

В. Б. Барахнин, И. С. Пастушков

*Институт вычислительных технологий СО РАН
пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия*

*Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия*

bar@ict.nsc.ru, pas2shkov.ilya@gmail.com

ТЕХНОЛОГИЯ АВТОМАТИЗИРОВАННОГО НАПОЛНЕНИЯ ОНТОЛОГИИ ФАКТОГРАФИЧЕСКОЙ ПОИСКОВОЙ СИСТЕМЫ *

В работе излагается технология автоматизированного наполнения онтологии фактографической поисковой системы. Суть технологии заключается в извлечении ключевых слов (словосочетаний) из корпуса текстов одно-родной тематики с целью дальнейшего использования извлеченных ключевых слов в качестве возможных значений атрибутов сущностей, описываемых в создаваемой онтологии предметной области, предназначенной для организации фактографического поиска в расширенном корпусе текстов соответствующей тематики. Предлагаемая технология основана на применении метода опорных векторов для разметки в текстах частей речи с последующим использованием метода случайных блужданий для извлечения семантически связанных ключевых слов (словосочетаний). К набору этих словосочетаний с целью отнесения конкретного словосочетания к определенному атрибуту описываемой в тексте сущности применяется обученная нейронная сеть со скрытым слоем. Таким образом, по набору семантически связанных пар слов строится онтология для конкретного документа, формирующаяся при работе нейронной сети, и далее с использованием СУБД на основе полученных данных организуется поиск.

Ключевые слова: факт, фактографический поиск, интеллектуальные системы, извлечение фактов, автоматизированное наполнение онтологий.

Введение

Важным этапом процесса функционирования фактографических систем является извлечение из текстов документов содержащихся в них фактов, то есть, в наиболее общем смысле, *«особого рода предложений, фиксирующих эмпирическое знание»*. Как показано в [1], при создании фактографических информационных систем разумно следующее понимание факта: *содержащаяся в тексте и метаданных документа совокупность связей между сущностями, описываемыми в онтологии информационной системы*. Такое понимание факта опирается на «Логико-философский трактат» Л. Витгенштейна [2], при этом атомарному факту, который, согласно [2], «есть соединение объектов (вещей, предметов)», соответствует единичная связь. Практическая реализация такого подхода опирается на модель множества сущностей [3], отличительные особенности которой заключаются в том, что, во-первых, в ней всё трактуется как объекты (в том числе, например, цвет), а, во-вторых, все связи в этой

* Работа выполнена при частичной поддержке РФФИ (проект 13-07-00258) и президентской программы «Ведущие научные школы РФ» (грант 5006.2014.9).

модели – бинарные, причем связи между объектами также рассматриваются как объекты, связанные, в свою очередь, с объектами – атрибутами связей. Таким образом, на практике в качестве атомарного факта можно рассматривать входящую в текст и в метаданные документа характеристику сущности, описываемую в онтологии информационной системы. Такая характеристика представляется как единичное значение данных.

Прежде всего, уточним, какого именно понимания термина «онтология» мы будем придерживаться в данной работе.

В работе [4] было проведено (применительно к рассматриваемой предметной области) установление определенности в понимании и разграничении использования терминов «тезаурус» и «онтология». Более или менее однозначное трактование термина «тезаурус» сложилось еще в конце 1960-х годов [5]: это «словарь-справочник, содержащий все лексические единицы информационно-поискового языка – дескрипторы (вместе с ключевыми словами, которые в пределах данной информационно-поисковой системы считаются синонимами этих дескрипторов), причем дескрипторы в словаре должны быть систематизированы по смыслу, а смысловые связи между ними эксплицитно выражены».

Что же касается термина «онтология», в настоящее время, как отмечено в [6], под онтологией нередко стали понимать широкий спектр структур, представляющих знания о той или иной предметной области с разной степенью формализации [7]:

- 1) словарь с определениями;
- 2) простая таксономия;
- 3) тезаурус (таксономия с терминами);
- 4) модель с произвольным набором отношений;
- 5) таксономия и произвольный набор отношений;
- 6) полностью аксиоматизированная теория.

В [4] было показано, что тезаурус становится онтологией тогда, когда связи между дескрипторами не просто эксплицированы (как это предусмотрено в классическом определении тезауруса), но и классифицированы универсальными зависимостями типа «общее – частное», «часть – целое», «причина – следствие» и т. п. (см., например, [8]). Конечно, это – лишь «нижняя граница» сложности онтологии. Для эффективной работы с фактами следует, чтобы сущности, относящиеся к предметной области, были представлены не только обозначающими их терминами, но и достаточно широким набором атрибутов, т. е. речь идет об онтологии, обладающей известными признаками модели предметной области.

Разумеется, на первоначальном этапе создания фактографической системы речь, как правило, идет о создании лишь каркаса онтологии, содержащего только краткие сведения о сущностях, а их более подробное описание будет происходить в процессе функционирования фактографической системы посредством извлечения из документов соответствующих фактов, выступающих в качестве тех или иных атрибутов сущностей.

Наиболее сложной проблемой, возникающей при создании фактографических систем, является разработка методик автоматизированного извлечения фактов из документов на естественном языке. Эта проблема до сих пор, по-видимому, не имеет сколько-нибудь общего решения, поскольку построение такого решения предполагает, в частности, достаточно точное моделирование когнитивной деятельности человека, а также наличие мощных средств как синтаксического, так и семантического анализа текстов, включая подробнейшие онтологии, тезаурусы которых учитывают, например, всё богатство синонимии естественного языка (не столько даже в части научной лексики, сколько в части лексики общеупотребительной).

Целью данной статьи является описание технологии извлечения ключевых слов (словосочетаний) из корпуса текстов однородной тематики с целью дальнейшего использования извлеченных ключевых слов в качестве возможных значений атрибутов сущностей, описываемых в создаваемой онтологии предметной области, предназначенной для организации фактографического поиска в расширенном корпусе текстов соответствующей тематики. Для достижения поставленной цели необходимо решить следующие задачи:

- 1) создать нейросеть, обученную извлечению лингвистических конструкций, в которые входят возможные значения атрибутов сущностей, описываемых в онтологии предметной области;

- 2) разметить текст, определяя, к какой части речи относится каждая лексема, входящая в текст;
- 3) извлечь из размеченного текста наборы семантически связанных слов;
- 4) используя коллекцию лингвистических конструкций и наборы семантически связанных слов, наполнить онтологию предметной области.

Особенностью предложенной технологии является возможность использования результатов её применения для автоматизированной добавления вики-разметки в слабоструктурированные документы, предназначенные для помещения в Википедию (или подобные ей системы).

Создание и обучение нейронной сети

Выполнение этого этапа сильно зависит от конкретной области, для которой создается онтология. Наиболее простой представляется ситуация, когда по тематике предметной области имеется корпус текстов с википодобной разметкой, используемый для обучения нейронной сети, извлекающей лингвистические конструкции, в которые входят возможные значения атрибутов сущностей, описываемых в онтологии предметной области. Мы остановили свой выбор на нейронной сети со скрытым слоем [9], однако реализация скрытого слоя не была инкапсулирована, в силу чего его можно использовать как справочную таблицу, в которой хранятся искомые лингвистические конструкции. Именно поэтому данный метод хорошо подходит для решения задачи извлечения фактов.

Сеть реализована на основе перцептронов. – простых нейронов, каждый из которых содержит элементы 3-х типов: S-элементы, A-элементы и один R-элемент. S-элементы – это сенсоры, каждый из которых может находиться в состоянии покоя или возбуждения, причем в последнем случае он передает сигнал в слой ассоциативных A-элементов, совокупность которых образует. Каждому A-элементу соответствует целый набор (ассоциация) S-элементов, причем A-элемент активизируется, как только количество сигналов от S-элементов на его входе превысит некоторую пороговую величину. Сигналы от активизированных A-элементов передаются (возможно, с весовыми коэффициентами) в сумматор R, подсчитывающий сумму значений входных сигналов. Элементарный перцептрон активизируется, если указанная сумма превышает заданный порог.

Обучение элементарного перцептрона состоит в изменении весовых коэффициентов на выходе A-элементов.

Приведем описание элементарного перцептрона на языке Python:

class Neuron:

```
def __init__(self, weights_type):
    self.weights_type = weights_type      #весовые коэффициенты
    self.a = 0                            #сигнал на выходе нейрона
    self.threshold = 0                    # пороговое значение
    self.error = 0                        #значение ошибки
    self.change = self.weights_type      #используется в процессе обучения
    self.t_change = self.weights_type    #аналогично
```

Таким образом, входной набор данных для обучения сети, подается на ее входной слой, и сеть функционирует в нормальном режиме (т. е. вычисляет выходные данные).

Полученные данные сравниваются с известными выходными данными для рассматриваемого входного набора. Известные входные данные – это лексические конструкции, например, для биографий: “родился *x*”, “появился на свет *x*” и т. п., где *x* – искомая характеристика (дата рождения). В итоге формируется булевозначный вектор ошибки, который используется для модифицирования весовых коэффициентов выходного слоя с тем, чтобы при повторной подаче того же набора входных данных вектор ошибки уменьшался. Стоит отметить, что в условиях данной задачи весовые коэффициенты являются булевозначными, равно как

и вектор ошибки. В рассматриваемой предметной области расстояние Левенштейна между словами не будет являться релевантным показателем улучшения работы алгоритма.

Весовые коэффициенты скрытого слоя модифицируются путем сравнения выходных сигналов нейронов скрытого слоя и входных сигналов нейронов выходного слоя.

Процесс обучения сети осуществляется посредством предъявления каждого входного набора данных и последующего распространения ошибки. Этот цикл повторяется n раз. В рассматриваемой задаче при распознавании синонимов сначала обрабатывается синоним первой характеристики, синоним второй и так далее, затем весь цикл повторяется. Не следует поступать иначе, то есть обучать сеть по отдельности сначала синонимам первой характеристики (n раз), потом второй, третьей и т. д., так как сеть постепенно настраивается на коэффициенты последних экспериментов, забывая более ранние серии.

В итоге обученная сеть способна извлекать лингвистические конструкции, в которые входят возможные значения атрибутов сущностей, описываемых в онтологии предметной области.

Отметим, что в текущей реализации скрытый слой реализован в виде коллекции в MongoDB.

Разметка текста

Для разметки текстов нами был использован национальный корпус русского языка (<http://www.ruscorpora.ru/>), точнее, синтаксический корпус – применительно к задачам информационного поиска под ним обычно понимается так называемый словарь-справочник СинТагРус, содержащий слова и их лексические характеристики. Синтаксический корпус представляет собой дерево зависимостей, в узлах которого стоят слова, входящие в предложение, а ветви помечены именами синтаксических отношений. В отличие от морфологически размеченного фрагмента Национального корпуса русского языка, СинТагРус (<http://www.ruscorpora.ru/instruction-syntax.html>) целиком состоит из структур со снятой морфологической и синтаксической омонимией. Это означает, что каждому слову текста сопоставляется ровно морфологическая структура, а каждому предложению ставится в соответствие ровно одна синтаксическая структура. СинТагРус представляет собой коллекцию текстов с разметкой для каждого слова, содержащей следующую информацию: часть речи, число, падеж, время глагола и т. д. В качестве алгоритма обучения нами был использован метод опорных векторов (SVM). Метод SVM функционирует как черный ящик, на вход которого подаются характеристики данных, а на выходе выдается классификация по заранее заданным категориям. В качестве характеристик в данной работе задаётся окончание слова, а в качестве категорий – части речи.

Алгоритм обучения имеет вид.

1. Считывается файл корпуса, и для каждого слова определяются его характеристики: само слово, окончание (две и три последние буквы), приставка (две и три первые буквы), а также части речи предыдущих слов.
2. Каждой части речи и характеристике присваивается порядковый номер и создается задача для обучения SVM.
3. Обучается модель SVM.
4. Обученная модель используется для определения части речи слов в предложении: для каждого слова нужно определить характеристики, описанные в п. 1., и подать на вход SVM-модели, которая подберет наиболее подходящий класс, т. е. часть речи. Таким образом, получая на вход предложение, программа определяет принадлежность слов к частям речи, и на выходе мы имеем размеченный текст для дальнейших манипуляций.

Извлечение наборов семантически связанных слов

Для извлечения онтологии необходимо, прежде всего, извлечь набор семантически связанных (ключевых) слов, при задании дескрипторов к которым и будет извлекаться онтология. Система извлечения семантически связанных слов должна основываться на обучаемом алгоритме без учителя, поскольку даже тексты авторов, работающих в одной предметной

области, могут обучить алгоритм противоположным вещам. В данной работе на основе морфологического анализа текста с помощью метода случайных блужданий извлекаются семантически связанные слова (словосочетания).

Одномерное дискретное случайное блуждание – это случайный процесс $\{Y_n\}_{n \geq 0}$ с дискретным временем, имеющий вид

$$\{Y_n\} = Y_0 + \sum_{i=1}^n X_i,$$

где Y_0 – начальное состояние; $X_i = \begin{cases} 1, p_i \\ -1, q_i \equiv 1 - p_i \end{cases}$, $p \in (0,1), i \in \mathbb{N}$; случайные величины

$Y_0, X_i, i = 1, 2, \dots$ совместно независимы.

Одномерное дискретное случайное блуждание является цепью Маркова с целыми состояниями, чье начальное распределение задается функцией вероятности случайной величины X_n а матрица переходных вероятностей имеет вид

$$P \equiv (p_{ij})_{i,j \in \mathbb{Z}} = \begin{pmatrix} \ddots & \ddots & \ddots & & & & \\ & q_{-1} & 0 & p_{-1} & & & \\ & & q_0 & 0 & p_0 & & \\ & & & q_1 & 0 & p_1 & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix},$$

то есть

$$\begin{aligned} p_{i,i+1} &\equiv P(X_{n+1} = i+1 | X_n = i) = p_i, \\ p_{i,i-1} &\equiv P(X_{n+1} = i-1 | X_n = i) = q_i, i \in \mathbb{Z}, \\ p_{i,j} &\equiv P(X_{n+1} = j | X_n = i) = 0, |i-j| \neq 1. \end{aligned}$$

Основываясь на данном подходе, будем рассматривать задачу извлечения наборов семантически связанных слов (подробное изложение алгоритма см. [10]).

Предположим, что у нас имеется текст, включающий в себя n документов:

$$D = (d_1, d_2, \dots, d_n)$$

В данном случае, в качестве документа при анализе текста целесообразно рассматривать отдельно взятое предложение. Обозначим через:

$$W = (w_1, w_2, \dots, w_n)$$

словарь нашего текста.

О каждом слове документа нам известны его морфологические характеристики (часть речи, род число, падеж и т. д.). Кроме того, для каждого документа мы можем подсчитать вес слова в данном документе (TF-IDF).

Анализируемый текст можно представить в виде гиперграфа $H = (V, E)$ в котором вершины $v \in V$ – это слова текста, гиперребра $e \in E$ – это документы нашего текста. Обозначим через $H \in R^{|V| \times |E|}$ матрицу смежности нашего гиперграфа:

$$h(v, e) = \begin{cases} 1, v \in e \\ 0, v \notin e \end{cases}$$

Тогда степень вершины v определяется как

$$d(v) = \sum_{e \in E} w(e)h(v, e).$$

Степень гиперребра e :

$$\delta(e) = \sum_{v \in V} (v, e) = |e|.$$

Пусть $w_e : E \longrightarrow R^+$ веса, приписываемые нашим гиперребрам (документам). В качестве такого веса может выступать, например, популярность документа. Каждой вершине гиперграфа мы также припишем набор весов, которые представляю собой веса TF-IDF для тех документов, в которые входит данное слово:

$$w(v_i | e) = \frac{tf(v, e)}{N(e)} \log\left(\frac{|D| + 0.5}{df(v_i) + 0.5}\right).$$

Вес вершины в данном случае представляет собой вектор:

$$w(v_i) = [w(v_i | e_{j_1}), w(v_i | e_{j_2}), \dots, w(v_i | e_{j_{d(v_i)}})].$$

Семантическую связь между вершинами (словами) нашего гиперграфа будем вычислять на основе модели случайных блужданий (random walk model) между вершинами гиперграфа. Предполагается, что случайный переход от вершины к вершине осуществляется в два этапа (на первом выбирается одно из гиперребер, смежных с нашей вершиной, на втором этапе случайным образом выбирается вершина данного гиперребра).

Таким образом, процесс случайных блужданий будет представлять собой дискретную цепь Маркова с конечным множеством состояний. Как известно, основным параметром такого процесса является матрица вероятностей переходов между состояниями цепи.

В нашем случае элементы матрицы переходов будут вычисляться по следующей формуле:

$$P(u, v) = \sum_{e \in E} w(e) \frac{h(u, e)}{\sum_{g \in E(x)} w(g)} \times \frac{h_w(v, e)}{\sum_{s \in e} h_w(s, e)}.$$

Здесь

$$h(v, e) = \begin{cases} 1, & v \in e \\ 0, & v \notin e \end{cases}$$

представляют собой элементы взвешенной матрицы смежности H_w .

Уравнение можно записать в векторно-матричном виде:

$$P = D_v^{-1} \cdot H \cdot W_e \cdot D_{ve}^{-1} \cdot H_w^T /$$

Здесь

D_v – диагональная матрица степеней вершин гиперграфа;

H – матрица смежности вершин гиперграфа;

D_{ve} – матрица коэффициентов TF-IDF;

H_w – взвешенная матрица смежности;

$(\cdot)^{-1}$ – операция обращения матрицы;

$(\cdot)^T$ – операция транспонирования матрицы.

Таким образом, на выходе получаем набор семантически связанных слов, на основе которого будет реализован алгоритм фактографического поиска.

Наполнение онтологии предметной области

Было решено опробовать алгоритм на биографиях учёных-математиков, так как биографии при произвольной структуре имеют сходную онтологию.

В качестве обучающей выборки были взяты статьи из Википедии при наличии в них источников данных, которые будут использоваться при обучении нейросети, так как из Википедии можно извлечь интересующую характеристику с помощью SPARQL-запроса, что позволяет легче обучить нейросеть. Для обучения нейронной сети со скрытым слоем было использовано около 1000 статей и примерно такое же количество извлечённых из соответствующих биографий наборов ключевых слов для обучения нейросети.

В качестве примера полного цикла работы алгоритма приводится биография С. Л. Соболева.

На первом этапе текст биографии размечается по частям речи. Затем с помощью метода случайных блужданий выделяются ключевые слова и словосочетания, ниже приведена часть из них:

соболев львович сергей, 6 октября 1908, Петербург, математическая физика, вычислительная математика, дифференциальные уравнения, академик, обобщённые функции, стеклов, Ленинградский университет

Далее нейросеть размещает данные по дескрипторам:

```
{
  "name": "сергей львович соболев",
  "date_of_birth": "6 октября 1908",
  "date_of_death": "3 января 1989",
  "alumni": "ЛГУ",
  "alumni_date": "1929",
  "publications": "Уравнения математической физики", "Некоторые применения функционального анализа в математической физике"
}
```

В таком виде документ попадает в базу данных MongoDB, которая ориентирована на хранение коллекций JSON-документов, в данном случае в коллекции объединены сущности одной предметной области.

Заключение

В работе предложена технология автоматизированного наполнения онтологии фактографической поисковой системы. Суть технологии заключается в извлечении ключевых слов (словосочетаний) из корпуса текстов однородной тематики с целью дальнейшего использования извлеченных ключевых слов в качестве возможных значений атрибутов сущностей, описываемых в создаваемой онтологии предметной области, предназначенной для организации фактографического поиска в расширенном корпусе текстов соответствующей тематики.

Предлагаемая технология основана на применении метода опорных векторов для разметки в текстах частей речи с последующим использованием метода случайных блужданий для извлечения семантически связанных ключевых слов (словосочетаний). К набору этих словосочетаний с целью отнесения конкретного словосочетания к определенному атрибуту описываемой в тексте сущности применяется обученная нейронная сеть со скрытым слоем. Таким образом, по набору семантически связанных пар слов строится онтология для конкретного документа, формирующаяся при работе нейронной сети, и далее с использованием СУБД (в данном случае MongoDB) на основе полученных данных организуется поиск. С помощью встроенных индексов данной СУБД возможен поиск по нескольким документам.

Список литературы

1. *Барахнин В. Б., Федотов А. М.* Построение модели фактографического поиска // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, вып. 4. С. 16–27.
2. *Wittgenstein L.* Logisch-Philosophische Abhandlung // *Annalen der Naturphilosophie*. Leipzig: Verlag Unesma, 1921. Vol. 14. Parts 3/4. P. 185–262.
3. *Chen P. P.* The entity-relational model. Toward a unified view of data // *ACM TODS*. 1976. № 1. P. 9–36.
4. *Барахнин В. Б., Федотов А. М.* Уточнение терминологии, используемой при описании интеллектуальных информационных систем, на основе семиотического подхода // *Изв. вузов. Проблемы полиграфии и издательского дела*. 2008. № 6. С. 73–81.
5. *Михайлов А. И., Черный А. И., Гиляревский Р. С.* Основы информатики. М.: Наука, 1968.

6. Добров Б. В., Лукашевич Н. В., Синицын М. Н., Шапкин В. Н. Разработка лингвистической онтологии по естественным наукам для решения задач информационного поиска // Тр. VII Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2005). Ярославль, 2005. С. 70–79.

7. Welty C., McGuinness D., Uschold M., Gruninger M., Lehmann F. Ontologies: Expert Systems all over again // AAAI-1999 Invited Panel Presentation. 1999.

8. Нариньяни А. С. Кентавр по имени ТЕОН: Тезаурус + Онтология // Тр. междунар. семинара «Диалог'2001» по компьютерной лингвистике и ее приложениям. Аксаково, 2001. Т. 1. С. 184–188.

9. Хайкин С. Нейронные сети. Полный курс. 2-е изд., испр.: Пер. с англ. М.: ООО «И. Д. Вильямс», 2006.

10. Bellaachia A., Al-Dhelaan M. HG-RANK: A Hypergraph-based Keyphrase Extraction for Short Documents in Dynamic Genre. // Making Sense of Microposts (# Microposts2014). http://ceur-ws.org/Vol-1141/paper_06.pdf

Материал поступил в редколлегию 10.06.2015

V. B. Barakhnin, I. S. Pastushkov

*Institute of Computational Technologies SB RAS
6 Acad. Lavrentjev pr., Novosibirsk, 630090, Russia*

*Novosibirsk State University,
2 Pirogov Str., Novosibirsk, 630090, Russia*

bar@ict.nsc.ru, pas2shkov.ilya@gmail.com

TECHNOLOGY OF AUTOMATED FACTOGRAPHIC RETRIEVAL SYSTEM'S ONTOLOGY FILLING

This work is about technology of automated factographic retrieval system ontology filling. This technology contains extracting keywords from corpus of texts with similar topic for following using these keywords as possible values of entity's attributes, that describes in created ontology of subject field for organizing of factographic retrieval in expanded corpus of text appropriated topics. Suggested technology based on support vector model for stemming text and following random-walk method for extracting keywords. After learned hidden layer neural network works with set of these keywords. So, ontology for document formed in neural network working builds by the set of semantic connected pairs of words and after with use of database search organizing.

Keywords: fact, factographic retrieval, intelligent systems, facts extraction, automated ontology filling.

References

1. Barakhnin V. B., Fedotov A. M. A model of factographic retrieval // Vestnik NSU: Information Technologies 2013. Vol. 11. N 4. P. 16–27. ISSN 1818-7900. (in Russian)

2. Wittgenstein L. Logisch-Philosophische Abhandlung // Annalen der Naturphilosophie. Vol. XIV. Parts 3/4. Leipzig: Verlag Unesma, 1921. P. 185–262.

3. Chen P. P. The entity-relational model. Toward a unified view of data // ACM TODS. 1976. № 1. P. 9–36.

4. Barakhnin V. B., Fedotov A. M. Clarification of the Terminology Used in the Description of Intellectual Information Systems, Based on the Semiotic Approach // Izvestiya VUZ: Problems of printing and publishing. 2008. № 6. P. 73–81.

5. Mikhailov A. I., Chernyi A. I., Gilyarevskiy R. S. Fundamentals of Informatics. Moscow: Nauka, 1968.
6. Dobrov B. V., Loukachevitch N. V., Sinitsyn M. N., Shapkin V. N. Development of Linguistic Ontology on Natural Sciences for Information Retrieval Purposes // Proceedings of the Seventh Anniversary of All-Russian Scientific Conference «Digital Libraries: Advanced Methods and Technologies, Digital Collections» (RCDL'2005). Yaroslavl, 2005. P. 70–79.
7. Welty C., McGuinness D., Uschold M., Gruninger M., Lehmann F. Ontologies: Expert Systems all over again // AAAI-1999 Invited Panel Presentation. 1999.
8. Narin'yani A. S. A centaur by name of «THEON»: Thesaurus + Ontology // Proceedings of the DIALOG'2001 International Workshop. Aksakovo, 2001. Vol. 1. P. 184–188.
9. Haykin S. Neural Networks: A Comprehensive Foundation, 2nd Edition. NJ: Prentice-Hall, 1998.
10. Bellaachia A., Al-Dhelaan M. HG-RANK: A Hypergraph-based Keyphrase Extraction for Short Documents in Dynamic Genre. // Making Sense of Microposts (# Microposts2014 http://ceur-ws.org/Vol-1141/paper_06.pdf)