

Существуют два достаточных условия разрешимости нечетко сбалансированных триаксиальных задач:

$$а) \tilde{x}_{ijk} = \frac{\tilde{a}_{ij}}{p} + \frac{\tilde{c}_{ik}}{n} + \frac{\tilde{b}_{jk}}{m} - \frac{\tilde{c}_k}{mn} - \frac{\tilde{b}_j}{mp} - \frac{\tilde{a}_i}{np} + \frac{\tilde{S}}{mnp} \gtrsim 0,$$

где

$$\tilde{a}_i = \sum_{j=1}^n \tilde{a}_{ij}, \tilde{b}_j = \sum_{k=1}^m \tilde{b}_{jk}, \tilde{c}_k = \sum_{i=1}^m \tilde{c}_{ik}, \tilde{S} = \sum_{i=1}^m \sum_{k=1}^p \tilde{c}_{ik}$$

выполняется для всех  $(i,j,k) \in E$ ;

$$б) \tilde{y}_{ijk} = \frac{\tilde{a}_{ij}}{\tilde{a}_i \tilde{b}_j} + \frac{\tilde{c}_{ik}}{\tilde{a}_i \tilde{c}_k} + \frac{\tilde{b}_{jk}}{\tilde{b}_j \tilde{c}_k} \gtrsim 0, (i,j,k) \in E.$$

В заключение отметим, что неразрешимость триаксиальной задачи с нечеткими параметрами не означает невозможность транспортировки грузов.

Неразрешимость означает, что не существует плана, компоненты которого неотрицательны. Однако оптимизация все равно имеет смысл и решается, если произвести замену переменных  $\tilde{y}_{ijk} = \tilde{x}_{ijk} + M (M > 0), (ijk) \in E$ .

Задача оптимизации сводится к отысканию набора  $Y = \{ \tilde{y}_{ijk} \}$ , минимизирующего функцию:

$$L(Y) = \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^p \tilde{c}_{ijk} \tilde{y}_{ijk},$$

и удовлетворяющего ограничениям:

$$\begin{aligned} \sum_{k=1}^p \tilde{y}_{ijk} &= \tilde{a}_{ij} + pM, \quad i \in I, \quad j \in J; \\ \sum_{i=1}^m \tilde{y}_{ijk} &= \tilde{b}_{jk} + mM, \quad j \in J, \quad k \in K; \\ \sum_{j=1}^n \tilde{y}_{ijk} &= \tilde{c}_{ik} + nM, \quad i \in I, \quad k \in K; \\ \tilde{y}_{ijk} &\gtrsim M, \quad (ijk) \in E. \end{aligned}$$

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Борисова Е.А., Финаев В.И.. Триаксиальная распределительная задача с нечеткими параметрами / Известия ТРТУ. Тематический выпуск «Интеллектуальные САПР». – Таганрог: Изд-во ТРТУ, 2006, №8(63). – С. 17-21.
2. Дюбуа Д., Прад А. Теория возможностей. / Пер. с французского В.Б.Тарасова. – М.: Радио и Связь, 1990. – 288 с.
3. Раскин Л.Г., Кириченко И.О. Многоиндексные задачи линейного программирования (теория, методы, приложения). – М.: Радио и связь, 1982. – 240 с.

**А.Н. Целых, Э.М. Котов**

#### МЕТОД ВЫДЕЛЕНИЯ ИЗ ТЕКСТОВОГО ДОКУМЕНТА СЕМАТИЧЕСКИ БЛИЗКИХ ПАР ПОНЯТИЙ

Принимая во внимание, что смысловое содержание документа является субъективной характеристикой текста и на сегодняшний день отсутствуют приемлемые математические методы описания смысловой нагруженности текста и его отдель-

ных понятий, возникает необходимость создания информационных средств описания смысла текстовых документов, путем выделения смысловых характеристик из текста на естественном языке, что в дальнейшем должно привести к ведению осмысленного поиска в больших массивах текстовой информации.

Рассмотрим метод выделения из текстового документа семантически близких пар понятий, что даст возможность использования информации о семантической близости понятий для решения задачи построения метода поиска текстовой информации.

Возможно выделить ряд проблем существующих при решении задачи поиска, состоящей в выдаче осмысленного результата со стороны ИПС в ответ на запрос пользователя, отражающий его представление о цели поиска [1]:

1. Морфология языка – слова русского языка имеют различные по написанию формы, но с одинаковым смыслом. Решением может являться реализация в ИПС морфологических анализаторов, включающих БД начальных форм слов и алгоритмы приведения слов, не содержащихся в данных БД к начальной словоформе. Благодаря данному подходу возрастает эффективность поиска в связи с реализацией поиска по всем формам слова запроса.

2. Синонимичность языка – существование множества слов с различным написанием, но общим смыслом. Возможным решением является использование для повышения надежности поиска словарей синонимов ориентированных на различные области знаний.

3. Семантика языка – пользователь ищет не тексты, а информацию, т.е. запрос пользователя состоит из понятий языка характеризующих информацию, которую необходимо найти. Возможным решением задачи учета семантики слов при поиске может являться построение экспертных поисковых систем определенной области знаний, с построением моделей баз знаний учитывающих семантические связи в узкой предметной области. Так же возможно применение механизмов классификации текстов, что, впрочем, дает учет только малой части семантических связей языка.

Рассмотрим возможность получения числовой оценки смысловой близости понятий с помощью построения метрики взаимозаменяемости в пространстве понятий [2]. Дадим следующие определения:

1. Термину «понятие» будем сопоставлять слово языка вместе с его семантикой, т.е. множеством смысла слова.

2. «Представлением» является осмысленное предложение представляющее конкретную реализацию понятия.

3. Каждое представление является реализацией всех понятий, содержащихся в нем.

4. Каждое понятие имеет множество представлений.

5. Множество понятий и их представлений находятся в отношении типа «многие ко многим», т.е. каждый экземпляр одной сущности связан с более чем одним экземпляром другой сущности и при реализации данного типа отношений необходимо ввести промежуточную сущность или сущность-пересечение не имеющую атрибутов, а только отношения.

6. Семантику языка определяют отношения множества понятий и их представлений.

Путем замены любого слова, обозначающего понятие, из предложения на некоторую переменную  $X$ , вместо которой может быть подставлено любое понятие, возможно любое предложение текста преобразовать в семантический предикат. Посредством подстановки понятия вместо переменной  $X$  возможно получить пред-

ложение. При этом значение семантического предиката, представляющего собой функцию, определенную на множестве понятий, при подстановки понятия может быть как семантически истинным, если совпадает с некоторым элементом множества представлений, так и семантически ложным в противном случае.

Пусть  $A$  – множество понятий естественного языка:  $A = \{a_1, a_2, \dots, a_i\}$ ,  $G$  – множество представлений:  $G = \{g_1, g_2, \dots, g_i\}$ , тогда множество семантических предикатов строится последовательной заменой слов, обозначающих понятия в каждом представлении  $g_i$  из  $G$ .

Допустим  $(A, G)$  – абстрактное семантическое пространство.

Можно сказать, что для  $(A, G)$  определено множество семантических предикатов  $P$ , если определена операция подстановки  $\pi$ , отображающая  $P \times A$  на множество  $U \supset G$ , причем для каждого  $g \in G$  существует такой элемент  $p \in P$ , что для некоторого  $a \in A$  выполняется:

$$\pi(p, a) = g. \quad (1)$$

Если для каждого понятия  $a$  построить множество  $Z_a$ , таких предикатов, подстановка в которые дает семантически истинное высказывание:

$$Z_a = \{p \mid p(a) \in G\}, \quad (2)$$

тогда возможно задать метрику взаимозаменяемости на множестве понятий в следующем виде:

$$\mu_{ab} = 1 - \frac{|Z_a \cap Z_b|}{|Z_a \cup Z_b|}. \quad (3)$$

Иными словами, чем больший процент занимают взаимозаменяемые случаи, от их возможных употреблений в языке, то тем меньше метрика взаимозаменяемости  $\mu_{ab}$ .

В этом случае метрика взаимозаменяемости  $\mu_{ab}$  может принимать следующие значения:

1.  $\mu_{ab} = 0$  – для полных синонимов, т.е. в предложении может использоваться любое из слов  $a$  и  $b$ .
2.  $\mu_{ab} = 1$  – в данном случае имеет место невозможность присутствия в одном контексте слов  $a$  и  $b$ .
3.  $\mu_{ab} \in (0, 1)$  – существует взаимозаменяемость понятий лишь на части семантических предикатов.

Таким образом, существует возможность использовать данную метрику  $\mu_{ab}$  для оценки близости семантики понятий, причем уменьшение значения меры  $\mu_{ab}$  отражает возможность взаимозаменяемости в тексте понятий  $a$  и  $b$  и наличие у них общего смысла, что приводит к возможности сравнения для нахождения близких понятий не самих понятий, а контекстов, в которых они встречаются.

Но, наряду с описанной возможностью анализа семантики понятий на основе анализа всех семантически истинных высказываний, возникает ряд проблем практической реализации данного подхода [3]:

- ◆ необходимость получения всех возможных высказываний языка;
- ◆ сложность оценки объема такого рода данных;

- ◆ сложность получения данных, в связи с отсутствием их представления в явном виде;
- ◆ временные затраты связанные с семантическим сравнением понятий, хранением предложений, их обработкой и подсчетом найденных общих и различных;
- ◆ трудность практического применения метода связанная с большим объемом вычислений.

Для решения данных проблем представляется необходимым уменьшить объем обрабатываемых данных, но без существенных потерь информации о смысле понятий языка, и видится возможным рассмотрение частей высказываний, отражающих информацию о смысле понятия, в отличие от рассмотрения семантически и синтаксически законченных предложений, что при сравнительно небольшом объеме обрабатываемой информации может дать сведения о близости понятий.

Но, с другой стороны, в связи с тем, что для выделения семантически законченных фрагментов из предложения существует необходимость обладания мощными средствами синтаксического анализа, выбор предиката описывающего понятие представляет собой определенную проблему.

Возможным решением является использование при построении предиката синтетического подхода, не учитывающего семантику или синтаксис языка. Назовем подобный предикат – синтетическим предикатом.

Синтетический предикат порядка  $r$  для понятия  $x$  напрямую связан с конкретным текстом, в отличие от семантического предиката, и имеет косвенное отношение к языку в целом и строится посредством выделения  $r$  элементов слева и справа от слова  $x$ . В результате, полученные  $2r$  элементов объединяются в предикат.

В данном случае введем новую метрику  $\mu'_{ab}$ , которая вычисляется аналогично метрике  $\mu_{ab}$  при использовании семантического предиката, но при использовании синтетического предиката множество синтетических представлений  $G'$  состоит из всех возможных наборов длины  $2r+1$  в рассматриваемом тексте слов.

При этом по причине фиксации позиции переменной в синтетическом предикате возникают следующие особенности:

1. Множество понятий и множество синтетических представлений находятся в отношении «один ко многим».
2. Множество понятий и множество представлений находятся в отношении «многие ко многим».
3. В семантическом предикате переменная может находиться в любом месте.
4. Понятие в синтетическом предикате может располагаться лишь в середине (на основе процедуры построения синтетического предиката).
5. Множество синтетических предикатов и множество понятий находятся в отношении «многие ко многим», аналогично отношению множества семантических предикатов и множества понятий, что дает возможность сходного построения метрики в пространстве понятий.
6. Отсутствие необходимости хранения позиции переменной упрощает представление синтетического предиката в отличие от семантического предиката.

В итоге, для подсчета семантической близости понятий, можно использовать метрику  $\mu'_{ab}$  и метод численной оценки близости понятий на основе реального текста будет состоять в следующем:

1. Из текста выделяется множество всех содержащихся в нем понятий  $A'$ .
2. Из текста выделяется множество синтетических представлений  $G'$ .

3. Осуществляется выбор из текста множества синтетических предикатов  $P'$ .
4. В связи с имеющимися функциональными зависимостями:  $G' \rightarrow A'$  и  $G' \rightarrow P'$  существует достаточность выделения из текста множества  $G'$ .
5. Множество  $G'$  представляет собой множество всех наборов длины  $2r+1$  элементов текста.
6. Множество  $A'$  получается из множества  $G'$  путем выделения центрального элемента  $\forall g' \in G'$ .
7. Множество  $P'$  получается путем исключения этого центрального элемента из множества  $G'$ .
8. Функция, задающая соответствие между парой любых понятий и вещественным числом из отрезка  $[0,1]$ , осуществляющая преобразование  $A \times A' \rightarrow [0,1]$ , задает численную оценку близости понятий.
9. Для преобразования  $A \times A' \rightarrow [0,1]$  используем метрику взаимозаменяемости при использовании синтетического предиката.
10. Для определения семантической близости понятий  $a$  и  $b$  будем рассматривать множество синтетических предикатов

$$Z_a' = \{p' \in P' \mid p'(a) \in G'\} \quad (4)$$

и множество

$$Z_b' = \{p' \in P' \mid p'(b) \in G'\}. \quad (5)$$

1. С учетом зависимостей  $G' \rightarrow A'$  и  $G' \rightarrow P'$  отметим, что  $Z_a'$  является множеством образов в  $P'$  прообразов элемента  $a$  из  $G'$ .
2. В результате численная характеристика семантической близости понятий  $a$  и  $b$  будет определена по формуле:

$$\mu_{ab}' = 1 - \frac{|Z_a' \cap Z_b'|}{|Z_a' \cup Z_b'|}. \quad (6)$$

Рассмотрим возможность применения семантической близости понятий для поиска текстовой информации.

Пусть для каждой пары понятий  $a$  и  $b$  задана метрика семантической близости  $\mu_{ab}$ . Задача поиска будет состоять в нахождении близкого по смыслу фрагмента текста или отсутствие подобного совпадения для запроса  $S = \{s_1, s_2, \dots, s_n\}$ .

Семантическое расстояние между запросом и некоторой выделенной строкой текста определим с помощью характеристики:

$$R(s, t) = \frac{1}{n} \sum_{i=1}^n \min_{j=1}^m \mu_{s_i, t_j}, \quad (7)$$

где  $n$  – количество слов запроса;

$m$  – количество слов в анализируемой текстовой строке;

$\mu_{s_i, t_j}$  – семантическое расстояние между  $i$ -м словом запроса и  $j$ -м словом анализируемой текстовой строки.

Рассматриваемая характеристика  $R(s, t)$ , определяющая семантическое расстояние, может принимать следующие значения:

$R(s, t) = 0$  – в анализируемой текстовой строке содержатся все слова запроса и расстояние между равными понятиями  $\mu_{aa} = 0$ , т.е. существует точное совпаде-

ние запроса с фрагментом текста, и характеристика  $R(s, t)$  отражает присутствие в тексте всех слов запроса.

$R(s, t) = 1$  – в анализируемой текстовой строке отсутствует любое семантически близкое словам запроса понятие, т.е. семантическое расстояние  $\mu_{s_i, t_j} = 1$  и в данном случае отсутствует совпадение запроса с фрагментом текста.

$R(s, t) < 1$  – в анализируемой текстовой строке присутствуют слова, семантически близкие словам запроса, и множество подобных строк составляют список найденных фрагментов, который ранжируется по степени соответствия запросу по принципу «чем меньше значение  $R(s, t)$ , тем текст точнее соответствует запросу».

Проблема выбора текстовой строки для анализа решается путем выбора в качестве строки-кандидата или всего документа в целом, или последовательного выбора фрагментов текста, например, абзацев, предложений и т.д. Если в результате анализа документа будут выбраны несколько строк  $t_k$ , отражающих смысловую нагруженность текста, то текстовому документу будет поставлена в соответствие строка, имеющая наименьшую характеристику  $R(s, t_k)$ .

Поиску на основе использования характеристики отражающей семантическое расстояние между запросом и некоторой строкой текста присущи следующие особенности:

1. Результаты поиска будут содержать документы близкие по смыслу к понятиям запроса даже при отсутствии в документе слов содержащихся в запросе.
2. Поиск, использующий семантическое расстояние, устойчив к ошибкам пользователя, т.е., если в запрос включено слово из другой предметной области, то документы, содержащие остальные слова запроса, будут включены в результат поиска, но с большими значениями  $R(s, t)$ .

Предложенный метод поиска отличается от традиционных алгоритмов поиска использованием информации о смысловой нагруженности отдельных понятий и связей между ними.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Solton G.*, The Performance of interactive information retrieval, Information processing letters, 1, No. 2, 1971.
2. *Флоренсов А.Н.* Метризация знаний в системах информации: Труды Международной научно-практической конференции KDS-2001 «Знание-Диалог-Решение», Т.1. – СПб.: «Лань», 2001. – С. 608-614.
3. *T. Landauer, P. Foltz, and D. Laham.* An introduction to latent semantic analysis. In *Discourse Processes*, v. 25, pp. 259-284.

**С.В. Скороход**

#### **ПРИМЕНЕНИЕ ФУНКЦИОНАЛЬНЫХ МОДЕЛЕЙ IDEFO ДЛЯ АНАЛИЗА КВАЛИФИКАЦИОННЫХ ХАРАКТЕРИСТИК РАБОЧИХ МЕСТ НА ОСНОВЕ НЕЧЁТКИХ ЦЕЛЕЙ**

**Постановка задачи.** Важным фактором эффективного управления персоналом является оптимальный подбор сотрудников, обеспечивающий получение наилучших результатов на рабочих местах. На начальной фазе такого подбора необхо-