

Н. И. Юсупова, А. В. Маркелова, О. Н. Сметанина

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ДЛЯ СОПОСТАВИТЕЛЬНОГО АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ НА ОСНОВЕ РЕГУЛЯРНЫХ ГРАММАТИК

Приведены результаты исследования инструментальных средств, используемых для сравнения образовательных программ подготовки информатиков университетами различных стран на основе регулярных грамматик. *Болонский процесс; ECTS; Федеральные государственные образовательные стандарты третьего поколения; образовательные программы, информатика; система поддержки принятия решений; семантический анализ текста; регулярные грамматики*

ВВЕДЕНИЕ

Создание общеевропейского пространства высшего образования, одной из целей которого является повышение мобильности граждан на рынке труда и усиление конкурентоспособности европейского высшего образования, позволило российским университетам выйти на международный рынок образовательных услуг. Для признания зарубежными университетами образовательных программ российских учебных заведений и для привлечения зарубежных студентов на обучение в Россию необходимо разрабатывать конкурентоспособные образовательные программы и правильно представлять их на рынке.

Цели Болонского процесса [5] – формирование единого рынка труда высшей квалификации в Европе, расширение доступа к европейскому образованию, расширение мобильности студентов и преподавателей, принятие сопоставимой системы ступеней высшего образования с выдачей узнаваемых во всех странах Европы приложений к дипломам.

Федеральные государственные образовательные стандарты третьего поколения ориентированы на компетентностный подход, использование системы зачетных единиц (ECTS-european credit transfer system), выдачу приложений к дипломам (Diploma Supplement), многоуровневую подготовку. Высшие учебные заведения в России при переходе на стандарты нового поколения получают достаточную свободу для определения объемов и содержания подготовки выпускников.

Новые образовательные программы подготовки информатиков следует разрабатывать

с учетом опыта ведущих зарубежных университетов в данной области.

Одной из проблем, которую призвана решить система кредитов в высшем образовании, является структуризация учебных планов вузов различных государств для обеспечения их совместимости [6]. Вопросы сопоставительного анализа учебных планов рассмотрены в работах многих специалистов как в России, так и за рубежом. Этой тематике посвятили свои труды ученые Уфимского государственного авиационного технического университета [7, 8].

В данной статье рассматривается подход с использованием регулярных грамматик при семантическом анализе текстовой информации, а именно, к обработке учебных планов подготовки информатиков университетами разных стран.

1. ЗАДАЧА СОПОСТАВИТЕЛЬНОГО АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ

Образовательная программа включает в себя комплект нормативных документов: учебный план, программы учебных дисциплин и производственных практик, а также требования к итоговой аттестации. Учебный план – основной документ, определяющий структуру учебного процесса на факультете, перечень и объемы учебных дисциплин, последовательность их изучения, названия и продолжительность практик, используемые виды занятий (лекции, лабораторные и практические занятия, семинары и др.), аттестации, формы контроля и т. д.

При сопоставительном анализе учебных планов университетов различных стран необходимо проводить сравнение дисциплин, наименования которых представлены на разных языках, и которые, к тому же, часто имеют одинаковый смысл, называются по-разному [7]. Название дисциплины состоит из набора слов и цифр, а нагрузка – из цифр. Также сопоставляются ключевые слова (при их наличии) из наимено-

ваний или содержания дисциплин, выделяются предварительно ключевые слова (при их отсутствии). Таким образом, задача сопоставительного анализа учебных планов сводится к семантическому анализу текстовой информации. Обработка такой информации для последующего создания базы знаний представляет довольно трудоемкий процесс.

Одним из эффективных методов решения выявленной проблемы сравнения неоднородных учебных планов является применение регулярных грамматик. Этот механизм позволяет задать шаблон для строки и осуществить поиск данных, соответствующих этому шаблону в заданном тексте. Дополнительные функции позволяют получить найденные данные в виде массива строк, произвести замену в тексте по шаблону, разбиение строки по шаблону и др. [1, 2].

2. РЕГУЛЯРНЫЕ ГРАММАТИКИ

Регулярные выражения представляют современную систему поиска текстовых фрагментов в электронных документах, основанную на специальной системе записи образцов для поиска. Образец (англ. pattern), задающий правило поиска, иначе называют «шаблоном», «маской».

На сегодняшний день регулярные выражения поддерживаются такими языками программирования, как php (preg_match), perl (expression perl regular), javascript (javascript regexp), java, delphi, c++, c и .NET. Регулярные выражения используются для сжатого описания некоторого множества строк с помощью шаблонов, без перечисления всех элементов этого множества. При составлении шаблонов используется специальный синтаксис (табл. 1).

Таблица 1

Примеры спецсимволов	
Символ	Обозначение
^	Начало строки
\b	Начало слова
\xNN	NN – шестнадцатеричный код ASCII-символа (\x20 – пробел, \x4A – J, \x6A – j и т. д.)
[^ ...]	Ни один из перечисленного набора символов. Внутри квадратных скобок не работают другие операторы, но можно пользоваться метасимволами. С помощью дефиса можно указывать наборы символов: от первого до последнего. Например, [^0-9] означает любой символы, кроме 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Операторы (табл. 2) не могут применяться сами по себе, без указания символа, на который они действуют. Оператор действует на опреде-

ленный перед ним символ (мета или обычный). Если какое-то выражение заключено в скобки, после которых стоит оператор, то он действует на всю скобку.

Таблица 2

Примеры операторов	
Оператор	Значение
(...)	Сгруппировать символы в один паттерн и запомнить
	Предыдущий или следующий паттерн (логическое "ИЛИ")
?	0 или 1 раз предыдущая маска
{n}	Повторять n раз

Таким образом, регулярные выражения можно применять для [5]:

а) поиска подстроки, удовлетворяющей шаблону регулярного выражения, в строке;

б) поиска и замены подстроки, удовлетворяющей шаблону регулярного выражения, в строке;

в) проверки на соответствие заданной строки шаблону;

г) извлечения подстроки, удовлетворяющей шаблону регулярного выражения, из строки.

Примеры регулярных выражений, применяемых при анализе учебных планов:

«(\w\s\d|'|/|\(|\)|)+(\$\n|r|t|;|,|.)*» – для получения отдельных ключевых слов из набора ключевых слов дисциплины;

«[^r\n\a]» – для получения названий дисциплины без символов форматирования.

Использование регулярных выражений позволяет проводить гибкий анализ текста, что дает ощутимые преимущества для экспертной системы.

3. АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ СОПОСТАВИТЕЛЬНОГО АНАЛИЗА УЧЕБНЫХ ПЛАНОВ

В начале работы алгоритма (рис. 1) из таблицы (FCurriculum), содержащей данные об анализируемом учебном плане зарубежного университета, в переменную FName запоминается название очередной иностранной дисциплины. Затем система обращается к таблице, содержащей правила-продукции, в этой таблице находятся поля:

– название дисциплины анализируемого учебного плана (rules_disc.FName),

– название дисциплины базового плана (rules_disc.RName).

rules_disc.FName вместе со значением поля представляет собой условие, а rules_disc.RName со значением поля представляет собой вывод.

Например: если rules_disc.FName = Physique, то rules_disc.RName = Физика.

Это означает, что если название сравниваемой дисциплины «Physique» нашлось в базе знаний, среди правил продукции, то ей соответствует дисциплина российского учебного плана с названием «Физика». Далее, если название дисциплины было найдено среди условий, то ее название заносится в таблицу результатов (Results) в поле FName (полное имя Results.FName), а название соответствующего вывода из правила, то есть наименование дисциплины российского учебного плана, – в поле RName (Results.RName). Затем проверяется, все ли правила были проверены, если нет, то выбирается следующее правило продукции, и снова проверяется поле условия. Это делается потому, что среди соответствующих дисциплин российского учебного плана для дисциплины учебного плана зарубежного вуза может встретиться не одна. Если при просмотре всех правил из базы знаний по наименованиям дисциплин не было найдено соответствующего условия, т. е. ни одно правило не сработало, то происходит обращение к базе знаний по ключевым словам (табл. rules_keywords). В таблице rules_keywords имеется поле FKeywords (rules_keywords.FKeywords), имеющее тип поля Memo. Если в поле есть данные, то выбирается первое ключевое слово и запоминается в переменной FKeyword. Затем начинается поиск правила, в условии которого находится поле с таким же значением. Если ключевое слово было найдено среди условий, то название FName помещается в поле Results.FName, а в поле Results.RName помещается значение вывода из правила в таблице rules_keywords.RName. Если совпадений не было, то дисциплина с названием FName помещается в таблицу несовпавших дисциплин (Lost).

На этапе сопоставления дисциплин формируются две таблицы: таблица схожих дисциплин (Results) и таблица дисциплин, для которых не было найдено соответствия (Lost). На втором этапе сначала производится проверка, существуют ли данные в таблице не совпавших дисциплин (Lost). Если таблица пуста, то данный этап завершается, если же имеются данные, то эксперт имеет возможность провести сопоставление в интерактивном режиме, т. е. соотносить не совпавшие дисциплины к соответствующим по наименованиям дисциплин блокам дисциплин или областям знаний вручную.

Если эксперт производит сопоставление по ключевому слову, то соотношенное наименование дисциплины базового учебного плана записывается в таблицу Results, в поле RName. В поле FName записывается наименование дисциплины учебного плана зарубежного вуза, ко-

торой принадлежит данное ключевое слово. Также производится пополнение базы знаний для ключевых слов (rules_keywords). В нее заносится наименование дисциплины базового учебного плана в поле RName и ключевое слово в поле FKeyword. Если же производится соотношение к наименованию дисциплины, то в поле Results.RName помещается наименование базовой дисциплины RName, а в Results.FName – FName. И этот факт заносится в базу знаний по наименованиям дисциплин (rules_disc).

Если производится сопоставление к блоку дисциплин (области знаний), то в таблицу Results в поле Block(Area) записывается наименование блока дисциплин (области знаний).

Сформированная на предыдущих этапах таблица Results позволяет определить общую учебную нагрузку в кредитах для плана зарубежного вуза и расчет процента от нагрузки всего учебного плана.

Для расчета количества вхождений дисциплины в таблицу Results проводится просмотр всей таблицы по столбцу Results.FName для дисциплин зарубежного учебного плана и по столбцу Results.RName для дисциплин российского учебного плана. Из таблицы FCurriculum для каждой различной дисциплины Results.FName выбирается значение ее учебной нагрузки и затем распределяется равномерно в столбце Results.FLoad, в зависимости от количества вхождений дисциплины в Results. Для дисциплины базового учебного плана (российского учебного плана), или блока дисциплин, значение учебной нагрузки выбирается из базы данных по направлению 230100 «Информатика и вычислительная техника», распределяется точно также в столбце Results.RLoad. Также учебная нагрузка FLoad на этом этапе при занесении в таблицу Results пересчитывается в часы и заносится в Results.FLoadConverted.

Затем производится расчет суммарной учебной нагрузки по блокам учебных дисциплин относительно учебного плана зарубежного университета. Вычисляется процент от общей учебной нагрузки, отводящийся под блоки гуманитарных и социально-экономических дисциплин (ГиСЭД), математических и естественно-научных дисциплин (МиЕНД), общепрофессиональных дисциплин (ОПД) и специальных дисциплин (СД). Затем вычисляется процент от общей нагрузки, приходящейся на несопоставленные дисциплины из таблицы Lost.

Разработанное алгоритмическое обеспечение является основой для разработки основных программных модулей.

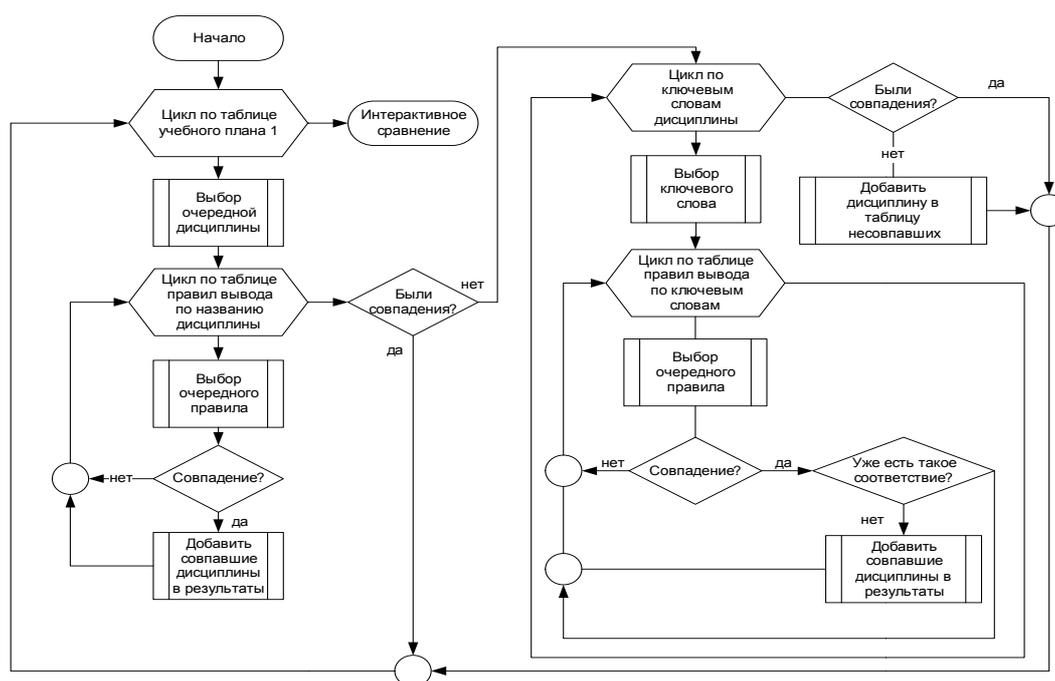


Рис. 1. Схема алгоритма сравнения

4. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ СОПОСТАВИТЕЛЬНОГО АНАЛИЗА УЧЕБНЫХ ПЛАНОВ

В качестве среды для разработки была выбрана MS Visual Studio, язык программирования MS Visual C#. Данный выбор обуславливается тем, что позволяет разрабатывать программный продукт, соответствующий современным стандартам. Также технология .NET обеспечивает гибкое и стабильное взаимодействие с базой данных (БД).

В диаграмме классов (рис. 2) выделяются следующие классы пользовательского интерфейса программы:

- MainForm: основное окно программы, содержащее меню, вкладки с элементами DataGridView для отображения и редактирования БД учебных планов, а также другие элементы;
- FormAdd: окно для добавления новых записей в БД;
- FormEdit: окно для редактирования существующих записей БД;
- FormCompare: окно для проведения экспертного анализа;
- FormDbDGView: является универсальным средством отображения и редактирования БД любой структуры. Используется для работы с внутренними БД.

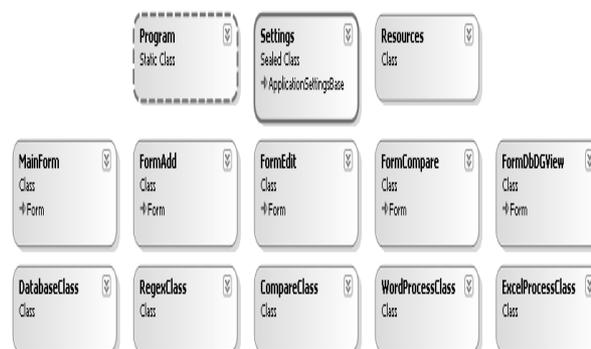


Рис. 2. Диаграмма классов

В главном окне программы сравнения учебных планов загружаются существующие в базе данных учебные планы либо добавляются новые. В этом же окне добавлена опция редактирования всех полей учебного плана.

С помощью окна для интерактивного сопоставления (рис. 3) пользователь обучает экспертную систему, относя дисциплины на иностранном языке к областям знаний на русском, таким образом, пополняя базу знаний.

Также на этапе проведения интерактивного сопоставления дисциплин, эксперт обрабатывает список тех дисциплин иностранного учебного плана, для которых автоматизированная информационная система не смогла найти соответствия.

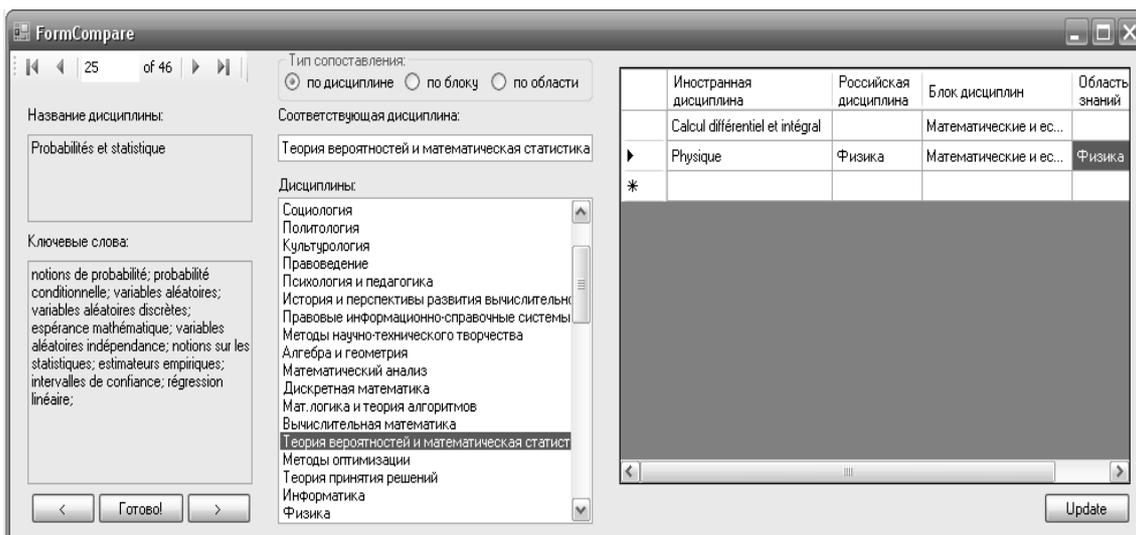


Рис. 3. Окно для интерактивного сравнения

5. РЕЗУЛЬТАТЫ СРАВНЕНИЯ УЧЕБНЫХ ПЛАНОВ ПРИ ИСПОЛЬЗОВАНИИ РЕГУЛЯРНЫХ ВЫРАЖЕНИЙ НА ПРИМЕРЕ РОССИЙСКИХ, ФРАНЦУЗСКИХ И ШВЕЙЦАРСКИХ УНИВЕРСИТЕТОВ

Результаты проведенного сопоставления учебных планов подготовки информатиков в российском (Уфимский государственный авиационный технический университет, УГАТУ) [4] и французском (Университет Марн-ля-Валле) [3] вузах по блокам дисциплин (табл. 3, рис. 4) и областям знаний (табл. 4, рис. 5) показали, что во французском университете, по сравнению с российским, на изучение гуманитарных и социально-экономических дисциплин отводится в 5 раз меньше времени. На изучение математических и естественно-научных дисциплин выделено столько же времени. На освоение общепрофессиональных дисциплин французы тратят больше времени, чем студенты из России. На изучение блока специальных дисциплин в российском вузе отводится почти вдвое меньше времени, чем во французском.

Сравнение по блокам дисциплин показывает, что во французском университете на изучение блока математических и естественно-научных дисциплин выделено немного больше времени чем в УГАТУ. В 5 раз больше времени отдано на изучение гуманитарных предметов сравнению с подготовкой в УГАТУ.

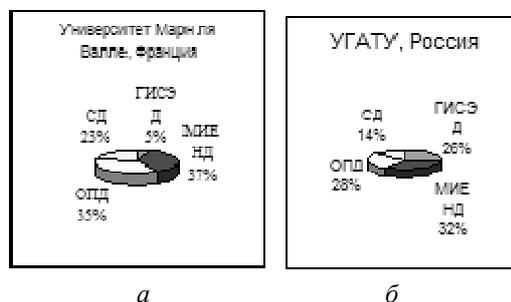


Рис. 4. Соотношение блоков дисциплин учебных планов подготовки информатиков в Университете Марн-ля-Валле, Франция (а) и УГАТУ, Россия (б)

Таблица 3

Фрагмент результатов сопоставления учебных планов подготовки информатиков в УГАТУ и Университете Марн-ля-Валле по блокам дисциплин

Блок	Университет Марн-ля-Валле, Франция (проц./кр.)	УГАТУ, Россия (проц./кр.)
ГиСЭД	5%/9	26%/ 50
МиЕНД	37%/66	32%/61

Таблица 4

Фрагмент результатов сопоставления учебных планов подготовки информатиков в УГАТУ и Университете Марн-ля-Валле по областям знаний

Области знаний	Университет Марн-ля-Валле, Франция (проц./кр.)	УГАТУ, Россия (проц./кр.)
Математика	32%/57	20%/39
Физика / химия / экология	5%/9	7%/14
Работа с графикой	2%/3	3%/7
Компьютерные сети	2%/3	3%/7

Сравнение по областям знаний показывает, что во французском университете математику изучают больше на 12%, но время, уделяемое физике, работе с графикой и компьютерным сетям, практически равно.



Рис. 5. Соотношение блоков дисциплин учебных планов подготовки информатиков в Университете Марн-ля-Валле, Франция (а) и УГАТУ, Россия (б)

Результаты проведенного сопоставления учебных планов подготовки информатиков в российском (УГАТУ) и швейцарском (Высшая политехническая школа Лозанны) [8] вузах по блокам дисциплин (табл. 5, рис. 6) показали следующее: в швейцарской высшей школе, по сравнению с УГАТУ, на изучение гуманитарных и социально-экономических дисциплин отводится минимум времени, на изучение математических и естественно-научных выделено примерно столько же времени.

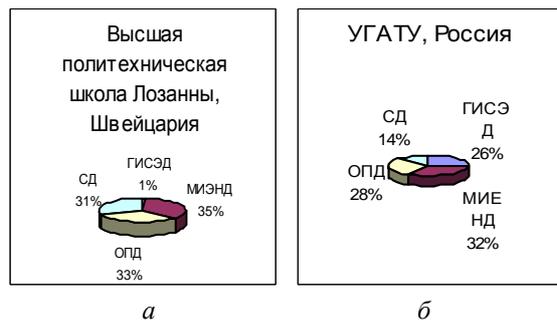


Рис. 6. Соотношение блоков дисциплин учебных планов подготовки информатиков в Высшей политехнической школе Лозанны, Швейцария (а) и УГАТУ, Россия (б)

Сопоставление по областям знаний (табл. 6, рис. 7) показывает, что в швейцарской высшей школе на изучение математики выделяется на 9% больше времени, чем в УГАТУ. Физика, работа с графикой и компьютерные сети изучаются примерно в таком же объеме.

Таблица 5
Фрагмент результатов сопоставления учебных планов подготовки информатиков в УГАТУ и Высшей Политехнической школе Лозанны по блокам дисциплин

Блок	Высшая политехническая школа Лозанны, Швейцария (проц./кр.)	УГАТУ, Россия (проц./кр.)
ГиСЭД	1%/2	26%/ 50
МиЭНД	35%/64	32%/61

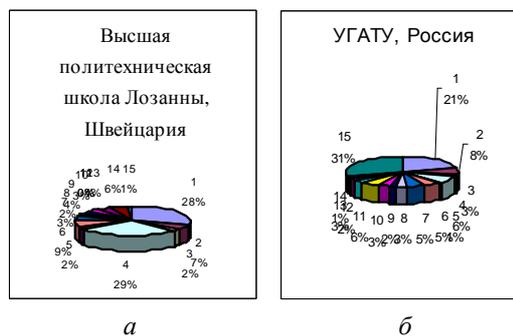


Рис. 7. Соотношение областей знаний учебных планов подготовки информатиков в высшей политехнической школе Лозанны, Швейцария (а) и УГАТУ, Россия (б)

Таблица 6
Фрагмент результатов сопоставления учебных планов подготовки информатиков в УГАТУ и Высшей Политехнической школе Лозанны по областям знаний

Области знаний	Высшая политехническая школа Лозанны, Швейцария (проц./кр.)	УГАТУ, Россия (проц./кр.)
Математика	29%/52	20%/39
Физика/химия/экология	7%/12	7%/14
Работа с графикой	2%/3	3%/7
Компьютерные сети	2%/4	3%/7

ВЫВОДЫ

В статье рассмотрена задача сопоставительного анализа образовательных программ различных стран. Выявлено, что сравнение учебных планов, как одного из основных документов образовательной программы, является задачей семантического анализа текстовой информации. Предложен подход к решению выявленной проблемы на основе применения механизма регулярных грамматик. Результаты исследований показали, что это позволяет эффективно и гибко анализировать текстовую информацию. Приведены наглядные примеры результатов сравнения учебных планов подготовки информатиков университетами разных стран.

Авторы продолжают накапливать статистику сопоставления зарубежных и российских образовательных программ подготовки информатиков. В дальнейшем предполагается проведение анализа совместного использования регулярных грамматик и онтологий. Таким образом, будет осуществлен эффективный поиск по шаблонам, а уже найденное соответствие будет анализироваться с помощью онтологий.

СПИСОК ЛИТЕРАТУРЫ

1. **Фридл Дж.** Регулярные выражения: учеб. пособие. СПб.: Питер, 2001. 352 с.
2. **Смит Б.** Методы и алгоритмы вычислений на строках (regex) = Computing Patterns in Strings : учеб. пособие. М.: Вильямс, 2006. 496 с.
3. **Куликов Г. Г., Набатов А. Н., Речкалов А. В.** Автоматизированное проектирование информационно-управляющих систем. Системное моделирование предметной области: учеб. пособие. Уфа, 2003. 104 с.
4. Российская Академия наук. 2010. Концепция создания Единой информационной системы РАН (ЕИС РАН). Вторая редакция. <http://www.ras.ru> (дата обращения: 17.04.2010).
5. The official Bologna Process website 2007-2010. 2010. <http://www.ond.vlaanderen.be> (дата обращения: 06.03.2010)

6. **Бадарч Д., Наранцега Я., Сазонов Б. А.** Организация индивидуально-ориентированного учебного процесса в системе зачетных единиц: учеб. пособие. М.: НИИВО, 2003. 76 с.

7. Университетские образовательные программы. Модели и методы для сопоставительного анализа./ М.Б. Гузаиров [и др.] // Изд-во МАИ, 2006. 117 с.

8. Сопоставительный анализ учебных планов по информатике университетов разных стран / М. Б. Гузаиров [и др.] // Мавлютовские чтения: сб. трудов. 2006. Т. 1, № 5. С. 77–83.

ОБ АВТОРАХ



Юсупова Нафиса Исламовна, проф., зав. каф. вычисл. мат. и киб., декан ФИРТ. Дипл. радиопизик (Воронежск. гос. ун-т, 1975).

Д-р техн. наук по упр-ю в техн. сист. (УГАТУ, 1998). Иссл. в обл. критич. сит. упр-я, информатики.



Маркелова Анастасия Вадимовна, асп. той же каф. Дипл. бакалавр техники и технологии (УГАТУ, 2006). Магистр техники и технологии (УГАТУ, 2008).



Сметанина Ольга Николаевна, доц. той же каф. Дипл. спец. по автоматизации процессов обработки и выдачи инф-ции (УАИ, 1985). Канд. техн. наук по АСУ (УГАТУ, 1999). Иссл. в обл. системы поддержки принятия решений в соц. и экон. системах.