

УДК 004.89:004.4

Д.Ю. Постановов

СИСТЕМА СЕМАНТИЧЕСКОГО ПОИСКА В ЗАДАЧЕ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

Описывается система информационного поиска с использованием семантического процессора и рассматривается возможность ее эффективного применения в задаче информационной поддержки инновационной деятельности, в частности, для поиска информации по естественно-языковым (ЕЯ) запросам так называемого нефактоидного типа. Приводятся также результаты оценки работы системы в сравнении с другими известными информационно-поисковыми и вопросно-ответными системами.

Введение

Эффективность инновационной деятельности как творческого процесса, направленного на преобразование знаний в новые продукты, услуги или технологии, в большой степени определяется уровнем развития ее информационной поддержки, к основным задачам которой можно отнести обеспечение изобретателя необходимыми знаниями на различных этапах инновационного процесса, в том числе на этапе фундаментальных исследований, прикладных исследований и разработок, опытно-конструкторских работ, а также при организации промышленного внедрения изобретений.

Как показали проведенные исследования, эффективная информационная поддержка инновационной деятельности предполагает доступ изобретателя к следующим основным типам знаний (табл. 1).

Таблица 1

Основные типы знаний в информационной поддержке инновационной деятельности
и примеры обобщенных ЕЯ-запросов для доступа к ним

Элемент искомого знания	Пример обобщенного запроса в случае ЕЯ-интерфейса
Объекты	What is an <object>?
Параметры объектов	What is a <parameter> of an <object>?
Функции объектов	How may an <object> be used?
Типы/функции объектов	What <object> may be used to <function>? (например, <function> = <action> + <object ₂ >)
Структура объектов	What does <object> include?
Технологии, методы, приемы	How to <function>?
Характеристики процессов	What is a <characteristic> of <process>?
Объектные взаимодействия (поиск объекта взаимодействия)	What may be <action-ed> by <subject>?
Объектные взаимодействия (поиск субъекта взаимодействия)	What can <action> the <object>?
Объектные взаимодействия (подтверждение факта взаимодействия)	Can the <subject> <action> the <object>?
Причинно-следственные закономерности	What is a cause of <event>?

Нетрудно убедиться, что перечисленные знания в своей основе соответствуют трем классическим типам знаний, рассматриваемым в теории искусственного интеллекта: объектам / классам объектов, фактам и правилам, отображающим закономерности предметной области в виде причинно-следственных отношений между фактами.

Как отмечается в работе [1], традиционный способ создания баз знаний посредством моделирования конкретных знаний специалистов-экспертов предметных областей себя не оправ-

дал. Автоматический же поиск этих знаний в силу их определения требует семантического анализа текстовых документов и запросов пользователя с целью создания их семантического индекса в отличие от традиционных поисковых систем, основанных на методах поиска по ключевым словам, а также большинства современных вопросно-ответных систем, которые используют статистические И/ИЛИ методы, основанные на так называемых шаблонах.

Несмотря на то, что некоторые производители поисковых систем заявляют об использовании технологий ЕЯ-обработки, это зачастую сводится к тому, что от пользователя не требуется вводить запросы с применением булевых операторов. При этом сама реализация поиска по-прежнему настроена на обработку одно- или двухсловных запросов, а лингвистический анализ, чаще всего начальных уровней [2], осуществляется только на этапе построения поискового образа запроса (например, с целью его расширения). Типичным улучшением поисковых алгоритмов являются способность системы игнорировать неинформативные слова, а также процедура автоматического усечения ключевых слов запроса с целью унификации поиска по числу, роду, падежу, времени и другим морфологическим признакам ключевых слов в зависимости от языка пользовательского запроса [3].

В свою очередь, значительная часть современных вопросно-ответных систем ориентирована исключительно на вопросы *фактоидного* типа (factoid questions). Так, например, открытый для широкого круга исследователей и разработчиков проект QA Track¹, осуществляемый с 1999 г. Национальным институтом стандартов и технологий США (NIST), ставит задачу разработки вопросно-ответных систем, способных отвечать на вопросы именно такого типа, т. е. вопросы, на которые ожидается короткий однозначный ответ, основанный на определенном факте, явно содержащемся в каком-либо документе [4]. На сегодняшний день существует ряд систем, довольно успешно решающих эту задачу, однако в общем случае такая постановка проблемы является существенным ограничением не только для применения подобных систем в качестве инструмента информационной поддержки при решении инновационных задач, но, как отмечено в [5], и для их использования индивидуальными пользователями. Все чаще подчеркивается необходимость дальнейшего развития методов точного информационного поиска, в том числе и по нефактоидным запросам.

В работе [6] в качестве решения этой задачи рассматривается вопросно-ответная система, основанная на статистическом анализе связи терминов, составляющих вопросы, с терминами, составляющими правильные ответы на них (например, слово *why* в запросе соответствует *because* в ответе). Источником для формирования корпуса вопросов и ответов были выбраны разделы интернет-сайтов с часто задаваемыми вопросами (FAQ) на различные темы.

В работе [5] предлагается похожая система, в которой вопрос пользователя анализируется с целью формирования специального запроса для поиска по ключевым словам. Результаты поиска по таким запросам, полученные от сторонних поисковых систем (например, Google), проходят стадию фильтрации, а также собственно стадию выделения ответов, реализованную на основе похожего статистического подхода. В качестве ответов определяется контекст документа из трех предложений.

В работах [7, 8] рассматриваются схожие вопросно-ответные системы, в основе которых лежит извлечение ответов на вопросы пользователя из результатов сторонних систем поиска по ключевым словам при помощи определенных шаблонов. После этого извлеченные ответы ранжируются путем так называемой *триангуляции* – процедуры подтверждения или отрицания ответа на основании подсчета семантически схожих ответов из избыточных источников информации. Как указано в [7], метод триангуляции ориентирован на подтверждение как фактоидных, так и нефактоидных ответов с большим контекстом за счет гибкого алгоритма определения семантической схожести.

Очевидно, что дальнейшее повышение эффективности рассматриваемого класса информационных систем возможно только за счет более глубокой лингвистической обработки текстовых документов и запросов пользователя, о чем особенно четко было высказано уже в [9]. В данной статье рассматривается именно такая задача и ее решение в виде информационно-поисковой системы, разработанной в рамках проекта Goldfire², с реализацией точного семанти-

¹ Question Answering Track Collections – <http://trec.nist.gov/data/qa.html>

² Goldfire – <http://www.goldfire.com>

ческого поиска перечисленных типов знаний в полнотекстовых базах данных (БД), в которых текстовые документы представлены на любом из четырех языков: английском, французском, немецком, японском. При этом предполагается использование ЕЯ-интерфейса пользователя, поскольку предлагаемый метод автоматического семантического анализа текстового документа и его индексирования применим и к ЕЯ-запросу, в том числе и нефактоидного типа, рассматриваемому в качестве текста некоторого специального вида.

1. Система точного семантического поиска в задаче информационной поддержки инновационной деятельности

Помимо основной процедуры поиска по ЕЯ-запросу пользователя и классификации ее результатов система поддерживает поиск в полнотекстовых БД определенных типов знаний:

- о причинно-следственных связях (поиск и классификацию причин/следствий для заданного события);
- о структуре объектов (поиск и классификацию компонентов заданной технической системы);
- об объектных взаимодействиях (поиск функциональных связей между компонентами технической системы, а также поиск параметров объекта, подверженных изменению под воздействием субъекта).

Функциональная часть системы информационного поиска Goldfire включает четыре компонента, типичных для поисковых систем [10]:

- процессор документов, основной функциональностью которого является обнаружение, анализ и индексация новых документов; под задачей индексации понимается составление поискового образа документа (ПОД) с занесением его в БД определенной структуры;
- процессор запросов, анализирующий пользовательский запрос и транслирующий его в структуру поискового образа запроса (ПОЗ);
- функцию соотнесения ПОЗ и ПОД с целью нахождения релевантных документов;
- процедуру ранжирования, располагающую более релевантные документы в начале списка результатов.

1.1. Процессор документов

Базовая функциональность процессора документов обеспечивается семантическим процессором, который включает:

- процессор знаний с лингвистическим и онтологическим процессорами [11];
- подсистему разрешения анафорических связей [12];
- подсистему автоматического реферирования [13];
- подсистему автоматической классификации документов [14].

Основной функцией процессора знаний в задаче семантического поиска является извлечение из ЕЯ-текста входного документа объектов, выраженных именными группами, семантических отношений между ними типа субъект–акция–объект (САО) и семантических отношений между объектами и самими САО типа причина–следствие, часть–целое, а также взаимодействие в технических системах с выделением параметра объекта, изменяемого воздействием субъекта. Первые три типа выделяемых отношений соответствуют трем классическим типам знаний: объектам / классам объектов, фактам и закономерностям предметной области.

Подсистема разрешения анафорических связей уточняет выделенные САО, в компонентах которых, выраженных именной группой (например, в субъекте), присутствует местоимение, в допустимых случаях заменяя его соответствующей именной группой-антецедентом.

Полученные семантические отношения далее подвергаются следующей переработке:

1. Составные именные группы разбиваются на простые, например: *evaporation or heating of exhaust liquids and gases* → *evaporation of exhaust liquids + evaporation of exhaust gases + heating of exhaust liquids + heating of exhaust gases*.

2. Определяются семантические роли слов в соответствии с табл. 2, а также фильтруются слова, неинформативные для поиска (артикли, местоимения, знаки пунктуации и т. д.), например: *a computer display of the invention* → *computer:[атрибут] display:[главное слово] invention:[атрибут]*.

3. Слова, составляющие компоненты отношений, приводятся к так называемой канонической форме в зависимости от части речи (глагол – к инфинитиву, существительное – к форме единственного числа), например: CAO = device – uses – components → CAO = DEVICE – USE – COMPONENT.

Таблица 2

Семантические роли слов в составе базовых семантических отношений

Отношение	Семантическая роль	Определение
1	2	3
Объект	«главное слово»	Главное лингвистическое слово простой именной группы, выражающей объект, т. е. слово, от которого зависят все остальные слова этой именной группы (например, <i>the heating device</i>)
	«атрибут»	Слово, зависящее от лингвистического главного слова именной группы (например, <i>computer system</i>)
CAO	«главное слово субъекта»	Главное лингвистическое слово именной группы субъекта
	«атрибут субъекта»	Атрибут именной группы субъекта
	«акция»	Глагол однословной позитивной акции (например, <i>heat</i>). Позитивная акция в нашем случае – это группа глагола вида [возвратная частица] + глагол + [послелог], в которой не присутствует отрицательная частица; негативная акция – группа глагола с отрицательной частицей вида <i>отрицательная частица</i> + [возвратная частица] + глагол + [послелог]
	«главное слово акции»	Глагольная часть составной позитивной акции вида <i>глагол</i> + <i>послелог</i> либо <i>возвратная частица</i> + <i>глагол</i> (например, для французского и немецкого языков). Данная семантическая роль позволяет различать слова, принадлежащие многословным акциям, при поиске в обратном индексе БД ПОД (например, <i>slow down</i>)
	«атрибут акции»	Послелог либо возвратная частица составной акции (например, <i>heat up</i>)
	«негативная акция»	Глагол негативной акции, например (<i>not heat</i>). Отрицательная частица (например, <i>not</i>) семантической роли не получает и не учитывается при поиске
	«главное слово негативной акции»	Глагольная часть составной негативной акции (например, <i>not slow down</i>)
	«атрибут негативной акции»	Послелог либо возвратная частица составной негативной акции (например, <i>not heat up</i>)
	«главное слово объекта»	Главное лингвистическое слово именной группы, выражающей объект CAO
	«прилагательное»	Прилагательное, выступающее в предложении в роли сказуемого (например, <i>water becomes warm</i>)
«атрибут объекта»	Атрибут именной группы объекта	

Продолжение табл. 2

1	2	3
	«предлог»	Предлог предложной группы (например, <i>to keep in memory</i>), относящейся к акции либо к объекту
	«главное слово непрямого объекта»	Главное лингвистическое слово именной группы непрямого объекта (например, <i>to keep in memory</i>)
	«атрибут непрямого объекта»	Атрибут именной группы непрямого объекта (например, <i>to keep in temporary memory</i>)
	«наречие»	Наречие, относящееся к акции
Причина–следствие	Аналогично ролям из отношений <i>именная группа</i> и <i>CAO</i> с дополнительным учетом принадлежности слов части причины либо следствия	

Полученные таким образом пары {слово, семантическая роль}, которым приписаны также номера исходных семантических отношений и предложений, составляют поисковый образ документа и сохраняются в БД ПОД. Аналогично, пары, полученные из семантических отношений «часть–целое» и «взаимодействие», сохраняются в соответствующих БД ПОД для дальнейшего поиска по ним.

3.2. Процессор запросов

Поддержка ЕЯ-интерфейса поисковой системы осуществляется обработкой запросов, выраженных на естественном языке модулем NLI (Natural Language Interface), производящим следующие операции:

- предобработку запроса с целью удаления неинформативных вводных частей, например: *is there a way to ..., please tell me ..., is it possible ...* и т. п.;
- извлечение из запроса объектов, представленных именными группами, а также семантических отношений типа CAO при помощи процессора знаний;
- расширение запроса введением в него альтернативных отношений, являющихся семантически эквивалентными его исходным семантическим отношениям, например: *water evaporation* → *water evaporation* ИЛИ CAO: *evaporate – water*;
- расширение запроса введением в него альтернативных слов/отношений, являющихся синонимичными исходным словам/отношениям, например: *decrease* → *decrease* ИЛИ *reduce*;
- переработку семантических отношений аналогично процедуре, выполняемой процессором документов, с целью получения информативных слов в канонической форме с приписанными им семантическими ролями;
- формирование ПОЗ в виде булевой формулы определенной структуры в зависимости от формы запроса и произведенных расширений, в которой в качестве термов выступают пары {слово, семантическая роль}.

3.3. Функция соотнесения ПОЗ и ПОД

Система поиска при помощи специального обратного индекса производит отбор поисковых образов документов из БД ПОД по наличию в них слов из ПОЗ. ПОД признается релевантным ПОЗ, если комбинация пар {слово, семантическая роль} из ПОД соответствует хотя бы одному конъюнкту из булевой формулы ПОЗ, приведенной к дизъюнктивной нормальной форме, т. е. выполняются следующие условия:

- 1) все слова из конъюнкта ПОЗ присутствуют в ПОД;
- 2) семантические роли слов из конъюнкта ПОЗ совпадают с семантическими ролями соответствующих слов из ПОД;
- 3) слова из конъюнкта ПОЗ, состоящие в одном семантическом отношении, составляют одно отношение того же типа в ПОД;
- 4) семантические отношения, принадлежащие одному предложению в конъюнкте ПОЗ, также принадлежат одному предложению в ПОД.

Поисковые образы документов, соответствующие одному из конъюнктов ПОЗ по условию 1, но не соответствующие ему по условиям 2, 3 и 4, формируют список результатов меньшей релевантности. Документы, поисковые образы которых не соответствуют ни одному из конъюнктов ПОЗ по условию 1, считаются нерелевантными запросу и не записываются в список результатов поиска.

3.4. Процедура ранжирования

Документы, соответствующие ПОД в списке результатов поиска, упорядочиваются с учетом следующих критериев:

- релевантности ПОД в соответствии с условиями 1–4: наибольшую релевантность (расположение в начале списка результатов) получают документы, удовлетворяющие условиям 1, 2, 3 и 4; далее – 1, 2 и 3; 1, 2 и 4; 1 и 2; 1, 3 и 4; 1 и 3; 1 и 4; 1;
- значимости предложений, в которых находятся искомые семантические отношения, определяемой по принадлежности предложения названию документа, а также реферату, построенному подсистемой автоматического реферирования;
- количества находений в ПОД семантических отношений из ПОЗ;
- соответствия семантических отношений из ПОД оригинальным либо расширенным семантическим отношениям из ПОЗ.

3.5. Способ представления результатов поиска

Результаты поиска представляются в виде страниц с упорядоченным списком цитат (по 10 на одной странице), в которых специальным образом выделены слова, соответствующие словам из ПОЗ. Список результатов явным образом разделен по степени релевантности запросу (наиболее релевантные – результаты, полностью соответствующие критериям 1-4). Поддерживается возможность просмотра полных текстов либо рефератов найденных документов.

Дополнительно список результатов поиска сопровождается их классификацией в виде подкатегорий именных групп запроса с соответствующими им частотами в наборе найденных документов. Так, например, на запрос «how to manufacture a vehicle?», будет предложен список объектов, уточняющих запрос: «automotive vehicle» (с частотой 15), «remotely controlled vehicle» (с частотой 10), «polymeric vehicle» (с частотой 5), «man-driven vehicle» (с частотой 3) и т. д.

3.6. Поиск в специализированных базах знаний

Для поиска в специализированных базах знаний типа «причина–следствие», «часть–целое» и «взаимодействия» пользователю предлагаются соответствующие диалоговые окна, в которых после ввода упрощенного запроса на естественном языке в качестве результатов выдаются искомые знания (табл. 3). В силу заведомо определенной функциональности диалоговых окон для поиска в специализированных базах знаний упрощенные запросы не требуют детального описания информационной потребности. Например, в окне поиска причин в базе знаний «причина–следствие» вместо полного ЕЯ-запроса «what is a reason of oil leakage?» предполагается использование упрощенного запроса «oil leakage».

Поиск в специализированных базах знаний осуществляется при помощи алгоритмов, схожих с основным алгоритмом семантического поиска.

4. Сравнительный анализ эффективности информационно-поисковой системы

Сравнительная оценка эффективности основной процедуры поиска в качестве решения задачи информационной поддержки инновационной деятельности была получена на основе сопоставления результатов поиска описанной системы Goldfire с результатами других известных современных поисковых и вопросно-ответных систем, поддерживающих поиск по ЕЯ-запросам в достаточно больших объемах документов и имеющих свободный пользовательский доступ (табл. 4).

Тестовый набор нефактоидных запросов был получен случайным выбором 25 записей из списка запросов, ранее задававшихся инженерами-пользователями продуктов серии Goldfire и являющихся запросами технического характера из различных предметных областей, сформу-

лированными в виде вопросительного предложения. Преобладающее большинство таких запросов имеет вид «how to ... ?» (например, *how to increase electrical conductivity of polymer?*).

Таблица 3

Примеры поиска в специализированных базах знаний

База знаний	Тип запроса	Тип ответов	Примеры
Причина–следствие	Событие–причина	Событие–следствие	Запрос: oil leakage Ответы: – high friction – oil contamination – ...
	Событие–следствие	Событие–причина	Запрос: oil leakage Ответы: – expansion of hot oil – rotted seals – ...
Часть–целое	Система–целое	Компоненты	Запрос: computer Ответы: – CPU – motherboard – ...
Взаимодействия	Элементы отношения субъект–акция–параметр–объект	Элементы отношения субъект–акция–параметр–объект	Запрос: C: piston – A: compress – П:? – O:? Ответы: – П:pressure, O:fuel mixture – ...

Таблица 4

Информационно-поисковые системы, задействованные в сравнительном анализе

Название системы	Адрес в Интернете
Answer Bus	http://www.answerbus.com
Ask Ed	http://asked.jp
Ask Jeeves	http://www.ask.com
Brain Boost	http://www.brainboost.com
CLAIR	http://tangra.si.umich.edu/clair/NSIR/html/nsir.cgi
Dmitry Roussinov's QA	http://qa.wpcarey.asu.edu
Hakia	http://www.hakia.com
ISI TextMap	http://brahms.isi.edu:8080/textmap
Language Computer	http://languagecomputer.com
Start	http://start.csail.mit.edu
Wee QA System	http://ciboulette.ath.cx/wee

Качество работы систем оценивалось по следующей методологии: по всем запросам из тестового набора были получены результаты поиска, выдаваемые на первой странице результатов от каждой системы (от 3 до 20 результатов) как наиболее релевантные. Далее два эксперта независимо оценивали эти результаты по семибалльной шкале от 0 до 6 баллов (табл. 5) на основании заданных критериев. В случаях когда на определенные запросы какая-либо из систем выдавала неудовлетворительные ответы, эксперту предлагалось осмысленно заменить слова в запросе, не нарушая его структуры, и оценить новые результаты поиска, после чего выбирался максимальный балл для данного класса запросов. Такая процедура необходима для того, чтобы исключить влияние на ход эксперимента объема либо предметной области доступной для систем информации, так как для оценки основной процедуры поиска системы Goldfire использовалась база знаний, составленная из текстов патентов и патентных заявок (более 15 млн докумен-

тов), а большинство других систем производили поиск в документах, доступных в сети Интернет (некоторые в том числе и в патентных базах).

Таблица 5

Критерии оценки результатов поиска

Балл	Критерии
0	Данный тип запросов не поддерживается системой. Не могут быть получены ни прямые ответы, ни результаты, отдаленно связанные с вопросом
1	Данный тип запросов не поддерживается в полной мере, однако могут быть получены результаты, отдаленно связанные с вопросом, среди которых могут встречаться релевантные
2	Данный тип запросов не поддерживается в полной мере, однако могут быть получены релевантные результаты, среди которых могут встречаться прямые ответы
3	Система не настроена на поддержку именно данного типа запросов, однако среди результатов встречается достаточно много прямых ответов
4	Система поддерживает поиск ответов на данный тип запросов, однако среди первых результатов поиска прямые ответы комбинируются с результатами, не являющимися релевантными
5	Система полностью поддерживает поиск ответов на данный тип запросов. Первые результаты всегда являются прямыми ответами, если таковые имеются в доступном наборе документов. Однако некоторые другие результаты могут быть получены ошибочно
6	Система полностью поддерживает поиск ответов на данный тип запросов. Все результаты всегда являются прямыми ответами, если таковые имеются в доступном наборе документов

На основании экспертных оценок наборов результатов от каждой системы были подсчитаны средний балл и его процентное отношение к максимально возможному баллу (а именно, 6) как итоговая оценка эффективности поиска (табл. 6).

Таблица 6

Результаты сравнительного анализа эффективности работы поисковых и вопросно-ответных систем

Название системы	Средний балл по запросам	Отношение к максимально возможному баллу, %
Goldfire	4,76	79,33
Hakia	3,88	64,67
Brain Boost	3,2	53,33
CLAIR	3	50,00
Dmintry Roussinov's QA	2,92	48,67
Language Computer	2,64	44,00
Ask Jeeves	2,48	41,33
Ask Ed	2,48	41,33
Answer Bus	2,44	40,67
Wee QA System	2,36	39,33
ISI	1,28	21,33
Start	0	0,00

Можно утверждать, что оценка результатов поиска с точки зрения пользователя, произведенная в данной работе, близка к оценке *pertinentности* как соотношения объема полезной для пользователя информации к общему объему полученной информации [15]. Такая оценка учитывает не только общую точность и полноту результатов поиска, но и способ представления найденных результатов пользователю, а также их ранжирование.

Информационно-поисковая система Goldfire, в реализации которой использован подход, предложенный в данной работе, показала довольно высокий результат (процентное отношение к максимально возможному среднему баллу – 79,33), опередив известные существующие системы. Полученные данные показывают, что вопросно-ответные системы, ориентированные на работу с запросами фактоидного типа, крайне неэффективны в их использовании для поиска информации по нефактоидным запросам. Так, например, известная вопросно-ответная система Start не смогла дать ответов ни на один из запросов тестового набора.

Заключение

В отличие от статистических и «шаблонных» методов, применяемых в большинстве современных вопросно-ответных систем, предлагаемый в данной работе метод информационного поиска с использованием процессора знаний для извлечения основных типов знаний из текста документов и ЕЯ-запросов пользователя в виде семантических отношений ориентирован на точный поиск релевантной информации по нефактоидным запросам, являющимся наиболее распространенным средством выражения информационной потребности пользователя на различных этапах инновационного процесса.

Сравнительный анализ вопросно-ответных и информационно-поисковых систем показывает возможность успешного промышленного применения системы Goldfire, реализующей описанный подход, в качестве инструмента информационной поддержки инновационной деятельности, а также ее явное преимущество перед другими известными системами.

Достаточно очевидными являются предполагаемые пути дальнейшего развития системы, а именно: распознавание класса запроса (в том числе и фактоидного типа) и комбинирование описанного метода с шаблонными методами вопросно-ответных систем, расширение правил рефразирования запроса в зависимости от его класса, подключение поддержки новых естественных языков помимо уже поддерживаемых (английского, французского, немецкого, японского). Также будут рассматриваться проблемы полноты поиска информации по длинным ЕЯ-запросам и возможность применения системы логического вывода.

Список литературы

1. Городецкий, В.И. Современное состояние технологии извлечения знаний из баз и хранилищ данных / В.И. Городецкий, В.В. Самойлов, А.О. Малов // *Новости ИИ*. – 2002. – № 3. – С. 3–12.
2. Liddy, E.D. Enhanced text retrieval using Natural Language Processing / E.D. Liddy // *Bulletin of the American Society for Information Science*. – Vol. 24, № 4. – 1998. – P. 74–76.
3. Liddy, E.D. Searching and Search Engines: When is Current Research Going to Lead to Major Progress? / E.D. Liddy // *Proceedings of the 2000 International Chemical Information Conference*. – Annecy, 2000.
4. Voorhees, E.M. Overview of the TREC 2002 question answering track / E.M. Voorhees // *Proceedings of the 11th Text REtrieval Conference (TREC 2002)*. – Gaithersburg, 2002. – P. 57–68.
5. Soricut, R. Automatic Question Answering: Beyond the Factoid / R. Soricut, E. Brill // *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*. – Boston, 2004. – P. 57–64.
6. Bridging the Lexical Chasm: Statistical Approaches to Answer-Finding / A. Berger [at al.] // *Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval*. – Athens, 2000. – P. 192–199.
7. Roussinov, D. Self-Learning Web Question Answering System / D. Roussinov, J. Robles // *Proceedings of 2004 World Wide Web Conference*. – New York, 2004.
8. Question Answering: CNLP at the TREC-9 Question Answering Track / A. Diekema [at al.] // *Proceedings of the 9th Text REtrieval Conference (TREC-9)*. – Gaithersburg, 2000.
9. Linguistic knowledge can improve information / W. Woods [at al.] // *Proceedings of the 6th conference on applied natural language processing*. – Seattle, 2000. – P. 262–267.

10. Penev, A. Shallow NLP techniques for Internet Search / A. Penev, R. Wong // Proceedings of 29th Australasian Computer Science Conference. – Hobart, 2006. – P. 167–176.
11. Совпель, И.В. Система автоматического извлечения знаний из текста и ее приложения / И.В. Совпель // Искусственный интеллект. – 2004. – № 3. – С. 668–679.
12. Поцепня, В.Н. Разрешение местоименной анафоры в многоязычных информационных системах / В.Н. Поцепня // Искусственный интеллект. – 2006. – № 4. – С. 619–626.
13. Воронков, Н.В. Реферирование как задача инженерии знаний / Н.В. Воронков // Искусственный интеллект. – 2004. – № 4. – С. 573–579.
14. Постаногов, Д.Ю. Автоматическая классификация документов с использованием семантического процессора / Д.Ю. Постаногов // Искусственный интеллект. – 2005. – № 4. – С. 627–633.
15. Ландэ, Д.В. Поиск знаний в Internet. Профессиональная работа / Д.В. Ландэ. – М.: Издательский дом «Вильямс», 2005. – 272 с.

Поступила 10.04.07

*Белорусский государственный университет,
Минск, пр. Независимости, 4;
ИП «Инвенцион Машин»,
Минск, Революционная, 11
e-mail: denpost@gmail.com,
dpostanogov@imb.invention-machine.com*

D.Y. Postanogov

SEMANTIC SEARCH SYSTEM FOR INFORMATION SUPPORT OF INNOVATIONS

The article describes an information retrieval system using a semantic processor and its efficient application to the task of information support of the innovation activity, in particular, to the task of information retrieval based on non-factoid type natural language queries. The performance of the system is evaluated in comparison with other well-known information retrieval and question answering systems.