

**ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЭМПИРИЧЕСКОЙ ПОДДЕРЖКИ  
МУЛЬТИЛИНГВИСТИЧЕСКОЙ ТЕХНОЛОГИИ В КОРПОРАТИВНОЙ  
ИНФОРМАЦИОННОЙ СИСТЕМЕ\***

*Рассмотрены пути реализации эмпирической поддержки мультилингвистической адаптивно-обучающей технологии, используемой в распределенных информационных системах, на примере работы комплекса программного моделирования КПМ v.1.0.*

*Ключевые слова: информационная система, мультилингвистическая технология, адаптивное обучение.*

В настоящее время распространенным способом упорядочения информационных потоков в рамках предприятий и организаций является создание и поддержание распределенной информационной системы. Такая система позволяет решать множество задач, начиная с отслеживания и электронного документирования финансовых потоков и заканчивая общением сотрудников данного предприятия. В перспективе корпоративная информационная система может стать более сложной структурой, так называемой ERP-системой. Такие системы несут в себе, наряду с информационными, функции планирования и управления предприятием.

При построении распределенных информационных систем разработчики неизбежно сталкиваются с целым рядом проблем. Учитывая тенденции к интеграции и созданию мультинациональных корпораций, одной из таких проблем является проблема языкового взаимодействия, когда специалисты разных национальностей (принадлежащие к разным языковым группам) должны в кратчайшие сроки прийти к пониманию, т. е. обучиться специализированной предметной лексике на незнакомом им языке. Другой проблемой является эффективное взаимодействие специалистов непосредственно с компьютерной системой, которое для людей, далеких от компьютерных и информационных технологий, не является тривиальной задачей [1].

В данной статье рассматривается реализация эмпирической поддержки мультилингвистической адаптивно-обучающей технологии, используемой в распределенных информационных системах, на примере работы комплекса программного моделирования КПМ v.1.0.

Итак, одним из новых путей решения проблемы интенсивного накопления иностранного профессионально ориентированного словарного запаса является применение мультилингвистической адаптивно-обучающей технологии (МЛ-технологии), информационным базисом которой являются частотные словари [2].

На сегодняшний момент МЛ-технология есть совокупность алгоритмов и моделей данных, направленных на эффективное, строго организованное изучение иностранной лексики [3]. Особенность МЛ-технологии состоит в искусственном построении

ассоциативных связей между языковыми аналогами некоторого множества языков непосредственно в процессе обучения. Это в некоторой степени снижает возможность возникновения «смешения языков» и ускоряет процесс запоминания.

Количество алгоритмов и моделей данных, которые включает в себя МЛ-технология, постоянно увеличивается. Возникает необходимость в сравнительном анализе алгоритмов, но поскольку все они ориентированы на данные, а эти данные зачастую должны быть многочисленны и разнообразны, проводить такой анализ становится затруднительным. Тем не менее необходимость в эмпирической проверке алгоритмов существует и с развитием МЛ-технологии только усиливается.

Проводить различного рода тесты и эксперименты в этой области сложно еще и потому, что данные, необходимые для этого, есть не что иное, как множество распределенных информационно-терминологических базисов (ИТБ) [3]. На разработку таких ИТБ уходит значительное количество ресурсов, включая работу экспертов-лингвистов и специалистов различных предметных областей. Решить эту проблему в некоторой степени возможно с помощью программного моделирования, при котором реальные ИТБ заменяются их адекватными моделями.

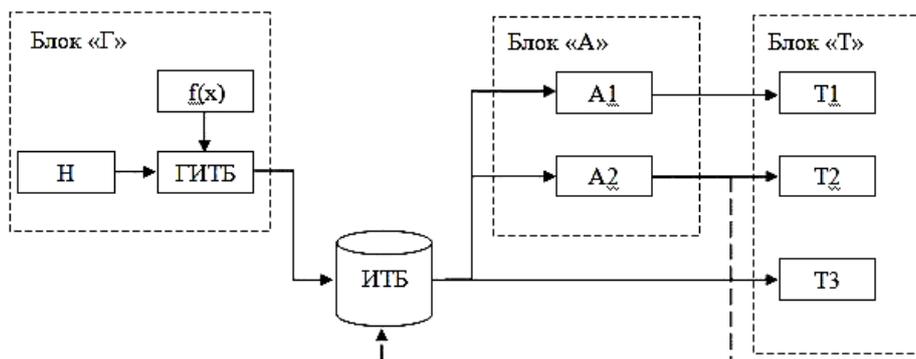
Структура таких моделей должна учитывать все возможные свойства ИТБ, включая абсолютную и относительную частоты, скрытые лексические связи и т. п.

После того как разработана и принята структура моделей, строится эмпирическая база для анализа алгоритмов, состоящая из множеств моделей заданной структуры с различными априорными характеристиками, такими как объем базиса, максимальные значения частот и количество лексических связей.

После того как эмпирическая база сформирована, на ней в режиме тестирования реализуются интересующие исследователя алгоритмы, затем анализируются результаты экспериментов.

Генерацию множества моделей и тестирование алгоритмов разумно осуществлять в рамках одной программной системы. Примером такой системы может служить комплекс программного моделирования КПМ v.1.0 [4] (см. рисунок).

\*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.



Структура программного комплекса КПМ v.1.0

Основой КПМ v.1.0 является блок генерации моделей ИТБ, представленный на приведенной схеме как блок «Г». Он состоит из генератора (ГИТБ), на вход которого подается список параметров (Н), необходимых для формирования текущей модели. Механизмы ГИТБ основаны на генерации псевдослучайных чисел, причем сама генерация может осуществляться по различным стохастическим законам ( $f(x)$ ). На сегодняшний момент КПМ v.1.0 поддерживает генерацию псевдослучайных чисел по нормальному и равномерному законам распределения вероятностей.

Список параметров Н в КПМ v.1.0 реализован в виде класса и состоит из набора переменных, определяющих следующие настройки формирования модели ИТБ:

- объем базиса в терминах;
- максимальное количество связей, приходящихся на одну лексему;
- максимальное значение абсолютной частоты лексем;
- максимальное значение частоты сочетаний лексем;
- объем материала, по которому произведен частотный анализ;
- коэффициент связанности лексем.

Коэффициент связанности лексем определяет максимальную долю от абсолютной частоты лексемы, которую занимает частота ее использования в лексических сочетаниях.

Используя список параметров как набор входных данных, ГИТБ генерирует модель согласно выбранному закону распределения.

Список параметров формирования моделей может быть сохранен в файл и в любой момент восстановлен.

Результаты генерации отображаются в соответствующем окне для их визуального анализа и также могут быть сохранены в файл и восстановлены впоследствии.

Блок «А» представляет собой набор алгоритмов для эмпирического анализа, реализуемых в виде отдельных программных модулей и библиотек.

Все алгоритмы этого блока делятся на два основных типа:

- алгоритмы, использующие ИТБ в процессе обучения (А1);
- алгоритмы, изменяющие структуру самого ИТБ на основе анализа его характеристик (А2).

На выходе алгоритмов А2-типа формируется ИТБ с новой структурой, который так же может быть использован алгоритмами обоих типов и проанализирован.

Блок «Т» – блок тестирования ИТБ и алгоритмов. Тесты, входящие в этот блок, логически подразделяются на три типа:

- тесты для анализа А1-алгоритмов (Т1);
- тесты для анализа А2-алгоритмов (Т2);
- тесты для анализа ИТБ как результата А2-алгоритмов (Т3).

Под тестами здесь следует понимать специальные аналитические алгоритмы, которые также могут быть добавлены в КПМ v.1.0 как отдельные программные модули и библиотеки. Необходимо заметить, что структура Т-алгоритмов зависит от интересов исследователя и может представлять собой как тривиальный счетчик итераций, так и многоступенчатые, многоатрибутивные алгоритмы с вычислением экстремумов разнообразных функций качества.

В качестве примера работы КПМ v.1.0 проведем сравнительный анализ двух алгоритмов А2-типа: алгоритма 1, алгоритма 2.

Приведем необходимый минимум информации об исследуемых алгоритмах. Итак, алгоритм 1 и алгоритм 2 имеют общие задачи, а именно являются алгоритмами формирования лексически связанных компонентов (ЛС-компонентов) [5]. Формирование ЛС-компонентов осуществляется на основе данных о скрытых лексических связях ИТБ, относительной частоте лексем и лексических сочетаний. Это означает, что модели, на которых мы будем производить тестирование, должны отражать эти свойства базиса в полной мере.

Суть формирования ЛС-компонентов состоит в выделении «наиболее подходящих» «лексических центров» и их связей с другими лексемами. Соответственно, все лексемы ИТБ, согласно приведенным алгоритмам, делятся на три группы:

- основные лексемы (представляют собой «лексические центры»);
- связанные лексемы (отражают лексические сочетания основных лексем);
- свободные элементы (лексемы, не попавшие ни под одну из категорий).

В качестве «лексических центров» выступают лексемы, имеющие наибольшую абсолютную частоту и наиболее сильные лексические связи.

Оба алгоритма состоят из трех общих фаз.

1. Упорядочение ИТБ.

Производится согласно критерию

$$Q_i = q_i e^{-\frac{0,7}{\sum_k q_k \mu_{ik} + 1}} \rightarrow \max,$$

где  $q_i$  – относительная частота, выражающая долю лексической единицы в тексте, подвергшемся статистической обработке при составлении частотного словаря,  $0 < q_i < 1$ ;  $\mu_{ik}$  – относительная частота сочетания  $i$ -й и  $k$ -й лексем, отражающая силу ассоциативной связи.

2. Перебор числа основных лексем.

2.1. Поиск связанных лексем.

2.2. Вычисление функции качества ИТБ как совокупности ЛС-компонентов от числа основных лексем:

$$Q(n) = \sum_i q_i e^{-\frac{0,7}{\sum_k q_k \mu_{ik} + 1}} \rightarrow \max.$$

3. Определение оптимального количества основных лексем и структуры ЛС-компонентов путем максимизации функции качества  $Q(n)$ .

Исследуемые алгоритмы имеют различия в фазе 2.1: в случае алгоритма 1 поиск связанных лексем осуществляется непосредственно при переборе основных лексем с учетом данных об их лексических связях; в случае алгоритма 2 поиск осуществляется путем перебора связанных лексем  $j$  и нахождения наилучших сочетаний:

$$\Delta Q_i = q_i e^{-\frac{0,7}{(\sum_k q_k \mu_{ik} + 1) + q_j \mu_{ij}}} - Q_i \rightarrow \max.$$

Теперь, когда мы имеем общее представление об исследуемых алгоритмах, можно переходить к фазе тестирования.

Итак, основными оцениваемыми характеристиками будут максимум функции качества как показатель эффективности алгоритма и количество итераций как показатель его ресурсоемкости.

Для более полного анализа исследуемых алгоритмов введем дополнительные характеристики, такие как количество основных лексем, количество свободных элементов и время исполнения как независимый показатель ресурсоемкости (для различных алгоритмов одна итерация может содержать различное число простых операций).

Настраиваемые параметры базиса:

- объем базиса в терминах (1 000);
- максимальное количество связей, приходящихся на одну лексему (10);
- максимальное значение абсолютной частоты лексем (100/50 000);
- максимальное значение частоты сочетаний лексем (20/50 000);
- объем материала, по которому произведен частотный анализ (50 000);
- коэффициент связанности лексем (1).

Результаты тестирования приведены в таблице.

Анализ данных тестирования показывает, что алгоритм 2 превзошел по значению максимума  $Q(n)$  алгоритм 1 на 0,000 000 004 402 925, что составляет  $\approx 10^{-6}$  %. Показатель качества увеличился незначительно, в то время как количество итераций и время исполнения алгоритма увеличились на несколько порядков. Тем не менее если рассматривать время исполнения алгоритма 2 независимо от аналогичного показателя алгоритма 1, то оно не так уж и велико, особенно если речь идет об этапе формирования ИТБ. Но это заключение является преждевременным, поскольку был проведен всего один эксперимент на одной модели.

Проведение ряда подобных экспериментов на моделях ИТБ с различными параметрами даст достаточную эмпирическую базу для анализа исследуемых алгоритмов, чтобы результаты этого анализа можно было считать достоверными.

Все это в полной мере относится не только к приведенным в примере алгоритмам, но и ко всем алгоритмам и структурам данных, которые включает в себя МЛ-технология.

**Библиографические ссылки**

1. Ковалев И. В., Кустов Д. В., Слободин М. Ю. Интеллектуализация информационных технологий в корпоративных системах // Информ. ресурсы России. 2006. № 3 (91). С. 31–34.
2. Ковалев И. В., Карасева М. В. Англо-немецко-русский частотный словарь по системному анализу / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2004.
3. Карасева М. В., Лесков В. О. Система формирования информационно-терминологического базиса мультилингвистической адаптивно-обучающей технологии // Вестник СибГАУ. 2007. Вып. 4 (17). С. 31–35.
4. Лесков В. О. Комплекс программного моделирования КПИМ v.1.0 / ВНИИЦ. М., 2008. № 50200802242.

**Тестирование алгоритмов**

Название алгоритма	Алгоритм 1	Алгоритм 2
max $Q(n)$	0,496 118 565 143 325	0,496 118 569 546 25
Основных лексем	188	188
Свободных элементов	320	324
Итераций	4 981 096	2 000 498 500
Время исполнения, с	3,41	130,75

M. V. Karaseva, A. V. Datsenko

## PROGRAM REALIZATION OF EMPIRICAL SUPPORT OF MULTILINGUAL TECHNOLOGY IN A CORPORATIVE INTELLIGENT SYSTEM

*The article considers ways of realization of empirical support of multilingual adaptive-training technology, used in distributed intelligent systems, on the example of program modeling complex КИМ v.1.0.*

*Keywords: intelligent system, multilingual technology, adaptive training.*

© Карасева М. В., Даценко А. В., 2012

УДК 621.355

Т. А. Леонова, А. И. Дудник, А. Е. Михеев, И. В. Осипова, Г. Н. Чурилов

## РАЗРЯДНЫЕ ХАРАКТЕРИСТИКИ ЛИТИЙ-ИОННЫХ АККУМУЛЯТОРОВ С УГЛЕРОДНЫМИ ЭЛЕКТРОДАМИ\*

*Исследованы литий-ионные аккумуляторы, в которых в качестве отрицательного электрода использовались фуллерит-литиевые композитные пленки, пластины гексагонального графита и пластины турбостратного графита (ТГ). Положительными электродами во всех литий-ионных аккумуляторах служили пластины металлического лития. При использовании в качестве анода турбостратного графита наблюдалось максимальное значение разрядного тока, в 1, 5 раза превышающее максимальный ток для анода из гексагонального графита. По результатам рентгеновской фотоэлектронной спектроскопии установлено, что толщина активного слоя (слоя анода интеркалированного ионами лития) для турбостратного графита составила 600 нм. Максимальное значение удельной емкости наблюдалось также при использовании турбостратного графита в качестве анода и достигало 244 мА·ч/г.*

*Ключевые слова: литий-ионные аккумуляторы, фуллерит, турбостратный графит, гексагональный графит, удельная емкость.*

Углерод является перспективным и популярным на сегодняшний день материалом для литиевых химических источников тока (ЛХИТ). Как известно, ЛХИТ широко используются в электронной технике благодаря высоким удельным характеристикам: емкости (180...400 мА·ч/г), плотности тока (0,1...100 мА/см<sup>2</sup>, в зависимости от материала) и доступности в цене, по сравнению с другими химическими источниками тока (ХИТ). Однако во всем мире продолжает рассматриваться маловероятная перспектива применения чистых фуллеренов в ЛИА, поскольку материалы на их основе не обладают ионной проводимостью. Известно, что элементарная ячейка ГЦК-решетки фуллерита содержит 8 тетраэдрических (2,2 Å) и 4 октаэдрических пустот (4,2 Å), поэтому эти пустоты способны вместить ионы лития (радиус иона 0,68 Å) при интеркаляции. По теоретическим расчетам отношение числа ионов лития к числу атомов углерода в фуллерите составляет 1/3, а в графите – 1/6, поэтому использование фуллерита в качестве анода может существенно увеличить емкость ЛХИТ [1].

Решение задачи по увеличению удельных характеристик ЛХИТ осложняется тем, что большинство материалов, имеющих высокие показатели по удельной емкости до 3 000 мА·ч/г, такие как кремний [2], на-

нотрубки и нановолокна на основе Si, Ge, TiSi<sub>2</sub>, ZnO, TiO<sub>2</sub>, SiC [3; 4] и графен [5; 6], при интеркаляции-деинтеркаляции ионов лития увеличивает свой объем это увеличение может составлять 400 %) и постепенно разрушается. Неоспоримым преимуществом углеродных материалов является изменение объема всего на 12 % при внедрении ионов лития. В графите расстояние между атомными плоскостями 3,35 Å, что позволяет проникать туда ионам лития. Как известно, в турбостратной структуре графита расстояние между графитовыми слоями больше и составляет 3,42 Å [7]. Поэтому интеркаляция ионов лития в ТГ будет приводить к меньшим необратимым изменениям структуры, чем в гексагональном графите, а значит, и количество циклов «заряд-разряд» аккумулятора увеличится. В увеличенную межплоскостную область ТГ может проникнуть больше ионов лития, что означает увеличение удельной емкости ЛИА с электродом из ТГ. В связи с этим использование турбостратного графита в качестве анода может повысить значение удельной емкости и разрядного тока аккумулятора, снизить деградацию емкостных параметров в процессе интеркаляции-деинтеркаляции ионов лития. Исследование анодов для ЛХИТ на основе углеродных материалов является актуальной задачей на сегодняшний день.

\*Исследование выполнено при поддержке Министерства образования и науки Российской Федерации, соглашения 14.В37.21.0163 и 8194, РФФИ в рамках научного проекта № 12-03-31439 проекта фундаментальных исследований НАН Беларуси и СО РАН № 24.