

В свою чергу адаптація функціоналу існуючих програмних продуктів під вимоги будівельної специфіки та унікальні потреби певної будівельної організації дає змогу розширити застосування інформаційних систем управління проектами та підвищити їх ефективність.

Управління проектами будівництва готелів «під ключ» за допомогою спеціально розроблених для окремих компаній інформаційних систем на всіх рівнях консолідації інформації зменшує складність та невизначеність процесів, що впливають на досягнення цілей проекту та бізнесу. Окремо при розробці та впровадженні спеціальних можливостей інформаційних систем слід приділяти значну увагу процесу валідації досягнутих результатів.

Література

1. Оберемок, И. И. Гибкий подход к внедрению корпоративной системы управления проектами [Текст] / И. И. Оберемок // Управление развитием складных систем. — 2014. — № 17. — С. 42–45.
2. Стутко, Н. Эффективность использования информационной системы управления проектами (ИСУП) [Электронный ресурс] / Н. Стутко. — Режим доступа: \www/URL: http://www.iteam.ru/publications/project/section_41/article_2709/
3. Чаинский, В. Эффективность компании и построение корпоративной системы [Электронный ресурс] / В. Чаинский // Журнал «Корпоративные системы». — Режим доступа: \www/URL: http://www.iteam.ru/publications/it/section_53/article_3686/
4. Мазур, И. И. Инвестиционно-строительный инжиниринг [Текст] / И. И. Мазур, В. Д. Шапиро, А. В. Гинзбург и др. — М.: Елима, 2007. — 1216 с.
5. Морозов, В. В. Розробка моделей визначення, формування та управління конфігурацією портфелів інвестиційних проектів будівельних корпорацій [Текст] / В. В. Морозов, К. М. Осетрин // Управління проектами та розвиток виробництва. — 2005. — № 4(16). — С. 38–45.
6. Морозов, В. В. Функціонально-ролевий підхід до опису життєвого циклу проектів проектно-орієнтованих корпорацій [Текст] / В. В. Морозов, О. В. Кальніченко // Управління розвитком складних систем. — 2011. — № 5. — С. 23–29.
7. Jurgen, A. Management 3.0: Leading Agile Developers, Developing Agile Leaders [Text] / A. Jurgen. — Boston: Addison-Wisley, 2010. — 451 p.

8. Cohn, M. Succeeding with Agile: Software Development Using Scrum [Text] / M. Cohn. — Addison-Wesley Professional, 2009. — 438 p.
9. Stack, J. The Great Game of Business. Oxford Oxfordshire [Text] / J. Stack. — Oxford University Press, 1994. — 373 p.
10. Дэвид, М. Стороны ESI Консалтинговых услуг «Гибкие методы или РМВОК и Вы можете использовать и то и другое!» [Текст]: презентация / М. Дэвид. — 2009. — 28 с.

РАЗРАБОТКА И ВНЕДРЕНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ПРОЕКТАМИ В СТРОИТЕЛЬНЫХ КОМПАНИЯХ

Рассмотрены ключевые вопросы, которые возникают перед генподрядными организациями на этапах подготовки и реализации проектов и их решение через прикладное использование специальных программных решений на базе современных информационных технологий, которые являются необъемлемой частью корпоративной системы управления проектами. Приведены результаты реализации дополнительных программных модулей в инвестиционно-строительных проектах на примере строительства отелей.

Ключевые слова: информационные технологии, проект, управление проектами, информационная система управления проектами.

Кальніченко Олена Володимирівна, кандидат технічних наук, доцент, кафедра бізнес-адміністрування та управління проектами, Університет економіки та права «КРОК», Київ, Україна, e-mail: kv_vl@ukr.net.

Чернова Марина Леонідівна, кандидат технічних наук, комерційний директор, ТОВ «Компанія по управлінню будівельними проектами «ГЕРЦ», Київ, Україна, e-mail: chml@gerc.kiev.ua.

Кальніченко Елена Владимировна, кандидат технических наук, доцент, кафедра бизнес-администрирования и управления проектами, Университет экономики и права «КРОК», Киев, Украина.

Чернова Марина Леонидовна, кандидат технических наук, коммерческий директор, ООО «Компания по управлению строительными проектами «ГЕРЦ», Киев, Украина.

Kalnichenko Elena, University of Economics and Law «CROC», Kyiv, Ukraine, e-mail: kv_vl@ukr.net.

Chernova Marina, LLC «Construction project management company «GERC», Kyiv, Ukraine, e-mail: chml@gerc.kiev.ua

УДК 004.912

DOI: 10.15587/2312-8372.2015.40895

**Кунгурцев А. Б.,
Поточняк Я. В.,
Силяев Д. А.**

МЕТОД АВТОМАТИЗИРОВАННОГО ПОСТРОЕНИЯ ТОЛКОВОГО СЛОВАРЯ ПРЕДМЕТНОЙ ОБЛАСТИ

Разработан метод автоматизированного построения толкового словаря, основанный на обработке множества текстов из заданной предметной области. Описана технология подбора и группировки исходных текстов, выделения терминов, поиска синонимов и значений терминов. Описана технология кластеризации документа для возможности подбора документов и создания общего словаря.

Ключевые слова: словарь, термин, предметная область, синоним, именные группы.

1. Введение

Под толковым словарем предметной области будем понимать специализированный словарь, дающий опреде-

ление множеству понятий, связанных с деятельностью некоторой организационной структуры. Существует множество задач, связанных с деятельностью любой организации, для решения которых существование такого

словаря было бы весьма желательно. Это задачи, связанные с созданием и развитием информационных систем [1], подготовкой кадров, созданием новой документации, четким распределением обязанностей между сотрудниками.

Предлагается под записью в словаре понимать одно слово или устойчивое для данной предметной области словосочетание (в дальнейшем будем использовать понятие «термин»), для которого приведено толкование, специфическое для данной предметной области, а также список синонимов.

2. Анализ литературных данных и постановка проблемы

Известны решения по автоматизации разработки словарей предметной области, ориентированные на англо-германскую группу языков — система jalingo [2]. Система облегчает работу эксперта в предметной области, однако требует участия специалиста в лингвистическом анализе. Кроме этого, система не автоматизирует поиск синонимов и не предлагает толкование терминов. В работах [3–5], ориентированных на группу славянских языков, отсутствует автоматизированная классификация текстов, определение синонимов и значений терминов.

Существует ряд вариантов выделения терминов из текста: выделение максимальных цепочек, содержащих термины [6], использование автоматических запросов к поисковой машине [7], синтаксический анализ (АОТ) [8, 9]. Анализ известных решений показал, что рассмотренные решения позволяют выделить общепотребляемые устойчивые словосочетания, которые удовлетворяют шаблонам или словосочетания, используемые в Интернете, однако проблема определения терминов в узкой конкретной предметной области остается нерешенной.

3. Объект, цель и задачи исследования

Объектом исследования является документация, используемая в любой организации, для которой создается либо модернизируется информационная система. Поскольку составление словаря предметной области «вручную» — чрезвычайно трудоемкий процесс, требующий участия квалифицированных специалистов, то целью данной работы является разработка метода автоматизации процессов составления словаря на всех этапах его создания.

Задачами исследования является автоматизация следующих видов работ, связанных с созданием словаря:

1. Подбор и группировка исходных текстов.
2. Выделение терминов.
3. Поиск синонимов для терминов.
4. Составление описания терминов.

Все перечисленные виды работ предусматривают привлечение эксперта в предметной области на определенных стадиях создания словаря для контроля и корректировки полученных в автоматизированном режиме решений.

4. Метод исследования текстов и автоматизированного построения толкового словаря предметной области

4.1. Подбор и группировка исходных текстов. Подготовка к составлению словаря предусматривает, прежде всего, определение границ предметной области, для

которой будет составляться словарь. Например, для некоторой организации проектируется автоматизированная информационная система (АИС). Планируется, что она охватит деятельность ряда подразделений и руководства организации. Тогда словарь АИС должен включать только те термины, которые используются в выделенных областях деятельности.

Деятельность любой организации отражается в ее документации. Это могут быть инструкции, отчеты, распоряжения и приказы, нормативные документы и т. д. Процесс выделения терминов из текстов основан на определении вероятности (частоты) их появления в документах. Если предметная область в свою очередь подразделяется на под-области, например, производственная деятельность и работа с кадрами, то вероятности использования терминов для этих подобластей должны быть определены раздельно.

Представим отдельный термин в виде кортежа:

$$M = \langle N, q \rangle, \quad (1)$$

где N — имя термина (слово или словосочетание), а q — количество появления термина в документе. В дальнейшем вместо q будем использовать частоту F появления термина в документе [10]:

$$M = \langle N, F \rangle. \quad (2)$$

Будем считать, что установлена некоторая нижняя граница значения частоты Fb , которая не позволяет включать термин в словарь. Тогда некоторый документ D может быть охарактеризован множеством терминов:

$$D = \{M\}. \quad (3)$$

Задача группировки документов сводится к задаче кластеризации.

Пусть D_1 и D_2 два документа, содержащие q_1 и q_2 слов соответственно.

Определим нормализованное расстояние d_{12} между документами. Для этого рассмотрим три ситуации:

- $D_1 = {}_N D_2$ — документы совпадают (знак $=_N$ означает, что все термины из документа D_1 входят в документ D_2 и наоборот. В этом случае будем считать расстояние $d_{12} = 0$).
- $D_1 \cap_N D_2 = \emptyset$ — документы без сомнения относятся к разным кластерам. Будем считать расстояние между кластерами $d_{12} = 1$.
- $R = D_1 \setminus_N D_2 \cup D_2 \setminus D_1 \neq \emptyset$, т. е. имеются термины, которые не являются общими для D_1 и D_2 и необходимо вычислить расстояние между документами для определения возможности отнесения их к одному кластеру.

Для вычисления расстояния следует определить количество терминов, которые не войдут в словарь в результате уменьшения их частоты ниже граничного значения Fb . Для этого строим два множества:

$$R_1 = \{M_i \mid N_i \in R \wedge N_i \in D_1\} i = 1, |R|, \quad (4)$$

$$R_2 = \{M_i \mid N_i \in R \wedge N_i \in D_2\} i = 1, |R|. \quad (5)$$

Пересчитываем частоты терминов с учетом объединения документов:

$$\forall M_i \in R_1 \text{ определяем } F'_i = \frac{F_i * q_1}{q_1 + q_2},$$

$$\forall M_i \in R_2 \text{ определяем } F'_i = \frac{F_i * q_2}{q_1 + q_2}.$$

Строим множества терминов, которые не войдут в объединенный документ:

$$R_1^1 = \{M_i | N_i \in R_1 \wedge F'_i \leq F_b\} | R_1 |,$$

$$R_2^1 = \{M_i | N_i \in R_2 \wedge F'_i \leq F_b\} | R_2 |.$$

Вычисляем расстояние между документами:

$$d_{12} = \frac{|R_1^1| + |R_2^1|}{|D_1| + |D_2|}. \quad (6)$$

Для практического применения (6) следует установить пороговое значение db для расстояния между документами.

Отдельный документ, подлежащий анализу, также может содержать «семантическую неоднородность», выраженную в явных или неявных разделах. Поэтому рекомендуется выделять явные разделы и параллельно с общей статистической обработкой производить статистическую обработку по разделам. Тогда можно предъявить эксперту термины значимые для определенного раздела (с высокой частотой появления именно в этом разделе), но редко встречающиеся во всем документе. Эксперт решает вопрос о включении этих терминов в общий словарь.

Для выявления неявно существующих разделов документа предлагается механизм «скользящей» кластеризации, основанный на последовательном анализе абзацев в анализируемом документе. Суть механизма изложена в следующем алгоритме:

1. Подсчитываются частоты терминов F_1 для абзаца. Результаты представляются как текущий раздел.

2. Подсчитываются частоты терминов F_2 для следующего абзаца.

3. Рассчитывается расстояние d_{12} между абзацем и текущим разделом. Если расстояние меньше порогового, то переход к пункту 4. В противном случае — к пункту 5.

4. Абзац включается в текущий раздел. Если имеется следующий абзац, то переход к пункту 2. В противном случае к пункту 6.

5. Если текущий раздел содержит несколько абзацев, то результаты его анализа сохраняются. Если имеется следующий абзац, то он представляется как текущий раздел. Переход к пункту 2. В противном случае переход к пункту 6.

6. Если текущий раздел содержит один абзац, то раздел ликвидируется. В противном случае он представляется как раздел.

4.2. Выделение терминов. Наиболее информативными словами в предложении являются существительные

и именные группы. Именно они составляют основу терминологической базы предметной области и должны быть включены в словарь. Термин — это слово, устойчивое словосочетание или сокращение, которое выражает и в известной мере классифицирует в данной предметной области определенное понятие или сущность, отражая в своей смысловой структуре характеристические признаки объекта терминования и взаимосвязи этого объекта с другими с достаточной для взаимного общения точностью [11].

Кандидаты в термины — это слова или словосочетания, которые извлекаются из обрабатываемого текста. Для получения кандидатов, текст проходит синтаксический анализ, с использованием Cognitive Dwarf [12]. Далее выполняется выбор существительных и словосочетаний (которые авторы данной работы будут называть именными группами) на основе синтаксического анализа. Именная группа — это существительное и связанная с ним другая часть речи (существительное и прилагательное или наречие). Отбираются все именные группы, не разделенные знаками препинания (кроме дефиса).

Рассмотрим пример: «Жесткий диск — это запоминающее устройство. Диск использует принцип магнитной записи».

В результате разбора синтаксический анализатор предлагает все возможные кандидаты слов и именных групп — табл. 1.

Таблица 1

Кандидаты слов и именных групп на основании разбора

| Слово (существительное — noun_es) | Именная группа (adj_es > noun_es) |
|-----------------------------------|-----------------------------------|
| диск | [жесткий] диск |
| устройство | [запоминающее] устройство |
| принцип | [магнитный] запись |
| запись | — |

Далее выполняется статистическая обработка результатов разбора, которая позволяет определить кандидатов для терминов словаря. Для этого рассчитывается частота вхождения термина-слова в документ:

$$F_{wi} = \frac{N_{wi}}{N_{ws}}, \quad (7)$$

где N_{wi} — количество вхождений слова в анализируемый текст, N_{ws} — общее количество слов в тексте.

Поскольку именных групп обычно значительно меньше чем терминов-слов, то предложено рассчитывать частоту вхождения именной группы в документ по следующей формуле:

$$F_{gi} = \frac{N_{gi}}{N_{gs}}, \quad (8)$$

где N_{gi} — количество вхождений именной группы в анализируемый текст; N_{gs} — общее количество обнаруженных именных групп в тексте.

Термины — слова и именные группы с низкой частотой вхождения не должны быть представлены в словаре. Для этого эксперт вводит пороговые значения частот

для терминов k_w и именных групп k_g . Предложена следующая последовательность действий эксперта.

1. Определение документа и ознакомление с его содержанием.

2. Запуск анализатора и получение списка терминов в соответствии с (2).

3. Установка значений k_w и k_g и анализ сокращенного списка. Если из списка исчезли значимые слова или именные группы, то соответствующие значения k_w и k_g нужно понизить. Если в списке сохраняются термины, которые не относятся к данной предметной области, то соответствующие значения k_w и k_g нужно повысить.

4.3. Определение синонимов. Для введения синонимов в словарь предметной области предложено использовать словарь синонимов SU соответствующего языка:

$$SY = \{N_i\} > i = 1, ns. \quad (9)$$

Представим запись такого словаря в виде множества:

$$M_j = \langle N_{s1}, \dots, N_{sj} \rangle > j = 1, ns, \quad (10)$$

где N_{sj} — один из терминов синонима.

Тогда представление термина с указанием синонимов приобретает следующий вид:

$$MS = \langle N_1, F, \{N_{sj} | N_{sj} \in M_j \ \& \ N_1 \in M_{sj}\} \rangle. \quad (11)$$

Условие, при котором термин $MS_1 = \langle N_1, F_1, \{N_{1i}\} \rangle$ можно предположить синонимом для термина $MS_2 = \langle N_2, F_2, \{N_{2j}\} \rangle$, формулируется следующим образом: $\{N_{1i}\} \cap \{N_{2j}\} \neq \emptyset$.

После формирования множества терминов с синонимами необходимо выявить термины словаря, которые оказались синонимами для других терминов. Назовем их «внутренние синонимы». Для повышения роли «внутренних синонимов» при использовании словаря они должны занимать первые позиции в списках синонимов, построенных на основании множеств $\{N_{ij}\}$.

Слова и словосочетания, определенные как синонимы для некоторого языка, не обязательно останутся таковыми в конкретной предметной области. Поэтому завершение данного этапа предусматривает проверку и корректировку экспертом результатов, полученных в процессе автоматического поиска синонимов.

4.4. Определение значений терминов. Для введения значений терминов в словарь предметной области предложено использовать толковый словарь соответствующего языка, либо некоторый специализированный словарь, перекрывающий анализируемую предметную область. Представим запись такого словаря в виде:

$$IN = \langle N, T \rangle, \quad (12)$$

где N — некоторое слово, а T — его значение (пояснительный текст).

Тогда представление термина с указанием синонимов приобретает следующий вид:

$$MSI = \langle N_1, F, \{N_i\}, T \rangle, \quad (13)$$

если $N = N_1$ либо $N \in \{N_i\}$.

Поэтому в общем случае запись в словаре имеет вид:

$$MSI = \langle N_1, F, \{N_i\}, T, T_a \rangle, \quad (14)$$

где T — толкование из толкового словаря.

Толкование слов, полученное из толкового словаря, не обязательно корректно в некоторой предметной области. Поэтому завершение данного этапа предусматривает проверку и корректировку, полученных в процессе подбора значения терминов.

5. Результаты экспериментов по автоматизированному построению словаря предметной области

Для построения словаря предметной области были проанализированы два документа D_1 и D_2 . Первый документ содержал параграф из книги Д. Э. Кнута [13] размером в 5603 слов; второй документ содержал параграф из книги К. Хорстманна [14] размером в 5564 слов.

В результате синтаксического и морфологического анализа документов D_1 и D_2 , выделяются все слова и именные группы, которые были обнаружены в документах. В табл. 2 представлен фрагмент результата разбора D_1 . Пороговые значения были определены как $k_w = 0,0006$ и $k_g = 0,00017$ опытным путем. Синонимы были определены из соответствующего словаря. Синонимы «перебор», «система», «налаженность» экспертом были удалены. «Внутренних синонимов» не было обнаружено. В результате «скользящей кластеризации» была установлена удовлетворительная однородность текста ($db = 0,22$).

Таблица 2

Разбор документа D_1

| Термин M | Число вхождений $N_{сл}$ | Частота появления F | Синоним SU |
|------------|--------------------------|-----------------------|-----------------------|
| сортировка | 46 | 0,0082 | упорядочение, перебор |
| порядок | 36 | 0,0064 | система, налаженность |
| инверсия | 35 | 0,0062 | — |
| ... | ... | ... | ... |

В табл. 3 представлено фрагмент результата разбора документа D_2 . Для этого документа пороговые значения также составили $k_w = 0,0006$ и $k_g = 0,00017$. Синонимы также были определены из словаря. Синоним «вода», «струя» экспертом был удален, а «алгоритм» был добавлен экспертом к «внутренним синонимам». В результате «скользящей кластеризации» была установлена удовлетворительная однородность текста ($db = 0,24$).

Произведена оценка возможности объединения выбранных исходных текстов D_1 и D_2 для создания общего

словаря. Для этого определено расстояние между документами (6). Его значение $db=0,8$ позволило эксперту сделать заключение о нецелесообразности создания единого словаря для двух документов.

Таблица 3

Разбор документа D_2

| Термин M | Число вхождений N | Частота появления F | Синоним SU |
|------------------------|---------------------|-----------------------|----------------------|
| поток | 315 | 0,0566 | вода, струя |
| метод | 135 | 0,0242 | алгоритм, технология |
| класс | 37 | 0,0066 | образ, урок |
| [операционный] система | 15 | 0,002 | — |
| ... | ... | ... | ... |

6. Обсуждение эффективности автоматизированного построения словаря предметной области

Для анализа эффективности рассматриваемого метода построения словаря предложена формула для определения времени t , необходимого на составление словаря:

$$t = (t_r + t_w + t_g + n_w * t_s + (n_w + n_g) * t_d) * n_p + t_k * n_p, \quad (15)$$

где t_r — время прочтения страницы экспертом; t_w — время выделения терминов-слов из страницы текста; t_g — время выделения именных групп из страницы текста; t_s — время поиска синонимов для термина-слова; n_w и n_g — количество терминов-слов и именных групп на странице; t_d — время поиска и редактирования толкования термина; $t_k * n_p$ — время кластеризации текста.

Автоматизация построения словаря позволяет значительно уменьшить значения t_w , t_g , t_s , t_d , $t_k * n_p$. Экспериментальные данные, полученные по результатам анализа ряда документов, приведены в табл. 4. Здесь время представлено в минутах.

Таблица 4

Оценка времени составления словаря

| Способ построения словаря | t_r | t_w | t_g | t_s | t_d | t_k |
|---------------------------|-----------|-----------|-----------|-------|-------|-----------|
| «Ручной» | 3 | 20 | 15 | 20 | 20 | 30 |
| Автоматизированный | 10^{-3} | 10^{-3} | 10^{-3} | 0,5 | 0,5 | 10^{-3} |

Время составления словаря для документа D_1 в режиме ручной работы составило около 10,5 часа, а в режиме автоматизированной — порядка 20 минут (с учетом времени на корректировку экспертом полученных результатов).

7. Выводы

Разработанный метод построения словаря предметной области позволяет автоматизировать процессы подбора и классификации исходных текстов, выделения терминов, поиска синонимов для терминов и поиска описания терминов, что существенно сокращает время

создания словаря. Проведенные эксперименты подтверждают эффективность предложенного метода построения словаря. Результаты показывают, что можно получить качественный словарь предметной области за очень незначительный период времени.

Основное приложение предложенного метода — разработка и модернизация информационных систем, различные интеллектуальные системы, использующие естественные языки.

Литература

1. Черткова, Е. А. Моделирование предметной области для проектирования компьютерных обучающих систем [Электронный ресурс] / Е. А. Черткова // Конгресс конференций «Информационные технологии в образовании». Секция VII. — 2005. — Режим доступа: \www/URL: http://ito.edu.ru/2005/Moscow/VII/VII-0-5032.html
2. JaLingo [Electronic resource]. — Available at: \www/URL: http://jalingo.sourceforge.net/
3. Кунгурцев, А. Б. Формирование словаря предметной области [Текст] / А. Б. Кунгурцев, И. В. Барыкина // Искусственный интеллект. — 2006. — № 1. — С. 144–151.
4. Кунгурцев, А. Б. Застосування мереж фреймів для побудови моделі видчуження фактів з текстів на природній мові [Текст] / А. Б. Кунгурцев, С. М. Бородавкін // Искусственный интеллект. — 2009. — № 4. — С. 202–207.
5. Кунгурцев, А. Б. Метод построения словарей предметных областей для извлечения фактов из текстов на естественном языке [Текст] / А. Б. Кунгурцев, С. Н. Бородавкин, А. П. Голуб // Восточно-Европейский журнал передовых технологий. — 2010. — № 1/4(43). — С. 32–36. — Режим доступа: \www/URL: http://journals.urau.ua/ejet/article/view/2550
6. Bourigault, D. Surface grammatical analysis for the extraction of terminological noun phrases [Text] / D. Bourigault // Proceedings of the 14th conference on Computational linguistics. — Association for Computational Linguistics (ACL), 1992. — P. 977–981. doi:10.3115/993079.993111
7. Baroni, M. Bootstrapping Corpora and Terms from the Web [Text] / M. Baroni, S. Bernardini // Proceedings of LREC. — Lisbon: ELDA, 2004. — P. 1313–1316.
8. Программный пакет синтаксический анализ. Проект АОТ [Электронный ресурс]. — Режим доступа: \www/URL: http://www.aot.ru/docs/synan.html
9. Шелов, С. Д. Терминоведение: семь вопросов и семь ответов по семантике термина [Текст] / С. Д. Шелов // НТИ. Сер. 2. Информационные процессы и системы. — 2001. — № 2. — С. 1–11.
10. Ляшевская, О. Н. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка) [Электронный ресурс] / О. Н. Ляшевская, С. А. Шаров. — М.: Азбуковник, 2009. — Режим доступа: \www/URL: http://dict.ruslang.ru/freq.php
11. Ожегов, С. И. Толковый словарь русского языка [Текст] / С. И. Ожегов, Н. Ю. Шведова. — М.: ОНИКС 21 век: Мир и Образование, 2004. — 1198 с.
12. Программный пакет синтаксического разбора и машинного перевода [Электронный ресурс]. — Режим доступа: \www/URL: http://cs.isa.ru:10000/dwarf/. — 2008.
13. Кнут, Д. Э. Искусство программирования [Текст]. Том 3. Сортировка и поиск / Д. Э. Кнут. — М.: Издательский дом Вильямс, 2007. — 800 с.
14. Хорстманн, К. Java [Текст]. Том 2. Библиотека профессионала / К. Хорстманн, Г. Корнелл. — М.: Издательский дом Вильямс, 2014. — 864 с.

МЕТОД АВТОМАТИЗОВАНОЇ ПОБУДОВИ ТЛУМАЧНОГО СЛОВНИКА ПРЕДМЕТНОЇ ОБЛАСТІ

Розроблено метод автоматизованої побудови тлумачного словника, заснований на обробці текстів із заданої предметної області. Описана технологія підбору і групування вихідних

текстів, виділення термінів, пошуку синонімів і значень термінів. Описана технологія кластеризації документа для можливості підбору документів і створення загального словника.

Ключові слова: словник, термін, предметна область, синонім, іменні групи.

Кунгурцев Алексей Борисович, кандидат технических наук, профессор кафедры системного программного обеспечения, Одесский национальный политехнический университет, Украина, e-mail: abkun@te.net.ua.

Поточник Яна Владимировна, аспирант, кафедра системного программного обеспечения, Одесский национальный политехнический университет, Украина, e-mail: yana_onpu@mail.ru.

Силаев Дмитрий Александрович, кафедра системного программного обеспечения, Одесский национальный политехнический университет, Украина, e-mail: dsilyaev@gmail.com.

Кунгурцев Олексій Борисович, кандидат технічних наук, професор кафедри системного програмного забезпечення, Одеський національний політехнічний університет, Україна.

Поточник Яна Володимирівна, аспірант, кафедра системного програмного забезпечення, Одеський національний політехнічний університет, Україна.

Сіляєв Дмитро Олександрович, кафедра системного програмного забезпечення, Одеський національний політехнічний університет, Україна.

Kungurtsev Alexei, Odessa National Polytechnic University, Ukraine, e-mail: abkun@te.net.ua.

Potochniak Iana, Odessa National Polytechnic University, Ukraine, e-mail: yana_onpu@mail.ru.

Silyaev Dmytro, Odessa National Polytechnic University, Ukraine, e-mail: dsilyaev@gmail.com

УДК 004.891.2:332.873:697.1

DOI: 10.15587/2312-8372.2015.40998

Дубинский А. Г.

РАЗРАБОТКА ПРОЕКТА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ПО УЛУЧШЕНИЮ ЭНЕРГОЭФФЕКТИВНОСТИ ЖИЛЫХ ПОМЕЩЕНИЙ

Рассмотрена задача повышения энергоэффективности жилых домов. Предложена структура онлайн-системы для получения рекомендаций по повышению тепло- и энергоэффективности. Выбраны программы и технологии, которые будут использованы на этапе реализации системы; для моделирования выбрана программная среда с открытым исходным кодом. Указаны основные необходимые шаги для создания предложенной системы.

Ключевые слова: энергоэффективность, энергосбережение, экономия энергопотребления, тепловой баланс, экспертная система, моделирование, стартап.

1. Введение

Одна из наиболее актуальных задач, стоящих перед Украиной — повышение энергоэффективности и снижение общего энергопотребления. Закупка энергоносителей составляет значительную часть совокупного импорта страны. Уменьшение энергопотребления должно позитивно повлиять на внешнеторговый баланс. В структуре энергопотребления значительную долю занимает обеспечение жилищно-коммунального хозяйства — отопление и обогрев. Повышение коммунальных тарифов на газ, отопление и электроэнергию привлекает еще больше внимания к тематике повышения энергоэффективности и делает эту сферу рентабельной для создания новых коммерческих проектов, в том числе в формате «стартапов» с использованием современных информационных технологий.

В марте 2015 года в шести городах Украины состоялся трехдневный хакатон энергетической свободы «Hack4energy» — форум для разработчиков в сфере энергосбережения, в ходе которого был представлен ряд проектов, предусматривающих внедрение как технологических, так и поведенческих решений, направленное на повышение энергоэффективности. В ходе мозгового штурма Днепропетровскими участниками была, в част-

ности, предложена идея создания калькулятора энергоэффективности с расчетом экономического эффекта для сравнения различных решений по утеплению жилых помещений. В данной статье рассматривается проект создания подобной информационно-рекомендательной системы.

2. Анализ литературных данных и постановка задачи

Общую информацию по проблематике энергосбережения можно найти в [1]. Согласно «Энергетической стратегии Украины на период до 2030 года», потенциал энергоэффективности и энергосбережения в Украине составляет около половины от ежегодного объема потребления [2]. По данным рейтинга Ukrainian Energy Index показатель энергоэффективности экономики Украины составляет 52 % от уровня ЕС. Таким образом, повышение энергоэффективности до европейского уровня позволит экономить порядка 11,8 млрд. евро ежегодно [3].

В целом по Украине в зданиях непроизводственного назначения (жилых домах, больницах, школах и т. п.) потребляется более 40 % энергоресурсов [4]. Ключевая часть расходуемой энергии используется для