

Метод динамического создания связей между информационными объектами базы знаний

© Обухова О.Л., Бирюкова Т.К., Гершкович М.М., Соловьев И.В., Чочиа А.П.

Институт проблем информатики РАН

support@intergallery.ru

Аннотация

В статье рассматривается формальная модель организации информационного поиска в базе знаний с использованием возможностей разработанной авторами адаптивной фасетной навигации. Поисковый образ объекта строится в форме фасетной формулы объекта. Построение поискового запроса представляет собой итерационный диалоговый процесс уточнения поискового запроса с учетом прямых и ассоциативных связей информационных объектов. Формальная модель положена в основу реализационной модели, в которой визуальный интерфейс представляет обзор содержания базы знаний и служит удобным механизмом построения поисковых запросов.

1 Введение

Интеллектуальный ресурс научно-исследовательского института находит свое отображение в электронной коллекции научных материалов по основным направлениям исследований, ведущихся научными работниками, и отраслям знаний, находящимся в сфере интересов сотрудников. Коллекция научных материалов может служить информационным источником для получения необходимых сведений о поставленных задачах и методах их решения в интересующей исследователя области знаний. Основную часть знаний аналитики получают в результате сравнения, анализа и синтеза информации, размещенной в текстах документов. Для доступа к требуемым материалам целесообразно использовать не только поиск по формальным признакам документов, таким как название, автор, год издания, но и поиск по смысловому содержанию документов. Для выполнения данной задачи авторы поставили целью разработать программную систему размещения научных материалов в базе знаний с возможностью информационного поиска.

«Когда удачно завершился поиск и нужная информация найдена - она становится знанием, необходимым для принятия решения или совершения действий.»[1] Своевременно предоставленные знания служат повышению эффективности деятельности научного сообщества. «Особое значение имеет повышение продуктивности интеллектуальной деятельности.»[2]

Задача разработки программной системы поддержки базы знаний является актуальной. Вопросам организации информационного поиска в базе знаний с использованием возможностей фасетной навигации в электронных коллекциях[3] посвящена данная статья.

2 Концептуальная модель

В основе нашего подхода к разработке информационного поиска лежит тот факт, что информация о научно-исследовательской работе научного института хранится в базе знаний в виде информационных объектов (ИО), сопровождаемых *поисковым образом документа (ПОД)* [4], в котором кратко выражается основное смысловое содержание документа. Информационными объектами являются темы НИР, публикации, зарегистрированные продукты и патенты, диссертации, доклады на конференциях. Применение гибкого механизма образования логически связанных цепочек информационных объектов позволяет получать те знания по выбранному направлению, которые интересуют конкретного пользователя. Для этого авторами предлагаются методы поиска, основанные на технологии адаптивной фасетной навигации.

Используя принципы фасетной классификации [5] и разработанную авторами технологию адаптивной фасетной навигации [6], мы предлагаем концептуальную модель создания ПОД в форме фасетной формулы объекта, представленной в виде совокупности: фасетный признак – список значений фасетного признака для данного объекта, и поискового запроса (строящегося итерационно в процессе диалога) в форме фасетной формулы запроса. Данный подход развивает идеи, изложенные авторами ранее [7], в том направлении, что для каждого информационного объекта по

каждому (в общем случае) фасетному признаку задается список значений фасетного признака, характеризующих объект. Набор фасетных признаков диктуется направленностью предметной области, к примеру, для коллекции электронных документов сайта научного института набор фасетных признаков определяется видами и характером научной деятельности. Количество фасетных признаков выбирается в соответствии с принципом, сформулированным американским психологом Миллером [8]: для того, чтобы выбор был эффективным, количество элементов в нем не должно быть больше семи-деяти.

Список значений для каждого фасетного признака формируется и обновляется в административном режиме в процессе работы программной системы поддержки базы знаний в момент занесения информационных объектов. Результирующий список доступных значений фасетного признака является объединением всех заданных значений данного фасетного признака для каждого ИО. Занесение нового или удаление существующего ИО приводит к обновлению списка доступных значений каждого фасетного признака. Анализ сочетаний фасетных признаков и их заданных значений для информационных объектов позволяет программным образом в процессе работы системы выстраивать логические связи между ними. Связи динамически создаются и модифицируются в зависимости от актуального состояния электронной коллекции научных материалов, размещенной в базе знаний.

Общий обзор всех фасетных признаков и полного списка их значений дает представление о сферах, характере и направлениях научной деятельности, которые представлены в работах, хранящихся в базе знаний.

3 Формирование фасетной формулы объекта

Введем определение понятия «Фасетная формула объекта».

$G = \{g_i \mid i = 1, \dots, k\}$ - коллекция информационных объектов, где g_i - информационный объект;

$\Phi(G) = \{\Phi_j(G) \mid j = 1, \dots, l\}$ - множество фасетов на G .

При рассмотрении конкретной коллекции G будем обозначать: $\Phi_j(G) = \Phi_j$.

$\varphi^{\Phi_j} = \{\varphi_{j1}, \dots, \varphi_{jm_j}\}$ - множество допустимых значений фасета Φ_j (здесь m_j - количество допустимых значений фасета Φ_j).

Если для объекта g_i задано хотя бы одно значение фасета Φ_j , то

$\varphi_j(g_i) = \{\varphi_{js}(g_i), s \in \{1, \dots, m_j\}\} \subset \varphi^{\Phi_j}$ - набор (множество) заданных значений фасета Φ_j для объекта g_i .

Если для объекта g_i значения фасета Φ_j не заданы, то $\varphi_j(g_i) = \emptyset$ (\emptyset - пустое множество).

$\varphi(g_i) = \{\varphi_j(g_i) \mid \forall j: j \in \{1, \dots, l\}, \varphi_j(g_i) \neq \emptyset\}$ - множество наборов заданных значений фасетов для объекта g_i .

Фасетная формула $FF(g_i)$ объекта g_i представляет собой множество совокупностей следующего вида:

$\{[\text{фасет } \Phi_j; \text{ набор заданных для объекта } g_i \text{ значений фасета } \Phi_j]\}$, то есть:

$$FF(g_i) = \{[\Phi_j, \varphi_j(g_i)] \mid \Phi_j \in \Phi(G), \varphi_j(g_i) \in \varphi^{\Phi_j} \mid \forall j: j \in \{1, \dots, l\}, \varphi_j(g_i) \neq \emptyset\}$$

Фасетная таблица содержит фасетные формулы для всех объектов коллекции:

$$FT(G) = \{FF(g_i) \mid g_i \in G, i = 1, \dots, k\} = \{[\Phi_j, \varphi^{\Phi_j}] \mid j = 1, \dots, l\} -$$

множество фасетов с их допустимыми значениями на объектах коллекции G . Фасетная таблица служит базисом для выявления взаимосвязей различных информационных объектов.

Для формирования результирующего запроса с целью получения искомой выборки объектов $RG \subset G$ необходимо в процессе пошагового алгоритма задать значения некоторого подмножества фасетов для построения фасетной формулы запроса.

4 Формирование фасетной формулы запроса

Поисковый запрос строится в форме фасетной формулы запроса в процессе выполнения пошагового алгоритма выбора фасетных признаков и их значений из списка допустимых значений. На каждом шаге пользователь уточняет фасетную формулу, выбирая одно или несколько значений для очередного фасетного признака из множества допустимых значений фасетного признака на данном шаге.

Определим фасетную формулу запроса RFF как множество совокупностей следующего вида:

{[фасет Φ_j ; набор значений фасета Φ_j , формирующих запрос]}

Операция выборки (*retrieve*) формирует подмножество объектов $RG \subset G$ в соответствии с фасетной формулой запроса:

$$RG = \text{retrieve}(G, RFF)$$

Такой способ построения поискового запроса обеспечивает получение на каждом шаге непустого подмножества объектов. Список допустимых значений для каждого фасета обновляется в соответствии с состоянием полученного подмножества объектов. Новые списки значений фасетов позволяют пользователям продолжить уточнение поискового запроса.

Авторами разработана модель фасетной навигации, в которой на каждом шаге актуальное подмножество информационных объектов определяется полным или частичным совпадением фасетной формулы объекта с фасетной формулой запроса.

В настоящей статье предлагается развитие этого подхода в направлении формирования расширенного актуального подмножества ИО, включающего объекты, объединенные прямыми и ассоциативными связями.

Определение 1. Информационные объекты g_1 и g_2 объединены прямой связью, если \exists хотя бы один фасетный признак Φ_j , для которого $\varphi_j(g_1) \cap \varphi_j(g_2) \neq \emptyset$

Определение 2. Информационные объекты g_1 и g_2 объединены ассоциативной связью, если для всех фасетных признаков Φ_j , $j=1, \dots, l$, $\varphi_j(g_1) \cap \varphi_j(g_2) = \emptyset$, но \exists хотя бы один информационный объект g_3 , у которого для хотя бы одного фасетного признака Φ_k ($k \in \{1, \dots, l\}$) выполняется условие: $\varphi_k(g_3) \cap \varphi_k(g_1) \neq \emptyset$ и $\varphi_k(g_3) \cap \varphi_k(g_2) \neq \emptyset$.

Данные определения используются в предлагаемом авторе подходе к формированию расширенного подмножества ИО в соответствии со следующими правилами: расширенное подмножество формируется из ИО, для которых фасетная формула объекта полностью или частично совпадает с фасетной формулой запроса, и расширяется информационными объектами, которые объединены с ними прямой или ассоциативной связью.

Список допустимых значений каждого фасета перегружается в соответствии с состоянием расширенного множества ИО.

Авторы поставили перед собой задачу разработки макета поисковой системы с целью апробации изложенного метода.

5 Визуальный интерфейс диалоговой модели информационного поиска

Для визуализации пользовательского интерфейса разработана технология динамического создания интерфейсных окон на основе технологии Adobe Flash со встроенным языком ActionScript.

Задача пользовательского интерфейса заключается в том, чтобы на каждом шаге выполнения алгоритма:

1. представить для просмотра все фасеты со множеством доступных значений;
2. обеспечить возможность выбора списка значений в определенном фасете;
3. получить текущую выборку ИО, анализ которой позволит пользователю определить свои последующие шаги по формированию фасетной формулы запроса.

Элементы визуального интерфейса формируются программным образом и определяются текущим состоянием подмножества информационных объектов и предысторией формирования фасетной формулы запроса.

Список допустимых значений для каждого фасетного признака обновляется в соответствии с состоянием полученного подмножества объектов. Новые списки значений фасетов позволяют пользователям продолжить уточнение поискового запроса. Текущее состояние визуального интерфейса позволяет получить представление о той части базы знаний, границы которой определяются выбранными значениями фасетных признаков на предыдущих шагах. Дополнительно к этой информации пользователь может проанализировать текущую выборку информационных объектов и сделать очередной шаг при выборе объектов, продолжив уточнение фасетной формулы запроса для достижения совпадения со своей информационной потребностью.

Динамически появляющаяся информация на каждом шаге доступа к объекту предоставляет пользователю данные, анализ которых позволяет ему сделать следующий выбор.

На рисунке 1 продемонстрирована работа с фасетом «Авторы». Список всех фасетов размещен в левой панели. Активация одного из фасетов приводит к генерации интерфейсного окна, содержащего множество доступных значений данного фасета с поддержкой возможности выбора списка значений, представляющих интерес для пользователя.

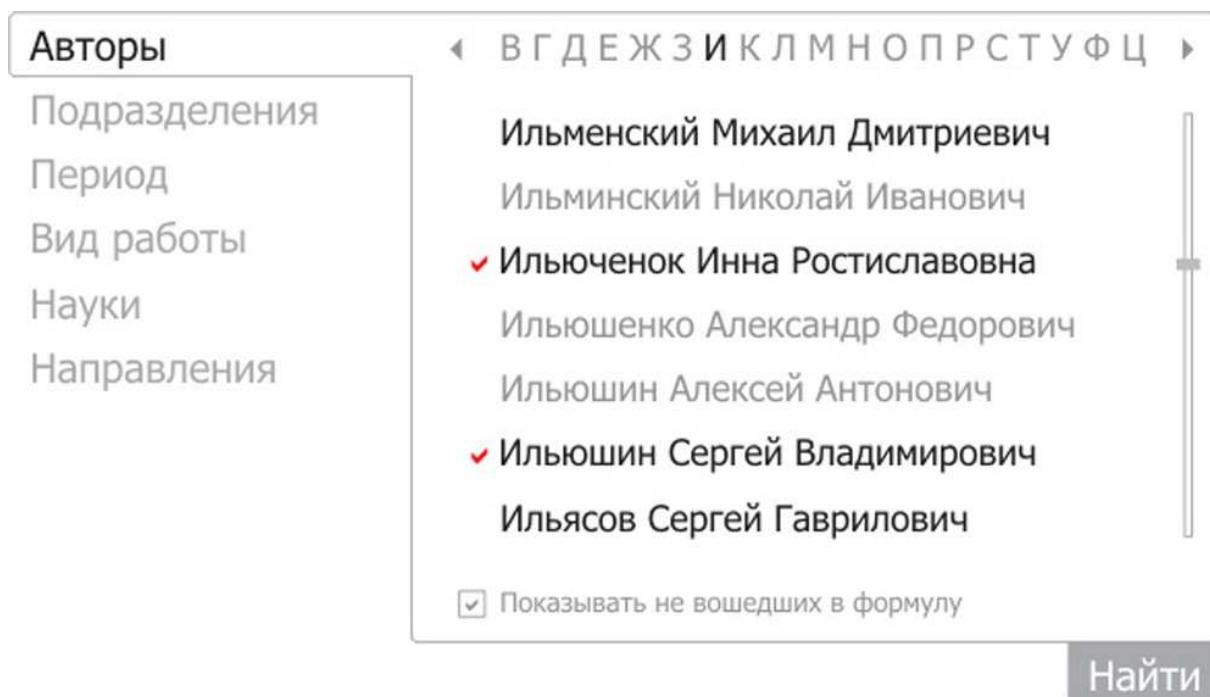


Рисунок 1

Выбранные значения формируют фасетную формулу запроса. Эти значения будут определять определенные границы подмножества ИО. Данное подмножество ИО на каждом шаге выполнения алгоритма будет обновляется с учетом прямых и ассоциативных связей объектов, соответственно, список доступных значений фасета на следующем

шаге будет определяться состоянием расширенной выборки. Но, возможно, пользователю интересно будет в какой-то момент посмотреть и те значения фасета, которые он «проигнорировал» в силу логики своей работы. Для этого введена функция «Показывать не вошедших в формулу» (рисунок 2).

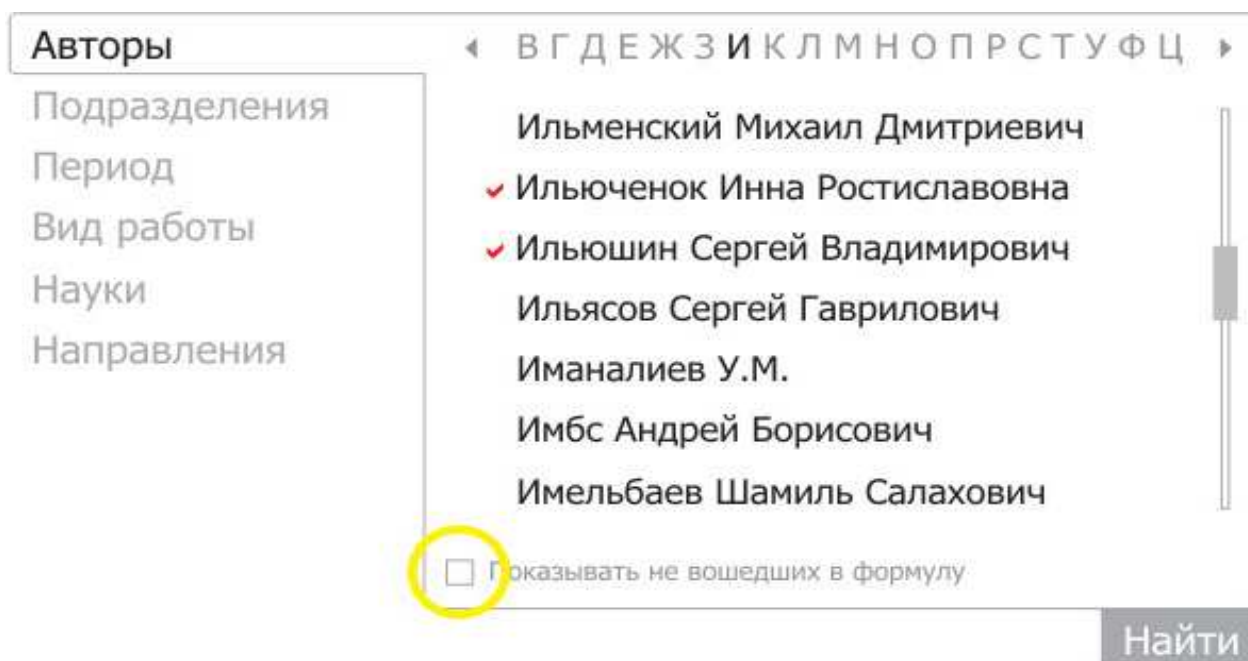


Рисунок 2

На рисунке 3 представлена работа с фасетом «Период».

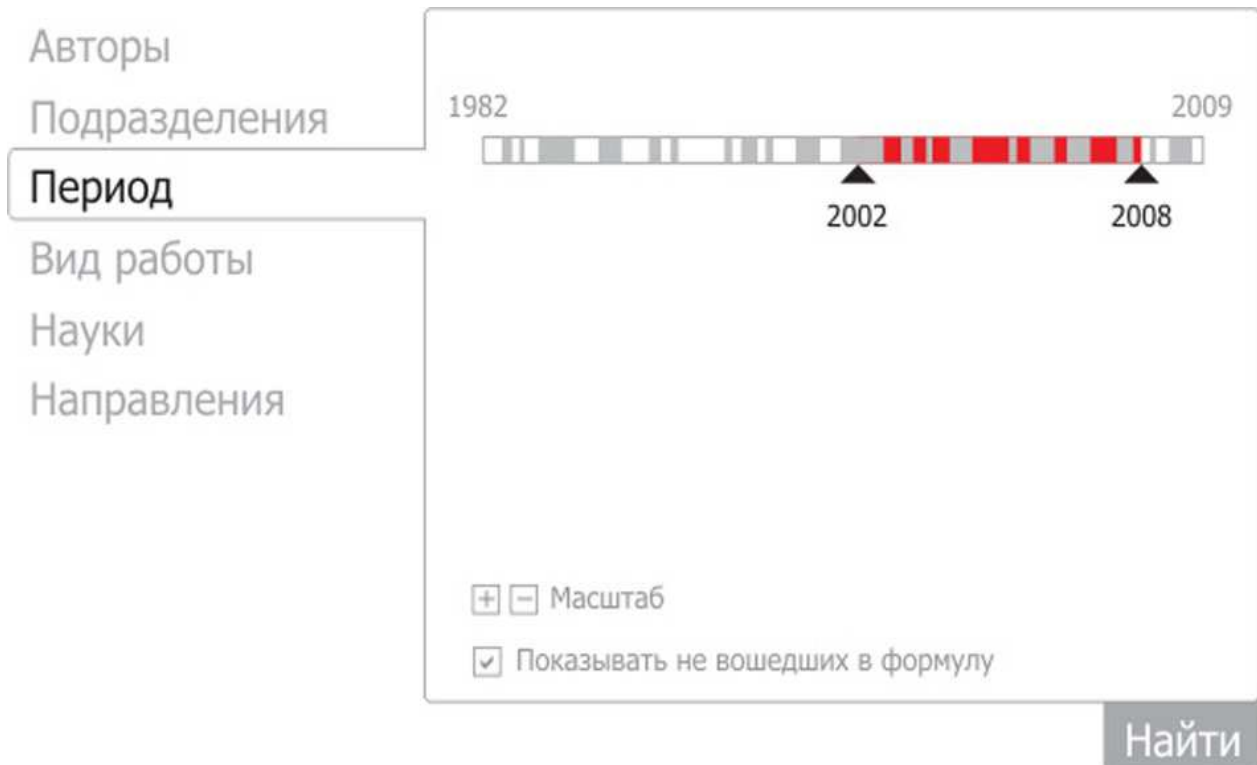


Рисунок 3

Для фасета «Период» введена дополнительная функция «Масштаб». Если пользователю необходимо более детальное уточнение периода появления той или иной

публикации, то он может использовать данную возможность (рисунок 4).

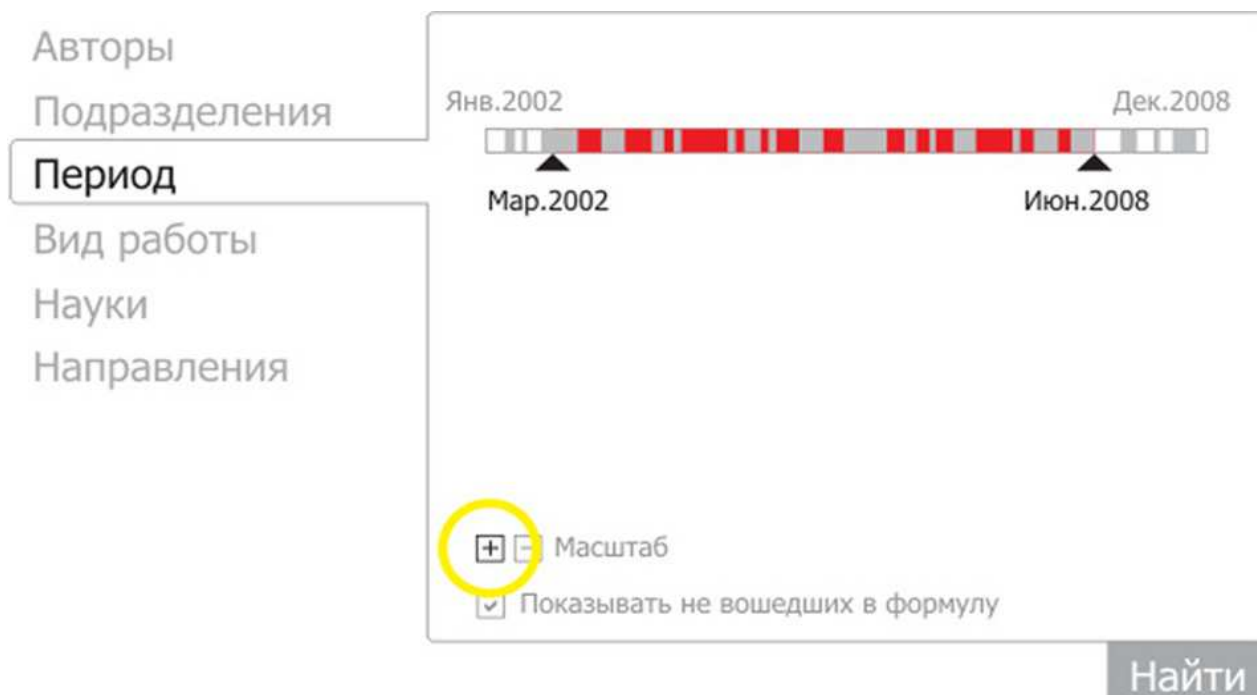


Рисунок 4

Особого подхода требуют фасеты, у которых список доступных значений представляет иерархию, к примеру, фасет «Науки» (рисунок 5).



Рисунок 5

После фиксации значения фасетного признака представлен второй уровень иерархии значений, что первого уровня генерируется окно, в котором показано на рисунке 6.

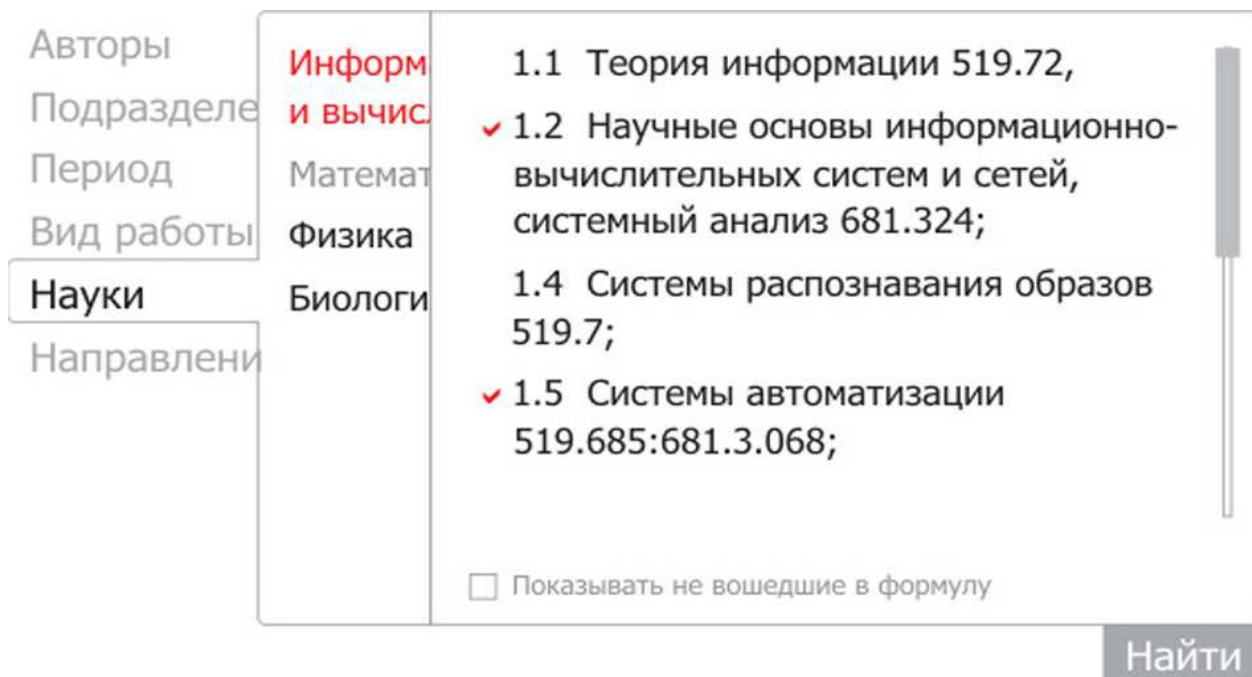


Рисунок 6

6 Задачи реализационной модели

Изложенные идеи авторами воплощаются в реализационной модели.

Используемые программные средства:

- реализация осуществляется в среде ОС: Microsoft Windows 2000/2003;

- веб-сервер: Microsoft Internet Information Server 5.0/6.0;

- контейнер серверного приложения: Microsoft ASP.NET 2.0;

- СУБД: Microsoft SQL Server 2005;

- в качестве средств разработки используется Microsoft Visual Studio 2005, ASP.NET 2.0;

- для построения пользовательского интерфейса используется Adobe Flash со встроенным языком ActionScript.

ASP.NET Web Services используют индустриальные стандарты:

- XML. Обмен данными между визуальным пользовательским интерфейсом и сервером

- SOAP. Протокол обмена сообщениями между Web службой и клиентом, основанный на XML .

- Web Services Description Language (WSDL). Описывает параметры сообщений Web-службы для взаимодействия с клиентами .

Опытная эксплуатация реализационной модели позволит вести файлы протокола и собирать статистическую информацию с тем, чтобы анализировать и обобщать возможные действия пользователей по формированию фасетной формулы запроса. В частности, только на практике можно проверить, интересен ли пользователям механизм учета прямых и ассоциативных связей информационных объектов. Авторы планируют сделать этот сервис опциональным, и сбор статистики об использовании этой опции позволит разработчикам определить, является ли востребованной эта возможность и стоит ли её поддержку перенести в действующую модель поисковой системы базы знаний.

7 Заключение

Авторы предложили свой взгляд на организацию информационного поиска в базе знаний научного института на базе адаптивной фасетной навигации. Реализационная модель позволит разработчикам исследовать вопрос - может ли предложенный подход повысить эффективность информационного поиска.

Литература

[1] Андрей Гребенюк, Сергей Киселев. Информационные системы управления знаниями компании / Журнал "Коммерческий директор" Москва, 2 апреля 2007 года <http://www.komdir.ru>

[2] Ильин В. Д., Соколов И. А. Символьная модель системы знаний информатики в человеко-автоматной среде // Информатика и её применения. 2007 г. том 1. выпуск 1.

[3] Чочиа А.П., Соловьев И.В., Обухова О.Л., Бирюкова Т.К., Гершкович М.М. Модель адаптивной фасетной навигации в открытых электронных коллекциях // "Системы и средства информатики", выпуск 18, , Москва, Наука, 2008 .

[4] ГОСТ 7.73 — 96. МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ. ПОИСК И РАСПРОСТРАНЕНИЕ ИНФОРМАЦИИ

[5] Ранганатан Ш.Р. Классификация двоеточием. Основная классификация: Пер. с англ. / Под ред. Т.С. Гомолицкой и др. - М., 1970. – 422 с.

[6] Обухова О.Л., Гершкович М.М., Бирюкова Т.К., Соловьев И.В., Чочиа А.П. Открытые электронные коллекции с адаптивным визуальным интерфейсом фасетной навигации // Труды 10-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2008, Дубна, Россия, 2008

[7] Н.А. Маркова, О.Л. Обухова, И.В. Соловьев, А.П. Чочиа. Эффективная фасетная навигация в электронных коллекциях // Системы и средства информатики, № 17, Москва, Наука, 2007, С. 214-222

[8] Миллер Дж. Магическое число семь, плюс или минус два. - В кн.: Инженерная психология. М. 1964

The method for dynamic association of informational objects in knowledge base

Olga Obuhova, Maxim Gershkovich, Tatiana Biryukova, Ivan Soloviev, Anton Chochia

We propose the formal model for the process of informational search in the knowledge base utilizing features of adaptive facet navigation. The 'search image' of the object is build as object's facet formula. Generation of the search request is conducted as a sequence of iterations, where search request being adjusted in dialogue mode using both direct and associative links between informational objects. The formal model is used to develop executive model, where visual interface shows content of the knowledge base and serves as a suitable tool for generation of the search requests.