибутов, взятых по одному, соединяются операцией конъюнкции;
- в запросах, содержащих несколько значений одного атрибута аспекта поиска, эти значения соединяются операцией дизъюнкции;
- при переформулировке запроса (для увеличения полноты выдачи информации) между синонимами, родовидовыми понятиями и ассоциативными терминами устанавливается операция дизъюнкции.

Ниже приведены примеры двух запросов, поисковые образы которых построены с учетом сформулированных правил:

1. Составить список монографий об использовании *электронных образовательных ресурсов в обучении математике* за 2015 год

Поисковый образ запроса

монографии **И** *электронные образовательные ресурсы* **И** математика **И** 2015

2. Составить список монографий и научных статей по теме: использование *электронных образовательных ресурсов (ЭОР) в обучении математике и физике* за 2015–2016 годы

Поисковый образ запроса

(монографии **ИЛИ** статьи) **И** (электронные образовательные ресурсы **ИЛИ** ЭОР) **И**
(математика **ИЛИ** физика) **И** (2015 **ИЛИ** 2016)

Каждая из автоматизированных ИПС использует свой *язык запросов*. В.П. Захаров определяет «языки запросов» как сложные объекты, которые «объединяют собственно ИПЯ (информационно-поисковые языки – А.В.) и критерий

смыслового соответствия, а также могут содержать в себе требования к интерфейсу выдачи» [6, с. 66]. Обобщенная структурная модель языка запросов, описанная Захаровым, включает, наряду с другими, поисковые (булевские) операторы.

Каждая из таких систем имеет свой индивидуальный интерфейс, который содержит поля для ввода (чаще – для выбора из готового списка) *атрибутов*, *значений* атрибутов и логических операций, а также средства (способы) задания критерия выдачи. На рисунке 1 показано окно автоматизированной информационно-библиотечной системы РНБ РК «Фолиант»⁶. Рассмотрим стратегию поиска информации по конкретному запросу «Составить список научных статей и монографий об использовании *электронных образовательных ресурсов в обучении математике и физике* за последние 5 лет» в электронном каталоге данной системы.

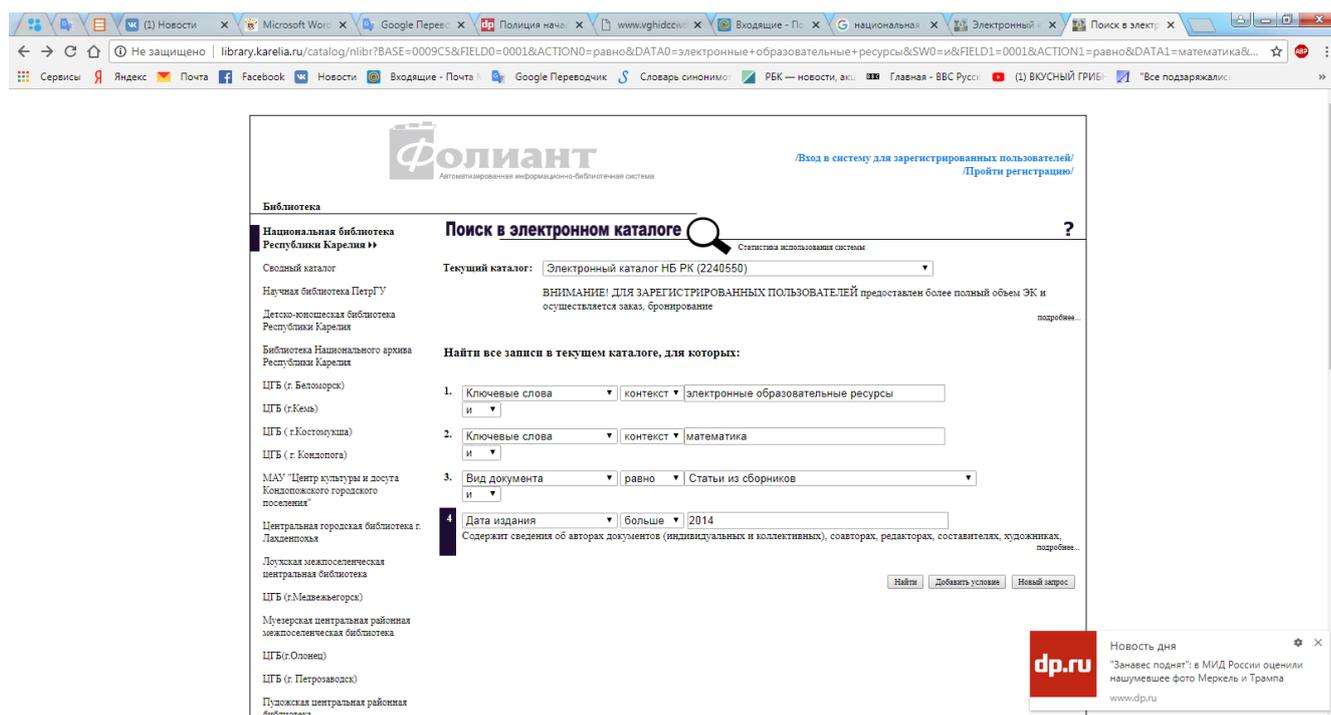


Рисунок 1. Поисковый образ запроса в электронном каталоге (на примере одного подзапроса)

⁶ <http://library.karelia.ru/catalog/nlibr>

Для нахождения релевантной информации мы заранее разделили его на 20 подзапросов⁷. Поскольку в рассматриваемом каталоге для числовой информации предусмотрено значение «больше», то у нас число подзапросов сократилось до четырех. Для ввода поискового образа запроса были использованы атрибуты: «Вид документа» (для объекта поиска), «Ключевые слова» и «Дата издания» (для аспектов поиска). Так как язык запросов данного электронного каталога (как и многих других) не предусматривает использование скобочной записи для реализации дизъюнкции конъюнкций, то поиск проводился последовательно, по каждому подзапросу. Для первого подзапроса в поле первого атрибута «Ключевые слова» введено ключевое слово «электронные образовательные ресурсы», во второе поле «Ключевые слова» введено слово «математика», в поле «Вид документа» выбрано значение «статьи из сборников» из заданного списка, а в поле «Дата издания» введен первый год из предусмотренного запросом интервала, а также критерий выдачи «больше». Между значениями всех атрибутов установлены логические операции конъюнкции.

КОМПЕТЕНЦИИ СТУДЕНТОВ В РЕАЛИЗАЦИИ СТРАТЕГИИ ПОИСКА ИНФОРМАЦИИ В КОНКРЕТНОЙ ИПС

Специалисты критически оценивают умения пользователей грамотно строить стратегию поиска. О том, что при поиске редко используются булевы операторы, свидетельствуют результаты исследований интерактивного поиска, проводившиеся американскими учеными во главе с профессором Амандой Спинк в 1997, 1999, 2001 и 2002 годах [цит. по 3].

Отечественный исследователь И.В. Тультаева делает вывод, что «подавляющее большинство интернет-пользователей не умеет пользоваться логическими операторами при построении поисковых запросов» [11, с. 113]. Проведенное Н.С. Жуковой несколько лет назад изучение уровня информационной компетенции студентов из России и Германии показало, что «при определении такой поисковой стратегии, как использование булевых операторов» для связки синонимов,

⁷ (Книги **ИЛИ** статьи) **И** Электронные образовательные ресурсы **И** (Физика **ИЛИ** Математика) **И** (2014 **ИЛИ** 2015 **ИЛИ** 2016 **ИЛИ** 2017 **ИЛИ** 2018)

правильный ответ дали только 26% студентов из России и 44% студентов из Германии [5, с. 552].

Многолетний опыт работы со студентами разных профилей обучения позволяет и нам сделать вывод о том, что многие студенты не владеют в достаточной степени информационно-поисковыми компетенциями, в частности, не умеют строить стратегию поиска по собственному информационному запросу. Последнее проявлялось особенно очевидно при обращении к электронным каталогам и электронным библиотекам на лабораторных занятиях по курсам «Информационным системы», «Информационные технологии» и «Теория и методика обучения информатике».

Для получения более точных данных об уровне информационно-поисковых компетенций, сформированных у недавних выпускников школы, нами было проведено анкетирование студентов первого и второго курсов, обучающихся по разным профилям (всего 139 человек). Методика анкетирования и результаты обработки данных, позволяющие судить о знании студентами видов информационно-поисковых систем и их способности делать правильный выбор ИПС при поиске информации, описаны в нашей первой публикации [2]. В ней были представлены результаты анкетирования, позволившие судить о слабом представлении студентов о многообразии информационно-поисковых систем и неспособности выбирать адекватные ИПС для поиска информации по предложенным запросам. Это, в свою очередь, позволило судить о том, что недавние выпускники школ не умеют правильно определять объект и аспект поиска информации, не знакомы с некоторыми видами информационно-поисковых систем [2].

Несколько вопросов анкеты было направлено на получение информации о том, способны ли студенты грамотно строить стратегию поиска, в частности, делить запрос на подзапросы и использовать логические операции при составлении поисковых образов запросов. Кроме этого, ставилась задача выяснить, изучили ли студенты язык запросов хорошо знакомой им информационно-поисковой системы.

Для получения информации об умении делить запрос на подзапросы студентам был предложен один запрос (они различались при анкетировании студентов разных профилей) со сложным объектом и аспектом поиска и предлагалось

указать правильное число подзапросов. Огорчает, что правильный ответ дали немногим более трети студентов (34,5%).

Поисковый образ запроса, обращенный к электронным системам, на первом этапе (а чаще всего и окончательно), создается на языке ключевых слов. Чтобы понять, как студенты вводят запрос в поисковую систему интернета, – создают ли они поисковый образ запроса или вводят текст запроса целиком на естественном языке, – мы поставили перед ними соответствующий вопрос. Оказалось, что около половины (48, 2%) опрошенных студентов обычно вводят весь текст запроса; несколько меньше студентов (около 42%) вводят только ключевые слова, выбранные из текста запроса. К сожалению, лишь 10% студентов отвечают, что используют логические операции при поиске информации в поисковых системах интернета.

На вопрос о том, какой логической операцией должны быть связаны ключевые слова запроса, соответствующего одной теме (подзапроса), правильный ответ дали 40% студентов, которые выбрали операцию *конъюнкции*; 14% дали неправильный ответ, указав *дизъюнкцию*; 22% студентов утверждают, что в логических операциях нет необходимости, а 20% не знают, что это такое. Вероятно, эти студенты просто не задумываются о том, что, вводя в реальную ИПС текст запроса (такое возможно в поисковых системах интернета) или ключевые слова, разделенные пробелом, они используют какую-либо логическую операцию.

Только 29% студентов посчитали, что для связи подзапросов нужно использовать логическую операцию *дизъюнкции*. Возможно, это уже можно расценивать как достижение, так как две трети студентов вообще не смогли правильно выделить подзапросы в предложенном сложном запросе (см. выше). Большинство студентов (38%) «установили» между подзапросами неправильную операцию *конъюнкции*, 20% опрошенных сказали, что они не знают, что такое логические операции.

42% студентов для связки синонимов (обращение к ним бывает необходимо при переформулировке запроса) совершенно верно посчитали нужным использовать логическую операцию *дизъюнкции*, что значительно больше, чем в исследовании Н.С. Жуковой. В свою очередь, 24% студентов признались, что не

знают, что собой представляют логические операции, а 12% посчитали, что в них нет необходимости.

Более половины опрошенных либо игнорируют логические операции при формулировке поискового образа запроса, либо принимают неправильные решения при их выборе в конкретном запросе. Обобщая ответы студентов на рассмотренные вопросы, можно было бы сделать вывод о том, что многие из них (20 – 24%) ничего не знают о логических операциях. Однако это не так: в школьном курсе математики изучаются «операции над высказываниями с использованием логических связок», а в курсе информатики ставятся задачи «проводить поиск информации в сети Интернет по запросам с использованием логических операций» [8]. Тогда можно предположить, что не во всех школах уделяется внимание адаптации этих знаний и умений при решении информационно-поисковых задач в разных учебных предметах.

По мнению И.В. Тультаевой, применение логических операторов для пользователей различных информационно-поисковых систем превращается в довольно сложный процесс, в т.ч. и потому, что в различных системах они записываются по-разному [11, с. 114]. Автор считает, что для решения этой проблемы в современных поисковых системах интернета используются специальные веб-интерфейсы расширенного поиска, который, по мнению автора, позволяет решать сложные поисковые задачи, не пользуясь языком запросов. Мы решили узнать, насколько популярным является это интерфейс для наших студентов. Выяснилось, что режим расширенного поиска в «своих» поисковых системах интернета не использует (и, видимо, ничего о нем не знает) около половины (48,9%) студентов. 41% студентов отвечает, что они используют его для отбора документов определенного типа (формата файлов), 10% студентов для указания региона или языка; несколько студентов указывают оба варианта ответов.

Известно, что в большинстве случаев студенты (и не только они) при поиске информации обращаются к поисковым системам интернета. Результаты нашего исследования, представленные ранее, показывают, что именно к ним готовы обратиться от 75 до 85% студентов для ответа на тематический, адресный и фактографический запросы [2]. При этом, как мы уже отмечали, лишь 10 % студентов ответили, что, обращаясь к поисковым системам интернета, они используют логические операции. Привыкая к простоте поиска, допустимой в поисковых

системах (ввод в поисковую строку всего текста запроса, игнорирование логических операций), пользователи не приобретают навыков формирования поисковых образов запросов, необходимых для обращения к электронным каталогам и электронным библиотекам. И тогда можно предположить, что в реальных условиях самостоятельного поиска информации в них студентов ожидают потери информации или даже нулевые результаты (при наличии релевантной информации в информационно-поисковой системе).

На возникшее противоречие между стратегией поиска в поисковых системах интернета и электронных каталогах (от себя добавим – и электронных библиотеках) обращает внимание Г.А. Скарук [9, 10]. В своих публикациях автор говорит о том, что «конкурировать на равных с поисковыми машинами интернета электронные каталоги и другие поисковые системы, создаваемые библиотеками, вряд ли смогут».

Очевидно, что необходима целенаправленная деятельность по формированию информационно-поисковых компетенций школьников и студентов. Возможны различные пути обеспечения этой деятельности, в частности:

- создание обобщающего курса по обучению информационно-поисковой деятельности;
- согласование и интеграция содержания различных учебных предметов в той части, которая направлена на формирование информационно-поисковых компетенций.

Второй путь представляется нам более реалистичным. В этом случае ведущую роль может сыграть «Информатика», в которой закладываются основы знаний об информационном поиске, формируются начальные умения грамотного формулирования и анализа информационного запроса, построения стратегии поиска информации как в комплексе ИПС, так и в конкретных системах. В других учебных предметах эти знания и умения будут получать развитие и практическое применение при решении информационно-поисковых задач в конкретных предметных областях.

ПРИЛОЖЕНИЕ

Таблица 3. Общая технология информационного поиска на примере поиска по информационному запросу

Процессы информационного поиска	Реализации технологии поиска информации по конкретному (намеренно усложненному) запросу
1. Формулировка информационного запроса, соответствующего информационной потребности	Составить список монографий и научных статей по теме: использование <i>электронных образовательных ресурсов (ЭОР) по математике и физике</i> за последние 5 лет
2. Анализ информационного запроса. Определение объекта и аспекта поиска	Объект: книги, научные статьи Аспект: <i>тема</i> (электронные образовательные ресурсы по математике и физике) и <i>годы издания</i> (2014–2018)
1. Деление запроса на подзапросы 2. Выявление в запросе тем, по которым нужно вести отдельный поиск 3. Формулировка соответствующих подзапросов	1. Монографии об <i>использовании ЭОР по математике</i> в 2014 году 2. Монографии об <i>использовании ЭОР по физике</i> в 2014 году 3. Научные статьи об <i>использовании ЭОР по математике</i> в 2014 году 4. Научные статьи об <i>использовании ЭОР по физике</i> в 2014 году ... В 2015 и т. д. годах (всего 20 подзапросов)
3. Отбор круга ИПС для поиска информации, релевантной информационному запросу. Соотнесение объекта и аспекта поиска с объектами и аспектами ИПС	<ul style="list-style-type: none"> • Систематический каталог библиотеки • Систематическая картотека статей • Электронный каталог библиотеки • Электронная библиотека • Поисковая система интернета

	• Библиографические пособия
4. Построение общей программы (стратегии) поиска информации по запросу в комплексе отобранных ИПС (с учетом возможностей доступа к ИПС, уровня информационно-поисковой и ИКТ-компетентности пользователя)	<p>Библиографические пособия</p> <p>Вариант 1</p> <ol style="list-style-type: none"> 1. Электронный каталог библиотеки (пример: ПетрГУ) 2. Электронная библиотека (пример: elibrary.ru) 3. Поисковая система интернета (пример: Google) <p>Вариант 2</p> <ol style="list-style-type: none"> 1. Систематический каталог библиотеки 2. Систематическая картотека статей 3. Поисковая система Интернета (пример: Google)
5. Составление поискового образа запроса (ПОЗ) на языке ключевых слов	<ul style="list-style-type: none"> - электронные образовательные ресурсы - ЭОР - математика - физика - монографии - научные статьи - 2014 - 2015 - 2016 - 2017 - 2018
6. Составление поискового образа запроса с использованием логических (булевских) операций конъюнкции, дизъюнкции и отрицания (при их необходимости)	<p><i>В свернутом виде:</i></p> <p>(Электронные образовательные ресурсы ИЛИ ЭОР) И (Физика ИЛИ Математика) И (2014 ИЛИ 2015 ИЛИ 2016 ИЛИ 2017 ИЛИ 2018)</p>
7. Обращение к конкретной ИПС из числа отобранных на этапе 5	<p>http://library.karelia.ru/catalog/nlibr</p>

<p>8. Составление поискового образа запроса на языке запросов данной ИПС:</p> <ul style="list-style-type: none"> • Анализ интерфейса ИПС; • Выбор ИПЯ; • Ввод лексических единиц (ключевых слов, индексов или рубрик) в соответствующие поля, выбор и ввод логических операций между лексическими единицами 	<p>См. рис. 1</p>
<p>9. Навигация по гиперссылкам электронного документа (в частности, на другие документы)</p>	
<p>10. Оценка результатов. Соотнесение полученных документов или фактографической информации с информационной потребностью</p>	
<p>11. Переформулировка запроса с использованием синонимичных терминов и/или ассоциативных, видовых (нередко и родовых) понятий Термин <i>электронные образовательные ресурсы</i></p> <p>Синонимы:</p> <ul style="list-style-type: none"> • ЭОР • цифровые образовательные ресурсы • ЦОР <p>Видовые понятия (выборочно):</p> <ul style="list-style-type: none"> • Обучающие программы • Тесты 	<ol style="list-style-type: none"> 1. Монографии об <i>использовании компьютера в процессе обучения математике с 2014 до 2018 гг.</i> 2. Монографии об <i>использовании компьютера в процессе обучения физике с 2014 до 2018 гг.</i> 3. Научные статьи об <i>использовании компьютера в процессе обучения математике с 2014 до 2018 гг.</i> 4. Научные статьи об <i>использовании ЭОР компьютера в процессе обучения физике с 2014 до 2018 гг.</i> 5. Научные статьи об <i>использовании ИКТ в процессе обучения математике с 2014 до 2018 гг.</i>

СПИСОК ЛИТЕРАТУРЫ

1. *Витухновская А.А.* Обучение технологии и стратегии информационного поиска на основе дифференциальных признаков информационно-поисковых систем // Информационное общество. 2013. № 1–2. С. 69–79. <http://emag.iis.ru/arc/infosoc/emag.nsf/BPA/0f0c1bb6d40ae0d544257bef002cb5dc>
 2. *Витухновская А.А.* Информационно-поисковая деятельность и информационно-поисковые компетенции студентов (по итогам опроса) // Электронные библиотеки. 2017. №3. С. 164–194. <http://ojs.kpfu.ru/index.php/elbib/article/view/406/81>
 3. *Галеева И.С.* Путеводитель библиографа по интернету. Санкт-Петербург, 2013.
 4. *Еременко Т.В.* Библиографический поиск в научной работе: учебно-методическое пособие / РГУ имени С.А. Есенина. Рязань, 2015. 87 с.
 5. *Жукова Н.С.* Сравнительный анализ уровня информационной грамотности студентов сетевого поколения в России и Германии // Образовательные технологии и общество (Educational Technology & Society). 2011. Т. 14. № 2. С. 539–565.
 6. *Захаров В.* Лингвистические средства информационного поиска в интернете // Библиосфера. 2005. № 1. С. 63–71.
 7. *Моргенштерн И.Г., Уткин Б.Т.* Занимательная библиография. М.: Изд-во «Книжная палата», 1987. 255 с.
 8. Примерная основная образовательная программа основного общего образования. URL: <http://mosmetod.ru/files/dokumenty/primernaja-osnovnaja-obrazovatel'naja-programma-osnovogo-obshchego-obrazovanija.pdf>
 9. *Скарук Г.А.* Электронные каталоги библиотек в борьбе за пользователя: «старые» и новые подходы // Библиосфера. 2016. № 2. С. 7–15.
 10. *Скарук Г.А.* Средства и методы помощи пользователям электронного каталога в самостоятельном поиске // Труды ГПНТБ СО РАН. 2015. № 8. С. 275–282.
 11. *Тультяева И.В.* Возможности применения поисковых систем для оптимизации сбора информации в сети Интернет // Бизнес. Образование. Право. 2014. № 4. С. 111–117.
-
-

THE STRATEGY OF INFORMATION RETRIEVAL IN INFORMATION RETRIEVAL SYSTEMS AND THE ASSESSMENT OF THE STUDENTS' COMPETENCIES

A. A. Vitukhnovskaya

Candidate of Pedagogical Sciences, assistant professor (veteran of Petrozavodsk State University)

alla.vit@inbox.ru

Abstract

The article is devoted to the consideration of information retrieval strategies in information retrieval systems (IRS). The preparatory stage of information retrieval is analyzed in detail, allocated simple and complex objects and aspects of search; the rule for dividing the request for subqueries is formulated. The role of logical operations for the construction of a search strategy is substantiated, the rules for designing and implementing search strategies in information retrieval systems are proposed. The results of the questioning of students are analyzed, and it's possible to state that many of them are not ready to use logical operations correctly in the process of information retrieval in information retrieval systems.

Keywords: information retrieval technology, information retrieval strategy, information retrieval systems, query language, logical operations, information retrieval competencies, students, questionnaires.

REFERENCES

1. Vituhnovskaya A.A. Obuchenie tehnologii i strategii informatsionnogo poiska na osnove differentsialnykh priznakov informatsionno-poiskovykh sistem // Informatsionnoe obshchestvo. 2013. # 1–2. S. 69–79. <http://emag.iis.ru/arc/infosoc/emag.nsf/BPA/0f0c1bb6d40ae0d544257bef002cb5dc>
2. Vituhnovskaya A.A. Informatsionno-poiskovaya deyatel'nost' i informatsionno-poiskovyye kompetentsii studentov (po itogam oprosa) // Elektronnyye biblioteki. 2017. #3. S. 164–194. <http://ojs.kpfu.ru/index.php/elbib/article/view/406/81>
3. Galeeva I.S. Putevoditel bibliografa po internetu. Sankt-Peterburg, 2013.
4. Eremenko T.V. Bibliograficheskiy poisk v nauchnoy rabote: uchebno-metodi-cheskoe posobie / RGU imeni S.A. Esenina. Ryazan, 2015. 87 s.
5. Zhukova N.S. Sravnitel'nyy analiz urovnya informatsionnoy gramotnosti studentov setevogo pokoleniya v Rossii i Germanii // Obrazovatelnyye tehnologii i obshchestvo (Educational Technology & Society). 2011. T. 14. # 2. S. 539–565.
6. Zaharov V. Lingvisticheskiye sredstva informatsionnogo poiska v internete // Bibliosfera. 2005. # 1. S. 63–71.
7. Morgenshtern I.G., Utkin B.T. Zanimatel'naya bibliografiya. M.: Izd-vo «Knizhnaya palata», 1987. 255 s.
8. Primernaya osnovnaya obrazovatel'naya programma osnovnogo obshchego obrazovaniya. URL: <http://mosmetod.ru/files/dokumenty/primernaya-osnovnaya-obrazovatel'naya-programma-osnovnogo-obshchego-obrazovaniya>.
9. Skaruk G.A. Elektronnyye katalogi bibliotek v borbe za polzovatelya: «staryye» i novyye podhody // Bibliosfera. 2016. # 2. S. 7–15.
10. Skaruk G.A. Sredstva i metody pomoschi polzovatelyam elektronnoy kataloga v samostoyatel'nom poiske // Trudy GPNTB SO RAN. 2015. # 8. S. 275–282.
11. Tultaeva I.V. Vozmozhnosti primeneniya poiskovykh sistem dlya optimizatsii sbora informatsii v seti Internet // Biznes. Obrazovanie. Pravo. 2014. # 4. S. 111–117.

СВЕДЕНИЯ ОБ АВТОРЕ



ВИТУХНОВСКАЯ Алла Александровна – кандидат педагогических наук, доцент. Длительное время преподавала информатику и дисциплины информационного цикла в Казанском государственном институте культуры, Карельском педагогическом и Петрозаводском государственном университетах. Область научных интересов: информатизация образования, информационные системы, теория и методика обучения информатике.

Alla Aleksandrovna VITUKHNOVSKAJA – Candidate of pedagogical sciences, associate professor. For a long time she taught informatics and other information disciplines at the Kazan State Institute of Culture, the Karelian Pedagogical University and the Petrozavodsk State University. Area of scientific interests: Informatization of education, information systems, theory and methodology of teaching informatics.

e-mail: alla.vit@inbox.ru

Материал поступил в редакцию 30 июня 2018 года

УДК 004.04

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ТЕКСТОВОЙ АНАЛИТИКИ ЮРИДИЧЕСКИХ ДОКУМЕНТОВ

¹Д. С. Зуев, ²М. Ф. Насрутдинов, ³А. Ф. Хасьянов

Казанский (Приволжский) федеральный университет, г. Казань, ул. Кремлевская, д. 35, 420008

¹dzuev11@gmail.com, ²marat.nasrutdinov@kpfu.ru, ³ak@it.kfu.ru

Аннотация

Обсуждено использование механизмов машинного обучения, анализа естественного языка и интеллектуального поиска в области юриспруденции. Основные ожидаемые результаты – методология применения алгоритмов текстовой аналитики и семантического анализа естественного языка (NLP) в задачах управления знаниями в судебном делопроизводстве, а также других видах юридической практики. Полученные результаты могут быть применены в области образования и управления знаниями в более широком контексте, поскольку исследование лежит на стыке юриспруденции, математической и компьютерной лингвистики.

Описан прототип многоагентной системы интеллектуального анализа текстов в юриспруденции, способной на имеющейся базе данных судебных документов выявлять общие зависимости, предоставлять для ознакомления юридические дела, близкие по тематике, рекомендовать наиболее вероятные исходы судебного рассмотрения или помечать важные места, на которые следует обращать внимание при процессуальных действиях с использованием инструментов текстовой аналитики.

Ключевые слова: аналитика и управление данными, интенсивное использование данных, электронные библиотеки, кластеризация, классификация судебных актов, рекомендательная система, микросервисная архитектура.

ВВЕДЕНИЕ

Целью работы является создание интеллектуальной программной системы с набором инструментов текстовой аналитики и приложений, ориентированной

на решение прикладных задач в области юриспруденции. Создание такой системы полностью лежит в русле развития и использования современных информационно-коммуникационных технологий (ИКТ) при обработке больших массивов информации и предусматривает развитие и практическую реализацию системы управления юридическими знаниями на основе семантических технологий и онтологий. Планируемые исследования соответствуют идеологии Инициативы открытых архивов (Open Archive Initiative), основное назначение которой – повысить доступность и объединить в распределенной системе взаимосвязанных хранилищ корпуса различных текстов, включая как современные источники, так и источники, ставшие историческими.

Любой документ имеет собственные особенности, и юриспруденция не является исключением. Здесь фундаментальной задачей является изучение особенностей юридических документов с целью применения известных моделей и методов текстового анализа и машинного обучения к специфическому классу документов, каковыми являются исковые заявления, судебные дела и решения. Реализуемый проект развивает названное направление исследований и предполагает разработку и внедрение современной информационной системы, ориентированной на работу с юридическими документами как с отдельным классом электронных документов.

1. АКТУАЛЬНОСТЬ ИССЛЕДОВАНИЯ

Как известно, сегодня мы являемся участниками перехода от традиционного типа общества к информационному – обществу, в котором большинство работающих занято производством, хранением, переработкой и реализацией информации, особенно высшей её формы – знаний (см., например, https://ru.wikipedia.org/wiki/Информационное_общество). Информационное общество характеризуется высоким уровнем развития ИКТ и их интенсивным использованием всеми и всюду, а целью формирования такого общества является, в частности, развитие на основе ИКТ научно-исследовательской деятельности в любых областях человеческой жизнедеятельности.

В основе ИКТ лежит информация, а сами они во многом определяют содержание, масштабы и темпы развития других технологий. В частности, развитие облачных технологий позволило принципиально изменить подходы к созданию

сложных программных систем практически для всех предметных областей. Одной из важнейших признана задача интеграции разнородных электронных ресурсов в мировое информационное пространство. Определяющей составляющей такой интеграции является процесс семантического структурирования контента. Для этого в последние годы консорциум W3C (www.w3.org) активно разрабатывает технологии Семантического Веба, в частности, на платформе XML создан широкий спектр языков разметки текста, позволяющих не только учесть специфику предметных областей, но и повысить эффективность структурирования при автоматизированной обработке информации.

Создаваемая система должна позволять участникам юридического процесса правильно проводить подготовку соответствующих судебных дел, а также осуществлять планирование судебной деятельности. Эта система ориентирована на арбитражные суды, занимающиеся рассмотрением споров, связанных с предпринимательской деятельностью. В целом наш проект направлен на развитие российского правового государства, обеспечение доступности, открытости и прозрачности правосудия, формирование у граждан правосознания, основанного на верховенстве Права.

На сегодняшний день судопроизводство в России является областью, постоянно наращивающей присутствие информационно-коммуникационных технологий в каждодневной деятельности. Сегодня в судах используется ряд программных систем, позволяющих вести документооборот в электронной форме. Тем не менее, информационная и производственная нагрузки на судей по-прежнему остаются недопустимо большими. Так, например, только в рамках инициативы по борьбе с коррупцией на судей арбитражного суда Республики Татарстан приходится более 50 судебных дел в месяц. В таких условиях без использования специализированных автоматизированных информационных систем качественное повышение эффективности работы судов просто невозможно.

2. СУЩЕСТВУЮЩИЕ РАЗРАБОТКИ

Одним из направлений разработки специализированных автоматизированных информационных систем в судопроизводстве является создание интеллектуальных систем, способных на имеющейся базе данных судебных документов выявлять общие зависимости, предоставлять судьям для ознакомления близкие по

тематике дела, рекомендовать наиболее вероятные исходы или пометать важные места, на которые судебным работникам следует обращать внимание при процессуальных действиях.

В [1] описаны основные онтологии и подходы к работе с юридическими документами с точки зрения их семантики. Известны успешные реализации подобных информационных систем за рубежом, например, система «Case Cruncher Alpha» (www.case-crunch.com), разрабатываемая в Sidney Sussex College, Cambridge и ориентированная на прогнозирование решений юридических задач. Однако имеющиеся решения не учитывают особенности русского языка и кириллической транскрипции.

В судах также имеется программное обеспечение, которое позволяет автоматизировать часть рутинных операций. Эти программные комплексы направлены либо на автоматизацию документооборота в целом, либо представляют собой широчайшие базы данных тематических документов, найти в которых необходимую информацию в сжатые сроки не всегда представляется возможным, и не используют весь спектр семантических технологий и инструментария текстовой аналитики.

Проведенные исследования по семантическому структурированию информации в других предметных областях (см., например, [2, 3]), анализ инструментов текстовой аналитики (см., например, [4]) показывают наличие всех типовых инструментов, которые можно было бы применить и в данной предметной области для реализации поставленной задачи.

ЦЕЛЬ И ЗАДАЧИ

Разрабатываемая информационная система должна позволять участникам юридического процесса правильно проводить подготовку судебных дел, а также осуществлять планирование судебной деятельности. На данном этапе система ориентирована на арбитражные суды, занимающиеся рассмотрением споров в сфере предпринимательства.

Сегодня судебная система России серьезно загружена, количество судебных споров неуклонно растет. С одной стороны, это свидетельствует о развитии правового государства, когда все большее количество споров решается в правовом поле. С другой стороны, суды являются одной из самых забюрократизиро-

ванных сфер жизнедеятельности. При этом рост количества судебных дел повышает нагрузку как на судей, так и на вспомогательных технических работников.

Судебная система – это область, где объем работы с текстовыми документами весьма значителен, а процесс принятия решения всегда должен быть понятным и прозрачным. Поэтому, особенно в условиях роста нагрузки на сотрудников судов, требуются инструменты, позволяющие осуществлять интеллектуальный анализ поступающего информационного массива. Автоматизированный текстовый анализ позволяет выделить важные признаки документов (подсудность, характер спора, участвующие стороны и т. д.), осуществить поиск в судебной базе данных и представить похожие документы, по которым уже приняты решения, или даже спрогнозировать вероятное решение суда по рассматриваемому делу. Именно на этот аспект работы судебной системы нацелена наша система.

Задача создаваемого программного комплекса – помочь определить характер спора, осуществить поиск и проверку действия правовых норм, регулирующих спорные правоотношения, оказывать содействие в установлении компетентного суда (подсудность, подведомственность), статуса участников спора (действующее, ликвидированное, банкрот), определении круга обстоятельств, имеющих значение для рассмотрения спора, характера спорного правоотношения, нормы права, подлежащей применению (действует ли данная норма), а также проверять достаточность и комплектность представляемых документов.

Для достижения поставленных целей поставлены следующие задачи и спроектирована архитектура системы [5]:

- создание портала для формирования шаблонов исковых заявлений с отслеживанием их жизненного цикла;
- разметка и анализ существующей базы судебных решений, исковых заявлений (классификация заявлений и решений, извлечение сущностей и фактов);
- подбор аналогичных дел и решений, рекомендательный сервис;
- сопоставление исковых заявлений и судебных решений;
- распределение судебных дел между судьями с учетом их специализации и текущей загрузки.

Фактически каждая из выделенных задач является автономным модулем разрабатываемой информационной системы, а сама система – практическая демонстрация совместного использования ряда семантических технологий и инструментов текстовой аналитики.

Важно отметить, что создание рекомендательной системы текстовой аналитики юридических документов с применением технологий искусственного интеллекта и использованием технологий Семантического Веба является новой задачей, аналоги подобных систем в России отсутствуют. Существующие мировые аналоги ориентированы на тексты на латинице и не работают с кириллическим текстами.

ПОДХОДЫ К РЕАЛИЗАЦИИ

Исследование свойств и возможностей специализированных языков разметки, построенных на базе расширенного языка разметки XML, является важным и актуальным, поскольку на этой базе постепенно строится концепция формального описания семантики различных понятий и явлений. То, что до сих пор не выработано универсального и всеобъемлющего подхода к описанию семантики, существенно ограничивает возможности обмена информацией, поскольку формализованным оказывается лишь синтаксис сообщений, а он лишен смысла: за обыкновенными и сходными по форме синтаксическими конструкциями может скрываться самая неожиданная семантика (смысл).

Предлагаемые нами подходы используют методы компьютерной и математической лингвистики. Анализ текста основан как на традиционных подходах, связанных с применением онтологий, так и на применении машинного обучения на базе обширного корпуса юридических документов, имеющих в открытом доступе, включая артефакты правоприменительной практики судов различных инстанций.

Информация в системе должна храниться в формализованном и понятном компьютеру виде, сформированном на основе технологий Семантического Веба. Такой способ управления знаниями позволит создать инструменты для работы непосредственно с объектами знания (средства агрегации, семантического поиска и идентификации тождественных объектов) (см. [6–8]).

Как известно, большинство современных электронных архивов документов представляет собой наборы неструктурированных документов, на базе которых трудно организовать семантический поиск, извлечение метаинформации и различные информационные сервисы. В этом смысле архивы судов и наборы иных юридических документов не являются исключением. Кроме того, в настоящее время наблюдается значительное увеличение объема данных, включаемых в репозитории, что в свою очередь создает дополнительные трудности при обработке информации. Поэтому в условиях непрерывного роста объемов, а также многообразия информации активно развиваются новые подходы, инструменты и методы обработки огромных объемов данных, обозначаемых термином «большие данные» (Big Data). При управлении репозиториями электронных документов больших данных в полной мере остаются актуальными, а также появляются новые задачи, в их числе: семантическая разметка, организация поиска, выделение метаданных, формирование тематических кластеров документов, определение зависимостей, поиск близких документов и др. Насущными становятся проблемы анализа и управления данными в различных областях с интенсивным использованием данных. Часть описанных теоретических проблем применительно к юридической информации исследуется в рамках настоящего исследования.

Переход к представлению внутренней структуры знания создает новую парадигму представления, в которой основные акценты смещаются на выделение элементов (классов) и их взаимосвязей, что позволяет создавать различные сетевые концептуальные структуры (например, граф цитирования, граф концептов и др.). Выделение классов объектов и организация соответствующих репозиторияв позволят создать новые вычислительные возможности по обработке данных, такие, как извлечение и обработка терминов, поиск близких результатов и т. п.

Семантический поиск позволит по введенному описанию объекта или после выделения термина в исковом заявлении получить дополнительную информацию (определение, свойства, связи с другими объектами, список документов, в которых встречается объект, указание источника, где он впервые введен). С помощью такого поиска, например, можно найти все судебные решения, в обосновании которых прямо или косвенно используются те или иные нормативные акты.

Важно отметить, что в различных документах объект может обозначаться различными терминами.

Базовые научные разработки, положенные в основу системы, разработаны его участниками в рамках таких проектов, как «Thinking and understanding» и др. (см. [9]).

Система построена с использованием сервис-ориентированной архитектуры (Service-oriented architecture, SOA), точнее, с использованием микросервисов. Напомним, что SOA – это построение среды, в которой отдается предпочтение слабым связям, абстрагированию низкоуровневой логики, гибкости, а также возможности многократного использования и обнаружения компонентов [10, 11]. Дальнейшим развитием парадигмы сервис-ориентированной архитектуры можно считать появление архитектуры микросервисов [12]. Термин «Microservice Architecture» получил распространение в последние несколько лет для описания способа проектирования приложений в виде набора независимо развертываемых сервисов.

Архитектурный стиль микросервисов – это подход, при котором единое приложение строится как набор небольших сервисов, каждый из которых работает в рамках собственного процесса и взаимодействует с остальными. Сервисы построены вокруг бизнес-потребностей и развертываются независимо с использованием полностью автоматизированной среды. Централизованное управление минимизировано, а сами сервисы могут быть написаны на разных языках программирования и использовать разные технологии хранения данных. Более того, внутри каждого микросервиса вполне может быть задействована собственная база данных (см. [12]).

С учетом достаточно большого количества модулей системы необходимо выбрать подход к организации всего приложения и минимизировать зависимости, связанные с изменениями внутри отдельных модулей. При этом очевидно, что модули текстовой аналитики со временем будут изменяться, возможна реализация различных алгоритмов классификации и аналитики в зависимости от массива обрабатываемых документов. К тому же итеративный процесс разработки системы требует простых механизмов независимого обновления различных модулей. Это обуславливает применение архитектуры микросервисов для создания системы.

Одной из важнейших задач формируемой информационной системы являются поиск и предоставление аналогичных решений по схожим судебным искам. Таким образом, необходим сервис поиска аналогичных документов, или рекомендательный сервис.

Существуют два основных типа рекомендательных систем: контент-ориентированные и социальные (см., например, [13]). Первые основаны на представлении предпочтений пользователей путем анализа содержимого рекомендательных элементов. Системы второго типа моделируют предпочтения, оценивая близость профилей пользователей. Ниже под рекомендательным сервисом будем понимать информационную систему, которая:

1) формирует модель предметной области на основе массива документов (включая подготовительные операции – приведение к векторному виду, кластеризацию и т. п.);

2) получает на вход документ и выдает список документов, близких к входному.

В разрабатываемой рекомендательной системе реализованы следующие основные этапы:

- извлечение ключевых слов из документов коллекции на основе онтологий;
- представление каждой публикации в виде вектора, компоненты которого соответствуют концептам онтологии;
- определение значения компоненты – это вес соответствующего понятия в данной статье (вычисляется с использованием количества его упоминаний в тексте статьи и количества упоминаний связанных понятий);
- использование в качестве меры близости между публикациями косинусной меры близости между их векторами.

Важным аспектом выполнения проекта является апробирование результатов на принципиально различных коллекциях электронных документов, таких, как математические, гуманитарные и юридические документы, что позволит построить универсальную модель системы, адаптировать разрабатываемые сервисы к различным документам. Полная реализация проекта позволит фактически построить программную систему управления электронными документами,

способную консолидировать управление разнородными электронными научными коллекциями, и в конечном итоге представить набор универсальных семантических сервисов анализа научных данных.

Подход, разрабатываемый в рамках настоящего проекта, полностью соответствует мировой практике перехода на электронные средства представления и обмена информацией.

ТЕКУЩИЕ РЕЗУЛЬТАТЫ

Описываемая система находится на начальном этапе разработки, на текущий момент предложена верхнеуровневая архитектура системы [5], проведен обзор функциональности имеющихся систем, работающих в данной предметной области, определен набор востребованных дополнительных сервисов для системы. Разработаны концептуальная и инфологическая модели системы. Определен жизненный цикл искового заявления как электронного документа, построены основные процессы системы – загрузка, редактирование, поиск и просмотр [5].

В рамках решения задачи классификации проведены предварительный анализ судебных документов, отбор значимых признаков для определенных категорий судебного спора, проведен латентно-семантический анализ для выявления общей структуры типовых документов. На тестовой выборке проверены алгоритмы байесовской классификации, k-ближайшего соседа и деревьев решений, разрабатывается модель на основе искусственной нейронной сети. На следующем этапе планируется увеличить выборку исковых арбитражных заявлений и рассмотреть большее число типов возможных судебных споров, а также разработать программные модули, выполняющие задачи отбора информативных признаков и классификации.

Работа выполнена за счет средств субсидии, выделенной Казанскому федеральному университету для выполнения государственного задания в сфере научной деятельности, проект 2.8712.2017/8.9.

СПИСОК ЛИТЕРАТУРЫ

1. *Peroni S.* Semantic Web Technologies and Legal Scholarly Publishing Law, Springer, Governance and Technology Series, 2014. V. 15. doi 10.1007/978-3-319-04777-5
2. *Елизаров А. М., Жижченко А. Б., Жильцов Н. Г., Кириллович А. В., Липачёв Е. К.* Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Доклады Академии наук. 2016. Т. 467, № 4. С. 392–395. doi: 10.1134/S1064562416020174
3. *Елизаров А. М., Липачёв Е. К., Невзорова О. А., Соловьев В. Д.* Методы и средства семантического структурирования электронных математических документов // Доклады Академии наук. 2014. Т. 457, № 6. С. 642–645. doi 10.7868/S0869565214240049
4. *Грант С. Ингерсолл, Томас С. Мортон, Эндрю Л. Фэррис.* Обработка неструктурированных текстов. Поиск, организация и манипулирование/ Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – 414 с.: ил.
5. *Зуев Д. С., Марченко А. А., Хасьянов А. Ф.* Применение инструментов интеллектуального анализа текстов в юриспруденции // CEUR Workshop Proceedings. 2017. V. 2022. P. 214–218. <http://ceur-ws.org/Vol-2022/paper35.pdf>
6. Digital Mathematics Library: a vision for the future. International Mathematical Union, 2006. http://www.mathunion.org/fileadmin/IMU/Report/dml_vision.pdf.
7. *Olver P. J.* What's happening with the World Digital Mathematics Library? http://www.math.umn.edu/~olver/t_wdmlb.pdf
8. Developing a 21st century global library for mathematics research. Washington, D.C.: The National Academies Press, 2014. 131 p. arxiv.org/pdf/1404.1905; <http://www.nap.edu/catalog/18619/developing-a-21st-century-global-library-for-mathematics-research>.
9. *Toshev A., Talanov M.* Thinking Lifecycle as an Implementation of Machine Understanding in Software Maintenance Automation Domain// Jezic G., Howlett R., Jain L. (eds) Agent and Multi-Agent Systems: Technologies and Applications. Smart Innovation, Systems and Technologies. 2015. Vol 38. Springer, Cham. doi: 10.1007/978-3-319-19728-9_25

10. *Gold N. et al.* Understanding Service Oriented Software. IEEE Software. 2004. V. 21, No. 2. P. 71–77.
11. *Jones S.* Toward an Acceptable Definition of Service. IEEE Software. 2005. V. 22, No. 3. P. 87–93.
12. *Fowler M.* Microservices a definition of this new architectural term. <https://martinfowler.com/articles/microservices.html>
13. *Ricci F., Rokach L., Shapira B., Kantor P.B.* Recommender Systems Handbook. N.Y.: Springer, 2011. 842 p.

RECOMMENDER SYSTEM OF TEXT ANALYTICS OF LEGAL DOCUMENTS

¹D. S. Zuev, ²M. F. Nasrutdinov, ³A. F. Khassianov

Kazan (Volga region) Federal University

¹dzuev11@gmail.com, ²marat.nasrutdinov@kpfu.ru, ³ak@it.kfu.ru

Abstract

The paper discusses the use of machine learning mechanisms, natural language analysis and intellectual search in the field of jurisprudence. The main expected results are the methodology for applying text-based analytics and semantic natural language processing (NLP) algorithms in knowledge management cases in different types of legal practice. The obtained results can be applied in the field of education and knowledge management in a wider context, since the study lies at the union of jurisprudence, mathematical and computer linguistics.

We describe a prototype of a multi-agent system of intellectual analysis of legal texts that is capable of identifying general dependencies on the existing database of legal documents, providing legal cases with similar topics, recommending the most likely outcomes of judicial review.

Keywords: *data analytics and data mining, data intensive domains, digital libraries, clustering, classification of judicial acts, recommender system, micro-service architecture.*

REFERENCES

1. *Peroni S.* Semantic Web Technologies and Legal Scholarly Publishing Law, Springer, Governance and Technology Series, 2014. V. 15. doi 10.1007/978-3-319-04777-5
2. *Elizarov A.M., Zhizhchenko A.B., Zhil'tsov N.G. Kirillovich A. V., Lipachev E.K.* Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics// Doklady Mathematics. 2016. V. 93, No 2. P. 231–233. <https://doi.org/10.1134/S1064562416020174>
3. *Elizarov A.M., Lipachev E.K., Nevzorova O.A., Solov'ev V.D.* Methods and means for semantic structuring of electronic mathematical documents // Doklady Mathematics. 2014. V. 90. No 1. P. 521–524. doi 10.7868/S0869565214240049
4. *Grant S. Ingersoll, Thomas S. Morton, and Andrew L. Farris* Taming Text: How to Find, Organize, and Manipulate It / Manning Publications, 2012. – 320 p. ISBN: 9781933988382
5. *Zuev D.S., Marchenko A.A., Khassianov A.F.* Text Mining Tools in Legal Documents // CEUR Workshop Proceedings. 2017. V. 2022. P. 214–218. <http://ceur-ws.org/Vol-2022/paper35.pdf>
6. Digital Mathematics Library: a vision for the future. International Mathematical Union, 2006. http://www.mathunion.org/fileadmin/IMU/Report/dml_vision.pdf.
7. *Olver P.J.* What's happening with the World Digital Mathematics Library? http://www.math.umn.edu/~olver/t_/wdmlb.pdf
8. Developing a 21st century global library for mathematics research. Washington, D.C.: The National Academies Press, 2014. 131 p. arxiv.org/pdf/1404.1905; <http://www.nap.edu/catalog/18619/developing-a-21st-century-global-library-for-mathematics-research>.
9. *Toshev A., Talanov M.* Thinking Lifecycle as an Implementation of Machine Understanding in Software Maintenance Automation Domain// Jezic G., Howlett R., Jain L. (eds) Agent and Multi-Agent Systems: Technologies and Applications. Smart Innovation, Systems and Technologies. 2015. Vol 38. Springer, Cham. doi: 10.1007/978-3-319-19728-9_25

10. *Gold N. et al.* Understanding Service Oriented Software. IEEE Software. 2004. V. 21, No 2. P. 71–77.

11. *Jones S.* Toward an Acceptable Definition of Service. IEEE Software. 2005. V. 22. No 3. P. 87–93.

12. *Fowler M.* Microservices a definition of this new architectural term. <https://martinfowler.com/articles/microservices.html>

13. *Ricci F., Rokach L., Shapira B., Kantor P.B.* Recommender Systems Handbook. N.Y.: Springer, 2011. 842 p.

СВЕДЕНИЯ ОБ АВТОРАХ



ЗУЕВ Денис Сергеевич – кандидат технических наук, заместитель директора по научной деятельности Высшей школы информационных технологий и интеллектуальных систем Казанского федерального университета.

Denis Sergeevich ZUEV – PhD, Deputy director for research, Higher Institute of Information Technology an Intelligent Systems

email: dzuev11@gmail.com



НАСРУТДИНОВ Марат Фаритович – кандидат физико-математических наук, заместитель директора по образовательной деятельности Высшей школы информационных технологий и интеллектуальных систем Казанского федерального университета.

Marat Faritovich NASRUTDINOV – Deputy Director for Education at Higher Institute for Information Technology and Information Systems of Kazan Federal University

marat.nasrutdinov@kpfu.ru



ХАСЬЯНОВ Айрат Фаридович – PhD, директор Высшей школы информационных технологий и интеллектуальных систем Казанского федерального университета.

Airat Faridovich KHASSIANOV – Phd, Director at Higher Institute for Information Technology and Information Systems of Kazan Federal University

ak@it.kfu.ru

Материал поступил в редакцию 31 июля 2018 года

УДК 004.91

СИСТЕМА АВТОМАТИЗИРОВАННОГО УЧЕТА ОПЛАТЫ ОБУЧЕНИЯ В ВУЗЕ

А. О. Федоров

Чувашский государственный университет им. И. Н. Ульянова, Московский просп.,
15, г. Чебоксары, Чувашская Респ., 428010

aofedorov@gmail.com

Аннотация

Рассмотрен вопрос возможности применения новых информационных технологий при организации учета оплаты обучения в вузе. Внедрение такой автоматизированной системы позволило уменьшить количество нарушений условий и сроков оплаты, а также увеличить стабильное поступление денежных средств в размере сумм, указанных в договоре об образовании.

Ключевые слова: образование, информационные технологии, университет, персональные данные, безопасность

ВВЕДЕНИЕ

Успех решения основных задач, стоящих перед крупным региональным вузом (обеспечение эффективного использования преподавательских кадров и материальных ресурсов, построение внутривузовской системы контроля качества, планирование и мониторинг достижения стратегических целей), напрямую зависит от тех инструментов и возможностей, которые предоставляет информационная система вуза для получения своевременной и полной информации.

Внедрение информационной системы учета оплаты обучения в вузе позволяет наиболее быстро и оперативно получать нужную информацию об оплате учебы студентами, повышать точность и оперативность работы с документацией, автоматизировать формирование различных отчетных документов, что значительно уменьшает временные, соответственно, и материальные затраты.

Представлен опыт Чувашского государственного университета им. И. Н. Ульянова (далее – Университет) по созданию информационной системы автоматизации учета оплаты обучения.

1. ЗАДАЧИ, СТОЯЩИЕ ПЕРЕД БУХГАЛТЕРИЕЙ ВУЗА

Одной из задач бухгалтерии вуза является учет оплаты договоров об обучении. Студенты, обучающиеся в вузе на платной основе, заключают с Университетом договор на платное обучение. Для каждого студента, с которым заключен договор, ведется лицевой счет, где отражаются все начисления и выплаты за обучение.

Сумма оплаты начисляется за определенный период, как правило, за один учебный год (с сентября по июнь). Стоимость обучения на последующих курсах начисляется через определенный период времени, обычно – перед началом каждого учебного года. Для каждого студента вычисляется сальдо расчетов, которое может быть дебетовым (начисления превышают оплаты) или кредитовым (оплаты превышают начисления).

При заключении договора студенту присваивается индивидуальный номер, который является уникальным. Этот код необходим для идентификации студента при всех расчетах.

Оплата учебы студентом осуществляется через кассу Университета или банковские перечисления. Плата за обучение может вноситься полностью (сразу) или частично (в несколько этапов).

Сведения о выплатах, проведенных студентами, формируются ежедневно. Каждый день в Университет из банка приходят выписки о движении денежных средств на расчетном счете за предыдущий день. В такой выписке содержится информация и о перечислениях денег студентами за обучение.

Использование автоматизированных методов учета движения денежных средств позволяет решать названные выше задачи функционирования бухгалтерии более эффективно и поддерживать ее работу с меньшими затратами труда, чем при использовании ручной обработки документации. Исходя из этого, возникает необходимость разработки системы, которая обеспечила бы автоматизацию учета оплаты обучения студентами при сохранении в ней всех функций, выполняемых при ведении учета бумажными методами.

2. ВЫБОР АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ

При проектировании автоматизированной информационной системы (ИС) учета оплаты обучения стоял непростой выбор:

- построение ИС на основе ERP-систем (Ахapta, SAP R3 и пр.);
- приобретение готовых программных разработок в данной области;
- выполнение собственных разработок [1].

Каждое из указанных направлений имеет как сильные, так и слабые стороны. Недостатками первого подхода, помимо крайне высокой стоимости лицензии на саму ERP-систему, является значительная трудоемкость в ее настройке и адаптации.

Стоимость готового продукта несоизмеримо ниже, однако отсутствие жестко регламентированных бизнес-процессов и специфические особенности вузов являются сдерживающим фактором для их внедрения.

Третий подход, который использует большинство вузов, – выполнение собственных разработок.

Использование готовых систем сторонних производителей было отклонено нами по той причине, что готовые системы требуют доработок в соответствии со спецификой Университета, а также ввиду необходимости непрерывной поддержки системы производителем, иначе система теряет свою актуальность через короткий срок. Поэтому Университет не стал исключением, и было принято решение заняться разработкой и внедрением собственной информационной системы.

Таблица 1. Взаимосвязь целей и вероятностей их реализации

№ п/п	Цели внедрения автоматизированной системы	Оценка вероятности по 10 бальной шкале	
		Новая автоматизированная система	Старая система, основанная на Excel
1.	Повысить скорость выполнения операций по сбору, передаче, обработке и выводу информации	9	3

2.	Выполнение задач, поставленных перед пользователем системы	8	4
3.	Проведение оперативного контроля хода работы пользователя в системе	9	0
4.	Применение современных методов рационализации использования информационных ресурсов	5	1
5.	Сокращение дублирующих функций	10	0
6.	Сокращение сроков формирования отчетных документов	10	2
7.	Планирование процессов учета оплаты	8	4
8.	Контроль и регулирование процесса учета оплаты со стороны пользователя	10	0
9.	Повышение качества учета оплаты обучения	8,6 (86%)	1,8 (18%)

Выбирая третий путь (самостоятельное проектирование информационной системы), мы учли следующие факторы.

Организационные аспекты

1. При проектировании информационной системы были тщательно проанализированы и формализованы все бизнес-процессы внутри Университета.

2. Выполнен комплекс работ по упорядочиванию документооборота, выработке и согласованию внутривузовских положений и стандартов.

3. Внедрение информационной системы носило поэтапный характер.

Экономические аспекты

Стоимость создания информационной системы складывалась из затрат на: проектирование и разработку самой системы; оборудование (серверное, сетевое, коммутационное), создание сетевой инфраструктуры; приобретение лицензионного программного обеспечения; внедрение информационной системы, включая обучение; сопровождение ИС.

Экономическая эффективность обуславливалась сокращением трудозатрат на организацию работы бухгалтерии по ведению документации и получению информации по необходимым формам. Помимо экономии времени улучшение показателей качества работы связано со своевременным получением информации, хранящейся в базе данных, повышением контроля правильности ввода информации, использования более информативных и наглядных документов, сокращением рутинных вычислений при получении выходных документов.

Технологические аспекты

1. Разработанная информационная система имеет полнофункциональный характер, автоматизируя основные функции бухгалтерского учета.

2. В основе лежит единая база данных, созданная на основе СУБД MySQL.

3. Особое внимание уделено вопросам информационной безопасности (разграничение прав доступа, использование защищенного протокола передачи данных).

4. Информационная система является кроссплатформенной и доступна всюду, где есть веб-браузер.

Технические аспекты

1. Учитывая существенные объемы информации, информационная система базируется на высокопроизводительном серверном оборудовании, обеспечивающем высокую надежность и резервирование.

2. Для обеспечения надежного хранения информации предусмотрена многоуровневая система архивирования данных с использованием современных технологий и аппаратного обеспечения.

3. Рабочие места обеспечены надлежащим комплектом периферийного оборудования (принтеры, сканеры и пр.), необходимого для эффективной работы.

3. ФУНКЦИОНАЛЬНОСТЬ СИСТЕМЫ

Разработанная система автоматизации учета оплаты обучения обеспечивает:

- учет студентов-контрактников (хранение записей в базе данных по каждому студенту);
- обеспечение изменения состояния записей в базе;

- учет оплаты студентами обучения;
- формирование отчетов о состоянии оплаты в течение семестра, отчетов о планируемом поступлении средств, отчетов по должникам;
- ежедневный обмен информацией с бухгалтерией Университета;
- расширяемость.

В схеме 1 представлены основные процессы, необходимые для учета эффективности внедрения автоматизированной системы учета оплаты обучения. Как результат, автоматизированная система позволила:

- повысить скорость обработки информации о студентах и скорость учета оплаты;
- сократить сроки формирования отчетов и дублирование функций по учету студентов;
- обеспечить контроль пользователя;
- сэкономить время.

Программный продукт предназначен для использования его в архитектуре клиент/сервер. Учет студентов-контрактников осуществляется через веб-интерфейс (рисунок 1).

Личная карточка студента

Иванова Ирина Сергеевна

Номер договора: 5 Фарм-18
Идентификатор договора (UID): 031330501180005

Факультет: Химико-фармацевтический
Форма обучения: очная
Курс: 1
Программа подготовки: 33.05.01 - Фармация *Специалитет*
Профиль: Фармация

Схема платежей: 4 раза в год (до 31.10, 10.12, 28.02, 25.04)

Платежи

Учебный год	Тариф	К оплате		Скидка	Оплачено (Ексел)	Переплата (Ексел)	Переплата / долг на начало	Платежи (бухгалтерия)	
		1 семестр	2 семестр					Всего:	2018-07-27
2018 / 2019	94020	47010	47010	6581				50000	50000.00 касса

Предоставленные отсрочки по платежам

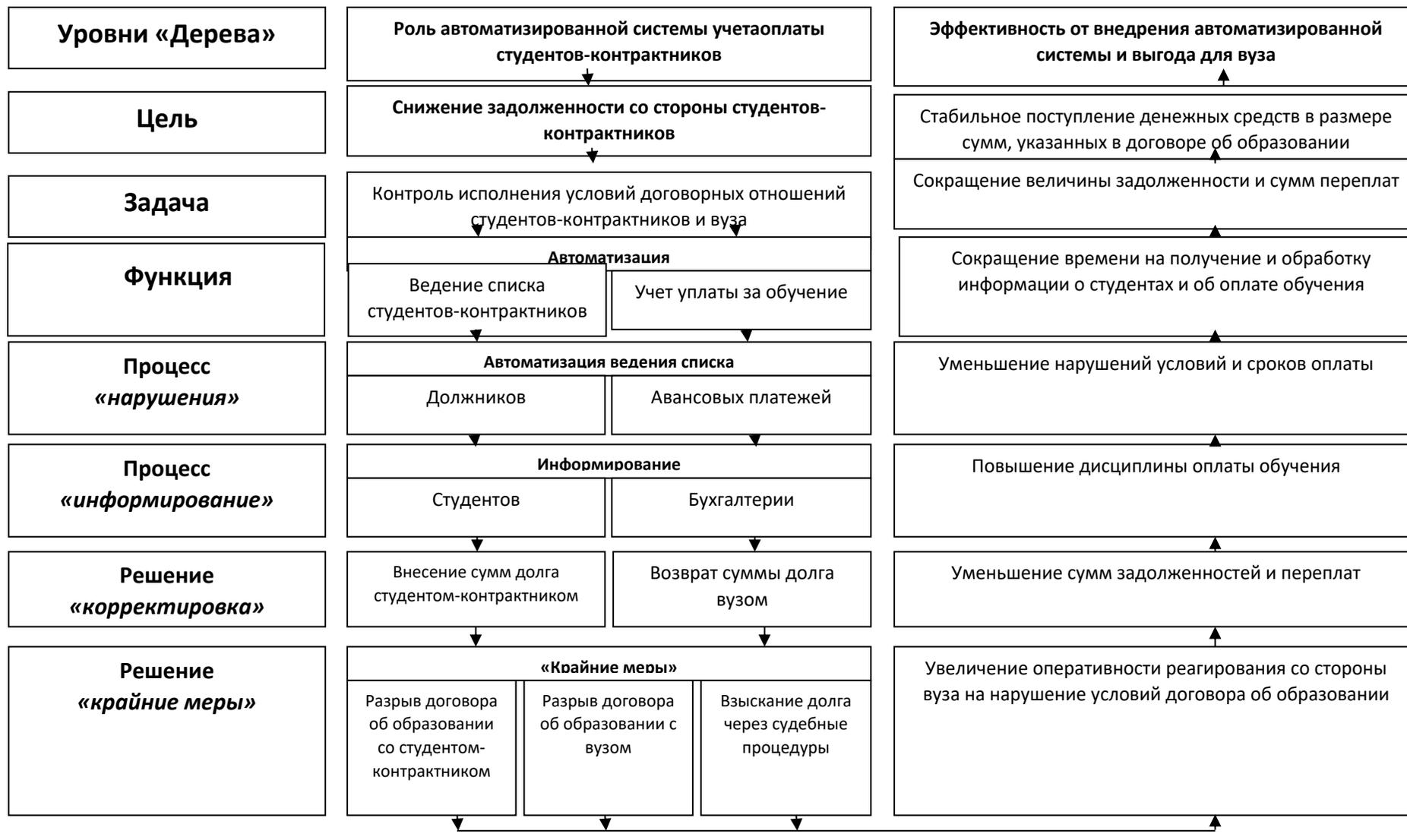
Учебный год	Семестр	Отсрочка до
Записи отсутствуют.		

В проектах приказов на отчисление

Учебный год	Семестр
Записи отсутствуют.	

Рисунок 1. Личная карточка студента

Схема 1. Процесс реализации автоматизированного учета оплаты обучения



Иванова Ирина Сергеевна (Id: 20060)

Факультет: Химико-фармацевтический
Форма обучения: очная
Уровень образования: (65) - Специалитет
Образовательная программа: Фармация

Учеба **Персональные данные** **Заказчик**

Договор № 5 - Фарм - 2018 от 2018-07-27
Идентификатор договора (UID): 031330501180005

Профиль: Фармация
Курс: 1
Базовое образование: СО
Срок обучения по УП: 5
Курс поступления: 1
Дата начала обучения: 2018-09-01
Основание: Зачислен на 1 курс

Движение:

Схема платежей: 4 раза в год (до 31.10, 10.12, 28.02, 25.04)
Текущий статус: Учится

Предыдущий контракт: 0

Сохранить Вернуться

Рисунок 2. Редактирование личной карточки студента

Учет оплаты студентами обучения осуществляется на основании данных, ежедневно выгружаемых бухгалтерией Университета.

Поиск платежа

Фамилия студента: Дата оплаты:
 Имя студента: Откуда выполнен платеж:
 № п/к ордера: Сумма оплаты:

№ п/к ордера	Дата оплаты	Номер договора	Дата заключения договора	ФНО, факультет, специальность, форма обучения	Откуда выполнен платеж	Сумма оплаты	Привязка	
							Студент	Учебный год
00020283	2018-08-13	26 МарЗЮр-18	0000-00-00	Иванова Ирина Леонидовна	касса	21500.00		
00019700	2018-08-06	32 Эж-18	0000-00-00	Иванова Ирина Геннадьевна	касса	20000.00		
иль1017	2018-07-31		0000-00-00	Иванова Ирина Васильевна	безналично	3000.00		
00019198	2018-07-27	Фарм-18	0000-00-00	Иванова Ирина Сергеевна	касса	50000.00	Иванова Ирина Сергеевна 5 Фарм-18 Историко-фармацевтический факультет 33.05.01 - Фармация очная	2018/2019
00001135	2018-06-09	ОРД-17	0000-00-00	Иванова Ирина Олеговна Клиническая ординатура Стоматология Дневная	эквайринг	24952.00		
мй1129	2018-05-31	45 ЗРЧ-Филол-16	2018-09-01	Еливанова Ирина Петровна ЗРЧ-Филол: Русской и чувашской ф Филология Зачочная	безналично	12500.00	Еливанова Ирина Петровна 45 ЗРЧ-Филол-16 Русской и чувашской филологии и журналистики факультет 45.03.01 - Филология очная	2017/2018

Рисунок 3. Поиск платежей

Формирование отчетов о состоянии оплаты в течение семестра, отчетов о планируемом поступлении средств, отчетов по должникам осуществляется за считанные секунды.

Платежи за период | Статистика по месяцам | Возвраты | Направления в кассу через ИК | Ожидаемые платежи | Платежи по странам | Платежи юр.лиц

Платежи за период с по

Рисунок 4. Формирование отчетов

Списки должников

Факультет: По состоянию на
 Учебный год:
 Дополнительные фильтры:
 Форма обучения: Образовательная программа:
 Уровень образования: Профиль:
 Курс: Программа обучения:
 Гражданство: Дистанционные технологии:

Форма обучения	Курс	Программа подготовки	№ контракта	Фамилия, имя, отчество	Год поступления	Профиль	Сумма просроченной задолженности
очная	2	43.03.02	3 ИГФ-16	Азанов Сергей Алексеевич	2016		16 644.00
заочная	1	21.03.02	32 ЗИГФ-17	Горбунов Анатолий Ильич	2017		22 000.00
заочная	1	21.03.02	29 ЗИГФ-17	Орлова Юлия Сергеевна	2017		12 000.00
заочная	3	05.03.06	27 ЗИГФ-15	Ильин Владимир Сергеевич	2015		10 749.00
заочная	4	05.03.06	64 ЗИГФ-14	Краснов Сергей Витальевич	2014		8 362.00
заочная	4	05.03.06	33 ЗИГФ-14	Никитин Алексей Валерьевич	2014		11.00

Рисунок 5. Пример отчета по должникам

С помощью разработанного базового инструментария на данный момент реализованы процессы формирования учета оплаты студентами обучения и обеспечение контроля самим студентом оплаты своего обучения. Разрабатываемая

информационная система носит общеуниверситетский характер и является развитием локальных наработок, созданных в Университете.

ЗАКЛЮЧЕНИЕ

Представлен разработанный программный продукт для учета оплаты обучения студентами Чувашского государственного университета им. И.Н. Ульянова, который может быть использован и другими вузами. Внедрение и использование этой информационной системы показали ее эффективность. В дальнейшем планируется усовершенствовать данный инструмент с помощью его внедрения в информационно-образовательную среду Университета.

СПИСОК ЛИТЕРАТУРЫ

1. *Афанасьев Ю.А., Гужов В.И., Стасышин В.М.* Некоторые аспекты разработки информационной системы автоматизации образовательного процесса в масштабах вуза. URL: <http://www.ict.edu.ru/ft/003844>

INFORMATION SYSTEM OF PAYMENT ACCOUNTING IN THE UNIVERSITY

A. O. Fedorov

I.N. Ulyanov Chuvash State University, Moscow Avenue, 15, Cheboksary, Chuvash Republic, 428010

aofedorov@gmail.com

Abstract

The experience of application of information technologies in the organization of accounting of payment in the university is presented. The introduction of the information system made it possible to reduce the number of violations of terms and conditions of payment, as well as to increase the steady flow of funds.

Keywords: *education, information technology, university, personal data, security*

REFERENCES

1. *Afanasyev Yu., Guzhov V., Stasyshin V.* Some aspects of the development of an information system for automation of the educational process at the scale of the university. URL: <http://www.ict.edu.ru/ft/003844>

СВЕДЕНИЯ ОБ АВТОРЕ



ФЕДОРОВ Андрей Олегович – кандидат педагогических наук, доцент, начальник отдела по работе со студентами, обучающимися на договорной основе, Чувашского государственного университета им. И.Н. Ульянова.

Andrey Olegovich FEDOROV – candidate of pedagogical sciences, associate professor, head of department for work with students studying on a contractual basis of the Chuvash State University.

email: aofedorov@gmail.com

Материал поступил в редакцию 31 июля 2018 года

УДК 004.023

АВТОМАТИЗАЦИЯ ПОСТРОЕНИЯ РАСПИСАНИЯ ЗАНЯТИЙ В ВУЗЕ: МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДЫ РЕАЛИЗАЦИИ

Р. И. Хабилов

Казанский (Приволжский) федеральный университет, г. Казань, ул. Кремлевская,
д. 35, 420008

paladin86@bk.ru

Аннотация

Построение расписания – это распределение дискретного множества событий в заданном временном интервале, с соблюдением заданных ограничений. Целью работы является описание математической модели автоматизации процесса составления расписания в образовательном учреждении. Рассмотрены также подходы к составлению оптимального расписания. В исследуемых задачах используется большое количество первоначальной исходной информации, различающейся по своему составу, и содержится большое количество требований, которые должны быть обязательно учтены. Поэтому составление расписания относится к классу NP-полных задач целочисленного программирования, который подразумевает, что с ростом числа значений заданных переменных сложность решения будет расти экспоненциально. Отметим, что качество сформированного расписания занятий напрямую влияет на эффективность учебного процесса вуза.

Описан процесс формирования расписания занятий: на первом этапе нужно сформировать начальное расписание на основе имеющихся студенческого контингента и преподавателей, учебных аудиторий, а также ряда дополнительных ограничений; на втором этапе происходит оптимизация первоначального расписания; на третьем этапе допускается корректировка полученного расписания сотрудниками вуза.

Ключевые слова: задача составления расписания, алгоритмы целочисленного линейного программирования, учебные планы, расписание занятий

ВВЕДЕНИЕ

Проблемой разработки методов и алгоритмов для решения задачи составления учебного расписания занимаются уже достаточно давно. Первые работы появились еще в 1970-е гг. [1, 2]. Наиболее близкими являются исследования [3]. В последние годы достаточно интенсивно развиваются приближенные и эвристические методы решения этой задачи [5].

Решение задач теории расписаний усложняется тем фактом, что большинство из них являются NP-трудными [1], и алгоритмы их решения, реализованные на ЭВМ, должны быть адаптированы для получения приемлемого варианта за предельно допустимое время.

Формирование оптимального расписания учебных занятий в вузе на учебный год требует от сотрудников, выполняющих эту работу, больших затрат времени, от нескольких дней до нескольких недель работы. Расписание занятий должно удовлетворять многочисленным требованиям организационного и методического характера, имеющим различные приоритеты, которые могут быть взаимно противоречивыми или даже взаимоисключающими. Такие требования могут значительно затруднить построение первоначального расписания, не говоря уже об оптимальном расписании.

При построении расписания в качестве исходных данных выступают учебная нагрузка, учебные группы, учебные аудитории, преподаватели и список дисциплин. Задача разработки оптимального расписания сводится к распределению максимального количества занятий в день в имеющихся учебных аудиториях для соответствующих студенческих групп, с учетом нагрузки и пожеланий преподавателей.

Существует несколько подходов [1–8] к построению расписания, связанных с применением:

- классических методов и алгоритмов целочисленного программирования;
- приближенных алгоритмов решения методом раскраски графов;
- алгоритма полного перебора, ветвей и границ, а также приближенных методов, которые основаны на генетических алгоритмах.

Используя данные алгоритмы и методы, можно получить решение задачи построения расписания, которое будет удовлетворять заданным критериям. Однако время, затраченное на решение этой задачи, может быть очень большим. Поэтому, чтобы найти оптимальное решение, применяют метод моделирования.

При таком методе алгоритм работает с уже заранее заданным списком занятий, которые должны быть включены в расписание. Процесс начинается с первого занятия, еще не отраженного в расписании, затем алгоритм заполняет соответствующий раздел расписания, используя максимальное количество занятий из имеющегося списка. Далее в расписание включается следующее занятие. Процесс продолжается до тех пор, пока не будет сформировано полное расписание или же не будет выполнено заданное число повторений.

Задача построения оптимального расписания для крупных вузов является достаточно сложной и трудоемкой, поэтому автоматизация такого процесса является актуальной. Кроме этого, чем крупнее вуз, тем актуальность в автоматизации названного процесса выше.

1. ПОСТАНОВКА ЗАДАЧИ

На основе имеющихся списка занятий, перечня преподаваемых дисциплин, контингента студентов, перечня учебных аудиторий и утвержденных учебных планов необходимо рассчитать учебную нагрузку всех кафедр вуза на текущий учебный год, а затем на основании рассчитанной учебной нагрузки необходимо сформировать расписание на этот период.

Выделим особенности решения задачи построения расписания:

- так как расписание традиционно составляется вручную, возникает необходимость в создании систем автоматизированного составления расписания, которые были бы внедрены в систему управления учебным процессом вуза;
- для обеспечения эффективности учебного процесса полученное расписание необходимо оптимизировать;
- для визуализации и оценки работы при составлении расписания необходимо разработать удобную для анализа форму его представления.

Алгоритм формирования расписания занятий реализуется в два этапа:

- на основе исходных данных, списка требований и ограничений формируется начальное расписание;
- полученное начальное расписание оптимизируется, и далее с помощью выбранных критериев оценивается его качество.

Основными составляющими элементами расписания являются:

- временной интервал – интервал времени для проведения учебного занятия, который соответствует определенному дню в учебном году; другими словами, это конкретный интервал времени, в который должно быть проведено занятие;
- аудитория – множество мест, возможных для проведения занятий.

Таким образом, начальное расписание считается сформированным, если каждому учебному занятию соответствуют определенный временной интервал и учебная аудитория.

2. МОДЕЛЬ РАСПИСАНИЯ

Обозначим множество учебных занятий $E = e_i, i=1..N$. Каждое из них характеризуется такими параметрами, как учебная дисциплина, аудитория, студенческая группа, преподаватель.

Учебные задания формируются для каждой дисциплины и каждой учебной группы. Множество учебных заданий можно представить как объединение двух подмножеств $E = E_T \cup E_N$, где E_T – учебные задания со специальными требованиями к аудиториям, например, для работы в компьютерных классах, E_N – учебные задания без специальных требований к аудиториям.

Введем множество временных интервалов T – это интервалы времени, которые однозначно определяются номерами дня недели и самой недели. Таким образом, множество временных интервалов получается с помощью декартового произведения двух временных параметров – недели и дня, иначе говоря, это дата–время: $E = E_T \cup E_N$.

Обозначим множество академических групп и соответственно множество атрибутов групп:

$$X = X_m, m=1..M, A_x = \{(a_{x1}, a_{x2}, a_{x3})\},$$

где a_{x1} – количество запланированных (согласно учебной нагрузке) групп, a_{x2} – количество уже распределенных по временным интервалам дисциплины в этой группе, a_{x3} – количество студентов в группе.

В некоторых случаях группы разбивают на подгруппы, что, как правило, связано с вместимостью аудиторий или спецификой преподаваемого предмета. Множества групп и подгрупп охватывают весь контингент студентов.

Введем следующие множества:

множество студентов

$$Z = X \cup Y, \quad Z = \{Z_p, p = 1..Q\},$$

где X – множество студентов женского пола, Y – множество студентов мужского пола;

множество преподавателей $PA_k = \{a_{p1}, a_{p2}\}$, где a_{p1} – количество в расписании дисциплин, преподаваемых данным преподавателем, a_{p2} – количество дисциплин, преподаваемых данным преподавателем и распределенных в расписании.

Пусть множество аудиторий $A = a_k, k=1..K$, а множество атрибутов для каждой аудитории $A_a = \{(a_{a1}, a_{a2}, a_{a3})\}$, где a_{a1} – общее количество дисциплин (согласно утвержденной учебной нагрузке), возможных для проведения в данной аудитории за рассматриваемый период, a_{a2} – количество распределенных в расписании дисциплин в данной аудитории, a_{a3} – вместимость аудитории.

Таким образом, конкретное учебное занятие в расписании можно представить как функцию параметров дисциплины, контингента студентов и преподавателя:

$$e_i = e(d_j, z_o, p_i); \quad d_j \in D, \quad z_o \in Z, \quad p_i \in P.$$

3. НАКЛАДЫВАЕМЫЕ ОГРАНИЧЕНИЯ

Основным критерием реализуемости сформированного расписания является его непротиворечивость: это означает, что в расписании нет ситуаций, когда разные занятия, требующие одинаковых ресурсов, размещены в одном и том же временном интервале расписания (аудитория, студенты, кафедра или преподаватель). Соблюдение обязательных ограничений при формировании расписания является достаточным условием для его непротиворечивости.

Для случая учебных занятий ограничения, касающиеся преподавателей, будут иметь следующий вид:

- 1) отсутствие наложений учебных групп

$$\forall z_o \in Z, \forall t_k \in T, \forall d_j \in D \\ d_j \neq d_i \leftrightarrow \omega(d_j, z_o, p, a, t_k) \neq \omega(d_i, z_o, p, a, t_k);$$

- 2) отсутствие совмещенных аудиторий

$$\forall a_p \in A, \forall t_k \in T, \forall d_i \in D, \forall d_j \in D \\ d_j \neq d_i \leftrightarrow \omega(d_j, g, p, a, t_k) \neq \omega(d_i, g, p, a, t_k);$$

- 3) отсутствие совмещенных по времени занятий у одного и того же преподавателя

$$\forall p_i \in P, \forall t_k \in T, \forall d_i \in D, \forall d_j \in D \\ d_j \neq d_i \leftrightarrow \omega(d_j, g, p, a, t_k) \neq \omega(d_i, g, p, a, t_k).$$

Таким образом, для каждой упорядоченной двойки (преподаватель, занятие) либо существует единственный цикл занятий, в котором «занят» данный преподаватель в указанное время, либо такого цикла не существует, следовательно, преподаватель свободен.

При формировании начального расписания и определении временного интервала и аудитории для занятия учитываются также следующие ограничения:

- обеспечение равномерности распределения занятий в каждой группе;
- обеспечение равномерности распределения занятий во временных интервалах.

Первое ограничение обусловлено необходимостью получения равномерных интервалов между занятиями, второе – рациональным использованием аудиторного фонда.

4. КРИТЕРИЙ ЗАГРУЖЕННОСТИ УЧЕБНОГО ЗАНЯТИЯ В РАСПИСАНИЕ

Каждое учебное занятие $E = e_i, i=1..l$, можно оценить, т. е. поставить ему в соответствие некоторый числовой критерий:

$$V^E = K_X \times K_P \times K_A \times K_Y, \quad 0 \ll K_X, K_P, K_A, K_Y \ll 1,$$

где K_X, K_P, K_A – частные критерии загруженности групп, преподавателей и аудитории, учета занятий в подгруппах, которые рассчитываются по формуле

$$K_i = \frac{a_1^i - a_2^i}{a^T - a_2^i}, \quad (1)$$

где a_1^i – количество запланированных занятий за период, a_2^i – количество распределенных критериев загруженности аудитории, определяемые следующим образом:

- учебные занятия, для которых требуются аудитории, например, дисплейные классы для проведения занятий по программированию; для них K_i рассчитываются по формуле (1);
- учебные занятия, для которых не требуются аудитории; частный критерий для учебных заданий этого подмножества рассчитывается следующим образом:

$$K_A^0 = \frac{a_1^0 - a_2^0}{a^0 - a_2^0},$$

где a_1^0 – количество учебных занятий без требуемых аудиторий, a_2^0 – количество включенных в расписание учебных заданий без требуемой аудитории, a^0 – количество временных интервалов доступных аудитории.

Критерий загруженности учебного занятия включает в себя частные критерии загруженности группы, преподавателя и аудитории и является итоговым критерием. Описанные критерии позволяют оценить средний интервал между занятиями для группы или подгруппы. Каждый из частных критериев является относительной величиной, возможность переопределения значения частного критерия обеспечивается изменением количества последовательно включаемых в расписание пар ресурса.

ЗАКЛЮЧЕНИЕ

Предложены математическая модель и метод создания начального расписания с применением критериев загруженности в выборе учебных заданий. Предлагаемый метод основан на базе данных вуза, в которой могут быть автоматически сформированы учебные задания для проведения занятий, и используются критерии загруженности, учитывающие требования пользователей и распределяющие ресурсы, что дает более оптимальный результат и приемлемое быстрое действие в пределах 10%, чем громоздкие методы, требующие больших вычислительных затрат.

СПИСОК ЛИТЕРАТУРЫ

1. Конвей Р.В., Максвелл В.Л., Миллер Л.В. Теория расписаний. М.: Наука, 1975. 360 с.
 2. Федотов А.Ф., Трунов Н.Н. Учебно-организационная работа в вузе. Л.: ЛПИ, 1980. 112 с.
 3. Безгинов А.Н., Трегубов С.Ю. Обзор существующих методов составления расписаний // Информационные технологии и программирование: межвуз. сб. ст. Вып. 2 (14). М.: МГИУ, 2005. 60 с.
 4. Кабальнов Ю.С., Шехтман Л.И., Низамова Г.Ф., Земченкова Н.А. Композиционный генетический алгоритм составления расписания учебных занятий // Вестник Уфимского государственного авиационного технического университета. 2006. Т. 7, № 2. С. 99–107.
 5. Гранков М.В., Аль-Габри В.М., Горлова М.Ю. Анализ и кластеризация основных факторов, влияющих на успеваемость учебных групп вуза // Инженерный вестник Дона, 2016, № 4, URL: ivdon.ru/ru/magazine/archive/n4y2016/3775 (дата обращения: 21.08.2018).
 6. Саати Т. Принятие решений. Метод анализа иерархий. М.: Радио и связь, 1993. 278 с.
 7. Семенов С.П., Татаринцев Я.Б. Сравнительный анализ подходов к автоматизации составления расписаний учебных занятий в образовательных учреждениях [Электронный ресурс] Режим доступа: <http://cyberleninka.ru/article/n/sravnitelnyy-analiz-podhodov-kavtomatizatsii-sostavleniya-raspisaniy-uchebnyh-zanyatiy-v-obrazovatelnyh-uchrezhdeniyah.pdf> (дата обращения: 10.07.2018).
 8. Танаев С.В., Шкурба В.В. Введение в теорию расписаний. М.: Наука, 1975. 257 с.
-

AUTOMATION OF THE CREATION OF THE SCHEDULE AT THE UNIVERSITY: THE MATHEMATICAL MODEL AND METHODS OF IMPLEMENTATION

R.I. Khabipov

Kazan (Volga region) Federal University

paladin86@bk.ru

Abstract

The construction of the schedule is the distribution of a discrete set of events in a given time interval, subject to the specified restrictions. The aim of the work is to describe a mathematical model of the automation of the scheduling process in an educational institution. Also the approaches to the compilation of the optimal schedule are considered. In the studied tasks a large amount of initial original information is used, which differs in its composition, and contains a large number of requirements that must be taken into account. Therefore, scheduling refers to the class of NP-complete integer programming problems, which implies that as the number of values of given variables increases, the complexity of the solution will grow exponentially. Note that the quality of the established lesson schedule directly affects the efficiency of the educational process of the university.

The article describes the process of forming the schedule of classes: at the first stage, you need to create an initial schedule based on the existing student contingent and teachers, audiences, as well as a number of additional restrictions; at the second stage, the initial schedule is optimized; at the third stage, it is allowed to adjust the received schedule by university staff.

Keywords: *the task of scheduling, algorithms of integer linear programming, educational plans, class schedule.*

REFERENCES

1. Conway R.V., Maxwell V.L., Miller L.V. Theory of schedules. M.: Science, 1975. 360 p.
2. Fedotov A.F., Trunov N.N. Educational and organizational work in high school. L.: LPI, 1980. 112 p.

3. *Bezginov A.N., Tregubov S.Yu.* Review of existing methods of scheduling // Information technology and programming: mezhvuz. Sat st. Issue 2 (14). M.: MGIU, 2005. 60 p.

4. *Kabalnov Yu.S., Shekhtman L.I., Nizamova G.F., Zemchenkova N.A.* Composite genetic algorithm for scheduling studies // Bulletin of Ufa State Aviation Technical University. 2006. V. 7, No. 2. P. 99–107.

5. *Grankov M.V., Al-Gabri V.M., Gorlova M.Yu.* Analysis and clustering of the main factors affecting the academic performance of university groups // Engineering Bulletin of the Don, 2016, No 4, URL: ivdon.ru/ru/magazine/archive/n4y2016/3775.

6. *Saaty T.* *Decision Making.* Hierarchy analysis method. M.: Radio and cCommunication, 1993. 278 p.

7. *Semenov S.P., Tatarintsev Ya.B.* Comparative analysis of approaches to automatization of scheduling studies in educational institutions [Electronic resource] Access mode: <http://cyberleninka.ru/article/n/sravnitelnyy-analiz-podhodov-kavtomatizatsii-sostavleniya-raspisaniy-uchebnyh-zanyatiy-v-obrazovatelnyh-uchrezhdeniyah.pdf>

8. *Tanayev S.V., Shkurba V.V.* Introduction to scheduling theory. M.: Science, 1975. 257 p.

СВЕДЕНИЯ ОБ АВТОРЕ



Хабипов Ришат Ильшатович – аспирант Высшей школы информационных технологий и интеллектуальных систем Казанского федерального университета (КФУ). Ведущий программист отдела разработок информационных систем КФУ.

Rishat Ilshatovich Khabipov – graduate student of Information Technologies and Intelligent Systems of the Kazan Federal University (KFU). Leading Programmer of the Information Systems Development Department of KFU.

email: paladin86@bk.ru

Материал поступил в редакцию 15 июня 2018 года