

пуклая поверхность мениска – эллипсоид вращения). Все поверхности – конического сечения, удобные в изготовлении и контроле.

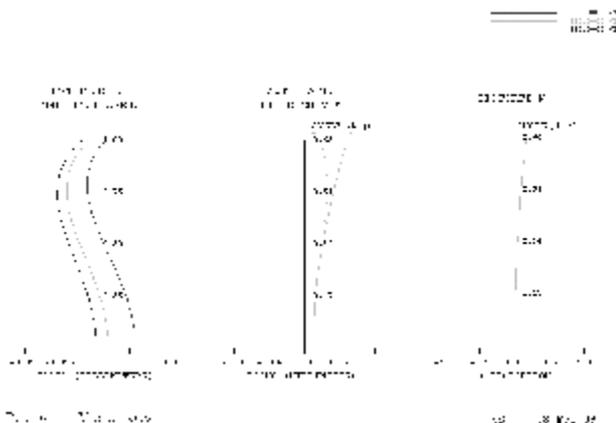


Рис. 6. Графики продольной сферической аберрации, астигматизма и дисторсии

В заключение отметим, что наибольшее распространение в мире (особенно США) получили системы Ричи–Кретьена с корректором Гаскойна, напоминающим линзу системы Шмидта [2], основным недостатком которого является хроматизм, вносимый корректором Гаскойна. На снимках, полученных спутником GeoEye-1 и опубликованных на различных интернет-ресурсах, также присутствует заметный хроматизм. Предложенная нами оптическая система имеет пренебрежимо малый хроматизм и плоское поле размером $1,4^\circ$. В качестве светоприемной аппаратуры можно использовать ПЗС линейки различной длины, при этом важно, чтобы размер пикселя соответствовал размеру кружка рассеяния, создаваемого оптической системой.

Библиографический список

1. Максудов, Д. Д. *Астрономическая оптика* / Д. Д. Максудов. Л.: Наука. Ленингр. отд-ние, 1979.
2. Попов, Г. М. *Современная астрономическая оптика* / Г. М. Попов. М.: Наука, 1988.

S. A. Veselkov

OPTICAL SYSTEM OF HIGH RESOLUTION FOR EARTH-IMAGING FROM SPACE

The space-photographic targets of the Earth with ultra-high resolution optical system are reviewed. The data received by detailed pre-calculation of the telescope in especial optimizing program are given.

Keywords: satellite, resolution, telescope, spectral range, orbit.

© Veselkov S. A., 2009

УДК 681.5

М. В. Карасева, М. А. Селиванова, П. В. Зеленков, Е. Е. Шукшина

ИСПОЛЬЗОВАНИЕ ТЕЗАУРУСОВ ПРИ ПОСТРОЕНИИ МУЛЬТИЛИНГВИСТИЧЕСКИХ МОДЕЛЕЙ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Предложена модификация существующих моделей распределенных информационных систем, основанных на тезаурусах. Модификация направлена на решение проблемы мультилингвистичности представления информации в современных системах. Показаны две модели: первая основана на мультилингвистическом тезаурусе, вторая – на частотном мультилингвистическом тезаурусе.

Ключевые слова: мультилингвистическая модель, тезаурус, частотный словарь.

В настоящее время разработано множество моделей для представления распределенных систем вычисления и (или) обработки информации. К ним, в частности, относятся информационные системы, корпоративные информационные системы и интенсивно развивающиеся системы поддержки принятия решения [1]. Однако большинство моделей распределенных систем строятся на основе одноязычного представления информации или учитывают многоязычие неявно [2].

Одним из перспективных направлений при разработке новых моделей является применение словарей, или тезаурусов. При этом необходимо отметить, что в современных системах подобные словари очень редко встречаются представленными в мультилингвистической реализации.

Авторский подход, отраженный в работах [2; 3], направлен в первую очередь на решение проблемы мультилингвистического представления информации в инфор-

мационно-управляющих системах, поскольку в современных условиях даже небольшие корпоративные информационные системы, как правило, работают в мультилингвистическом режиме.

Рассмотрим следующее определение мультилингвистического тезауруса T слов T_{ij} , сформированное на базе положений, представленных в работе [4], где I – языковое множество, а j – терм в языковом множестве, отвечающие следующим условиям:

1) имеется непустое подмножество $T_0 \subset T$, называемое *множеством дескрипторов*;

2) имеется симметричное, транзитивное рефлексивное отношение $R \subset T \times T$, такое что

$$t_{i1} \neq t_{i2} \wedge t_{i1} R t_{i2} \Rightarrow (t_{i1} \in T_0) \vee (t_{i2} \in T_0),$$

$$t_{i1} \in T_0 \Rightarrow (\exists t_i \in T_0)(t_i R t_{i1}).$$

При этом отношение R называется *синонимическим отношением*, а слова t_{i1}, t_{i2} , отвечающие этому отношению, – *синонимическими дескрипторами*;

3) имеется транзитивное и несимметричное отношение $K \subset T_0 \times T_0$, называемое *обобщающим отношением*.

В случае если два дескриптора t_{i1}, t_{i2} удовлетворяют отношению $t_{i1} K t_{i2}$, то полагается, что дескриптор t_{i1} более общий, чем дескриптор t_{i2} .

Элементы множества $T \setminus T_0$ называются *множеством аскрипторов*.

Таким образом, информационную систему с применением тезауруса можно представить в виде четверки (T, D_i, M_i, δ) , где T – мультилингвистический тезаурус с дескрипторным множеством T_0 ; D_i – коллекция разноязычных документов; M_i – множество разноязычных вопросов; $\delta: \rightarrow 2^D$ – отображение, сопоставляющее каждому независимому от языковой принадлежности вопросу множество разноязычных документов.

Пусть описание любого документа $d_i \in D_i$ может быть представлено в виде

$$t(d_i) = \{t_{i1}, t_{i2}, \dots, t_{ik}\}$$

и удовлетворяет условию: никакие два дескриптора не встречаются в одном $t_i(d_i)$, если они удовлетворяют отношению K .

Можно также считать, что каждый вопрос $m_i \in M_i$ представляется в форме, аналогичной описанию разноязычных документов.

Множество описаний вопросов и документов частично упорядочено отношением включения (\leq) следующим образом:

$$\forall d_{i1}, d_{i2} \in D_i,$$

$$t_i(d_{i1}) < t_i(d_{i2}) \Leftrightarrow (\forall t'_i \in t_i(d_{i1})) \times (\exists t''_i \in t_i(d_{i2})) (t'_i K t''_i) \vee (t'_i - t''_i).$$

т. е. каждый дескриптор из $t_i(d_{i1})$ представляет собой обобщение дескриптора из $t_i(d_{i2})$ или идентичен дескриптору из $t_i(d_{i2})$.

Отношение \leq позволяет сформулировать ответ Q на вопрос $m_i \in M_i$ в виде

$$Q = \delta(m_i) = \{d_i : d_i \in D_i \wedge m_i \leq t_i(d_i)\}.$$

Пусть S_1, \dots, S_n – локальные информационные системы, где $S_j = (T_j, D_{ij}, M_{ij}, \delta_j), j = 1, \dots, n$. Объединим системы

S_1, \dots, S_n в одну систему S , которую в дальнейшем будем называть *распределенной информационной системой*, базирующейся на глобальном тезаурусе $T = \bigcup_j T_j$.

Итак, пусть у нас имеется распределенная информационная система $S = (T, D_i, M_i, \delta)$ с синонимическим отношением R и обобщающим отношением K . Тогда последовательность информационных систем можно определить следующим образом:

$$S_j = (T_j, D_{ij}, M_{ij}, \delta_j),$$

где $T_j \in T, D_{ij} \in D_i, M_{ij} \in M_i; \delta_j$ – сужение δ на $M_{ij}; \forall j = 1, n$. Кроме того, отношения $R_j = R \cap (T_j \times T_j)$ и $K_j = K \cap (T_{0j} \times T_{0j})$, где T_{0j} – множество дескрипторов системы S_j .

Необходимо отметить, что множества T_j и D_{ij} могут пересекаться.

Также покажем возможность существования подсистемы, которая позволяет формализовать отношение предпочтения одной системы перед другой.

Пусть $S_1 = (T_1, D_{i1}, M_{i1}, \delta_1)$ и $S_2 = (T_2, D_{i2}, M_{i2}, \delta_2)$ – информационные системы. Система S_1 является подсистемой S_2 ($S_1 \subset S_2$), если:

- 1) $(T_1 \subset T_2) \wedge (K_1 = K_2 \cap (T_{01} \times T_{01})) \wedge (R_1 = R_2 \cap (T_1 \times T_1))$;
- 2) $D_{i1} \subset D_{i2}$;
- 3) $(M_{i1} \subset M_{i2}) \wedge \leq_1 = \leq_2 (M_{i1} \times M_{i1})$;
- 4) $\delta_1(m_i) = \delta_2(m_i) \cap D_{i1}, m_i \in M_{i1}$.

Очевидно, что локальные информационные системы являются подсистемами распределенной информационной системы.

Ввиду того что множество разноязычных документов, например, в распределенной системе поддержки принятия решения является объединением множеств разноязычных документов локальных информационных систем, ответ на вопрос к распределенной системе можно выразить как результат некоторых операций над ответами от локальных систем.

Пусть $\overline{m_i} = \{m_i\}$ – вопрос, содержащий один дескриптор, а $\delta(m_i)$ – ответ на вопрос $\overline{m_i}$. В подсистеме S_j ответ на локальный вопрос $\overline{m_i} = \{m_i\}$ согласовывается следующим образом:

$$\delta_j(\overline{m_i}) = \{d_i : d_i \in D_{ij} \wedge m_i \leq_{ij} t_i(d_i)\}.$$

В свою очередь в распределенной системе S глобальный ответ на вопрос $\overline{m_i} = \{m_i\}$ будет объединением локальных ответов $\delta(\overline{m_i}) = \bigcup_j \delta_j(\overline{m_i})$.

Ответ на произвольный вопрос $\overline{m_i} = \{m_{i1}, \dots, m_{ik}\}$ для любого языкового множества i в распределенной системе выражается формулой

$$\delta(\overline{m_i}) = \bigcap_{l=1}^k \delta(\overline{m_{il}}), \quad \overline{m_{il}} = \{m_{il}\}.$$

В [4] показана справедливость этой формулы и ее согласованность с определением последовательности информационных систем в рамках произвольного языкового множества i , однако необходимо отметить, что данная формула может быть модифицирована и для случая, когда имеется произвольное количество языковых множеств.

Итак, распределенная информационная система обладает свойством включения, если множество вопросов частично упорядочено отношением и выполняется условие

$$\{\bar{m}_{i1}, \bar{m}_{i2}\} \subset M_i \wedge \bar{m}_{i1} \leq \bar{m}_{i2} \Rightarrow \delta(\bar{m}_{i1}) \supset \delta(\bar{m}_{i2}).$$

Свойство включения позволяет формулировать цепь ответов на цепь вопросов, уточняя ответы более специализированными вопросами.

При реализации систем поддержки принятия решений находит место и более интересная модель распределенной системы, основанная на частотных мультилингвистических тезаурусах.

Для работы с обобщенной распределенной мультилингвистической информационной системой, а также для работы с мультилингвистическими системами поддержки принятия решений, вводятся веса, которые будут описывать информационное содержание термина в каждом языковом множестве [4].

Покажем работу предлагаемой модели для произвольного языка j (количество языков в обработке зависит от имеющегося мультилингвистического словаря и представленного в нем языкового многообразия) [5]. Вес w_{ij} соответствует объему информации, относящейся к дескриптору t_{ij} , т. е. $t_j(d_j) = \{ \langle t_{1j}, w_{1j} \rangle, \langle t_{2j}, w_{2j} \rangle, \dots, \langle t_{kj}, w_{kj} \rangle \}$ и выполняются условия:

- 1) $t_{ji} \neq t_{qj} \wedge t_{ij} K t_{qj} \Rightarrow (\forall w_{ij} \in [0,1]) (t_{ij}, w_{ij}) \in t_j(d_j) \vee (\forall w_{qj} \in [0,1]) (t_{qj}, w_{qj}) \in t_j(d_j)$;
- 2) $\sum_{i=1}^k w_{ij} = 1$.

Описание мультилингвистических документов также удовлетворяет свойству включения. Дадим определение подобия вопросов и описаний документов.

Пусть $t_j(d_{1j}), t_j(d_{2j})$ – описания мультилингвистических документов, представленных в j -м языковом множестве:

$$\begin{aligned} t_j(d_{1j}) &= \{ \langle t_{11j}, w_{11j} \rangle, \\ &\langle t_{12j}, w_{12j} \rangle, \dots, \langle t_{1k_1j}, w_{1k_1j} \rangle \}, \\ t_j(d_{2j}) &= \{ \langle t_{21j}, w_{21j} \rangle, \langle t_{22j}, \\ &w_{22j} \rangle, \dots, \langle t_{2k_2j}, w_{2k_2j} \rangle \}. \end{aligned}$$

Описание $t_j(d_{1j})$ подобно описанию $t_j(d_{2j})$ с точностью ρ ($0 \leq \rho \leq 1$), если

$$t_j(d_{1j}) \leq t_j(d_{2j}) \wedge \forall_{\substack{\langle t_{1ij}, w_{1ij} \rangle \in t_j(d_{1j}) \\ \langle t_{2qj}, w_{2qj} \rangle \in t_j(d_{2j})}} w_{1ij} \rho \leq w_{2qj},$$

где $I_j(t_{ij}) = \{ t_{2qj} : t_{ij} K t_{2qj} \wedge 1 \leq q \leq k_2 \}$ при $1 \leq i \leq k_1$. Если $t_j(d_{1j})$ и $t_j(d_{2j})$ удовлетворяют этому условию, то пишется $t_j(d_{1j}) \triangleleft_c t_j(d_{2j})$.

Отношения подобия, доказательство которых очевидно, также можно представить в мультилингвистическом виде:

- если $(t_j(d_{1j}) \triangleleft_{\rho_1} t_j(d_{2j})) \wedge (\rho_1 \geq \rho_2)$, то $t_j(d_{1j}) \triangleleft_{\rho_2} t_j(d_{2j})$;
- если $(t_j(d_{1j}) \triangleleft_{\rho_1} t_j(d_{2j})) \wedge (t_j(d_{2j}) \triangleleft_{\rho_2} t_j(d_{3j}))$, то $t_j(d_{1j}) \triangleleft_{\rho} t_j(d_{3j})$, ($\rho \leq \rho_1 \rho_2$).

Представляет интерес величина, характеризующая коэффициент подобия, или меру корреляции пар мультилингвистических векторов, удовлетворяющих отношению \leq .

Для векторов $t_j(d_{1j}), t_j(d_{2j})$, таких что $t_j(d_{1j}) \leq t_j(d_{2j})$, мера корреляции будет

$$\begin{aligned} \mu(t_j(d_{1j}), t_j(d_{2j})) &= \\ &= \sum_{i \in \{1, \dots, k_1\}} \min(w_{1ij}, \sum_{q \in \{1, \dots, k_2\}} w_{2qj}) \end{aligned}$$

где $I(t_{ij}) = \{ t_{2sj} : t_{ij} K t_{2sj} \wedge 1 \leq s \leq k_2 \}$ при $1 \leq i \leq k_1$.

Мера μ фактически оценивает пересекающуюся информацию, заключенную в описании документов $t_j(d_{1j}), t_j(d_{2j})$.

Таким образом, распределенная мультилингвистическая информационная система на тезаурусе с весами представляется четверкой (T, D_j, M_j, δ) , где T – тезаурус с дескриптовым множеством $T_0 \subset T$; $\delta: M_j \times [0,1] \rightarrow 2^{D_j \times [0,1]}$, сопоставляющей каждой паре (вопрос, точность подобия) множество пар (документ, мера корреляции).

Ответом системы на вопрос m_j с требуемой точностью $\rho = c$ называется множество

$$\begin{aligned} Q &= \delta(\bar{m}_j, c) = \\ &= \left\{ (d_j, \alpha) : d_j \in D_j \wedge \bar{m}_j \triangleleft_c t_j(d_j) \wedge \alpha = \right. \\ &\quad \left. = \mu(\bar{m}_j, t_j(d_j)) \right\} \subseteq \\ &\subseteq D_j \times [0,1]. \end{aligned}$$

В определении ответа $N = \delta(\bar{m}_j, c)$ отношение подобия \triangleleft_c осуществляет выбор мультилингвистических документов, точность подобия которых не менее c . Мера корреляции $\mu(\bar{m}_j, t_j(d_j))$ показывает, какая часть информации в документе соответствует ответу на вопрос.

В заключение рассмотрим одно из фундаментальных свойств ответа, связанное с точностью ρ и мерой корреляции μ вопроса к описанию мультилингвистических документов, включенных в ответ. Покажем его следующим образом: если документ $d_j \in D_j$ с мерой корреляции α включен в ответ, т. е. $(d_j, \alpha) \in \delta(\bar{m}_j, c)$, то верно неравенство $\alpha \geq c$. Обратное утверждение неверно. Также необходимо отметить, что если $\rho_1 \geq \rho_2$, то $\delta(\bar{m}_j, \rho_1) \subset \delta(\bar{m}_j, \rho_2)$.

Таким образом, по сравнению с исходными моделями распределенных информационных систем [6; 7] представленные мультилингвистические модели позволяют составлять мультилингвистические ответы даже на одноязычные запросы более гибко и с учетом неопределенности описания как мультилингвистических документов, так и самих запросов [8].

Библиографический список

1. Мультилингвистическая технология поиска данных для подготовки и принятия решения в информационно-управляющих системах / П. В. Зеленков, П. М. Лохмаков, И. В. Ковалев, С. С. Огнерубов // Прогр. продукты и системы. 2007. № 2. С. 32–33.
2. Карасева, М. В. Метапоисковая мультилингвистическая система / М. В. Карасева, П. В. Зеленков, И. Н. Карцан // Вестник СибГАУ. 2007. Вып. 3 (16). С. 130–131.
3. Зеленков, П. В. Современные поисковые системы в сети Интернет: анализ принципов работы и классификация / П. В. Зеленков, И. Н. Карцан, И. Н. Кожевников // Вестник НИИ СУВПТ : сб. науч. тр. / НИИ систем упр., волновых процессов и технологий. Вып. 23. Красноярск, 2006. С. 221–227.
4. Дмитриев, А. К. Основы теории построения и контроля сложных систем / А. К. Дмитриев, П. А. Мальцев. Л. : Энергоатомиздат. Ленингр. отд-ние, 1988.
5. Оптимизация обработки данных в распределенных образовательных средах / И. В. Ковалев, П. В. Зеленков,

С. А. Яркова, С. Ф. Шевчук // Прогр. продукты и системы. 2007. № 3. С. 70–71.

6. Зеленков, П. В. Инструментарий поисковых систем сети Интернет / П. В. Зеленков, И. Н. Карцан., П. М. Лохмаков // Вестник НИИ СУВПТ : сб. науч. тр. / НИИ систем упр., волновых процессов и технологий. Вып. 23. Красноярск, 2006. С. 103–118.

7. Метапоисковая мультилингвистическая система поиска узкоспециализированной информации / П. В. Зеленков, М. В. Карасева, И. Н. Карцан и др. М., 2007. Зарег. во Всерос. науч.-техн. центре, № 50200701673, рег. № ОФАП 8891.

8. Мультилингвистическая модель распределенной системы на основе тезауруса / С. В. Рогов, П. Н. Зеленков, И. В. Ковалев, М. В. Карасева // Вестник СибГАУ. 2008. № 1 (18). С. 26–28.

M. V. Karaseva, M. A. Selivanova, P. V. Zelenkov, E. E. Shukshina

THE USE OF THESAURUSES FOR BUILDING MULTILINGUAL MODELS OF THE DISTRIBUTED INFORMATIONAL SYSTEMS

The paper offers the modification of the existed models of the distributed informational systems based on thesauruses. The modification is aimed at problems of multilingual informational presentation in modern systems solutions. Two models are shown: the first one is based on multilingual thesauruses, the second one is based on frequency multilingual thesauruses.

Keywords: multilingual model, frequency dictionary, thesaurus.

© Карасева М. В., Селиванова М. А., Зеленков П. В., Шушкина Е. Е., 2009

УДК 004.42

Е. М. Товбис

OLAP-ТЕХНОЛОГИЯ КАК СРЕДСТВО МОНИТОРИНГА ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Без тщательного мониторинга образовательного процесса в высшем учебном заведении невозможно всесторонне оценить достижения студента. В целях мониторинга предложено использовать технологию многомерного анализа данных OLAP. Рассмотрен пример внедрения OLAP-технологии в учебный процесс на уровне дисциплины.

Ключевые слова: мониторинг, OLAP, учебный процесс.

Мониторинг является средством оценки качества образования и повышения эффективности учебного процесса. Обобщая различные определения, под педагогическим мониторингом в данной статье будем понимать отслеживание результатов обучения путем организации системы контроля, сбора и обработки данных об учебном процессе, необходимых для анализа и прогноза ситуации. Введение мониторинга в учебный процесс позволяет получать актуальную информацию о его протекании, своевременно принимать меры в случае отклонений от планируемой траектории и предпринимать шаги по сохранению численности студентов.

Анализ публикаций [1–3] позволил выделить несколько основных способов мониторингового исследования образовательного процесса:

- опросные способы (экспертное оценивание, анкетирование с помощью специально подобранных вопросов);
- анализ результатов учебной деятельности, в том числе и статистический анализ;

– способы, основанные на тестировании.

Чем качественнее проводится мониторинг, тем большее количество данных накапливается в результате, а значит, тем труднее обходиться без автоматизированной системы. Компьютерный мониторинг позволяет значительно сократить время на обработку и анализ результатов наблюдений за учебной деятельностью. При этом удается избавиться от так называемых невыборочных ошибок, возникающих при традиционных методах получения информации вследствие неполучения ответа, получения ложного ответа или неверной регистрации ответов респондента. При условии корректного построения системы мониторинга и ее интеграции в учебный процесс удастся избежать одной из основных проблем измерения – воздействия способа измерения на испытуемого.

Однако далеко не все автоматизированные решения способны полностью интегрироваться в сложившийся технологический процесс обучения. Их недостатком является и то, что большинство систем мониторинга и под-