

РАНЖИРОВАНИЕ АЛЬТЕРНАТИВНЫХ ИСТОЧНИКОВ ЗНАНИЙ НА ОСНОВЕ ТЕХНОЛОГИИ ГРУППОВОЙ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

М. П. Пашкин

Санкт-Петербургский институт информатики и автоматизации РАН
199178, Россия, Санкт-Петербург, 14 линия, 39
Тел. +7-(812)-328-8071
michael@iiias.spb.su

УДК 681.3

М. П. Пашкин. Ранжирование альтернативных источников знаний на основе технологии групповой поддержки принятия решений // Труды СПИИРАН. Вып. 1, т. 3. — СПб: СПИИРАН, 2003.

Аннотация. *Рассмотрены принципы использования технологии групповой поддержки принятия решений при ранжировании доступных альтернативных источников знаний для повышения эффективности работы систем, занимающихся быстрой интеграцией знаний из распределённых источников. — Библ.13 назв.*

UDC 681.3

M. P. Pashkin. Using group decision support technology for ranking of alternative knowledge sources // SPIIRAS Proceedings. Issue 1, v. 3. — SPb: SPIIRAS, 2003.

Abstract. *This paper contains description of principles of using of group decision support technology for alternate knowledge sources ranking. This technology leads to increase of effectiveness of rapid knowledge fusion systems. —Bibl.13 items.*

1. Введение

Широкое использование Internet привело к тому, что огромный объём информационных ресурсов стал доступен широкому кругу пользователей, и возникла необходимость организации поиска источников данных. В процессе развития Internet набор информационных ресурсов расширялся, и модифицировались средства поиска информации. Наряду с этим решались и другие задачи — проблемы унификации стандартов хранения информации, проблемы поиска информации в различных слабосвязанных распределённых источниках и т.д. Большую популярность получили системы, занимающиеся поиском и сбором данных. Разрабатывались алгоритмы поиска в больших базах данных, средства доступа к разнородным данным, способы хранения и представления и т.д. Но по мере развития вычислительной техники и информационных технологий возникла потребность в системах, работающих со знаниями, т.е. занимающихся созданием, классификацией, синтезом, анализом, хранением, поиском и отображением знаний. Сформировалась необходимость разработки подхода для быстрой интеграции знаний с целью глобального понимания происходящих процессов и явлений, динамического планирования и глобального обмена знаниями [1]. Один из способов представления знаний, для которого применим интеллектуальный анализ, предлагают теории онтологий - смысловые теории о разнородностях, свойствах объектов и связях между ними [2].

Для того чтобы обеспечить быструю интеграцию знаний из различных слабосвязанных источников необходимо не только найти необходимые источники, но и определить их полезность для решения тех или иных проблем. Настоящая статья посвящена описанию возможностей использования технологии группо-

вой поддержки принятия решений для ранжирования альтернативных источников знаний при решении сформулированных выше задач.

2. Принципы построения систем для быстрой интеграции знаний из распределённых источников

Под интеграцией знаний понимается процесс синтеза существующих знаний с целью получения нового.

К основным функциям систем для быстрой интеграции знаний относятся [3]:

- изучение проблемной области и формирование набора онтологий, описывающих проблемную область, в которой они используются, поиск источников знаний, касающихся данной проблемной области;
- предоставление интерфейса конечному пользователю — потребителю знания для ввода запроса на поиск знаний;
- подбор источников знаний;
- выбор, объединение и проверка знаний из распределённых источников;
- предоставление итогов поиска пользователю;
- сохранение результатов работы.

Как правило, изучением проблемной области занимается группа экспертов. Результатами её работы являются:

- Онтология приложения (application ontology) — базовая онтология проблемной области, полученная в результате интеграции знаний из существующих библиотек онтологий, охватывающих предметную область и средства для её обработки;
- Специальным образом организованная картограмма знаний (knowledge map) - информация об имеющихся источниках знаний, касающихся данной проблемной области;
- Построенные над ресурсами знаний — онтологии ресурсов в терминологии онтологии приложения (application oriented source ontologies).

После того, как пользователь вводит запрос в систему, возникает онтология запроса, построенная в результате перевода запроса пользователя в терминологию базовой онтологии (application oriented request ontology). Запрос декомпозируется в упорядоченный набор подзапросов, знания для которых могут быть обнаружены без дальнейшей декомпозиции [4]. По мере развития системы формируется картограмма знаний - специальным образом организованное представление информации о расположении источников знаний, используемых для решения проблем, сформулированных в подзапросах [5], [6]. По картограмме знаний формируется дерево поиска знаний. На рис. 1 изображена укрупнённая схема обработки запроса пользователя.

Для эффективной работы системы при оптимизации формирования дерева поиска знаний [7] необходимо учитывать следующие характеристики источников знаний:

- Количественные:
 1. Временные интервалы, в которые источник доступен;
 2. Стоимость получения знаний из платных источников;
 3. Полнота — степень перевода понятий источника знаний в понятия базовой онтологии [8];

4. Результат экспертной оценки источника знания по итогам работы экспертной группы, получаемый в процессе формирования проблемной области при ранжировании альтернативных источников знаний по заданным критериям.
- Качественные:
 1. Наличие /отсутствие требуемых знаний;
 2. Новизна — дата последнего обновления источника [9];
 3. Степень удовлетворения пользователя ответом системы [10] — оценка, формируемая по результатам работы системы.
 4. Релевантность — степень соответствия источника знания введённому запросу по определённому шаблону (template).

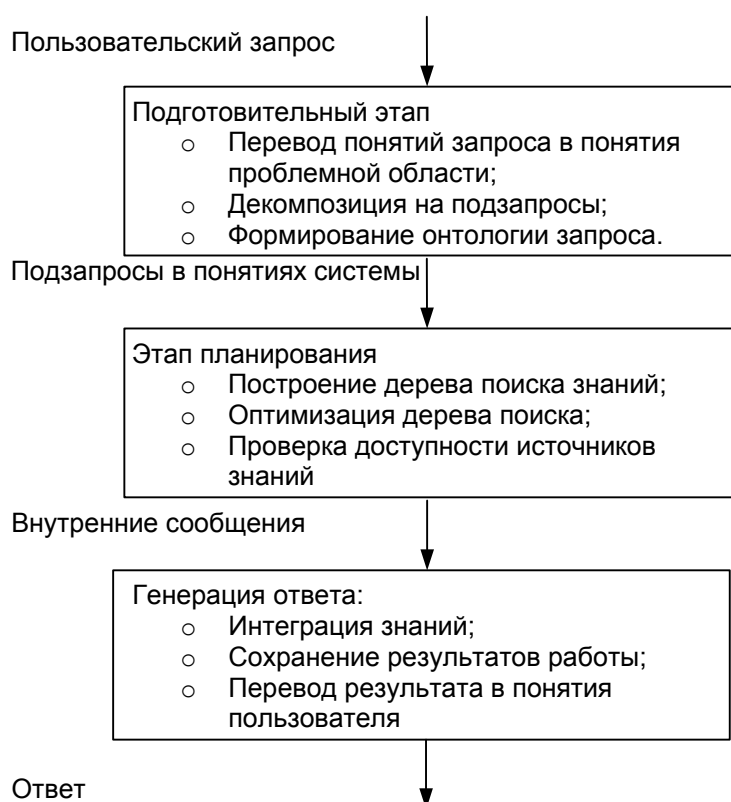


Рис. 1. Укрупнённая схема обработки пользовательского запроса

Количественные характеристики 1–2 и качественные характеристики 1–2 источников знаний известны заранее и приписываются им на этапе формирования проблемной области. Остальные характеристики получают различными способами. Полнота (количественный показатель 3) — величина, вычисляемая на этапе формирования проблемной области. Степень удовлетворения пользователя ответом системы (качественный показатель 3) будет формироваться по мере использования системы. Качественные характеристики приобретают числовые значения при помощи вербально-цифровых шкал.

Введём понятие альтернативных источников знаний в проблемной области. Пусть $AO = \langle O, Q, D, C \rangle$ — онтология приложения проблемной области в нотации сетей ограничений, где O — классы, Q — атрибуты, D — домены и C — ограничения проблемной области, включающие отношения между классами. Каждый источник знаний S_i включает в себя онтологию источника знания

$A(S_i) = \langle O(S_i), Q(S_i), D(S_i), C(S_i) \rangle$ в терминах онтологии приложения и содержащее, которое может быть переведено в систему интеграции знаний полностью или частично, в зависимости от запроса пользователя или потребностей экспертов при описании проблемной области. Результат перевода есть множество экземпляров классов (объектов) $I(S_i) = \{i_k\}_{k=1}^N$, где: $i_k = \langle o_k, o_value_k, attr_k \rangle$ — тройка, состоящая из имени класса из онтологии источника знания, значения объекта и множества, состоящего из положительного числа N_k пар атрибутов и их значений: $attr_k = \{q_m, q_value_m\}_{m=1}^{N_k}$. $o_k \in O(S_i)$, $q_m \in Q(S_i)$ соответственно имя класса и атрибута из онтологии источника знания, а o_value_k, q_value_m - значения соответствующих объектов и атрибутов.

Например, в результате выбора информации из базы данных может быть получен результат: $\langle \text{Автомобиль}, \text{"Форд"}, \langle \langle \text{Цвет}, \text{"Белый"} \rangle, \langle \text{Привод}, \text{"Передний"} \rangle, \dots \rangle$, где имя класса - "Автомобиль", значение объекта - "Форд", имена атрибутов - "Цвет" и "Привод", а значения атрибутов — "Белый" и "Передний" соответственно.

Рассмотрим два экземпляра классов, полученные в результате перевода содержимого источников знаний S_i и S_j в понятия проблемной области: $i_k \in I(S_i)$ и $i_l \in I(S_j)$. Введём понятие близости множеств пар атрибутов μ_{attr} . Пусть $attr' = attr_k \cap attr_l$ и задано число, $\varepsilon_{attr} \in]0;1[$, зависящее от проблемной области.

Тогда $\mu_{attr} = \frac{|attr'|}{\frac{|attr_k| + |attr_l|}{2}}$. Если $\mu_{attr} \in [\varepsilon_{attr}; 1]$ то множества $attr_k$ и $attr_l$ называются близкими и обозначаются $attr_k \approx attr_l$.

Экземпляры классов $i_k \in I(S_i)$ и $i_l \in I(S_j)$ называются семантически близкими, если $o_k = o_l, o_value_k = o_value_l, attr_k \approx attr_l$ и обозначаются $i_k \cong i_l$.

Выделим множество экземпляров классов $I' = \{i_k : i_k \in I(S_i), \exists i_l \in I(S_j), i_k \cong i_l\}_{k=1}^N$, введём меру $\mu_I = \frac{|I'|}{\frac{|I(S_i)| + |I(S_j)|}{2}}$ и выберем

число $\varepsilon_I \in]0;1[$, зависящее от проблемной области. Если $\mu_I \in [\varepsilon_I; 1]$, то источники знаний S_i и S_j называются альтернативными.

Источники могут быть «альтернативными относительно проблемной области» — когда в понятия системы полностью переводится всё их содержимое или «альтернативными относительно шаблонов пользователей», когда в понятия системы полностью переводится только часть содержимого.

Для того чтобы эффективно использовать такие источники полезно определить, какие из них необходимо использовать в первую очередь при обработке пользовательских запросов. Рейтинг альтернативных источников может быть определён, в частности, при помощи систем групповой поддержки принятия решений.

Для более эффективной работы системы может быть определена степень релевантности источников знаний хранящимся шаблонам. Данная величина также может быть вычислена экспертной группой при помощи систем групповой поддержки принятия решений.

Таким образом, сформулированы две задачи, для решения которых ниже предлагается воспользоваться технологией групповой поддержки принятия решений:

1. определение рейтинга альтернативных источников знаний;
2. определение степени релевантности источников знаний хранящимся шаблонам.

Рассмотрим сценарий решения только для первой задачи, так как сценарий решения второй задачи будет отличаться только описанием цели.

3. Использование систем групповой поддержки принятия решений для оценки рейтинга альтернативных источников знаний

Общая схема принятия решений состоит из следующих шагов: (1) спецификация требований, (2) генерация решений, (3) оценка альтернатив, (4) выбор эффективного решения. При формировании проблемной области эксперты выполняют первый и второй шаги данного процесса - изучают проблемную область, определяют источники знаний. Третий и четвёртый шаги данного процесса и предназначены для решения задач, поставленных в предыдущем пункте.

Для поддержки принятия решений разработано множество систем, использующих различные математические методы и алгоритмы. Воспользуемся разработанной нами ранее системой «Multiexpert», ориентированной на оценку количественных и качественных показателей или критериев, их взвешивание и ранжирование с последовательной выработкой согласованного решения группой экспертов и включающая в себя два типа пользователей — руководителя и эксперта [11].

Выделим следующих участников процесса принятия решений: руководитель, аналитик и эксперты. Руководитель экспертной группы является лицом, ответственным за принимаемое окончательное решение. Для всестороннего анализа предлагаемых к рассмотрению альтернативных источников знаний руководитель привлекает экспертов различных проблемных областей. В задачу аналитика входит проверка согласованности предоставленных экспертами решений и выработка рекомендаций по дальнейшей работе экспертной группы.

Исходные данные для работы группы экспертов подготавливаются руководителем в виде задания, в состав которого входит перечень оценочных и классификационных показателей (критериев оценки источников знаний) и описание источников знаний. Перечень показателей может быть стандартным и модифицированным. Стандартный набор используется для облегчения работы руководителя — ускорения ввода данных. По мере использования данный набор может модифицироваться и сохраняться в качестве альтернативного набора для будущего использования.

Предлагается следующий базовый набор показателей для оценки источников знаний:

- Объём хранимых знаний [9];
- Доступность в режиме on-line [9];
- Корректность [9];
- Качество объяснения [9];
- Понятность ответов [9];
- Скорость доступа;

- Скорость обработки запросов;
- Популярность;
- Возможность внесения изменений в содержимое;
- Возможность повторного использования ранее извлечённых знаний.

На рис. 2 представлена одна из экранных форм для работы эксперта на основном этапе формирования иерархии показателей для оценки альтернативных источников знаний. На ней представлен пример организации иерархии предложенных показателей. Эксперты сгруппировали их по определённым критериям в 3 группы: (1) дополнительные характеристики, (2) полезность, (3) производительность (упорядочены по алфавиту).

Каждому подготовленному руководителем заданию соответствует календарь его выполнения, содержащий контрольные даты выполнения основных этапов, наименование этапов и основные результаты работы на каждом из этапов.

В системе «MultiExpert» реализована следующая последовательность процедур [12]:

- Метод парных сравнений для выяснения предпочтений экспертов по каждому показателю, основанный на методе анализа иерархий [13];
- Метод многокритериального выбора с относительной нормировкой для расчёта формальным образом весов заданных источников знаний;
- Методы формирования согласованных решений.

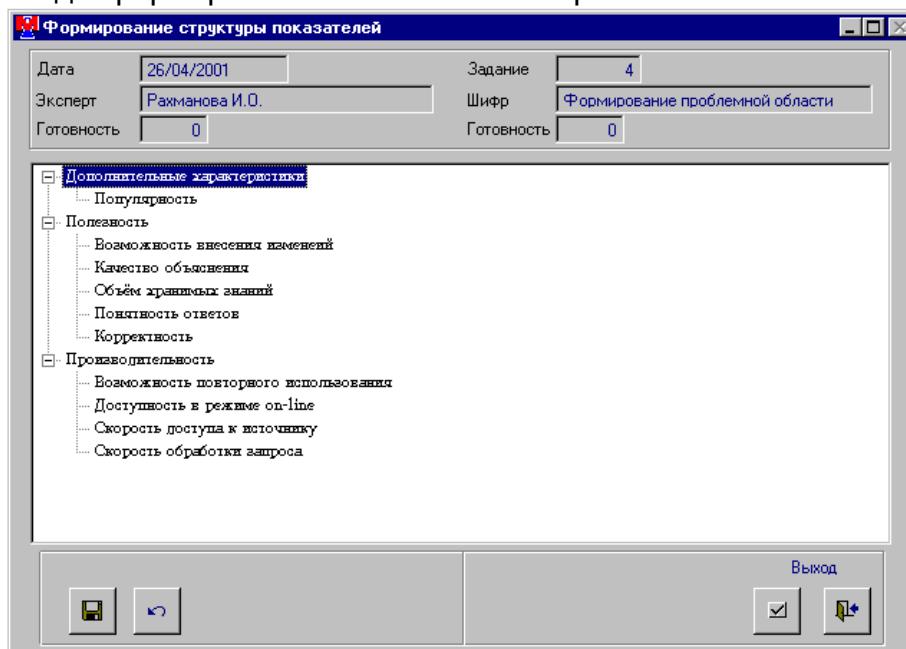


Рис. 2. Этап формирования структуры показателей в системе «MultiExpert»

Работа над заданием происходит по следующим этапам:

1. Формирование иерархии показателей качества;
2. Согласование структуры показателей качества;
3. Оценка качественных показателей;
4. Согласование оценок качественных показателей;
5. Парное сравнение показателей
6. Согласование полученных весов показателей
7. Ранжирование проектов

В предложенном алгоритме работы над заданием можно выделить два типа этапов: основные (1, 3, 5, 7) и согласования (2, 4, 6). Основные этапы предоставляют возможность экспертам высказывать свое мнение по предлагаемым вопросам. С каждым из таких этапов связан соответствующий интерфейс. Этапы согласования позволяют приостановить процесс выполнения задания для оценки сгенерированных системой групповых решений.

В качестве примера рассмотрим проблемную область, связанную с поддержкой принятия решений при внедрении нововведения на предприятии. В качестве источников знаний для поддержки работы пользователей были предложены: (i) база данных, хранящая ответы на часто задаваемые вопросы, (ii) экспертная система, помогающая пользователям решать возникающие трудности и (iii) видеокурсы по нововведению. В табл. 1 указаны значения показателей для выбранных источников знаний, полученные в результате экспертной оценки их сторонним консультантом. Оценивалась возможность эффективного использования данных источников для быстрого освоения пользователями нововведения.

Таблица 1. Исходные данные — значения показателей для альтернативных источников знаний.

Источник	База данных	Экспертная система	Видеокурсы
Объем хранимых знаний	Большой	Средний	Средний
Доступность в режиме on-line	Постоянная	час в день	Нет
Корректность	удовл.	хорошая	приемлемая
Качество объяснения	доступное	доступное	доступное
Понятность ответов	средняя	высокая	низкая
Скорость доступа к источнику	10 МБит	1 МБит	56К
Скорость обработки запроса	Низкая	высокая	высокая
Популярность	известный	известный	мало известный
Возможность повторного использования	отсутствует	внутренние алгоритмы	отсутствует
Возможность внесения изменений	отсутствует	неявная	отсутствует

На рис. 3 представлено распределение, изображающее веса предложенных показателей после оценки экспертами качественных показателей и проведения парных сравнений в группах показателей, а в табл. 2 приведены вычисленные значения весов показателей.

Таблица 2. Значения весов показателей после оценки

№	Показатель	Вес
1	Доступность в режиме on-line	0.031335
1	Скорость доступа	0.070067
1	Скорость обработки запросов	0.149993
1	Возможность повторного использования	0.014637
2	Объем хранимых знаний	0.046707
2	Корректность	0.095103
2	Качество объяснения	0.193645
2	Понятность ответов	0.374352
2	Возможность внесения изменений	0.024161

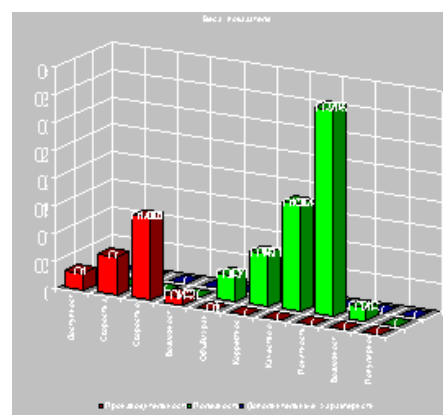


Рис. 3. Веса показателей после оценки

После прохождения по всем этапам поставленная задача оказывается выполненной, то есть всем источникам приписываются определённые веса. Источник знаний, получивший наибольший вес в дальнейшем будет обрабатываться системой в первую очередь. Если несколько источников получили одинаковые веса, при обработке пользовательских запросов могут либо проверяться дополнительные критерии оптимизации, либо источники могут выбираться случайным образом. На рис. 4 представлен график, изображающий веса источников знаний, полученных в результате экспертной оценки. По итогам экспертной оценки наибольший вес имеет экспертная система и она будет использоваться в первую очередь при построении картограммы знаний для обработки пользовательских запросов, связанных с особенностями внедрения нововведения.

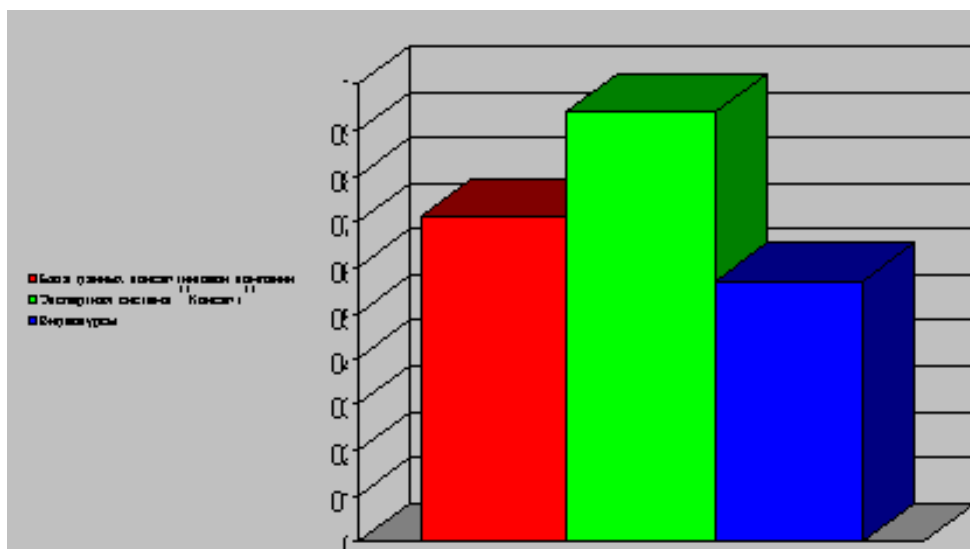


Рис. 4. Источники знаний, отранжированные в результате экспертной оценки

5. Заключение

В настоящей статье рассмотрены принципы использования технологии групповой поддержки принятия решений при ранжировании доступных альтернативных источников знаний для повышения эффективности работы систем, занимающихся быстрой интеграцией знаний из распределённых источников. Предложены критерии, по которым необходимо оценивать доступные источники знаний, и предложено использовать системы групповой поддержки принятия решений для определения значений таких критериев, как рейтинг альтернативных источников знаний и степень релевантности источников знаний хранящимся шаблонам

Литература

- [1] *Smirnov A. V.* Rapid Knowledge Fusion into the Scalable Infosphere: A Concept and Possible Manufacturing Applications. — Proceedings of the International NAISO Congress on Information Science Innovations (ISI'2001), Symposium on Intelligent Automated Manufacturing (IAM'2001), Dubai, U.A.E., 2001.

- [2] *Guarino N.* Some organizing principles for a unified top-level ontology // Working Notes of AAAI Spring Symposium on Ontological Engineering. — Stanford, 1997.
- [3] *Смирнов А. В., Пашкин М. П., Левашова Т. В., Шилов Н. Г.* Основные принципы организации систем быстрой интеграции знаний // Труды СПИИРАН. Вып. 1, т. 2 — СПб: СПИИРАН, 2002.
- [4] *Goyal M., Parameswaran N.* Attitudes in Team Problem Solving. — Proceedings of the ICSC Symposia on Intelligent Systems & Applications (ISA'2000). Wollongong, Australia, 2000
- [5] *Gordon J. L.* Creating knowledge maps by exploiting dependent relationships // Knowledge-Based Systems № 13, 2000. — p. 71.
- [6] *Goñi A., Illaramendi A., Mena E., Blanco J. M.* An Ontology connected to several data repositories: query processing steps // Journal of Computing and Information (JCI), Volume 3, № 1, 1998.
- [7] *Шилов Н. Г.* Математические модели конфигурирования сетей источников знаний в бизнес среде. — Труды СПИИРАН. Вып. 1, т. 3 — СПб: СПИИРАН, 2003.
- [8] *Mena E., Kashyap V., Sheth A., Illaramendi A.* Estimating Information Loss for Multi-ontology Based Query Processing. — Technical Report, 13th Biennial European Conference on Artificial Intelligence (ECAI'98) workshop on Intelligent Information Integration (III98), Brighton (UK), August 1998.
- [9] *Chaudhri V. K., Lowrance J. D., Stickel M. E., Thomere J. F., Wadlinger R. J.* Ontology Construction Toolkit. — Technical Note Ontology, AI Center. Report, January 2000. SRI Project No. 1633.
- [10] *Chidlovskii B., Glance N. S., Grasso M. A.* Collaborative Re-Ranking of Search Results. — Proceedings of the AAAI-2000 Workshop on AI for Web Search. Austin, Texas, 2000.
- [11] *Смирнов А. В., Пашкин М. П., Рахманова И. О.* Многоагентная среда групповой поддержки принятия решений в распределённых экспертных командах. — Проблемы информатизации № 1, 1998. — 41 с.
- [12] *Smirnov A. V., Pashkin M. P., Rakhmanova I. O.* Multiexpert: an Inthernet-based Support Environment for Solutions Evaluation. — Proceedings of the Special Day on pan-European Cooperation and Technology Transfer. - Zakopane, Poland, 1996. — p. 73.
- [13] *Saaty T. L., Kearns K. P.* Analytical Planning. The Organization of Systems. — Graduate School of Business, University of Pittsburg, Pennsylvania, USA, 1985.