

Квазисемантический поиск текстовых данных. Способы модификации запроса

Предложены формальные модели процедур внесения семантики в поисковый запрос при реализации квазисемантического поиска. Модификация запроса проводится с использованием лингвистической онтологии предметной области. Рассмотрена специфика отображения поискового запроса из поля терминов в поле синсетов онтологии. Предложены и исследованы три стратегии коррекции поискового запроса: горизонтальная, вертикальная и ассоциативная. Проанализировано влияние процедур модификации поискового запроса на поисковый отклик.

Запропоновано формальні моделі процедур внесення семантики у пошуковий запит при реалізації квазісемантичного пошуку. Модифікація запиту проводиться з використанням лінгвістичної онтології предметної галузі. Розглянуто специфіку відображення пошукового запиту із поля термінів в поле синсетів онтології. Запропоновано та досліджено три стратегії корекції пошукового запиту: горизонтальну, вертикальну та асоціативну. Проаналізовано вплив процедур модифікації пошукового запиту на пошуковий відгук.

Ключевые слова: поиск, квазисемантика, лингвистическая онтология, модификация запроса.

Введение

В условиях постиндустриального общества необходимым условием общественно-полезной деятельности практически любого характера становится использование высокоэффективных средств поиска необходимой информации в глобальном электронном информационном пространстве. На сегодняшний день для обеспечения информационных потребностей пользователей всех категорий используются, в основном, традиционные полнотекстовые поисковые машины [1], которые широко представлены в глобальной сети Internet. Отдавая должное весомому вкладу таких поисковых систем в процесс усовершенствования информационной деятельности, необходимо отметить их основной недостаток, а именно: отсутствие возможности непосредственного учета семантики в поисковом процессе. В тоже время нужно заметить, что, с точки зрения перспектив реального прорыва в области информационного поиска, классическая концепция полнотекстового поиска на сегодняшний день практически себя исчерпала [2, 3], и

сегодня необходимо использование высокоинтеллектуальных методов доступа к распределенным по глобальной компьютерной сети данным и знаниям [4, 5]. Однако, развитие собственно семантических машин происходит недостаточно быстро в связи с высокой сложностью практической реализации эффективного семантического анализа естественно-языковых текстов [6].

Именно поэтому существенный теоретический и практический интерес представляет концепция квазисемантической поисковой системы, согласно которой привнесение смысловой компоненты в процесс поиска осуществляется собственноручно пользователем путем итерационной модификации начальной версии запроса, с использованием интеллектуального редактора запроса, который построен на основе лингвистической онтологии предметной области [7]. Реализация процесса функционирования такого редактора, по аналогии с экспертной системой [8], позволяет анализировать содержание, фактически заложенное пользователем в текущий поисковый запрос, и предлагать на выбор такую модификацию последнего, которая бы в конечном итоге позволяла найти именно те информационные объекты, которые максимально отвечают настоящему поисковому интересу пользователя.

Согласно упомянутой концепции, для формирования и модификации поискового запроса пользователя могут применяться следующие три основные процедуры, которые используют логическое отображение поискового запроса в поле понятий онтологии.

А) Процедура горизонтальной коррекции запроса.

Выполняется в три этапа: поиск понятий онтологии, которые отвечают терминам запроса; разрешение неоднозначности многозначных терминов при участии пользователя на основе определений соответствующих понятий; расширение запроса за счет синонимических рядов каждого из понятий. Эта процедура позволяет представить поисковый запрос, изначально заданный терминами естественного языка, в виде понятий, имеющих вполне определенное значение, адекватное поисковым интересам пользователя. В последующих процедурах используется именно такое понятийное представление поискового запроса. При необходимости такое представление можно легко транслировать в форму терминов, уже более точно учитывая смысл понятий, который они выражают.

Б) Процедура вертикальной коррекции поискового запроса.

Характерной чертой любой онтологии является ее иерархическая упорядоченность [9]. Понятия объединяются в общую структуру, представляющую собой дерево, в корне которого находится наиболее общее понятие. Чем ниже спускаться по иерархии онтологии, тем более конкретными будут понятия. Используя эту особенность, можно

предложить модификацию поискового запроса путем навигации по вертикальным связям онтологии и включения в запрос понятий, которые логически связаны с терминами первоначальной версии запроса, но более или менее конкретными в зависимости от информационного интереса пользователя. Таким образом, можно получить соответственно более конкретные или более общие результаты в поисковой выборке, по сравнению с поисковым отзывом на немодифицированный поисковый запрос.

В) Процедура ассоциативной коррекции поискового запроса.

Эта процедура призвана среди всех понятий онтологии отобрать для модификации поискового запроса лишь те, которые потенциально близки по смыслу к поисковому запросу с учетом контекста. Такие понятия будут составлять своеобразный семантический срез онтологии по поисковому запросу. Связи онтологии в данном случае рассматриваются как ассоциации между объектами, а сама онтология – как многомерная семантическая сеть. Найденный семантический срез может быть выборочно использован пользователем для модификации поискового запроса с учетом возможного уточнения информационного интереса.

Ввиду вышеизложенного, целью данной работы является формализация ключевых процедур квазисемантического поиска текстовых данных, которая направлена на обеспечение высокой эффективности программных реализаций. Среди задач исследования можно выделить следующие:

- формализация описания информационной инфраструктуры, на которой базируется квазисемантический поиск текстовых данных;
- разработка математической интерпретации процедур формирования, коррекции и реализации поискового запроса в процессе квазисемантического поиска;
- исследование влияния процедур коррекции поискового запроса на характер результатов поисковой выборки.

В ходе исследования будем считать, что в общем случае поисковый индекс текстовых информационных объектов может быть как классическим, то есть, организованный на основе терминов естественного языка [10], так и семантически-ориентированным или концепт-индексом, то есть, построенным на основе понятий [11]. Поэтому в данной работе на формально-логическом уровне рассматривается и сравнивается реализация квазисемантического поиска в текстоориентированных базах и хранилищах данных, проиндексированных согласно обоих упомянутых подходов. При этом особенности программно-технического воплощения поиска информационных объектов подробно не

рассматриваются, поскольку этот вопрос должен исследоваться в рамках отдельной задачи построения соответствующей поисковой машины.

1. Особенности инфраструктуры квазисемантического поиска

Для анализа всех процедур квазисемантического поиска необходимо ввести формальное описание данных, с которыми эти процедуры взаимодействуют. Поэтому следует определить основные понятия: служебный лингвистический ресурс, который используется в алгоритмах поиска, множество текстовых информационных объектов (документов) и поисковый запрос.

В качестве служебного лингвистического ресурса для реализации квазисемантического поиска будем использовать лингвистическую онтологию. Согласно [9], лингвистический подход к формированию и исследованию баз знаний онтологического типа основан на изучении значительных массивов (корпусов) естественного языка. Получаемая таким образом лингвистическая онтология проецирует соответствующую область знаний в естественно-языковое поле с одной стороны, и в систему семантических отношений - с другой. Согласно [12], онтология может быть представлена тройкой

$$O = \langle X, R, \Phi \rangle, \quad (1)$$

где X - это конечное множество концептов (понятий). В контексте рассмотрения процедур квазисемантического поиска под множеством концептов X будем понимать конечное множество синсетов $S = \{s_i : i = \overline{1, I}\}$, где I - количество синсетов в онтологии. Синсет - это уникальная смысловая единица, которая интегрирует в себе набор близких синонимов, выраженных терминами (словами или словосочетаниями) естественного языка, а также однозначное определение, позволяющее отличить его от других синсетов. Таким образом, каждый синсет s описывается двойкой

$$s = \langle T^s, Desc \rangle, \quad (2)$$

где $Desc$ - дескриптор, описательное представление понятия, которое выражает синсет, T^s - конечное непустое множество терминов-синонимов, которые составляют синсет. Каждый синсет $s \in S$ отображается в поле терминов

$$T^s = \{t_0, t : t_0 \in T, t \in T, Syn(t_0, t)\}, \quad (3)$$

где T - множество терминов естественного языка, t_0 - основной термин синсета (обычно определяется как наиболее употребляемый термин, среди тех, которые определяют соответствующее понятие), t - остальные термины, которые выражают понятие s , а Syn - функция синонимии двух терминов. Далее под синсетом s , если не указано иного, будем

понимать его отображение в поле терминов T^s . В общем случае $s_i \cap s_j \neq \emptyset$, при $i \neq j$, поскольку вследствие омонимии произвольный термин может одновременно принадлежать синсетам s_i и s_j .

R – это конечное множество семантических отношений. В состав R могут входить семантические отношения любых видов, но в общем случае для задачи квазисемантического поиска будем различать два их вида $R = \{r_h, r_a\}$, где r_h – отношение гиперонимии, а r_a – отношение ассоциации. Отношение r_h в связке с множеством синсетов в нашем случае определяет таксономию понятий онтологии.

Φ – конечное множество функций интерпретации. Содержание этого множества в контексте проблематики квазисемантического поиска определим как функцию интерпретации отношений онтологии, определяющую вес семантического отношения w между синсетами. Чем больше значение веса, тем сильнее выражена соответствующая семантическая связь между двумя понятиями.

Выбор данного типа онтологии в качестве служебного лингвистического ресурса определен следующими соображениями:

- эффективность лингвистической онтологии в качестве универсального средства описания понятийного аппарата различных предметных областей;
- широкая распространенность подобных онтологий (например, англоязычная лингвистическая онтология WordNet и ее локализации для других естественных языков [9]) и возможность приведения к данному виду других лингвистических баз знаний онтологического типа;
- сравнительная простота автоматической машинной обработки лингвистической онтологии в случае использования последней на различных этапах информационного поиска;
- потенциальная возможность автоматизации процедур синтеза лингвистической онтологии (или наращивания уже существующей) за счет применения программного инструментария интеллектуального анализа текстовых массивов.

Множество документов рассматривается как конечное подмножество коллекции естественно-языковых текстовых информационных объектов $\{d_j : j = \overline{1, J}\} \subseteq D$, где J – количество документов в коллекции. Каждый документ d из коллекции текстовых информационных объектов D классически индексируется конечным множеством терминов $\{t_g : g = \overline{1, G}\} \subset T$, где G – количество терминов, с помощью которых индексируется документ. Однако квазисемантический подход к поиску текстовых

информационных объектов предусматривает также возможность индексации множеством синсетов $\{s_i : i = \overline{1, I_d}\} \subset S$, где I_d - количество синсетов, с помощью которых индексируется документ.

В первом случае

$$T^d = \{t : t \in T, F^t(t) > 0\}, \quad (4)$$

где T^d - конечное множество терминов естественного языка, которые описывают (индексируют) документ d , $T^d \subset T^D$, где T^D - множество терминов полнотекстового индекса общей коллекции документов D ; $F^t(t)$ - функция присутствия термина в документе, который индексируется. В самом простом случае эта функция принимает ненулевой результат, когда термин t присутствует в документе: $\{t\} \cap T^d \neq \emptyset$, где $t \in T$, то есть $\exists t \in T : t \in T^d$.

Во втором случае

$$S^d = \{s : s \in S, F^s(s) > 0\}, \quad (5)$$

где S^d - конечное множество синсетов онтологии, которые описывают (индексируют) документ d , $S^d \subset S^D$, где S^D - множество синсетов концепт-индекса полной коллекции документов D . $F^s(s)$ - функция семантической принадлежности синсета документу, который индексируется. В самом простом случае эта функция принимает ненулевое значение, если $s \cap T^d \neq \emptyset$, где $s = \{t : t \in T\}$, то есть $\exists t \in T : t \in T^d, t \in s$.

Поисковый запрос – это формализованный способ выражения информационного интереса пользователя. Поэтому чаще всего в начальной форме поисковый запрос дается на естественном языке. В общем случае поисковый запрос пользователя можно рассматривать как

$$Q^t = \{t : t \in T\}, \quad (6)$$

где Q^t - поисковый запрос, который выражается на конечном множестве терминов естественного языка.

Большинство текстовых поисковых систем позволяют подавать поисковый запрос в виде логического выражения. Основные логические операторы, которые при этом используются, - *AND, OR, NOT*.

Запрос вида $Q = t_i \text{ AND } t_j, i \neq j$, как результаты поиска выдает лишь те документы $\{d_k : k = \overline{1, K}\} \subseteq D$, где K - количество выбранных документов, которые содержат и термин t_i , и термин t_j одновременно.

Запрос вида $Q = t_i OR t_j, i \neq j$, как результат поиска выдает объединение выборок документов по термину t_i и термину t_j .

Запрос вида $Q = NOT t_i$ в качестве результатов поиска выдает лишь те документы $\{d_k : k = \overline{1, K}\} \subseteq D$, где K - количество выбранных документов, которые не содержат t_i . Но фактически итоговая выборка документов по такому поисковому запросу может быть получена путем реализации дополнения множества документов, которые содержат термин t_i до общей коллекции документов D .

Учитывая тот факт, что в большинстве случаев поисковый запрос формулируется на естественном языке, и в соответствии с классическим синтаксисом запросов поисковых систем рассматривается как термины, которые объединены оператором AND [13], поисковый запрос далее будем считать множеством терминов естественного языка, каждый из которых должен входить в искомый текстовый информационный объект. В случае более сложного запроса с вхождением операторов OR или NOT , весь запрос может быть разбит на подзапросы с последующим объединением результатов или исключением лишних в соответствии с семантикой конкретного логического оператора. Поэтому для упрощения формального представления процедур квазисемантического поиска далее в статье запрос будет рассматриваться как набор терминов, связанных оператором AND . В случае необходимости все поданные в статье выкладки могут быть формально получены для любого варианта сложного поискового запроса.

2. Формализация процедуры проекции поискового запроса на онтологию и горизонтальной коррекции запроса

Процедуры квазисемантического поиска работают не с терминами, а с понятиями, или в контексте онтологии – с синсетами. Поэтому для их работы вначале нужно связать исходный поисковый запрос Q^t с онтологией. Введем функцию отображения терминов запроса в множество синсетов

$$\varphi(t) : Q^t \rightarrow Q^s, \quad (7)$$

где $Q^s = \{s : s \in S\}$ - поисковый запрос, выраженный конечным множеством синсетов онтологии предметной области, $\varphi(t)$ - отображение множества терминов запроса в множество синсетов. Правило отображения можно сформулировать следующим образом:
 $\forall t \in Q^t \rightarrow s \in S : t \in s$.

Запрос, поданный в виде Q^s , позволит:

а) в случае, когда документы коллекции D описываются концепт-индексом S^D , выполнять поиск результатов, определив функцию подобия документа и запроса, как

пересечение множеств $Q^s \cap S^d$, то есть выполнять поиск на уровне концептов, а не на уровне терминов;

б) в случае, когда документы коллекции D описываются множеством T^D , для выполнения процедуры поиска использовать не начальный запрос Q^t , а модифицированный $Q^{t'}$, полученный путем преобразования Q^s , а именно:

$$Q^{t'} = \bigcup_j (s_j : s_j \in Q^s) = \bigcup_j (\bigcup_i (t_i : t_i \in s_j) : s_j \in Q^s). \quad (8)$$

Таким образом, $Q^{t'}$ в соответствии с (3) будет автоматически расширен за счет синонимических связей (возможно, с последующей селекцией пользователем), а поиск документов будет происходить с использованием функции подобия, в основе которой будет лежать пересечение множеств терминов модифицированного запроса и множества терминов каждого документа $Q^{t'} \cap T^d$.

Чтобы определить влияние процедуры связывания поискового запроса с онтологией на результаты поиска по классическому полнотекстовому индексу, то есть, когда процедура поиска функционирует на уровне терминов, рассмотрим следующую теорему.

Теорема 1. Использование поискового запроса $Q^{t'}$, полученного за счет использования синонимических связей онтологии предметной области, во время поиска в термин-ориентированном индексе увеличивает полноту поискового отклика в сравнении с немодифицированным запросом Q^t .

Доказательство. Пусть начальный запрос состоит из одного термина $Q^t = \{t_q\}$, где $t_q \in T$. Тогда поисковый отклик будет формироваться на основе пересечения индекса коллекции документов T^d с запросом следующим образом

$$T^d \cap Q^t = T^d \cap \{t_q\}. \quad (9)$$

Обозначим множество документов поискового отклика по термину t_q как D_1 , а множество релевантных документов в отклике как $D_1^{relevant} \subseteq D_1$. Согласно [14] полнота поискового отклика рассчитывается как

$$Recall = \frac{|D_{relevant_item_retrieved}|}{|D_{relevant_items}|}, \quad (10)$$

где $D_{relevant_item_retrieved}$ - множество полученных в поисковом отклике релевантных документов, $D_{relevant_items}$ - общее множество релевантных документов в коллекции.

Для начального запроса Q^t полнота $Recall_1$ будет иметь вид

$$Recall_1 = \frac{|D_1^{relevant}|}{|D_{relevant_items}|}. \quad (11)$$

После преобразования (7), Q^s будет иметь вид $Q^s = \{s_q\}$, где $s_q \in S$ и $s_q = \{t_0, t_1, \dots, t_q, \dots, t_n\}$, в соответствии с (3), где n - количество синонимов для t_0 в синонимическом ряду синсета s_q .

Исходя из (8), получим Q^t

$$\begin{aligned} Q^t &= \bigcup_j (s_j : s_j \in Q^s) = \bigcup_j (\bigcup_i (t_i : t_i \in s_j) : s_j \in Q^s) = \\ &= \bigcup_i (t_i : t_i \in s_q) = (t_0 \cup t_1 \cup \dots \cup t_q \cup \dots \cup t_n). \end{aligned} \quad (12)$$

Поисковый отклик будет формироваться на основе пересечения индекса коллекции с запросом следующим образом

$$\begin{aligned} Q^t \cap T^D &= (t_0 \cup t_1 \cup \dots \cup t_q \cup \dots \cup t_n) \cap T^D = \\ &= (t_0 \cap T^D) \cup (t_1 \cap T^D) \cup \dots \cup (t_q \cap T^D) \cup \dots \cup (t_n \cap T^D). \end{aligned} \quad (13)$$

Каждое выражение $(t_i \cap T^D)$ дает некоторую выборку документов, назовем ее D_i , а соответствующие множества релевантных документов $D_i^{relevant} \subseteq D_i$. Для модифицированного запроса Q^t полнота $Recall_2$ будет рассчитываться как

$$Recall_2 = \frac{|D_0^{relevant} \cup D_1^{relevant} \cup \dots \cup D_q^{relevant} \cup \dots \cup D_n^{relevant}|}{|D_{relevant_items}|}. \quad (14)$$

Поскольку все термины $t_0, t_1, \dots, t_q, \dots, t_n$ являются синонимами, то общее количество релевантных документов по поисковому запросу будет равно мощности объединенного множества релевантных документов по каждому из терминов.

Очевидно, что $Recall_2 \geq Recall_1$. Следовательно, полнота поискового отклика по модифицированному поисковому запросу в общем случае возрастает, в сравнении с поисковым запросом по немодифицированному запросу, что и нужно было доказать. Результат доказательства теоремы несложно расширить на запрос, который состоит из m слов, где $m \geq 1$.

В процедуре $\varphi(t)$, где происходит преобразование Q^t в Q^s , учитывая (3), возможна ситуация, когда $t \in Q^t$ удовлетворяет двум и более неравенствам одновременно

$$\left. \begin{array}{l} t \cap s_i \neq \emptyset \\ t \cap s_j \neq \emptyset \\ \dots \\ t \cap s_k \neq \emptyset \end{array} \right\} \quad (15)$$

где $i \neq j \neq \dots \neq k$.

Теорема 2. Если в начальном поисковом запросе Q^t присутствует термин, который имеет несколько значений, то без разрешения проблемы омонимии при отображении множества терминов запроса во множество синсетов точность поисковой выборки документов по запросу Q^t уменьшится.

Доказательство. Пусть $Q^t = \{t^b\}$, где $t^b \in T$ - многозначный термин, или, иначе говоря, верным является выражение (15).

Тогда в соответствии с правилом отображения (7)

$$Q^s = \{s_i, s_j, \dots, s_k\}, \text{ где } i \neq j \neq \dots \neq k. \quad (16)$$

Таким образом, модифицированный поисковый запрос, который будет использован для поиска документов, сформированный по формуле (8), буде иметь следующий вид

$$Q^t_1 = \bigcup_l (s_l : s_l \in Q^s) = s_i \cup s_j \cup \dots \cup s_k = (t : t \in s_i) \cup (t : t \in s_j) \cup \dots \cup (t : t \in s_k). \quad (17)$$

Поисковая выборка документов из коллекции D по запросу Q^t_1 будет формироваться как пересечение индекса документов и запроса следующим образом

$$\begin{aligned} Q^t_1 \cap T^D &= ((t : t \in s_i) \cup (t : t \in s_j) \cup \dots \cup (t : t \in s_k)) \cap T^D = \\ &= ((t : t \in s_i) \cap T^D) \cup ((t : t \in s_j) \cap T^D) \cup \dots \cup ((t : t \in s_k) \cap T^D). \end{aligned} \quad (18)$$

Поскольку пользователь при задании поискового запроса под термином t имел в виду одно конкретное значение, то будем считать, что это значение фиксируется за синсетом s_i . Тогда согласно процедуре разрешения омонимии

$$Q_s = \{s_i\}, \quad (19)$$

тогда как модифицированный поисковый запрос имеет вид

$$Q^t_2 = \bigcup_l (s_l : s_l \in Q^s) = s_i = \{t : t \in s_i\}. \quad (20)$$

И, соответственно, отбор документов будет происходить по формуле

$$Q^t_2 \cap T^D = (t : t \in s_i) \cap T^D. \quad (21)$$

По определению из [13] точность поиска определяется как

$$Precision = \frac{|D_{relevant_item_retrieved}|}{|D_{retrieved_items}|}, \quad (22)$$

где $D_{retrieved_items}$ - общее множество полученных документов.

Каждое выражение $(t : t \in s_l) \cap T^D$, где $l \in \{i, j, \dots, k\}$, дает определенную выборку документов D_l . Пусть $D_l^{relevant} \subseteq D_l$ - множество релевантных поисковым интересам пользователя документов. Тогда расчет точности для случая Q^t_1 будет, соответственно, следующим

$$Precision_1 = \frac{|D_i^{relevant} \cup D_j^{relevant} \cup \dots \cup D_k^{relevant}|}{|D_i \cup D_j \cup \dots \cup D_k|}. \quad (23)$$

Поскольку в запросе лишь значение синсета s_i отвечает поисковым интересам пользователя, то $|D_i^{relevant}| \geq 0$, а $|D_j^{relevant}| = \dots = |D_k^{relevant}| = 0$.

Отсюда точность можно представить как

$$Precision_1 = \frac{|D_i^{relevant} \cup D_j^{relevant} \cup \dots \cup D_k^{relevant}|}{|D_i \cup D_j \cup \dots \cup D_k|} = \frac{|D_i^{relevant}|}{|D_i \cup D_j \cup \dots \cup D_k|}. \quad (24)$$

Для случая Q^t_2 расчет точности будет иметь следующий вид

$$Precision_2 = \frac{|D_i^{relevant}|}{|D_i|}. \quad (25)$$

Очевидно, что $Precision_2 \geq Precision_1$. Таким образом, без разрешения омонимии при отображении $Q^t \rightarrow Q^s$ точность поиска будет снижаться, что и нужно было доказать.

Исходя из последней теоремы, при наличии в начальном запросе многозначных терминов, обязательно должна быть разрешена проблема омонимии, как это описано в [7], при этом пользователем фиксируется определенный синсет $s \in \{s_i, s_j, \dots, s_k\}$, где s - синсет, который наиболее отвечает поисковым интересам самого пользователя. Принятие решения в отношении фиксации определенного значения s основывается на описательной части синсета $Desc$ из (2), которая предлагается пользователю интеллектуальным редактором запроса [8]. Таким образом, на выходе процедуры отображения $\varphi(t)$ будем иметь однозначное преобразование Q^t в Q^s .

3. Формализация процедуры вертикальной коррекции поискового запроса

В основе онтологии предметной области, наподобие WordNet [9], лежит упорядоченная иерархия синсетов. Ее базовую структуру можно рассматривать как дерево, в корне которого находится наиболее общий синсет, выражающий наиболее

абстрактное понятие для данной онтологии, а с каждым следующим уровнем синсеты конкретизируются. Таким образом, терминальные узлы дерева будут наименее общими, описывая соответственно очень узкие понятия предметной области. Исходя из этой особенности структуры онтологии, в [7] было предложено для модификации начального поискового запроса использовать специальный модуль редактора запроса, который предназначен для навигации по иерархии понятий онтологии. Этот модуль позволяет провести визуализацию части дерева синсетов и, используя иерархичность отношений между ними, дать возможность пользователю изменять уровень абстракции запроса, передвигаясь по иерархии вниз или вверх, что непосредственно влияет на результирующую поисковую выборку документов.

Данная процедура квазисемантического поиска использует отношение $r_h \in R$ онтологии для модификации поискового запроса Q^s . Рассмотрим фрагмент иерархии некой онтологии, приведенный на рис. 1.

Рис. 1

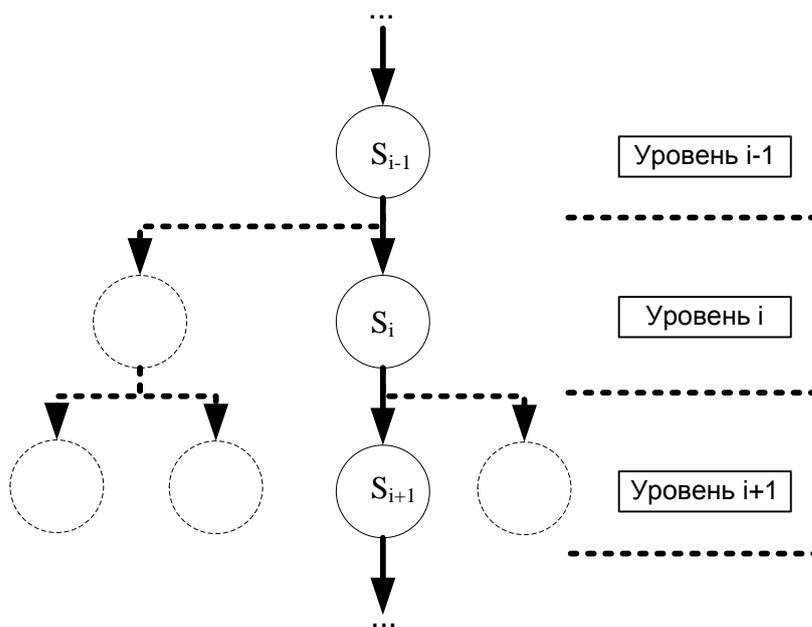


Рис. 1. Фрагмент иерархии онтологии

Пусть, для наглядности в данном примере, номер синсета соответствует уровню общности или абстрактности понятия. Чем ниже уровень, тем более общим (абстрактным) является понятие. Поскольку на каждой итерации процедуры модификации поискового запроса механизм навигации может быть использован лишь относительно одного из понятий начального поискового запроса, пусть после выполнения процедуры преобразования (7) получим $Q^s = \{s_i\}$. Согласно правилам выполнения навигации по иерархии, модификация запроса происходит за счет перехода от понятия s_i к более

общему понятию s_{i-1} , или к более конкретному s_{i+1} . Q^s при этом может быть модифицирован соответственно следующим образом

$$Q^s' = (Q^s \setminus \{s_i\}) \cup \{s_{i-1}\}, \quad (26)$$

$$Q^s' = (Q^s \setminus \{s_i\}) \cup \{s_{i+1}\}. \quad (27)$$

И (26) и (27) подразумевают модификацию поискового запроса за счет синсета (s_{i-1} или s_{i+1}), путем замены текущего синсета s_i в запросе новым (s_{i-1} или s_{i+1} соответственно). Таким образом, выбор любой из стратегий приведет к вхождению нового синсета s_{i-1} или s_{i+1} в состав поискового запроса в зависимости от того, куда будет двигаться во время навигации по дереву онтологии пользователь – в направлении гиперонимизации (генерализации) или в направлении гипонимизации (конкретизации).

Поскольку, согласно (4) и (5), возможны два варианта поискового индекса документов – классический полнотекстовый индекс и концепт-индекс, когда документы индексируются не терминами, а синсетами, необходимо рассмотреть механизм работы квазисемантической процедуры вертикальной модификации для каждого из случаев.

Для случая, когда документы индексируются согласно (4), процесс поиска документов происходит на основе терминов. Гипонимизация синсета $s_i \in Q^s$ за счет синсета s_{i+1} , согласно условиям формирования иерархических связей в онтологии, будет конкретизировать термины $t \in s_i$ из запроса терминами $t \in s_{i+1}$. Таким образом, пересечение множеств индексов документов T^D и терминов синсета s_{i+1} в составе модифицированного поискового запроса будет давать гипонимизированную, с более семантически узкими рамками, выборку документов, нежели пересечение того же множества T^D и терминов синсета s_i . Иными словами, поскольку полнотекстовый индекс документов в общем случае строится без учета связей онтологии, иерархическое отношение между терминами синсетов прямо не отображается на таком индексе. Если термин гипонима будет более употребляемым в контексте данной коллекции документов, то, в соответствии с правилами функционирования полнотекстовой поисковой системы, количество документов по такому модифицированному запросу будет большей, чем по немодифицированному. Однако, такая выборка в большей мере будет отвечать поисковым интересам пользователя. Таким образом, гипонимизированная поисковая выборка характеризуется повышенным уровнем пертинентности по отношению к поисковым потребностям пользователя.

Из этого вытекает и обратное свойство: если провести модификацию поискового запроса путем гиперонимизации синсета, поднимаясь по иерархии онтологии вверх, то пересечение множества T^D и терминов синсета s_{i-1} будет давать соответственно гиперонимизированную поисковую выборку документов, которая будет включать документы, которые будут отвечать более общей интерпретации начального немодифицированного поискового запроса.

Если же рассматривать вариант поисковой системы на основе концепт-индекса, где индексация документов проводится при помощи синсетов онтологии [11], согласно (5), то и процедура поиска документов в таком случае будет происходить на основе синсетов без перехода на уровень терминов. Поскольку процедура создания такого индекса с самого начала тесно связана с использованием онтологии предметной области, семантика связей между синсетами может непосредственно влиять на результаты поиска. Так отношения гипоним/гипероним может быть использовано при формировании результатов поиска, когда полная выборка по синсету, согласно требованиям пользователя, включает в свой состав выборку по самому синсету, объединенную с выборкой по его дочерним элементам. Выборка по дочерним элементам может иметь определенные ограничения по глубине. Ранг результатов, которые выбираются по дочерним синсетам, может уменьшаться соответствующим коэффициентом, который зависит от глубины (количества переходов) соответствующей сложенной иерархической связи. Кроме того, присоединение результатов по дочерним синсетам, вероятно, целесообразно выполнять только на достаточно глубоких ветках иерархии, где иерархическая связь проявляется сильнее. Конкретное решение по этому вопросу должно приниматься при проектировании реальной поисковой системы.

Теорема 3. При условии выполнения поиска результатов в концепт-индексе модификация квазисемантического поискового запроса путем гипонимизации синсета согласно $Q^s' = (Q^s \setminus \{s_i\}) \cup \{s_{i+1}\}$ обеспечивает выборку документов более конкретного характера, а модификация путем его гиперонимизации в соответствии с $Q^s' = (Q^s \setminus \{s_i\}) \cup \{s_{i-1}\}$ обеспечивает ее более абстрактный характер.

Доказательство. Пусть D_i - выборка документов лишь по синсету s_i , а D_{i-1} и D_{i+1} соответственно только по синсетам s_{i-1} и s_{i+1} , а $D(s)$ - множество полной выборки документов по некому синсету s с учетом автоматического включения в запрос соответствующих поддеревьев. Тогда,

$$\left. \begin{aligned} D(s_{i+1}) &= D_{i+1} \\ D(s_i) &= D_i \cup D(s_{i+1}) = D_i \cup D_{i+1} \\ D(s_{i-1}) &= D_{i-1} \cup D(s_i) = D_{i-1} \cup D_i \cup D_{i+1} \end{aligned} \right\} \quad (28)$$

Таким образом, модифицируя текущий запрос синсетом s_{i+1} (другими словами, меняя синсет из текущего запроса на его гипоним), от результирующей выборки документов автоматически отсекается та часть документов D_i , которая соответствовала лишь синсету s_i , увеличивая концентрацию документов, пертинентных поисковому интересу пользователя.

И наоборот, модификация запроса синсетом s_{i-1} , т.е. используя гипероним синсета из текущего запроса, обеспечивает расширение результирующей выборки за счет документов D_i , содержание которых может потенциально отвечать поисковым интересам пользователя, но в более абстрактном или общем виде благодаря гиперониму синсета s_i , что и нужно было доказать.

Согласно Теореме 3, механизм вертикальной модификации поискового запроса за счет навигации по иерархическим связям онтологии, при поддержке редактора запроса, дает возможность управлять качественными (а иногда и количественными) характеристиками поисковой выборки по модифицированному поисковому запросу. При высоком уровне разрозненности результатов поиска или его большом объеме пользователь имеет возможность сузить поле поиска и соответственно выделить пертинентные документы. Когда коллекция документов не дает ожидаемого отклика (небольшое количество результатов), у пользователя есть возможность ослабить условия поиска, расширив итоговое множество документов, которые соответствуют поисковому запросу.

4. Формализация процедуры ассоциативной коррекции поискового запроса

Основная цель этой процедуры квазисемантического поиска – предложить пользователю возможные варианты нелинейной модификации поискового запроса на основе ассоциативных связей между синсетами. В этом случае онтология может рассматриваться как семантическая сеть понятий (рис.2). Выделенные из текущего поискового запроса понятия фиксируются в этой сети, и выполняется процедура поиска других синсетов онтологии, которые семантически близки сразу всем или части понятий из запроса, то есть, находится их семантическое окружение или семантический срез онтологии. Варианты модификации – это взвешенное множество синсетов онтологии, которые с определенной пропорциональностью семантически близки множеству синсетов из текущего поискового запроса. Запрос при этом рассматривается как отображение терминов запроса во множестве синсетов Q^s . Пусть, для примера, на рис. 2 это синсеты

Рис. 2

$\{s_1, s_2, s_3\}$. А семантическая близость определяется по определенному алгоритму на основе семантических связей в онтологии.

Нелинейность модификации поискового запроса достигается за счет того, что семантический срез может представлять собой области онтологии, напрямую не связанные с понятием из начального запроса, но вследствие активизации ассоциативных связей между синсетами могут выявлять скрытые пути модификации поискового запроса в рамках поисковых интересов пользователя.

Пусть $Q^s = \{s_1, \dots, s_n\}$, где n - количество синсетов в поисковом запросе. Для решения задачи поиска среза онтология рассматривается как семантическая сеть, поэтому во внимание берутся все связи r_h и r_a . Причем, каждый тип связи может иметь собственный вес - w_h и w_a соответственно. Более сложный и обобщенный вариант ассоциативной коррекции возможен в случае, когда каждое конкретное отношение между двумя синсетами s_i и s_j онтологии в рамках определенного типа отношения определяется собственным весом w_{ij} (рис.2). Значение этого веса зависит от силы семантической связи между соответствующими понятиями в онтологии.

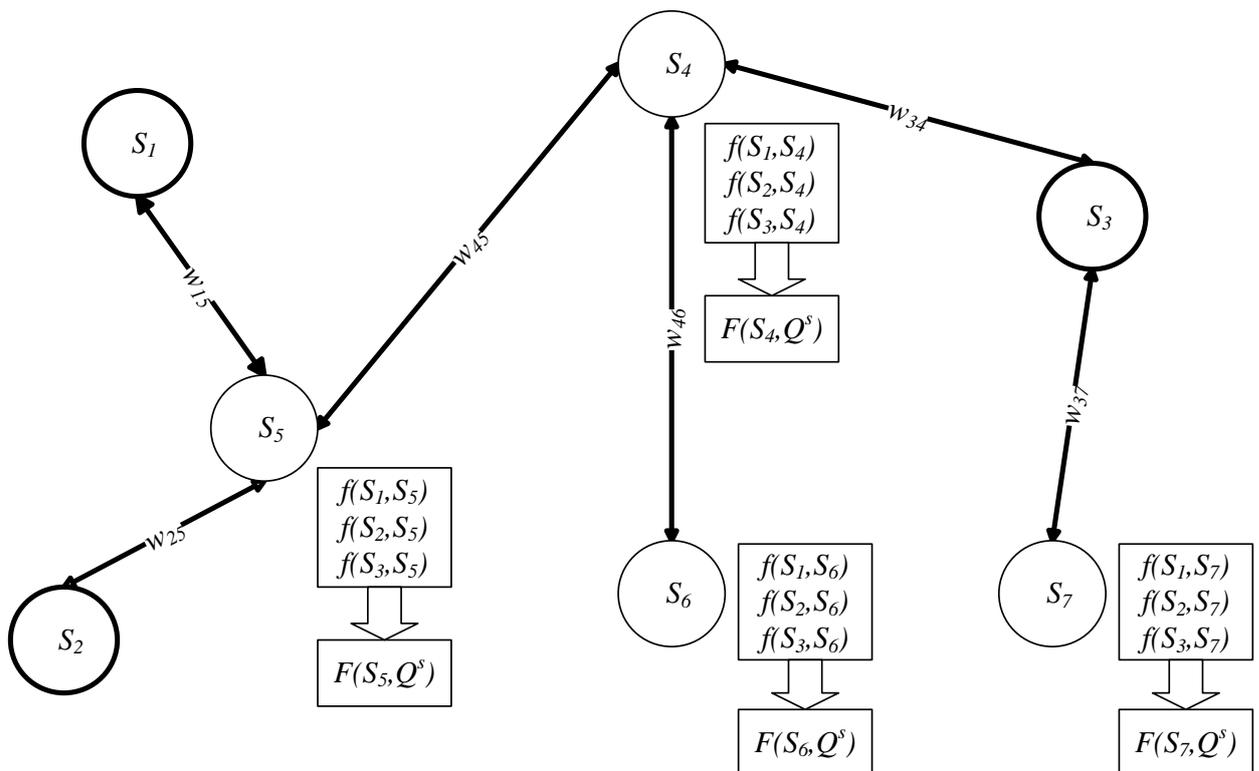


Рис.2. Процедура поиска семантического среза онтологии

Процедуру нахождения семантического среза можно рассматривать как последовательность трех этапов (рис. 3).

Рис. 3

1) Для каждого $s_i \in Q^s$ и каждого $s_j \in S$, где $i \neq j$, рассчитывается функция семантической близости синсетов $f(s_i, s_j)$. Известен ряд подходов к решению этой задачи. Так, в работе [15] рассмотрены пять различных методик поиска семантической близости двух понятий на основе связей онтологии. В контексте квазисемантического поиска возможным является использование только одной из них, а именно:

$$f(s_i, s_j) = C - path(s_i, s_j) - d \cdot k, \quad (29)$$

где C - наибольший возможный радиус семантического влияния узла s_i , $path(s_i, s_j)$ - самая короткая длина пути от s_i к s_j , d - количество смен типов семантического отношения (так называемых поворотов) на пути от s_i к s_j , k - некая константа, которая определяется экспериментальным путем.

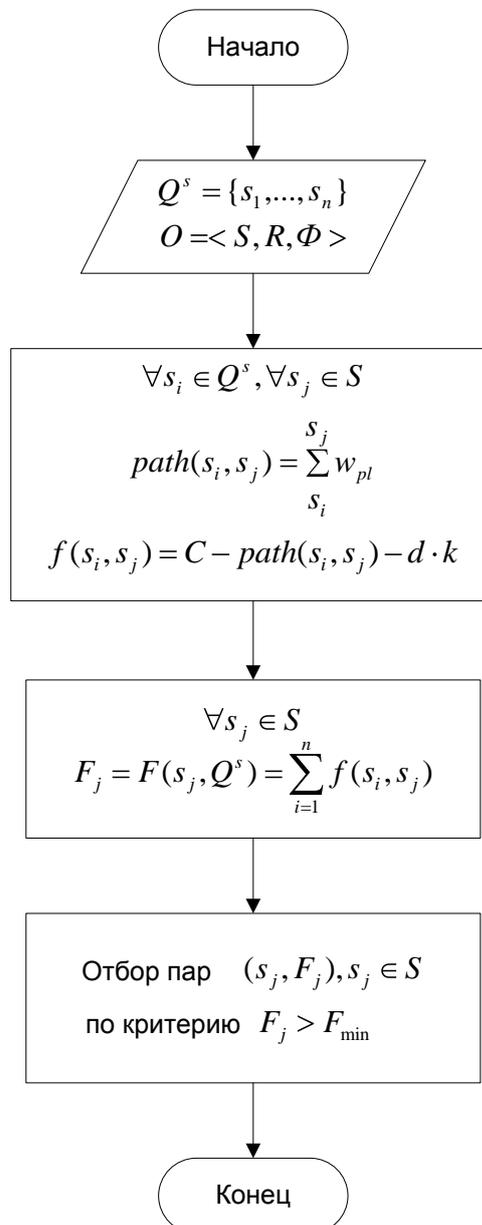


Рис.3. Блок-схема алгоритма процедуры поиска семантического среза

Хотя формула (29) была признана как «самая медленная» из всех пяти рассмотренных в [15] методик, тем не менее, необходимо отметить, что лишь выбранная функция использует все типы связей в онтологии. Высокое быстродействие других методик достигается за счет использования только лишь отношения гипонимии, что неприемлемо в связи с поставленной задачей поиска среза с учетом всех типов связей онтологии.

Длина пути рассчитывается по следующей обобщенной формуле

$$path(s_i, s_j) = \sum_{s_i}^{s_j} w_{pl}, \quad (30)$$

где w_{pl} - вес отношения между стоящими рядом синсетами s_p и s_l , которые являются промежуточными узлами на самом коротком пути между синсетами s_i и s_j .

2) Для каждого $s_j \in S$ рассчитывается интегральная величина F_j близости синсета s_j к поисковому запросу Q^s

$$F_j = F(s_j, Q^s) = \sum_{i=1}^n f(s_i, s_j). \quad (31)$$

3) Третий и заключительный этап реализуется путем сортировки всех синсетов онтологии $s_j \in S$ по значению их интегральной величины F_j . Из полученной таблицы пар (s_j, F_j) отбираются пары, для которых $F_j > F_{min}$, где F_{min} - некий минимальный порог. Результатом этого этапа есть множество пар (s, F) , наиболее семантически близких к поисковому запросу Q^s , причем F определяет меру семантической близости соответствующего синсета s .

Редактор запроса использует найденное множество пар (s, F) для визуализации и предложения пользователю семантического среза онтологии по текущему поисковому запросу. Величина F для каждого отобранного синсета s указывает на вес семантической близости к запросу Q^s и позволяет выполнить визуальное ранжирование выбранных понятий. На основе этих данных пользователь, при необходимости опираясь на описание понятий, принимает решение относительно модификации запроса синсетами из предложенного семантического среза, что позволит в случае необходимости соответствующим образом изменить направление поиска и перейти в ту область результатов, которая больше отвечает поисковым интересам пользователя.

5. Исследование влияния процедур модификации запроса на характеристики квазисемантической поисковой выборки

Все три рассмотренные процедуры направлены на формирование оптимизированного запроса на квазисемантический поиск и позволяют априори итерационно управлять поисковым откликом. Можно выделить две основные характеристики поискового результата, на которые упомянутые процедуры влияют: полнота и точность результатов поиска. В таблице 1 приведены иллюстрированные примерами соображения относительно влияния каждой из рассмотренных в статье процедур модификации квазисемантического поискового запроса на указанные характеристики.

Табл. 1

Таблица 1. Влияние способа модификации запроса на характеристики поисковой выборки

Процедура модификации запроса	Характеристика поисковой выборки	
	Полнота	Точность
Проекция начального запроса на онтологию и процедура горизонтальной коррекции поискового запроса	Расширение результирующей поисковой выборки достигается за счет добавления в начальный поисковый запрос синонимических рядов синсетов онтологии. Пример: Начальный запрос: <i>база данных</i> Модифицированный запрос в терминах: <i>база данных, бд, хранилище данных</i>	Разрешение проблемы омонимии на основе онтологии и при непосредственном участии пользователя позволяет исключить присутствие непертинентных объектов в поисковой выборке документов. Пример: Начальный запрос: <i>экология</i> Модифицированный запрос: <i>экология (раздел биологии)</i> или <i>экология (состояние окружающей среды)</i>
Процедура вертикальной коррекции поискового запроса	Модификация понятием, которое стоит выше по иерархии онтологии, чем текущее понятие запроса, позволяет включать в поисковую выборку более широкий по уровню обобщения круг результатов, поскольку они отвечают более обобщенным (в противовес конкретным) терминам. Это дает возможность повышать содержательную полноту результатов за счет добавления обобщенных (абстрактных) документов по отношению к начальному поисковому запросу. Пример: Начальный запрос: <i>естественная наука</i> Модифицированный запрос: <i>наука</i>	Модификация поискового запроса гипонимом к синсету из текущего поискового запроса делает результирующую выборку семантически более узкой, то есть способной повысить пертинентность результатов поиска в соответствии с поисковыми намерениями пользователя, или такой, которая включает лишь некоторый подкласс результатов из перечня документов по более абстрактному (начальному) поисковому запросу. Пример: Начальный запрос: <i>естественная наука</i> Модифицированный запрос: <i>география</i>
Процедура ассоциативной коррекции	Поскольку такая процедура обеспечивает возможность нелинейной модификации поискового запроса, расширение	Повышается пертинентность результатов поиска, поскольку поиск семантических срезов позволяет найти скрытые латентные

Процедура модификации запроса	Характеристика поисковой выборки	
	Полнота	Точность
поискового запроса	<p>содержательной полноты результатов поиска достигается за счет потенциально семантически связанных документов. Семантическая связанность достигается при помощи семантических связей между понятиями онтологии, которыми модифицируется поисковый запрос. Пример: Начальный запрос: <i>лингвистика, база знаний</i> Модифицированный запрос: <i>лингвистика, база знаний, онтология</i></p>	<p>семантические связи и предложить пользователю понятия, которые могут помочь охарактеризовать более точно его поисковый интерес. Пример: Начальный запрос: <i>лингвистика, программное обеспечение</i> Модифицированный запрос: <i>лингвистика, программное обеспечение, программа-переводчик</i></p>

Примечание. В случае ассоциативной коррекции поискового запроса анализировать точность поисковой выборки в классическом ее понимании нельзя. Основная задача этой процедуры – нелинейный переход в область документов, семантически связанных (через онтологию) с поисковой выборкой по немодифицированному поисковому запросу. Такой переход способен обеспечить точность результатов с точки зрения ожидания пользователя. Именно поэтому в данном случае повышение точности можно рассматривать как повышение пертинентности. В тоже время полнота поисковой выборки в этом случае также не является классической полнотой, поскольку здесь она достигается за счет охвата других областей онтологии во время формирования поискового запроса, что можно рассматривать как скрытую содержательную полноту. Поиск по модифицированному запросу, который сформирован согласно приведенному принципу, позволяет включить релевантные документы из соседних областей предметной области.

Выводы

Таким образом, использование предложенных способов реализации квазисемантического поиска при взаимодействии как с полнотекстовым индексом, так и с концепт-индексом, позволяет целенаправленно повышать пертинентность результатов поиска. Кроме того, за счет включения в поисковую выборку семантически связанных документов дополнительно расширяется поле потенциально пертинентных документов. Такой эффект достигается за счет реализации трех стратегий коррекции поискового запроса: горизонтальной (за счет синонимических связей онтологии), вертикальной (с использованием иерархических связей онтологии) и ассоциативной (с применением оригинального алгоритма поиска семантического среза онтологии на основе текущего поискового запроса). А рассмотренная специфика отображения поискового запроса из поля терминов в поле синсетов онтологии позволяет решать проблему омонимии интерпретации поискового запроса на этапе его формирования. Активную роль в этом

процесі грає інтерактивне взаємодія з користувачем, який виступає основним джерелом пошукового інтересу.

Formal models of the procedures to involve a semantic component into search query used by quasi-semantic search are proposed. The query is modified using a linguistic ontology of the given subject field. The specific of the search query mapping from the terms field to the ontology synsets field is considered. The three search query correction strategies (horizontal, vertical and associative) are proposed and investigated. The impact on the search retrieval results characteristics by the search query modification procedures is analyzed.

Литература

1. Вдовіченко А.В. Інтелектуалізовані пошукові системи. Класифікація та порівняння / А.В. Вдовіченко // Штучний інтелект. — 2002. — №3. — С.61-70.
2. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа / Д.В. Ландэ. — М.: Издательский дом «Вильямс», 2005. — 272с.
3. Ландэ Д.В. Интернетика. Навигация в сложных сетях: модели и алгоритмы / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов. — М.: Книжный дом «ЛИБРОКОМ», 2009. — 264с.
4. Додонов О.Г. Інформаційні потоки в глобальних комп'ютерних мережах: монографія / О.Г. Додонов, Д.В. Ланде, В.Г. Пуятін; Ін-т пробл. реєстрації інформації НАН України. — К.: Наук. думка, 2009. — 294 с.
5. Згуровський М.З. Інформаційні мережеві технології в науці і освіті / М.З. Згуровський, Ю.І. Якименко, В.І. Тимофєєв // Систем. дослідж. та інформ. технології. — 2002. — № 3. — С. 43-56.
6. Марченко О.О. Алгоритми семантичного аналізу природомовних текстів : автореф. дис. на здобуття наук. ступеня канд. фіз.-мат. наук: спец. 05.13.11 "Математичне та програмне забезпечення обчислювальних машин, комплексів та комп'ютерних мереж" / О.О. Марченко. — Київ, 2005. — 17 с.
7. Квазісемантичний пошук текстових даних в електронному інформаційному ресурсі / А.Ю. Михайлюк, О.В. Пилипчук, М.В. Сніжко, В.П. Тарасенко // Радиоэлектроника и информатика. — Харьков: ХНУРЭ. — 2009. — №3. — С.61-67.
8. Спосіб структурно-алгоритмічної організації інтелектуального редактора запиту на квазісемантичний пошук / А.Ю. Михайлюк, О.В. Пилипчук, М.В.Сніжко, В.П. Тарасенко // Міжнародна науково-технічна конференція «Інтелектуальні технології

- лінгвістичного аналізу»: тези доповідей. — К.: Видавництво Нац. авіац. ун-ту «НАУ-друк», 2009. — С.43.
9. Онтологии и тезаурусы: Учебное пособие / Соловьев В.Д., Добров Б.В., Иванов В. В., Лукашевич Н.В. — Казань, Москва., 2006. — 190 с.
 10. Clarke C. Dynamic Inverted Indexes for a Distributed Full-Text Retrieval System / C. Clarke, G.Cormack; TechRep MT-95-01, University of Waterloo. — 1995. — 13 p.
 11. Добров Б.В. Тезаурус и автоматическое концептуальное индексирование в университетской информационной системе РОССИЯ / Б.В. Добров, Н.В. Лукашевич // Третья Всероссийская конференция по Электронным Библиотекам "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". — Петрозаводск, 2001 — С.78-82.
 12. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. — СПб: Питер, 2000. — 384с.
 13. Дорнфест Р. Секреты Google. Трюки и тонкая настройка / Р. Дорнфест, П. Бош, Т.Калишейн. — М.: «Русская редакция», 2008. — 512с.
 14. Manning C. An Introduction to Information Retrieval / Christopher Manning, Raghavan Prabhakar, Hinric Schutze. — Cambridge: Cambridge University Press., 2009. — 581p.
 15. Budanitsky A., Hirst G. Evaluating WordNet-based measures of lexical semantic relatedness / A. Budanitsky, G. Hirst // Computational Linguistics. — 2006. — Vol.32. — №1. — P.13-47.

Замятин Денис Станиславович, канд. техн. наук, доцент кафедры специализированных компьютерных систем Национального технического университета Украины «Киевский политехнический институт». Киев, Украина. Почт. адр. орг.: 03056, г. Киев, просп. Победы, 37, НТУУ «КПИ», кафедра СКС. Тел. служ. (044) 454-90-31. E-mail: dsz@ukr.net. Дом. адрес: 01135, Украина, г. Киев, ул. Исаакяна, д. 2, кв.36. Дом. тел. 236-31-58. Окончил Национальный технический университет Украины «Киевский политехнический институт» в 2000г. Научные интересы: Методы и средства поиска и логической систематизации распределенных данных.

Михайлюк Антон Юрьевич, канд. техн. наук, заведующий научно-исследовательской лабораторией информатизации образования Киевского университета имени Бориса Гринченко. Киев, Украина. Почт. адр. орг.: 04212, г. Киев, ул. Тимошенко, 13-Б, к. 13, НДЛ ПИ. Тел. служ. (044) 426-84-05. E-mail: mau-62@ukr.net. Дом. адрес: 33057, Украина, г.Киев, ул. Академика Янгеля, д. 4, кв. 122. Дом. тел. 456-88-68. Окончил Киевский

политехнический институт в 1985г. Научные интересы: Методы и средства интеллектуального анализа данных.

Михайлюк Елена Станиславовна, научный сотрудник кафедры специализированных компьютерных систем Национального технического университета Украины «Киевский политехнический институт». Киев, Украина. Почт. адр. орг.: 03056, г. Киев, просп. Победы, 37, НТУУ «КПИ», кафедра СКС. Тел. служ. (044) 454-94-92. E-mail: mes@scs.ntu-kpi.kiev.ua. Дом. адрес: 33057, Украина, г. Киев, ул. Академика Янгеля, д. 4, кв. 122. Дом. тел. 456-88-68. Окончила Киевский политехнический институт в 1989г. Научные интересы: экспертные системы, их применение в задачах анализа данных.

Петрашенко Андрей Васильевич, канд. техн. наук, доцент кафедры специализированных компьютерных систем Национального технического университета Украины «Киевский политехнический институт». Киев, Украина. Почт. адр. орг.: 03056, г. Киев, просп. Победы, 37, НТУУ «КПИ», кафедра СКС. Тел. служ. (044) 454-90-31. E-mail: petrashenko@gmail.com. Дом. адрес: 03179, Украина, г.Киев, ул. Пушиной, д. 49, кв.152. Дом. тел. 452-78-65. Окончил Национальный технический университет Украины «Киевский политехнический институт» в 2000г. Научные интересы: Методы поиска информации в несистематизированном гетерогенном ресурсе.

Пилипчук Алексей Васильевич, аспирант кафедры специализированных компьютерных систем Национального технического университета Украины «Киевский политехнический институт». Киев, Украина. Почт. адр. орг.: 03056, г. Киев, просп. Победы, 37, НТУУ «КПИ», кафедра СКС. Тел. служ. (044) 406-84-76. E-mail: ilexcorp@ukr.net. Дом. адрес: 03062, Украина, г. Киев, ул. Кулибина, д. 5, кв. 198. Дом. тел. 422-36-27. Окончил Национальный технический университет Украины «Киевский политехнический институт» в 2008г. Научные интересы: Поиск информации, квазисемантический поиск текстовых данных.

Тарасенко Владимир Петрович, д-р техн. наук, профессор, заведующий кафедрой специализированных компьютерных систем Национального технического университета Украины «Киевский политехнический институт». Киев, Украина. Почт. адр. орг.: 03056, г. Киев, просп. Победы, 37, НТУУ «КПИ», кафедра СКС. Тел. служ. (044) 236-32-02. E-mail: vtarasen@scs.ntu-kpi.kiev.ua. Дом. адрес: 02225, Украина, г. Киев, ул. Закревского, д. 19, кв. 123. Дом. тел. 515-72-57. Окончил Киевский политехнический институт в 1968г.

Научные интересы: повышение эффективности обработки ресурсов глобального электронного информационного пространства.