

Математические модели оптимальной систематизации фактографической информации в электронных библиотеках

С.А. Пиявский, В.В. Ворошилов

Самарский государственный архитектурно-строительный университет
spiyav@mail.ru, vitekio@yandex.ru

Аннотация

Рассматривается новый подход к вопросу систематизации и использования фактографической информации в информационном пространстве. Данный подход предлагается как альтернатива существующим разработкам в области представления и систематизации знаний, в частности информационно-поисковым системам. Описывается технология накопления, обработки, систематизации и выдачи информации. Описываются математические модели оптимальной систематизации фактографической информации.

1. Постановка проблемы фактографического поиска

В условиях информационного общества основным источником информации для специалиста в сфере науки, техники и экономики является информация из интернета, при этом все большую роль играет фактографическая информация. Понятие фактографической информации в настоящее время находится в стадии становления и допускает различные близкие интерпретации. К примеру, в [1] дается следующее определение фактографической информации: «К фактографическим данным относятся сведения, извлеченные из документов, как первичных, так и вторичных и получаемые непосредственно из источников их возникновения». В [2] под фактографической информацией называется информация об объектах в тексте на естественном языке, которые однозначно соотносятся с объектами, существующими в действительности. В зарубежных словарях дается более широкое понятие фактографической информации, например в словаре «Macmillan» [3] под фактографической информацией понимается информация, основанная на фактах или содержащая только факты, а не теории или мнения.

пополнение БД с выявлением тавтологической и

Интернет и современное общество: сборник научных статей XVI Всероссийской объединенной конференции IMS-2013, Санкт-Петербург, 9 - 11 октября 2013 г.

Под фактографической информацией будем понимать информацию, представленную в виде числа (с указанием размерности) с набором терминов, поясняющих объект, характеризуемый этим числом. Такое понимание имеет важное практическое значение, позволяет достаточно просто решить самостоятельную сложную задачу выделения самого факта наличия фактографической информации в тексте на естественном языке. Кроме того, разработанные нами алгоритмы не используют существенно числовой характер фактографической информации и поэтому могут быть использованы и в рамках других интерпретаций.

Между тем специальные средства для поиска такой информации лишь разрабатываются (например, wolfram [4]) и тем более не включены в учебный процесс в вузах.

В настоящей работе авторы поставили перед собой задачу разработать специальный инструментальный для поиска и систематизации фактографической информации.

2. Технология фактографического поиска на основе формирования атомарной базы данных

В основе фактографической информации лежат количественные данные описания фактов, сгруппированных по определенным системообразующим признакам. Таким образом, элементом, единицей фактографической информации является «атом», состоящий из набора числового значения, размерности и набора лексем. Лексема представляет собой словоформу (термин, понятие) используемое из семантического тезауруса.

Таким образом, в основе автоматизированного фактографического поиска должна лежать специальным образом организованная база данных фактографической информации (АтБД), представляющая собой совокупность атомов фактографической информации [5]. Система управления такой АтБД должна обеспечивать

противоречивой информации, а также выдачу, по определенному запросу пользователя, релевантной фактографической информации как в не систематизированном виде (набор атомов упорядочения по релевантности), так и

систематизированной (в виде наиболее плотно заполненных двумерных таблиц). Использование такой формы фактографического поиска требует специального обучения студентов, а это обучение может проходить достаточно эффективно лишь с использованием специальной информационной системы, позволяющей контролировать работу студентов, как на аудиторных занятиях, так и во внеучебное время.

На основе вышеизложенной структуры данных была реализована атомарная база знаний. Имеются три канала пополнения АТБД. Первый из них - пополнение службой сопровождения БД, оно включает в себя формирование и пополнение базовых наборов лексем характеризующие различные предметные области, т.е. прямое пополнение тезауруса системы базовым набором лексем характеризующие предметную область.

Второй канал пополнения — это автоматизированный поиск фактографической информации, ее индексирование и ввод в АТБД, который состоит из двух основных этапов. На первом этапе проводится отбор информационных ресурсов из интернета (документов, файлов, веб-страниц) по прямым ссылкам, указанным методологической службой сопровождения, составленных на основе поисковой деятельности пользователей системы, а также отбор с использованием известных поисковых систем и сервисов, по поисковым запросам, составленным методологической службой сопровождения и состоящих из набора лексем описывающих предметную область. На втором этапе проводится анализ информационного ресурса полученного в результате поиска, а затем выделение атомов фактографической информации их индексирование и сохранение в АТБД.

Третий канал - это пополнение АТБД в процессе ее использования: пользователями в процессе своей профессиональной поисковой деятельности, студентами по различным дисциплинам в процессе учебной поисковой деятельности, а также студентами при обучении фактографическому поиску на основе АТБД. Пополнение происходит на основе формирования запросов к методологической службе сопровождения о расширении базы поиска в виде указания: поисковых запросов, предметной области, прямых ссылок на документы или веб-страницы. А также в виде указания набора атомов фактографической информации, с указанием источника и дополнительной метаинформацией, составленных студентами в процессе обучения фактографическому поиску на основе АТБД.

На рисунке 1 показана схема использования атомарной базы знаний для поиска фактографической информации.

На основе изложенного выше подхода была разработана и реализована информационная система организации и управления фактографической поисковой деятельностью. Данная система реализована в виде web-приложения с клиентской частью, обеспечивающей дополнительные возможности.

Разработанная система логически разделена на три подсистемы:

- подсистема формирования и пополнения АТБД;
- подсистема поиска и выдачи фактографической информации по запросу;
- подсистема мониторинга и управления учебно-поисковой деятельности студентов в процессе обучения.



Рис. 1. Использование атомарной базы знаний для поиска фактографической информации

Целесообразность включения третьей подсистемы определяется тем, что понятие фактографического поиска недостаточно включено в функционал обычной компьютерной грамотности. Между тем, сбор и анализ фактографической информации, особенно для направлений подготовки специалистов в области естественных, технических и экономических наук приобретает все большее значение. Использование при этом математического аппарата и соответствующих информационных систем в условиях развитой инфокоммуникационной среды вуза является прогрессивной тенденцией [6-10].

Физически информационная система организации и управления фактографической поисковой деятельностью студентов в интернете состоит из трех модулей и двух баз данных:

- клиентская часть поиска фактографической информации – используется непосредственно для поиска и анализа информации в интернете, а также фактографического поиска в АТБД;
- веб-сайт для работы с фактографической информацией:
 - всеми пользователями для поиска фактографической информации в АТБД;
 - методологической и административной службой для формирования и пополнения АТБД;
 - преподавателями и студентами для контроля и управления поисковой деятельностью в учебном процессе вуза;
- атомарная база данных;
- база данных учебного процесса.

Подсистема формирования и пополнения АТБД предоставляет следующие возможности:

- предоставляет интерфейс ввода фактографической информации, как в виде одного, так и множества атомов;
- предоставляет интерфейс ввода фактографической информации в виде файлов, поддерживаемые типы файлов: файлы MS Excel, MS Word, Html, PDF.

Использование АТБД происходит по средствам использования следующих функций:

- поиск и выдача фактографической информации по запросу:
 - предоставляет интерфейс поиска фактографической информации в виде указания поисковой строки.
 - предоставляет интерфейс поиска фактографической информации в виде указания набора поисковых лексем.
- вывод результата поиска в виде последовательного набора атомов (сортировка по имени, дате, релевантности).
- вывод результата поиска в виде набора наиболее плотных двумерных таблиц.

Обычная технология использования системы состоит в следующем.

Вначале пользователю необходимо авторизоваться в системе.

Затем указать поисковый запрос в виде набора лексем из тезауруса системы или в виде строки, при этом строка поискового запроса автоматически преобразуется в набор лексем из тезауруса системы. При указании поискового запроса существует возможность указать запрос один из сохраненных, ранее используемых пользователем.

Далее следует задать лимитные параметры выдачи фактографической информации из АТБД, такие как:

- максимальное значение количества атомов в выдаваемом результате;
- минимальное значение количества атомов в выдаваемом результате;
- максимальное значение количества лексем при поиске информации в АТБД;
- минимальное значение количества лексем при поиске информации в АТБД.

После указания параметров выдачи проводится поиск в АТБД и в результате представляется фактографическая информация в виде набора максимально заполненных двумерных таблиц. Эти таблицы содержат наиболее плотно-заполненные кластеры информации (под кластером информации понимается набор атомов). Для количественной оценки степени заполненности кластера введем понятие его информативности. Под информативностью лексемы относительно кластера будем понимать отношение числа атомов, в описании которых содержится данная лексема к общему числу атомов кластере. Информативностью кластера будем называть среднюю информативность относительного данного кластера всех лексем входящих в описание хотя бы одного из атомов этого кластера. Кластер будем называть замкнутым, если входящие в его атомы лексемы не входят в описание других атомов, т.е. не входящих в кластер.

Затем пользователем имеет возможность провести анализ информации в таблицах на предмет их наполненности. Выявив недостающую информацию, по пустотам в таблицах, рекомендуется провести ее поиск с использованием известных поисковых средств и сервисов и добавить найденную новую информации в АТБД для его пополнения.

3. Математические модели оптимальной систематизации фактографической информации.

3.1 Математическая модель формирования кластера наибольшей информативности

Введем следующие обозначения:

N – число атомов в БД;

M – число лексем в БД;

U_j – признак включения лексемы в описание кластера;

V_i – признак включения атома в кластер.

Тогда:

$\sum_{i=1}^N V_i$ – число атомов в кластере;

$\sum_{j=1}^M U_j$ – число лексем в описании кластера;

n_j – число атомов, в описание которых входит лексема j ;

$\sum_{i=1}^N V_i a_{ij}$ – число атомов в кластере, в описание которых входит лексема j .

Задача: найти кластер максимальной удельной информативности на один атом при заданной длине его описаний m .

Это задача сводится к задаче нелинейной оптимизации со следующими целевой функцией и ограничениями:

$$F = \frac{1}{\sum_{j=1}^M U_j \sum_{i=1}^N V_i} \sum_{i=1}^N \sum_{j=1}^M a_{ij} V_i U_j \rightarrow \max$$

$U_j \geq a_{ij} V_i$ – условие, что если лексема не выбрана, ее не должно быть ни в одном атоме кластера.

Лимитные условия на размер получаемого кластера:

$$\sum_{j=1}^M U_j \leq \max_{lex}, \quad \sum_{j=1}^M U_j \geq \min_{lex},$$

$$\sum_{i=1}^N V_i \leq \max_{atom}, \quad \sum_{i=1}^N V_i \geq \min_{atom}.$$

Рассмотрим пример данной модели.

Исходные данные:

$N = 10$ – число атомов;

$M = 8$ – число лексем.

Распределение значений между атомами и лексемами показано на таблице 1, где строка это соответствующий атом ($i \in [1, N]$), а столбец соответствующая лексема ($j \in [1, M]$), т.е. например, атом с порядковым номером 1 описывается лексемами 1, 4, 7 и 8, а атом 10 описывается лексемами 3, 5 и 8. Средняя информативность исходной таблицы равна 0,39.

В результате решения данной задачи получился кластер, показанный в таблице 2, с максимальной удельной информативностью на один атом равной 0,58, вошедшими атомами 1, 3, 7, 8 и лексемами 1, 2, 4, 5, 7, 8.

Варьируя максимальными и минимальными ограничениями на размер результирующего кластера можно получить кластеры разной размерности и разной удельной информативности. В таблице 3 показана зависимость полученной в результате удельной информативности кластера от ограничений на размер результирующего кластера, для данной задачи.

Таблица 1. Исходная таблица описания атомов набором лексем, информативность 0,39

		Лексемы							
		1	2	3	4	5	6	7	8
Атомы		Петя	Вася	Лев	Вес	Рост	IQ	Юность	Журнал
1	60 кг	1			1			1	1
2	120 ед	1					1	1	
3	50 кг		1		1			1	1
4	110 ед		1				1	1	
5	70 кг			1	1			1	
6	180 см	1				1			
7	170 см	1				1			1
8	150 см		1			1		1	
9	130 ед			1			1		1
10	190 см			1		1			1

Таблица 2. Полученный замкнутый кластер наибольшей информативности, информативность 0,58

		Лексемы						
		1	2	4	5	7	8	
Атомы		Петя	Вася	Вес	Рост	Юность	Журнал	
1	60 кг	1		1		1	1	
3	50 кг		1	1		1	1	
7	170 см	1			1		1	
8	150 см		1		1	1		

Таблица 3. Зависимость информативности кластера от ограничений

Ограничения				Информативность
L_{min}	A_{min}	L_{max}	A_{max}	
2	2	4	4	0,83
2	2	6	6	0,83
4	4	6	6	0,58
4	4	6	8	0,54
4	6	6	8	0,43

Из таблицы 3 видно, что чем меньше кластер, тем больше информативность, это объясняется тем, что больший, по размерности, кластер имеет больше лексем, соответственно для того чтобы его информативность стремилась к единице необходимо существование большего количества значений между всеми этими лексемами. Отметим, также, что наибольшая информативность, равная единице, будет достигаться только для маленьких, по размерности, кластеров, потому что для получения данного результата необходимо, чтобы для каждого атома из кластера существовали все лексемы включенные в кластер.

3.2 Математическая модель выделения 2-мерной таблицы наибольшей информативности

Для удобства представления информации содержащейся в АТБД большое значение имеет возможность использования табличной формы описания кластера. В такой таблице в заголовке строк и столбцов должны содержать название лексем, а клетки – числа, входящие в атом информации, описание которого состоит из лексем отвечающих строке и столбцу на пересечении которых находятся клетки. Ясно, что такими таблицами могут описываться лишь атомы, в описании которых входит не более двух лексем.

За основу возьмём модель, описанную выше, и добавим ограничение:

$$V_i \sum_{j=1}^M a_{ij} \leq 2 - \text{условие что выбираемые атомы}$$

содержат не более двух лексем.

Рассмотрим решение этой задаче в условиях предыдущего примера. В таблице 4 показан полученный с использованием математической модели замкнутый кластер наибольшей информативности, атомы которого содержат ровно две лексемы.

Таблица 4. Полученный замкнутый кластер наибольшей информативности, информативность 1

		Лексемы	
		1	5
Атомы		Петя	Рост
6	180 см	1	1

Только один атом из исходной таблице (таблица 1) имеет ровно две лексемы. Заметим, что он не является подмножеством максимально информативного кластера выделенного из АТБД без ограничения на число лексем в описание атомов, поскольку условия этих двух задач различны, хотя математическая модель решения второй задачи получается добавлением к математической модели первой задачи дополнительного условия, что сужает множество допустимых вариантов.

В таблице 5 показано описание полученного кластера в виде двумерной таблицы.

Таблица 5. Описание полученного кластера в виде двумерной таблицы

	Рост
Петя	180 см

3.3 Математическая модель выделения 2-мерной таблицы наибольшей информативности с уточняющими лексемами

Требование представления фактографической информации в виде двумерных таблиц достаточно жестко, поскольку распространяется только на атомы, описание которых содержит не более двух лексем. Его можно существенно смягчить, если допустить чтобы в клетках двумерной таблицы указывалось не только численное значение, но и перечислялись лексемы входящие в описание соответствующего атома помимо лексем, записанных в строке, столбце клетки таблицы. Естественно что общее число таких уточняющих лексем в таблице должно быть ограничено для того чтобы она носила «удобочитаемый» вид.

Пусть:

Z_j – признак включения лексемы в кластер как уточняющей лексемы (т.е. она не используется в заголовках строк, столбцов, но является поясняющей). Тогда, для удобства описания лексемы, включаемые признаками U_j назовём базовыми.

$U_j + Z_j \leq 1$ – условие того что уточняющая лексема не может входить в название строки или столбца таблицы

$$U_j + Z_j \geq \frac{1}{N} \sum_{i=1}^N a_{ij} V_i - \text{условие, что если}$$

описание хотя бы одного атома из кластера содержит лексему j , то она должна быть включена в кластер, как базовая или уточняющая.

$$\sum_{j=1}^M Z_j \leq Z_{max} - \text{условие что общее количество}$$

уточняющих лексем ограничено величиной Z_{max} .

Беря за основу модель, описанную выше, необходимо заменить ограничение количества лексем в атомах условием, что выбираемые атомы содержат не более двух базовых лексем:

$$V_i \sum_{j=1}^M a_{ij} U_j \leq 2$$

Рассмотрим решение этой задаче в условиях предыдущего примера. В таблице 6 показан результирующий набор атомов и лексем замкнутой двумерной таблицы, полученных при решении задачи со следующими начальными условиями: $L_{min} = 5$, $A_{min} = 5$, $L_{max} = 8$, $A_{max} = 10$ и $Z_{max} = 3$. Отметим, что минимальные и максимальные ограничения на размер атома (L_{min} , A_{min} , L_{max} , A_{max}) имеют средние значения. В полученной таблице курсивом отмечены уточняющие лексемы.

Таблица 6. Полученный замкнутый кластер наибольшей информативности

		Лексемы						
		1	2	3	4	6	7	8
Атомы		Петя	Вася	Лев	Вес	IQ	Юность	Журнал
1	60 кг	1			1		1	1
2	120 ед	1				1	1	
3	50 кг		1		1		1	1
4	110 ед		1			1	1	
5	70 кг			1	1		1	

Из этого набора можно составить двумерную таблицу, где уточняющие лексемы будут вынесены непосредственно в ячейки, вместе со значениями, таблица 7.

Таблица 7. Описание полученного кластера в виде двумерной таблицы

	Вес	IQ
Петя	60 кг (<i>Юность, Журнал</i>)	120 ед
Вася	50 кг (<i>Юность, Журнал</i>)	110 ед
Лев	70 кг (<i>Юность</i>)	

Заметим, что на полученный результат оказывает сильное влияние ограничение на максимальное количество уточняющих лексем в кластере, параметр Z_{max} . Рассмотрим решение этой задаче в условиях различных значений этого ограничения.

При разрешённой одной уточняющей лексеме ($Z_{max} = 1$), и не менее одного атома в кластере ($A_{min} = 1$) мы получили кластер с максимальной информативностью равной единице, при этом полученный кластер и двумерная таблица имеют вид, показанный в таблицах 8, 9.

Таблица 8. Полученный замкнутый кластер наибольшей информативности при $Z_{max} = 1$ и $A_{min} = 1$

		Лексемы		
		1	6	7
Атомы		Петя	IQ	Юность
2	120 ед	1	1	1

Таблица 9. Описание полученного кластера в виде двумерной таблицы при $Z_{max} = 1$ и $A_{min} = 1$

	IQ
Юность	120 ед (<i>Петя</i>)

При разрешённой одной уточняющей лексеме ($Z_{max} = 1$), и не менее двух атомов в кластере ($A_{min} = 2$) мы получили кластер с максимальной информативностью равной $0,67$. Результаты показаны в таблицах 10 и 11.

Таблица 10. Полученный замкнутый кластер наибольшей информативности при $Z_{max} = 1$ и $A_{min} = 2$

		Лексемы			
		3	5	6	8
Атомы		Лев	Рост	IQ	Журнал
9	130 ед	1		1	1
10	190 см	1	1		1

Таблица 11. Описание полученного кластера в виде двумерной таблицы при $Z_{max} = 1$ и $A_{min} = 2$

	Рост	IQ
Лев	190 см (<i>Журнал</i>)	130 ед (<i>Журнал</i>)

Отметим, что информативность полученного кластера составляет $0,67$, при этом получилась полностью заполненная двумерная таблица.

При значениях максимального количества уточняющих лексем в кластере равном двум или более ($Z_{max} \geq 2$), это ограничение не оказывает влияние на результирующий набор атомов и лексем. Но варьируя минимальными параметрами ограничениями на количество атомов и лексем в результирующем кластере (L_{min} , A_{min}) можно получить таблицы, как разной размерности и удельной информативности, так и другим содержанием лексем и атомов. В таблице 12 показана зависимость построенных двумерных таблиц полученных в результате расчета удельной информативности кластера от минимальных ограничений на размер результирующего кластера, для данной задачи, при $Z_{max} = 3$.

Заметим, что максимальные ограничения на количество атомов и лексем в результирующем кластере (L_{max} , A_{max}) не оказывают существенного влияния на результирующий размер кластера.

Таблица 12. Описание полученного кластера в виде двумерной таблицы при разных параметрах модели

Ограничения	Информативность	Результирующая двумерная таблица										
$L_{min}=2, A_{min}=2$	0,5	<table border="1"> <tr> <td></td> <td>Юность</td> <td>Журнал</td> </tr> <tr> <td>Вася</td> <td>110 ед (IQ)</td> <td></td> </tr> <tr> <td>Лев</td> <td></td> <td>190 см (Рост)</td> </tr> </table>		Юность	Журнал	Вася	110 ед (IQ)		Лев		190 см (Рост)	
	Юность	Журнал										
Вася	110 ед (IQ)											
Лев		190 см (Рост)										
$L_{min}=3, A_{min}=2$	0,67	<table border="1"> <tr> <td></td> <td>IQ</td> <td>Рост</td> </tr> <tr> <td>Петя</td> <td>120 ед (Юность)</td> <td>180 см, 170 см (Журнал)</td> </tr> </table>		IQ	Рост	Петя	120 ед (Юность)	180 см, 170 см (Журнал)				
	IQ	Рост										
Петя	120 ед (Юность)	180 см, 170 см (Журнал)										
$L_{min}=4, A_{min}=3$	0,67	<table border="1"> <tr> <td></td> <td>Журнал</td> <td>Юность</td> </tr> <tr> <td>Юность</td> <td>60 кг, (Вес, Петя) 50 кг (Вес, Вася)</td> <td></td> </tr> <tr> <td>Рост</td> <td>170 см (Петя)</td> <td>150 см (Вася)</td> </tr> </table>		Журнал	Юность	Юность	60 кг, (Вес, Петя) 50 кг (Вес, Вася)		Рост	170 см (Петя)	150 см (Вася)	
	Журнал	Юность										
Юность	60 кг, (Вес, Петя) 50 кг (Вес, Вася)											
Рост	170 см (Петя)	150 см (Вася)										
$L_{min}=4, A_{min}=4$	0,5	<table border="1"> <tr> <td></td> <td>Рост</td> </tr> <tr> <td>Петя</td> <td>180 см, 170 см (Журнал)</td> </tr> <tr> <td>Вася</td> <td>150 см (Юность)</td> </tr> <tr> <td>Лев</td> <td>190 см (Журнал)</td> </tr> </table>		Рост	Петя	180 см, 170 см (Журнал)	Вася	150 см (Юность)	Лев	190 см (Журнал)		
	Рост											
Петя	180 см, 170 см (Журнал)											
Вася	150 см (Юность)											
Лев	190 см (Журнал)											

3.4 Математическая модель выделения замкнутой 2-мерной таблицы наибольшей информативности с вынесением уточняющих лексем в шапку

Введем q_j – признак вынесения уточняющей лексемы в шапку.

$a_{ij}V_i \geq Z_jq_j$ – условие что для любого атома существует лексема, выносимая в шапку.

Рассмотрим решение данной модели на примере описанном выше. В результате решение данной задачи при различных начальных условиях, ограничениях на размер результирующего кластера ($L_{min}, A_{min}, L_{max}, A_{max}, Z_{max}$) было получено только одно решение, с информативностью кластера равной 0,49, показанное в таблицах 13 и 14. В таблицах курсивом отмечены уточняющие лексемы, жирным лексемы, вынесенные в шапку таблицы

Таблица 13. Полученный замкнутый кластер наибольшей информативности

		Лексемы						
		1	2	3	4	6	7	8
Атомы		Петя	Вася	Лев	Вес	IQ	Юность	<i>Журнал</i>
1	60 кг	1			1		1	1
2	120 ед	1				1	1	
3	50 кг		1		1		1	1
4	110 ед		1			1	1	
5	70 кг			1	1		1	

Таблица 14. Описание полученного кластера в виде двумерной таблицы

Юность		
	Вес	IQ
Петя	60 кг (<i>Журнал</i>)	120 ед
Вася	50 кг (<i>Журнал</i>)	110 ед
Лев	70 кг	

Литература

[1] Фактографический поиск, Википедия свободная энциклопедия, 2012 г. URL: <http://ru.wikipedia.org/> (дата обращения: 07.03.2012).

[2] Извлечение фактографической информации, 2012 г. URL: <http://lksystems.ru/factography.aspx> (дата обращения: 03.04.2012).

[3] Фактографическая информация, словарь Macmillan, 2012 г. URL: <http://www.macmillandictionary.com/> (дата обращения: 12.03.2012).

- [4] Поисковая система WolframAlpha, 2012 г. URL: <http://www.wolframalpha.com/> (дата обращения: 14.03.2012).
- [5] Gallagher S. Using The Knowledge Management Maturity Model (KM3) As An Evaluation Tool, 2010 г. URL: <http://bprc.warwick.ac.uk/km028.pdf> (дата обращения: 29.03.2010).
- [6] Пиявский С.А. Математическое моделирование управляемого развития научных способностей // Известия Академии наук. Серия «Теория и системы управления». 2000. №3. С. 100—106.
- [7] Пиявский С.А., Савельева Г.П. Система управления формированием универсальных компетенций студентов высших учебных заведений. Самара, М.: Исследовательский центр проблем качества подготовки специалистов. 2009. 105 с.
- [8] Пиявский С.А. Исследовательская деятельность студентов в инновационном вузе: учебник. / СГАСУ. Самара: 2011. 198 с.
- [9] Гаврилова Т.А. Онтологический подход к управлению знаниями при разработке корпоративных информационных систем // Новости искусственного интеллекта. 2003. № 2. С. 24—30.
- [10] Бабанин Л.Н. Организация дистанционного обучения на основе ресурсов Интернета / Л.Н. Бабанин, А.Е. Войскунский, С.А. Козловский. М.: Гос. Университет русского языка им. А.С.Пушкина, 2005.

Mathematical models of factual information optimal systematization in electronic libraries

Pijavskij S.A., Voroshilov V.V.

Consider a new approach to organize and use factual information in the information space. This approach is proposed as an alternative to existing development in the knowledge representation and systematization, in particular information search systems. The article describes the special technology of information storage, processing, systematization and representation. The article describes the mathematical models of factual information optimal systematization.