

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ

УДК 026.6

П.С. Жуков, С.Е. Савотченко

МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННОГО ПОИСКА В БАЗЕ ДАННЫХ С УЧЕТОМ СЕМАНТИЧЕСКИХ СВЯЗЕЙ

Жуков Павел Сергеевич, аспирант кафедры библиотековедения, библиографоведения и книговедения Белгородского государственного института искусств и культуры. Имеет статьи в области автоматизированных библиотечных информационных систем. [e-mail: pavel_zhukov@hotmail.com].

Савотченко Сергей Евгеньевич, доктор физико-математических наук, доцент, окончил физический факультет Харьковского государственного университета, в настоящее время профессор кафедры информатики и информационно-аналитических ресурсов Белгородского государственного института искусств и культуры. Имеет статьи в области математического моделирования, информационных технологий и автоматизированных библиотечных информационных систем. [e-mail: savotchenko@hotmail.ru].

Аннотация

В работе предложена модель информационного поиска с использованием дескрипторов, которая дает возможность рассматривать границы между множествами как неопределенно выраженные. В работе определены функции принадлежности понятия запросам. Рассмотренный подход посредством применения семантических связей позволяет сформировать граничные множества как документов, соответствующих дескрипторам, так и документов, не совсем соответствующих запросу, но которые могут быть задействованы в дальнейшем.

Ключевые слова: дескриптор, информационный поиск, семантические связи, парадигматические отношения, информационно-поисковые системы.

Pavel Sergeevich Zhukov, Post-graduate Student at the Department of Library science, Bibliography and Bibliology of Belgorod State Institute of Art and Culture; author of articles in the field of automated library systems. e-mail: pavel_zhukov@hotmail.com.

Sergei Evgenievich Savotchenko, Doctor of Physics and Mathematics, Professor; graduated from the Faculty of Physics of Kharkov State University; Professor at the Department of Computer Science and Information and Analytical Resources of Belgorod State Institute of Art and Culture; author of articles in the field of mathematical modeling, information technologies, and automated library systems. e-mail: savotchenko@hotmail.ru.

Abstract

The article presents a an information search model based on using descriptors which enables to consider the boundaries of sets as being inexplicit. The queries concept membership functions are defined. The approach considered by the use of semantic relations enables to form boundary sets of both documents corresponding to the descriptors and documents not quite corresponding to a query, but which may be enabled in the future.

Key words: descriptor, information search, semantic relations, paradigmatic relations, information retrieval systems.

ВВЕДЕНИЕ

Основная задача информационного поиска – из имеющегося множества информации отобразить подмножество, соответствующее информационной потребности пользователя. В свою очередь, одна из центральных задач, от решения которых зависит эффективность модели информационного поиска, – это расширение анализируемой предметной области.

Потребность проектировщиков информационных поисковых систем в более удобных и мощных средствах моделирования предметной области вызвала к жизни направление семантических моделей данных. Главным назначением семантических моделей является обеспечение возможности выражения семантики данных, под которой понимается смысл информации посредством символов. Чаще всего на практике семантическое моделирование используется на первой стадии проектирования базы данных (БД). При этом в терминах семантической модели производится концептуальная схема БД, которая затем вручную преобразуется к реляционной схеме.

В статье рассматривается информационный поиск с учетом семантических связей, суть которого состоит в нахождении множеств документов, сопоставимых определенным дескрипторам и имеющих числовую оценку вхождения запроса пользователя в найденные документы.

В [1] была рассмотрена модель информационно-поисковой системы (ИПС) «Пусто-непусто». Данная ИПС выдает результаты блоками: сначала выдаются документы, наиболее подходящие запросу пользователя, а затем документы в порядке снижения соответствия запросу, включая видовые и родовые понятия. В отличие от нее, модель ИПС, предлагаемая в настоящей работе, позволяет получить более расширенный набор документов в ответ на запрос пользователя, которому предоставляется возможность получить доступ к структурированному результату поиска данных, учитывающему различные типы парадигматических отношений между лексическими единицами запроса.

1 Модель информационно-поисковой системы

Для моделирования информационного поиска в БД с учетом семантических связей можно использовать теорию нечетких множеств, в которой отношение элементов множеству описывается при помощи функции принадлежности. Вначале введем определение дескриптора.

Дескриптор – лексическая единица (слово, словосочетание) информационно-поискового языка, служит для описания основного смыслового содержания документа или формулировки запроса при поиске документа (информации) в ИПС.

Введем основные определения, необходимые для моделирования информационного поиска. Пусть $P = \{P_{ep}, P_{ec}, P_{nc}, P_{nc}, P_c, P_a\}$ – множество, состоящее из понятий, входящих в семантическое хранилище данных, где P_{ep} – множество дескрипторов, описывающих вышестоящие родовые понятия; P_{ec} – множество дескрипторов, описывающих вышестоящие понятия, обозначающие целое; P_{nc} – множество дескрипторов, описывающих ни-

жестоящие видовые понятия; P_{nc} – множество дескрипторов, описывающих нижестоящие понятия, обозначающие часть; P_c – множество дескрипторов, описывающих синонимы понятия; P_a – множество дескрипторов, описывающих ассоциации к понятию; $F^p = \{f_1, f_2, \dots, f_m\}$ – множество ключевых слов, описывающих понятия. Каждое понятие хранилища определяется набором из нескольких ключевых слов [2]. Стоит отметить, что между понятиями предметной области и ключевыми словами существуют некоторые множественные отношения: одно понятие может быть выражено с помощью нескольких ключевых слов, одно ключевое слово может быть задействовано в нескольких описаниях понятий.

Разделим данную модель поиска на подразделы $Z = \{z_1, z_2, \dots, z_l\}$, где подраздел – это часть предметной области, отвечающей за определенную тематику. Каждый подраздел привязан к одному или большему количеству понятий предметной области.

Формально модель поисковой системы представляется в следующем виде: $M = \langle M_p, M_z \rangle$, где $M_p = \{M_p^{ep}, M_p^{ec}, M_p^{nc}, M_p^{nc}, M_p^c, M_p^a\}$ – множество отношений «понятия - ключевые слова», а $M_z = \{M_z^{ep}, M_z^{ec}, M_z^{nc}, M_z^{nc}, M_z^c, M_z^a\}$ – множество отношений понятий к подразделам Z ,

$$M_p^{ep} = \begin{bmatrix} m_{11}^{ep} & m_{12}^{ep} & \dots & m_{1n}^{ep} \\ \dots & \dots & \dots & \dots \\ m_{g1}^{ep} & m_{g2}^{ep} & \dots & m_{gn}^{ep} \end{bmatrix}, \tag{1,a}$$

где $m_{ij}^{ep} = 1$, если i -е ключевое слово содержится в описании, иначе $m_{ij}^{ep} = 0, i = \overline{1, g}, j = \overline{1, n}$,

$$M_p^{ec} = \begin{bmatrix} m_{11}^{ec} & m_{12}^{ec} & \dots & m_{1l}^{ec} \\ \dots & \dots & \dots & \dots \\ m_{s1}^{ec} & m_{s2}^{ec} & \dots & m_{sl}^{ec} \end{bmatrix}, \tag{1,b}$$

где $m_{ij}^{ec} = 1$, если i -е ключевое слово содержится в описании, иначе $m_{ij}^{ec} = 0, i = \overline{1, s}, j = \overline{1, l}$,

$$M_p^{nc} = \begin{bmatrix} m_{11}^{nc} & m_{12}^{nc} & \dots & m_{1p}^{nc} \\ \dots & \dots & \dots & \dots \\ m_{r1}^{nc} & m_{r2}^{nc} & \dots & m_{rp}^{nc} \end{bmatrix}, \tag{1,v}$$

где $m_{ij}^{nc} = 1$, если i -е ключевое слово содержится в описании, иначе $m_{ij}^{nc} = 0, i = \overline{1, r}, j = \overline{1, p}$,

$$M_p^{nc} = \begin{bmatrix} m_{11}^{nc} & m_{12}^{nc} & \dots & m_{1t}^{nc} \\ \dots & \dots & \dots & \dots \\ m_{h1}^{nc} & m_{h2}^{nc} & \dots & m_{ht}^{nc} \end{bmatrix}, \tag{1,r}$$

где $m_{ij}^{nc} = 1$, если i -е ключевое слово содержится в описании, иначе $m_{ij}^{nc} = 0, i = \overline{1, h}, j = \overline{1, t}$,

$$M_P^c = \begin{bmatrix} m_{11}^c & m_{12}^c & \dots & m_{1k}^c \\ \dots & \dots & \dots & \dots \\ m_{o1}^c & m_{o2}^c & \dots & m_{ok}^c \end{bmatrix}, \quad (1,д)$$

где $m_{ij}^c = 1$, если i -е ключевое слово содержится в описании, иначе $m_{ij}^c = 0, i = \overline{1, o}, j = \overline{1, k}$,

$$M_P^a = \begin{bmatrix} m_{11}^a & m_{12}^a & \dots & m_{1v}^a \\ \dots & \dots & \dots & \dots \\ m_{q1}^a & m_{q2}^a & \dots & m_{qv}^a \end{bmatrix}, \quad (1,е)$$

где $m_{ij}^a = 1$, если i -е ключевое слово содержится в описании, иначе $m_{ij}^a = 0, i = \overline{1, q}, j = \overline{1, v}$.

В представленных матрицах каждой строке отношений (1, а, б, в, г, д, е) ставится в соответствие ключевое слово из модели, а столбцу – понятие [3].

Отношения понятий к подразделениям представим как:

$$M_Z^{ep} = \begin{bmatrix} z_{11}^{ep} & z_{12}^{ep} & \dots & z_{1n}^{ep} \\ \dots & \dots & \dots & \dots \\ z_{d1}^{ep} & z_{d2}^{ep} & \dots & z_{dn}^{ep} \end{bmatrix}, \quad (2,а)$$

где $z_{ij}^{ep} = 1$, если для описания i -го подразделения используется j -е понятие. В противном случае $z_{ij}^{ep} = 0$,

$$M_Z^{eu} = \begin{bmatrix} z_{11}^{eu} & z_{12}^{eu} & \dots & z_{1l}^{eu} \\ \dots & \dots & \dots & \dots \\ z_{f1}^{eu} & z_{f2}^{eu} & \dots & z_{fl}^{eu} \end{bmatrix}, \quad (2,б)$$

где $z_{ij}^{eu} = 1$, если для описания i -го подразделения используется j -е понятие. В противном случае $z_{ij}^{eu} = 0$,

$$M_Z^{ne} = \begin{bmatrix} z_{11}^{ne} & z_{12}^{ne} & \dots & z_{1p}^{ne} \\ \dots & \dots & \dots & \dots \\ z_{g1}^{ne} & z_{g2}^{ne} & \dots & z_{gp}^{ne} \end{bmatrix}, \quad (2,в)$$

где $z_{ij}^{ne} = 1$, если для описания i -го подразделения используется j -е понятие. В противном случае $z_{ij}^{ne} = 0$,

$$M_Z^{nu} = \begin{bmatrix} z_{11}^{nu} & z_{12}^{nu} & \dots & z_{1t}^{nu} \\ \dots & \dots & \dots & \dots \\ z_{l1}^{nu} & z_{l2}^{nu} & \dots & z_{lt}^{nu} \end{bmatrix}, \quad (2,г)$$

где $z_{ij}^{nu} = 1$, если для описания i -го подразделения используется j -е понятие. В противном случае $z_{ij}^{nu} = 0$,

$$M_Z^c = \begin{bmatrix} z_{11}^c & z_{12}^c & \dots & z_{1k}^c \\ \dots & \dots & \dots & \dots \\ z_{b1}^c & z_{b2}^c & \dots & z_{bk}^c \end{bmatrix}, \quad (2,д)$$

где $z_{ij}^c = 1$, если для описания i -го подразделения используется j -е понятие. В противном случае $z_{ij}^c = 0$,

$$M_Z^a = \begin{bmatrix} z_{11}^a & z_{12}^a & \dots & z_{1v}^a \\ \dots & \dots & \dots & \dots \\ z_{y1}^a & z_{y2}^a & \dots & z_{yv}^a \end{bmatrix}, \quad (2,е)$$

где $z_{ij}^a = 1$, если для описания i -го подразделения используется j -е понятие. В противном случае $z_{ij}^a = 0$.

Множество документов описывается множеством значений, выделенным из исследуемого списка документов, исключая шумовые слова. Обозначим множество значений как $T = \{T^{ep}, T^{eu}, T^{ne}, T^{nu}, T^c, T^a\}$, для которого выполняется следующее соотношение [3]: $F^p \subset T$. Таким образом, ключевые слова для описания понятий предметной области определяются из полученного множества значений документов.

Множество документов выразим следующим отношением:

$$O = \begin{bmatrix} o_{11} & \dots & o_{1s} \\ \vdots & \ddots & \vdots \\ o_{k1} & \dots & o_{ks} \end{bmatrix},$$

где o_{ij} – вес j -го значения для i -го документа, $i = [1, k]$, k – количество документов в хранилище, $j = [1, s]$. Для расчета весов значений можно использовать различные формулы (абсолютных и частотных мер, «сигнал-шум» и другие).

2 МОДЕЛИРОВАНИЕ ПРОЦЕССА ПОЛУЧЕНИЯ РЕЗУЛЬТАТОВ ВЫПОЛНЕНИЯ ЗАПРОСОВ

Запрос пользователя выразим в виде последовательности $S = \langle F^S \subset F^p, z_i \in Z \rangle$, которая включает в себя указание на подразделение, для которого необходимо получить результат, и ключевые слова.

Для нахождения документов, удовлетворяющих запросам пользователя, будем использовать приближенные множества Павлака [4].

Вначале необходимо определить классы эквивалентности понятий. Для этого компонент запроса F^S формирует на отношении M^p классы эквивалентности $F(p)$, куда входят понятия предметной области, которые неразличимы относительно элементов множества F^S .

Получив f_i из отношений (2, а, б, в, г, д, е), определим подмножества $\hat{P}_{ep}, \hat{P}_{eu}, \hat{P}_{ne}, \hat{P}_{nu}, \hat{P}_c, \hat{P}_a \subset P$, состоящие из тех p_j для которых в (2, а, б, в, г, д, е) $z_{ij} = 1$. Каждому подмножеству $\hat{P}_{ep}, \hat{P}_{eu}, \hat{P}_{ne}, \hat{P}_{nu}, \hat{P}_c, \hat{P}_a$ ставится в соответствие пара множеств, которые называются F -нижняя и F -верхняя аппроксимации соответствующего множества:

1) $F_*(\hat{P}_{ep}) = \{p \in P : F(p) \subseteq \hat{P}_{ep}\}$ – F -нижняя аппроксимация множества \hat{P}_{ep} ,

$F^*(\hat{P}_{ep}) = \{p \in P : F(p) \cap \hat{P}_{ep} \neq \emptyset\}$ – F -верхняя аппроксимация множества \hat{P}_{ep} ;

2) $F_*(\hat{P}_{eu}) = \{p \in P : F(p) \subseteq \hat{P}_{eu}\}$ – F -нижняя аппроксимация множества \hat{P}_{eu} ,

$F^*(\hat{P}_{eu}) = \{p \in P : F(p) \cap \hat{P}_{eu} \neq \emptyset\}$ – F -верхняя

аппроксимация множества $\hat{P}_{вц}$;

3) $F_*(\hat{P}_{вц}) = \{p \in P : F(p) \subseteq \hat{P}_{вц}\}$ – F -нижняя аппроксимация множества $\hat{P}_{вц}$,

$F^*(\hat{P}_{вц}) = \{p \in P : F(p) \cap \hat{P}_{вц} \neq \emptyset\}$ – F -верхняя

аппроксимация множества $\hat{P}_{нв}$;

4) $F_*(\hat{P}_{нв}) = \{p \in P : F(p) \subseteq \hat{P}_{нв}\}$ – F -нижняя аппроксимация множества $\hat{P}_{нв}$,

$F^*(\hat{P}_{нв}) = \{p \in P : F(p) \cap \hat{P}_{нв} \neq \emptyset\}$ – F -верхняя

аппроксимация множества $\hat{P}_{нч}$;

5) $F_*(\hat{P}_{нч}) = \{p \in P : F(p) \subseteq \hat{P}_{нч}\}$ – F -нижняя аппроксимация множества $\hat{P}_{нч}$,

$F^*(\hat{P}_{нч}) = \{p \in P : F(p) \cap \hat{P}_{нч} \neq \emptyset\}$ – F -верхняя

аппроксимация множества \hat{P}_c ;

6) $F_*(\hat{P}_c) = \{p \in P : F(p) \subseteq \hat{P}_c\}$ – F -нижняя аппроксимация множества \hat{P}_c ,

$F^*(\hat{P}_c) = \{p \in P : F(p) \cap \hat{P}_c \neq \emptyset\}$ – F -верхняя

аппроксимация множества \hat{P}_a .

F -границные области множеств определяются соответственно выражениями:

$$1) BN_F(\hat{P}_{вп}) = F^*(\hat{P}_{вп}) - F_*(\hat{P}_{вп}),$$

$$2) BN_F(\hat{P}_{вц}) = F^*(\hat{P}_{вц}) - F_*(\hat{P}_{вц}),$$

$$3) BN_F(\hat{P}_{нв}) = F^*(\hat{P}_{нв}) - F_*(\hat{P}_{нв}),$$

$$4) BN_F(\hat{P}_{нч}) = F^*(\hat{P}_{нч}) - F_*(\hat{P}_{нч}),$$

$$5) BN_F(\hat{P}_c) = F^*(\hat{P}_c) - F_*(\hat{P}_c),$$

$$6) BN_F(\hat{P}_a) = F^*(\hat{P}_a) - F_*(\hat{P}_a).$$

При $BN_F(\hat{P}_{вп}) \neq \emptyset, BN_F(\hat{P}_{вц}) \neq \emptyset, BN_F(\hat{P}_{нв}) \neq \emptyset, BN_F(\hat{P}_{нч}) \neq \emptyset, BN_F(\hat{P}_c) \neq \emptyset, BN_F(\hat{P}_a) \neq \emptyset$ совокупность множеств $\hat{P}_{вп}, \hat{P}_{вц}, \hat{P}_{нв}, \hat{P}_{нч}, \hat{P}_c, \hat{P}_a$ является приближенным множеством относительно компонента запроса F^S .

Мерой соответствия запроса системе предметной области, выражаемой в виде набора множеств понятий и связанных с ними ключевых слов, будет считаться точность аппроксимации для множества, которая определяется формулами для соответствующих множеств:

$$1) \delta_F(\hat{P}_{вп}) = \frac{N(F_*(\hat{P}_{вп}))}{N(F^*(\hat{P}_{вп}))},$$

$$2) \delta_F(\hat{P}_{вц}) = \frac{N(F_*(\hat{P}_{вц}))}{N(F^*(\hat{P}_{вц}))},$$

$$3) \delta_F(\hat{P}_{нв}) = \frac{N(F_*(\hat{P}_{нв}))}{N(F^*(\hat{P}_{нв}))},$$

$$4) \delta_F(\hat{P}_{нч}) = \frac{N(F_*(\hat{P}_{нч}))}{N(F^*(\hat{P}_{нч}))},$$

$$5) \delta_F(\hat{P}_c) = \frac{N(F_*(\hat{P}_c))}{N(F^*(\hat{P}_c))},$$

$$6) \delta_F(\hat{P}_a) = \frac{N(F_*(\hat{P}_a))}{N(F^*(\hat{P}_a))},$$

где N – мощность множества.

Меры соответствия описания понятий пользовательскому запросу будем определять приближенными функциями принадлежности понятия запросу, которые выразим следующим образом:

$$1) \mu_{\hat{P}_{вп}}^F(p) = \frac{N(\hat{P}_{вп} \cap F(p))}{N(F(p))}, \quad \mu_{\hat{P}_{вп}}^F(p) = \overline{0,1},$$

$$2) \mu_{\hat{P}_{вц}}^F(p) = \frac{N(\hat{P}_{вц} \cap F(p))}{N(F(p))}, \quad \mu_{\hat{P}_{вц}}^F(p) = \overline{0,1},$$

$$3) \mu_{\hat{P}_{нв}}^F(p) = \frac{N(\hat{P}_{нв} \cap F(p))}{N(F(p))}, \quad \mu_{\hat{P}_{нв}}^F(p) = \overline{0,1},$$

$$4) \mu_{\hat{P}_{нч}}^F(p) = \frac{N(\hat{P}_{нч} \cap F(p))}{N(F(p))}, \quad \mu_{\hat{P}_{нч}}^F(p) = \overline{0,1},$$

$$5) \mu_{\hat{P}_c}^F(p) = \frac{N(\hat{P}_c \cap F(p))}{N(F(p))}, \quad \mu_{\hat{P}_c}^F(p) = \overline{0,1},$$

$$6) \mu_{\hat{P}_a}^F(p) = \frac{N(\hat{P}_a \cap F(p))}{N(F(p))}, \quad \mu_{\hat{P}_a}^F(p) = \overline{0,1}.$$

Имея значения приближенных функций принадлежности для каждого понятия, можно определить тот набор понятий, который будет принимать участие в нахождении подходящих документов. Такое множество понятий будем определять согласно выражениям:

$$\tilde{P}_{вп} = \{p_j \in P : \mu_{\hat{P}_{вп}}^F(p) \geq \alpha\},$$

$$\tilde{P}_{вц} = \{p_j \in P : \mu_{\hat{P}_{вц}}^F(p) \geq \alpha\},$$

$$\tilde{P}_{нв} = \{p_j \in P : \mu_{\hat{P}_{нв}}^F(p) \geq \alpha\},$$

$$\tilde{P}_{нч} = \{p_j \in P : \mu_{\hat{P}_{нч}}^F(p) \geq \alpha\},$$

$$\tilde{P}_c = \{p_j \in P : \mu_{\tilde{P}_c}^F(p) \geq \alpha\},$$

$$\tilde{P}_a = \{p_j \in P : \mu_{\tilde{P}_a}^F(p) \geq \alpha\},$$

где $\alpha = [0, 1]$ – пороговое значение, с которым сравнивается значение приближенной функции принадлежности для каждого понятия.

Для того, чтобы найти необходимые документы, можно использовать ключевые слова, которые входят в состав понятий из множеств $\tilde{P}_{ep}, \tilde{P}_{ec}, \tilde{P}_{nc}, \tilde{P}_{nc}, \tilde{P}_c, \tilde{P}_a$:

$$\tilde{F}_{ep} = \{f_i \in F^P \mid \forall i, \exists j : p_j \in \tilde{P}_{ep} \cap m_{ij}^{ep} = 1\},$$

$$\tilde{F}_{ec} = \{f_i \in F^P \mid \forall i, \exists j : p_j \in \tilde{P}_{ec} \cap m_{ij}^{ec} = 1\},$$

$$\tilde{F}_{nc} = \{f_i \in F^P \mid \forall i, \exists j : p_j \in \tilde{P}_{nc} \cap m_{ij}^{nc} = 1\},$$

$$\tilde{F}_{nc} = \{f_i \in F^P \mid \forall i, \exists j : p_j \in \tilde{P}_{nc} \cap m_{ij}^{nc} = 1\},$$

$$\tilde{F}_c = \{f_i \in F^P \mid \forall i, \exists j : p_j \in \tilde{P}_c \cap m_{ij}^c = 1\},$$

$$\tilde{F}_a = \{f_i \in F^P \mid \forall i, \exists j : p_j \in \tilde{P}_a \cap m_{ij}^a = 1\}.$$

Определим каждому документу из хранилища степень возможности включения его в найденные документы:

$$\lambda_{ep}(o_{ij}) = \langle o_{ij} \rangle, \forall j : f_j \in \tilde{F}_{ep},$$

$$\lambda_{ec}(o_{ij}) = \langle o_{ij} \rangle, \forall j : f_j \in \tilde{F}_{ec},$$

$$\lambda_{nc}(o_{ij}) = \langle o_{ij} \rangle, \forall j : f_j \in \tilde{F}_{nc},$$

$$\lambda_{nc}(o_{ij}) = \langle o_{ij} \rangle, \forall j : f_j \in \tilde{F}_{nc},$$

$$\lambda_c(o_{ij}) = \langle o_{ij} \rangle, \forall j : f_j \in \tilde{F}_c,$$

$$\lambda_a(o_{ij}) = \langle o_{ij} \rangle, \forall j : f_j \in \tilde{F}_a,$$

где $\langle o_{ij} \rangle$ – оператор обобщенного усреднения, который вычисляется по формуле:

$$\langle o_{ij} \rangle = \frac{1}{k} \sum_{i=1}^k \frac{1}{s} \sum_{j=1}^s o_{ij},$$

где $k \times s$ – размерность матрицы, определяющей множество документов.

В множество документов, удовлетворяющих запросу пользователя, будем включать документы со значениями:

$$\tilde{O}_{ep} = \{o_i \in O : \lambda_{ep}(o_i) > 0\},$$

$$\tilde{O}_{ec} = \{o_i \in O : \lambda_{ec}(o_i) > 0\},$$

$$\tilde{O}_{nc} = \{o_i \in O : \lambda_{nc}(o_i) > 0\},$$

$$\tilde{O}_{nc} = \{o_i \in O : \lambda_{nc}(o_i) > 0\},$$

$$\tilde{O}_c = \{o_i \in O : \lambda_c(o_i) > 0\},$$

$$\tilde{O}_a = \{o_i \in O : \lambda_a(o_i) > 0\}.$$

В результате мы получаем выборку документов, которые соответствуют запросу пользователя и имеют числовые оценки их вхождения в множества, соответствующие дескрипторам, которые могут быть сформированы в виде списков выдаваемых пользователю.

ЗАКЛЮЧЕНИЕ

Задача информационного поиска представляется в виде разбиения предметной области на дескрипторы, описывающие различные морфологические признаки запросов пользователей. Описанная модель поиска дает возможность рассматривать границы между указанными множествами, соответствующими дескрипторам, как неопределенно выраженные, – это предоставляет возможность сформировать граничные множества документов, соответствующих дескрипторам, и документов, не совсем соответствующих запросу, но которые могут быть задействованы в дальнейшем.

Полученные результаты в дальнейшем можно будет использовать при построении пользовательского интерфейса поисковой системы, который улучшит восприятие и повысит производительность системы. В последующих исследованиях планируется автоматизировать процесс поиска и протестировать полученный программный продукт.

СПИСОК ЛИТЕРАТУРЫ

1. Бернштейн Э.С., Лахути Д.Г., Чернявский В.С. Вопросы теории поисковых систем. – М. : М-во электротехн. пром-сти СССР, ВНИИЭМ, 1966. – 130 с.
2. Стулов А. Особенности построения информационных хранилищ // Открытые системы [Электронный ресурс]. – 2003. – № 4. – URL: <http://www.osp.ru/os/2003/04/182942>.
3. Селяев А.Г. Взвешивание терминов в процессах индексирования электронных информационных ресурсов // Автоматизация процессов управления. – 2007. – № 2 (10). – С. 93–96.
4. Pawlak Z. Rough relations // Pr. IPI PAN. 1981. № 435. P. 10.