

I. N. Zyabrev, K. N. Vasilenko, O. V. Pozharkov, I. N. Pozharkova

## THE COMPARATIVE ANALYSIS OF SPECTRAL LANGUAGE MODEL AND OTHER INFORMATION RETRIEVAL MODELS

*Comparison of spectral language model (SLM) with models most widely used in information retrieval, such as DFR (Difference From Randomness) and BM25 by quality estimation of the search problems decision.*

*Keywords: information retrieval, probability model, SLM.*

© Зябрев И. Н., Василенко К. Н., Пожарков О. В., Пожаркова И. Н., 2011

УДК 658.512.001.56

М. В. Карасева, Д. В. Кустов

## ИНФОРМАЦИОННО-ТЕРМИНОЛОГИЧЕСКИЙ БАЗИС В МУЛЬТИЛИНГВИСТИЧЕСКОЙ АДАПТИВНО-ОБУЧАЮЩЕЙ ТЕХНОЛОГИИ\*

*Рассмотрено функциональное назначение и структура программной системы TuMLas v.1,0, реализующей программно-алгоритмическое обеспечение мультилингвистической адаптивно-обучающей технологии. Представлена реляционная модель структуры информационно-терминологического базиса, а также структура первичного информационно-терминологического базиса в виде концептуальной ER-диаграммы.*

*Ключевые слова: мультилингвистическая технология, программная система, ER-диаграмма.*

Одним из ключевых моментов в обучении иностранному языку является изучение иностранной лексики. В этом случае направленная методика обучения подразумевает создание профессионально ориентированных словарей. Если словарь исполнен в классическом виде, то термины в нем располагаются в алфавитной последовательности. Это упрощает их поиск, но никоим образом не повышает эффективность обучения при использовании такого словаря.

С этой целью разработана мультилингвистическая адаптивно-обучающая технология [1], которая рассматривает свойства лексем применительно к конкретным предметным областям, что позволяет для каждой области сформировать информационный терминологический базис (ИТБ) [2]. На основе данных о его элементах становится возможным принять решение о той или иной структуре профессионально ориентированного словаря или же самого ИТБ.

Вопрос об эффективной организации ИТБ нашел решение в ряде методов оптимизации [3], но ни один из этих методов до сих пор не учитывал ассоциативные связи между лексемами (понятиями) того или иного языка.

Такие методы были разработаны и реализованы в программной системе TuMLas v.1,0 [4].

Функциональное назначение программной системы заключается в осуществлении программно-алгоритмической поддержки мультилингвистической адаптивно-обучающей технологии. TuMLas v.1,0 является высококачественным инструментом построения мультилингвистического ИТБ на основе новейших методов оптимизации его структуры [5].

Система работает в двух взаимосвязанных режимах, каждый из которых соответствует ее отдельной функции и обеспечивается отдельным функциональным блоком. Это режим анализа текстов и последующей генерации первичного информационно-терминологического базиса; режим формирования целевого (пригодного для использования в процессе обучения) информационно-терминологического базиса путем реорганизации структуры первичного базиса, в том числе с помощью методов оптимизации его структуры.

Структура разработанной программной системы представлена на рис. 1.

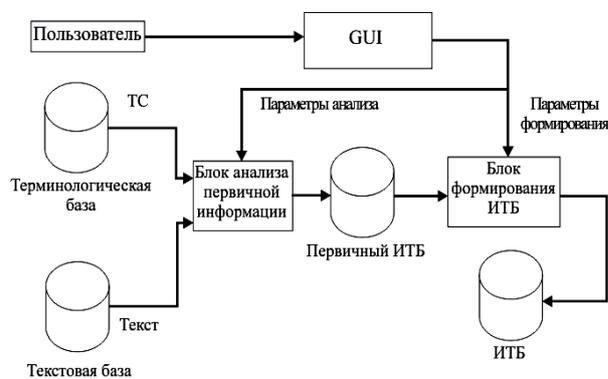


Рис. 1. Структура программной системы TuMLas v.1,0

\* Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 гг.» и ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» 2011-1.9-519-005.

Блок анализа первичной информации (БАПИ) и блок формирования ИТБ составляют основу системы и обеспечивают ее работу.

Кратко рассмотрим каждый из них.

Согласно рис. 1, на вход БАПИ подается текст интересующей нас предметной области и терминологический словарь (ТС) этой же области.

Управляющее воздействие осуществляется пользователем, который через графический пользовательский интерфейс (GUI) задает соответствующие параметры. Такими параметрами могут быть следующие [4]:

- язык текста;
- режим обработки лексических конструкций (по умолчанию зависит от языка текста и служит для нахождения семантического соответствия);
- режим генерации базиса (замена/дополнение) и т. п.

Выходом БАПИ является первичный ИТБ. Его структура в виде концептуальной ER-диаграммы изображена на рис. 2.



Рис. 2. Концептуальная ER-диаграмма структуры первичного информационно-терминологического базиса

Таким образом, для каждого языка формируется набор лексем некоторой предметной области, учитывая их частотные [2] и ассоциативные характеристики [6]. Каждой полученной лексеме ставится в соответствие набор ассоциативно связанных с ней лексем – лексем реляционного ряда [5], также принадлежащих данной предметной области.

Полученный ИТБ не является непосредственно объектом изучения при обучении иностранной лексике, но он содержит достаточную информацию, для того чтобы такой базис построить.

Построение целевого ИТБ осуществляется блоком формирования ИТБ (см. рис. 1).

На вход блока подается сгенерированный БАПИ первичный ИТБ. Управляющее воздействие, аналогично рассмотренному в БАПИ, осуществляется пользователем через GUI (задаются параметры формирования ИТБ).

Такими параметрами могут быть следующие:

- языковой состав целевого ИТБ;
- методы формирования структуры базиса и т. д.

Блок формирования ИТБ содержит различные вариации методов:

- частотного построения ИТБ [2];
- оптимального разбиения ИТБ на блоки и модули [3];
- реализации реляционной модели структуры ИТБ [5].

Применение этих методов как отдельно друг от друга, так и в различных сочетаниях позволяет значительно снизить трудоемкость прохождения целевого базиса и повысить эффективность использующей его системы обучения в целом [5].

В отличие от подобных, ранее разработанных систем, TuMLas v.1,0 позволяет при формировании ИТБ учитывать ассоциативные связи между элементами базиса. Структура ИТБ в данном случае формируется согласно реляционной модели (рис. 3) [5].

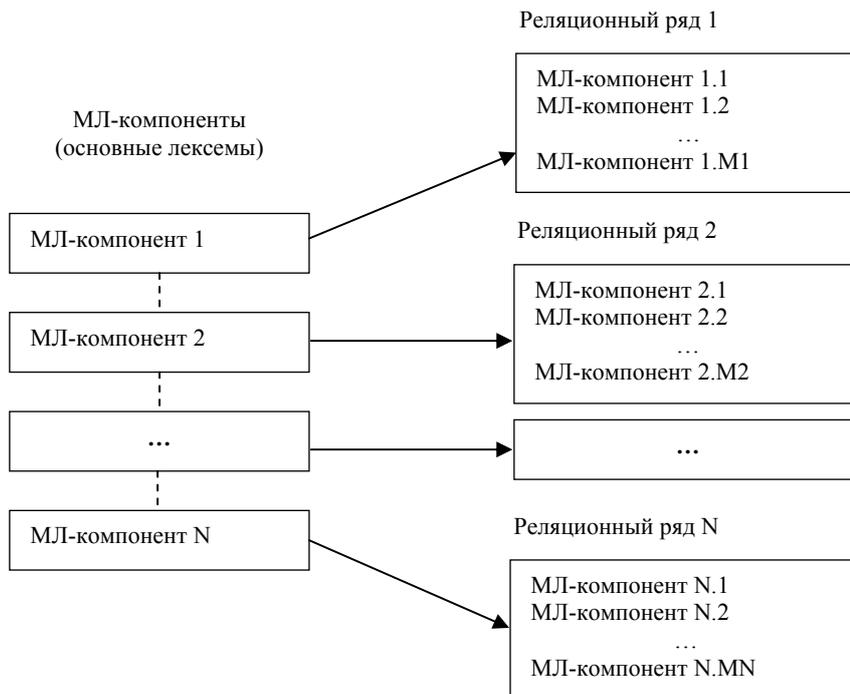


Рис. 3. Реляционная модель структуры ИТБ

По своим частотным характеристикам все лексемы ИТБ разделяются на две группы: 1) лексемы, имеющие наибольшую частоту (основные лексемы); 2) лексемы, частота которых не превышает некоторого числового порога. Из них впоследствии формируются реляционные ряды, каждый из которых однозначно соответствует одной из основных лексем.

При этом структура МЛ-компонента [7], отражающего основную лексему, дополняется соответствующими переходными вероятностями.

Построение ИТБ согласно реляционной модели позволяет значительно снизить трудоемкость его прохождения и увеличить эффективность системы обучения в целом. При этом применение реляционной модели хорошо сочетается с другими методами оптимизации структуры ИТБ, в частности с методами оптимального разбиения ИТБ на блоки и модули [5].

#### Библиографические ссылки

1. Ковалев И. В., Усачев А. В. Мультилингвистический метод изучения иностранной терминологической лексики на базе мнемотехнического подхода // Первая Всерос. науч. интернет-конф. «Соц.-психол. пробл. развития личности». Вып. 4. Тамбов : ТГУ, 2001. С. 57.

2. Карасева М. В., Ковалев И. В., Суздалева Е. А. Модель архитектуры мультилингвистической адаптивно-обучающей технологии // Новые информ. тех-

нологии в унив. образовании : тез. междунар. науч.-метод. конф. Кемерово : КемГУ и ИДМИ, 2002. С. 204–205.

3. Огнерубов С. С. Алгоритмы оптимизации блочно-модульной структуры информационно-терминологического базиса // Вестн. унив. комплекса. Вып. 20. 2006. С. 104–118.

4. Программа анализа и формирования информационного мультилингвистического терминологического базиса, на основе реляционной модели оптимизации, TuMLas v.1,0 : программа для ЭВМ : свидетельство об офиц. регистрации № 50200701261 / В. О. Лесков, И. В. Ковалев. Зарегистр. ВНИИЦ. 2010.

5. Лесков В. О., Огнерубов С. С. Реляционная модель и алгоритмы оптимизации модульной структуры мультилингвистического информационно-терминологического базиса // Вестн. унив. комплекса. Вып. 21. 2009. С. 116–133.

6. Лесков В. О. Система предварительной обработки текстов, основанная на скрытой марковской модели, в контексте мультилингвистической адаптивно-обучающей технологии // Вестн. унив. комплекса. Вып. 23. 2006. С. 110–119.

7. Ковалев И. В., Карасева М. В., Суздалева Е. А. Системные аспекты организации и применения мультилингвистической адаптивно-обучающей технологии // Образоват. технологии и общество – Educational Technology & Society. Вып. 5(2). 2002. С. 198–212.

M. V. Karaseva, D. V. Kustov

#### INFORMATIONAL AND TERMINOLOGICAL BASIS IN MULTILINGUAL ADAPTIVE-TRAINING TECHNOLOGY

*The paper considers functional assignment and structure of the algorithmic software of the multilingual adaptive-training technology. It is realized as a program system TuMLas v. 1,0. The relation model of the informational and terminological basis structure and the initial informational and terminological basis structure are presented as a conceptual ER-diagram.*

*Keywords: multilingual technology, software system, ER-diagram.*

© Карасева М. В., Кустов Д. В., 2011

УДК 517.552

А. С. Кацунова

#### О ПОВТОРНОМ ОСОБОМ ИНТЕГРАЛЕ КОШИ–СЕГЕ

*Получен аналог формулы Пуанкаре–Бертрана для особого интеграла Коши–Сеге в шаре. Главное значение интеграла рассмотрено в смысле Керзмана–Стейна.*

*Ключевые слова: интеграл Коши–Сеге, главное значение особого интеграла в смысле Керзмана–Стейна, формула перестановки повторного интеграла.*

Будем рассматривать  $n$ -мерное комплексное пространство  $\mathbf{C}^n$  ( $n > 1$ ) переменных  $z = (z_1, \dots, z_n)$ . Если  $z, w \in \mathbf{C}^n$ , то  $\langle z, w \rangle = z_1 w_1 + \dots + z_n w_n$ , а  $|z| = \sqrt{\langle \bar{z}, z \rangle}$ , где  $\bar{z} = (\bar{z}_1, \dots, \bar{z}_n)$ .

Пусть  $B_z(\varepsilon)$  – шар в  $\mathbf{C}^n$  с центром в точке  $z$  радиуса  $\varepsilon$ , т. е.

$$B_z(\varepsilon) = \{ \zeta \in \mathbf{C}^n : |\zeta - z| < \varepsilon \}.$$