

УДК 004.832.3

МОДУЛЬ ПОДГОТОВКИ КОЛЛЕКЦИИ РАБОЧИХ ПРОГРАММ ДЛЯ ЛАТЕНТНО-СЕМАНТИЧЕСКОГО АНАЛИЗА

Дышенов Б.А., Найханова Л.В., Ширапов Д.Ш.

*ФГБОУ ВПО «Восточно-Сибирский государственный университет технологий и управления»,
Улан-Удэ, e-mail: obeka_nlv@mail.ru, dyshenov@gmail.com*

В данной статье рассматривается одна из проблем установления соответствия между компетенциями образовательного стандарта и дисциплинами учебного плана – формирование коллекции рабочих программ дисциплин. Коллекция рабочих программ используется при выполнении латентно-семантического анализа. Процесс формирования коллекции рабочих программ является достаточно трудоемким, поэтому в статье рассматривается задача автоматизации процесса поиска и формирования коллекции из сети Интернет. Основная проблема поиска информации по компетенции в интернете заключается в том, что компетенции представляют собой текст на естественном языке, состоящий из нескольких сложных терминов. Поэтому исходный текст компетенции анализируется с применением лингвистических методов и затем строится иерархическая структура терминов. Из полученных терминов формируются поисковые запросы. Результатами поисковых запросов являются ссылки на документы, которые содержат рабочие программы дисциплин. В последующем из полученных ссылок формируем искомую коллекцию рабочих программ. Эксперименты проведены для общепрофессиональных компетенций федерального государственного образовательного стандарта (ФГОС3+) направления подготовки 02.04.03 «Математическое обеспечение и администрирование информационных систем».

Ключевые слова: образовательный стандарт, компетенция, учебный план, дисциплина, иерархия терминов компетенции, поисковый запрос, коллекция рабочих программ

MODULE OF PREPARING OF THE COLLECTION OF WORKING PROGRAMS FOR THE LATENT AND SEMANTIC ANALYSIS

Dyshenov B.A., Naykhanova L.V., Shirapov D.Sh.

*Federal State Educational Institution of Higher Professional Education East-Siberian State University
of Technology and Management, Ulan-Ude, e-mail: obeka_nlv@mail.ru, dyshenov@gmail.com*

In this article one of problems of establishment of compliance between competences of the educational standard and disciplines of the curriculum – formation of a collection of working programs of disciplines is considered. The collection of working programs is used at implementation of the latent and semantic analysis. Process of formation of a collection of working programs is rather labor-consuming. In article the problem of automation of process of search and formation of a collection of the Internet. The main problem of information search of competence of the Internet is that competences represent the text in a natural language consisting of several difficult terms. Therefore the source text of competence is analyzed with application of linguistic methods and then the hierarchical structure of terms is under construction. Search queries are formed of the received terms. Results of search queries are references to documents which contain working programs of disciplines. In the subsequent of the received references we form a required collection of working programs. Experiments were carried out for the preparation of competencies direction 02.03.03 «Software and administration of information systems».

Keywords: educational standard, competence, curriculum, discipline, hierarchy of terms of competence, search query, collection of working programs

В образовательном процессе после выхода новой версии образовательного стандарта требуется разработка нового учебного плана в короткие сроки. Трудоемкость процесса разработки учебного плана, короткие сроки, а также и многие другие факторы обуславливают необходимость автоматизации данного процесса. Одним из простых аспектов разработки учебного плана является формирование перечня дисциплин учебного плана, соответствующих компетенциям образовательного стандарта.

Для решения задачи «Установление соответствия между компетенциями и дисциплинами» используется метод латентно-семантического анализа [1, 5] и нейро-нечеткие сети [3]. Метод латентно-семантического анализа применяется для получения грубых оценок соответствия

дисциплин и компетенций. А аппарат нейро-нечетких сетей [3, 4] позволяет уточнить эти оценки. Латентно-семантический анализ – это метод обработки информации на естественном языке, анализирующий взаимосвязь между коллекцией документов и терминами в них встречающимися, сопоставляющий некоторые факторы всем документам и терминам [2]. Таким образом, для применения данного метода требуется сформировать коллекции документов.

Формирование коллекции документов

Для выполнения латентно-семантического анализа формируются две коллекции документов. Первая коллекция строится на основе анализа текста компетенции. Для каждой компетенции строится ее иерархическая структура, пример показан на рис. 1. Ли-

сточку дерева (иерархии) содержат термины (например, «основные методы испытаний программного обеспечения»), по которым можно осуществлять запросы в Wikipedia. Трассировка статей Wikipedia позволяет выделить нужную информацию об объекте поиска (термине, содержащемся в листе дерева). Коллекция статей с найденной информацией по всем терминам листочков составляет содержание компетенции. Так, например, рассмотрим общепрофессиональную компетенцию ОПК-5 «Владение основными методами и средствами автоматизации проектирования, производства, испытаний и оценки качества программного обеспечения» по направлению подготовки 02.04.03 «Математическое обеспечение и администрирование информационных систем». В ней можно выделить две составляющие:

1) основные методы автоматизации проектирования, производства, испытаний и оценки качества программного обеспечения;

2) основные средства автоматизации проектирования, производства, испытаний и оценки качества программного обеспечения.

Каждая из приведенных составляющих разбивается на субкомпоненты. Например, первая имеет следующие:

а) методы автоматизации проектирования программного обеспечения;

б) методы автоматизации производства программного обеспечения;

в) методы автоматизации испытаний программного обеспечения;

г) методы автоматизации оценки качества программного обеспечения.

Компоненты $a-g$ являются листочками дерева иерархии, на основе которых и строится запрос, в первую очередь к энциклопедии Википедия. На основе анализа найденных ресурсов строится вторая коллекция, содержащая тексты, описывающие компетенции.

Вторая коллекция состоит из набора рабочих программ по различным дисциплинам. По многим направлениям подготовки в интернете имеются искомые документы. Надо отметить, что в коллекцию рабочих программ по рассматриваемому направлению подготовки будем включать рабочие программы дисциплин и по смежным направлениям подготовки.

Формирование коллекции рабочих программ

Для составления запросов на поиск рабочих программ используются элементы второго уровня иерархической структуры компетенции. Запрос начинается с термина «Рабочая программа» + элемент второго уровня иерархии. Например, запрос: «Рабочая программа» + «Методы автоматизации проектирования программного обеспечения». Результаты данного запроса показаны в табл. 1.



Рис. 1. Пример иерархической структуры компетенции

Таблица 1

Результаты поискового запроса

№ п/п	Наименование рабочей программы	Направление подготовки
1	РАБОЧАЯ ПРОГРАММА учебной дисциплины (модуля) – CASE-средства проектирования программного обеспечения	010400.68 – Прикладная математика и информатика
2	РАБОЧАЯ ПРОГРАММА модуля – Технология разработки программного обеспечения	230100 Информатика и вычислительная техника
3	РАБОЧАЯ ПРОГРАММА дисциплины – Технология разработки программного обеспечения	230700 Прикладная информатика
4	РАБОЧАЯ ПРОГРАММА дисциплины – Технология разработки программного обеспечения	010500.62 «Математическое обеспечение и администрирование информационных систем»
5	РАБОЧАЯ ПРОГРАММА дисциплины – Разработка и стандартизация программных средств и информационных технологий	080801 «Прикладная информатика (по областям)»
6	РАБОЧАЯ ПРОГРАММА по дисциплине – Проектирование Информационных систем	230201 – Информационные системы и технологии
7	РАБОЧАЯ ПРОГРАММА дисциплины Технологии и инструментальные средства автоматизации разработки программного продукта	09.06.01 Информатика и вычислительная техника
8	УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС дисциплины «Автоматизированное проектирование средств и систем управления»	220400 «Программное обеспечение вычислительной техники и автоматизированных систем»
9	Рабочая программа по дисциплине «Автоматизация проектирования систем и средств управления»	230100 Информатика и вычислительная техника
10

Таблица 2

Результаты поискового запроса

№ п/п	Наименование рабочей программы	Направление подготовки
1	УМКД дисциплины «Архитектура вычислительных систем и компьютерных сетей»	02.03.03 – Математическое обеспечение и администрирование информационных систем
2	РАБОЧАЯ ПРОГРАММА учебной дисциплины «Архитектура ЭВМ и вычислительных систем»	010503 – Математическое обеспечение и администрирование информационных систем
3	РАБОЧАЯ ПРОГРАММА дисциплины «ЭВМ и периферийные устройства, ЭВМ и ПУ»	02.03.03 – Математическое обеспечение и администрирование информационных систем
4	РАБОЧАЯ ПРОГРАММА дисциплины «Операционные системы и оболочки»	010500.62 «Математическое обеспечение и администрирование информационных систем»
5	РАБОЧАЯ ПРОГРАММА дисциплины «Администрирование информационных систем»	010500.62 – Математическое обеспечение и администрирование информационных систем
6	РАБОЧАЯ ПРОГРАММА дисциплины «Архитектура вычислительных систем и компьютерных сетей»	010500.62 – Математическое обеспечение и администрирование информационных систем
7	УМКД дисциплины «Администрирование в информационных системах»	230201.65 «Информационные системы и технологии»
8	УЧЕБНАЯ ПРОГРАММА дисциплины «Администрирование информационных систем»	010500.62 – Математическое обеспечение и администрирование информационных систем
9	РАБОЧАЯ ПРОГРАММА дисциплины «Вычислительные машины, системы и сети»	220400 – Программное обеспечение вычислительной техники и автоматизированных систем
10	РАБОЧАЯ ПРОГРАММА модуля «Организация ЭВМ дисциплины ЭВМ и периферийные устройства»	230100 Информатика и вычислительная техника
11	РАБОЧАЯ ПРОГРАММА учебной дисциплины «Архитектура ЭВМ И ВС»	010501.65 Прикладная математика и информатика
12

Анализ Формирование коллекции РБ

Модуль подготовки коллекции рабочих программ

Входные данные

Источник	Google.com
Категория ресурса	Рабочая программа
Способ нормализации	Лемматизация
Входной xml файл компетенции	C:\Resource\pc11.xml
Тип файлов	doc

Выходные данные

Имя выходного файла	result.xml
Директория результирующей коллекции документов	C:\Resource

Поиск

Рис. 2. Основное окно пользовательского интерфейса модуля

Для компетенции ОПК-11 «Владение навыками выбора архитектуры и комплексирования современных компьютеров, систем, комплексов и сетей системного администрирования» составлен запрос: «Рабочая программа» + «Выбор архитектуры и комплексирования современных компьютеров, систем, комплексов и сетей системного администрирования». На этот запрос получены результаты, показанные в табл. 2, строки с 1-ой по 8-ую.

Документ в 9-ой строке получен на запрос с выполненным стеммингом словоформ: «Рабоч программ» + «Метод автоматизац проектирован программн обеспечен».

Данный пример показывает, что для полноты комплектования коллекции рабочих программ следует рассматривать все виды запросов по всем элементам иерархии. В качестве левого элемента шаблона запроса можно использовать три элемента: рабочая программа, учебная программа и УМКД.

Компонентная структура модуля. Основными компонентами модуля являются: интерфейс пользователя, анализ и формализация данных, формирование запросов к поисковым системам, формирование коллекции рабочих программ.

Пользовательский интерфейс. Данный компонент представляет собой совокупность средств и методов, при помощи которых пользователь взаимодействует

с модулем подготовки коллекции рабочих программ.

Анализ и формализация данных. На вход программы подается массив текста компетенции. Наша задача заключается в том, чтобы выделить именных субстантивных словосочетаний и затем построить иерархическую структуру компетенции. Надо отметить, что на верхних уровнях иерархии располагаются термины, представляющие собой сложные составные именные субстантивные словосочетания. Декомпозиция терминов осуществляется посредством морфологических шаблонов.

Формирование запросов к поисковым системам. Из полученного набора терминов формируются поисковые запросы. Каждая поисковая система предоставляет свой API (набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением/библиотекой/сервисом) для формирования автоматических запросов. Исходя из этого, для каждой поисковой системы формируются свои индивидуальные методы извлечения информации.

Формирование коллекции рабочих программ. Из найденных ссылок, полученных в результате выполнения поисковых запросов, формируем коллекцию документов. При формировании данной коллекции в качестве критерия отбора принято включать документы, в названии которых имеются ключевые слова: *рабочая программа, учебная программа, УМКД.*

Для иллюстрации работы программы рассмотрим пользовательский интерфейс, представленный на рис. 2. В показанном окне ввода входной информации предполагаются следующие действия:

1. Выбирается источник запроса – это основные поисковые системы, такие как *google.com*, *yandex.ru*. У каждого поискового сервера есть свои особенности работы. Результирующие данные разных поисковых серверов могут сильно отличаться.

2. Выбирается категория ресурсов – это левая часть запроса: *рабочая программа*, *учебная программа*, *УМКД*.

3. Выбирается способ нормализации словоформ – это лемматизация или стемминг. Лемматизация приводит словоформу к инфинитиву. Стемминг оставляет основу слова.

4. Выбирается входной xml-файл компетенции.

5. Выбирается результирующий тип документов *doc* или *pdf*.

6. Вводится имя выходного xml-файла, в котором будут содержаться данные по найденным документам – это ссылка на источник, имя файла, путь к директории сохранения.

Пример выходного файла xml:

```
<?xml version="1.0" encoding="windows-1251"?>
<frames>
  <file name = "ПП ТРПО 3 МОАИС.pdf"
  link="https://esstu.ru/uportal/test/subject.htm?code=0052147&dirCode
  =010500" directory="C:\Resource"></file>
  ...
</frames>
```

7. Выбирается директория сохранения.

Основным критерием отбора документов в коллекцию служит наличие указанных пользователем ключевых слов в названии документа: *рабочая программа*, *УМКД*, *учебная программа*.

Заключение

В работе исследована возможность автоматизации процесса построения коллекции рабочих программ дисциплин. Исследования показали возможность построения достаточно полной коллекции. Конечно, качество рабочих программ различно, но это не самый важный фактор установления соответствия между компетенциями образовательного стандарта и дисциплинами учебного плана.

Список литературы

1. Единов С.В. Хабрахабр [Электронный ресурс] // Разработка. Латентно-семантический анализ: сайт. – URL: <https://habrahabr.ru/post/110078> (дата обращения 08.08.2016).
2. Интернет-энциклопедия: Википедия. Латентно-семантический анализ. [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/Латентно-семантический_анализ (дата обращения 5.10.2016).
3. Найханова Л.В. Применение нейро-нечеткой системы вывода для установления дисциплин, соответствующих

компетенции образовательного стандарта / Л.В. Найханова, Б.А. Дышенов Б.А. // Фундаментальные исследования. – 2016. – № 5. – С. 47–52.

4. Седова Н.А. Система поддержки принятия решений для определения безопасной скорости судна / Н.А. Седова, В.А. Седов // Информационные технологии. Радиоэлектроника. Телекоммуникации. – Москва, 2016. – № 6–2. – С. 204–210.

5. Deerwester S. Indexing by Latent Semantic Analysis // Journal of the American Society for Information Science. – 1990. – 41 (6): 391–407.

References

1. Edinov S.V. Habrahabr [Elektronnyj resurs] // Razrabotka. Latentno-semanticheskij analiz: sajt. URL: <https://habrahabr.ru/post/110078> (data obrashhenija 08.08.2016).
2. Internet-jenciklopedija: Vikipedija. Latentno-semanticheskij analiz. [Elektronnyj resurs]. Rezhim dostupa: https://ru.wikipedia.org/wiki/Latentno-semanticheskij_analiz (data obrashhenija 5.10.2016).
3. Najhanova L.V. Primenenie nejro-nechetkoj sistemy vyvoda dlja ustanovlenija disciplin, sootvetstvujushih kompetencii obrazovatel'nogo standarta / L.V. Najhanova, B.A. Dyshenov B.A. // Fundamentalnye issledovanija. 2016. no. 5. pp. 47–52.
4. Sedova N.A. Sistema podderzhki prinjatija reshenij dlja opredelenija bezopasnoj skorosti sudna / N.A. Sedova, V.A. Sedov // Informacionnye tehnologii. Radioelektronika. Telekomunikacii. Moskva, 2016. no. 6–2. pp. 204–210.
5. Deerwester S. Indexing by Latent Semantic Analysis // Journal of the American Society for Information Science. 1990. 41 (6): 391–407.