УДК 004.912
Чала Л.Э., Чижевский А.В., Волощук Е.Б.
Харьковский национальный университет радиоэлектроники

МЕТОД ПОИСКА ПЕРТИНЕНТНЫХ СВЯЗЕЙ МЕЖДУ КОНЦЕПТАМИ ПРОЕКТИРУЕМЫХ ОНТОЛОГИЙ

Чала Л.Е., Чижевський А.В., Волощук О.Б. Метод пошуку пертинентних зв'язків між концептами онтологій, що проектуються. У статті запропоновано метод визначення найбільш пертинентних зв'язків між концептами онтологічних моделей, що формуються. Обчислювальна схема методу, яка основується на модифікованому алгоритмі Гінзбурга, дозволяє поліпшити якість автоматично створюваних онтологій. Метод може бути ефективно використано для задач семантичного пошуку в системах інтелектуального аналізу електронних текстів та формування онтологічних моделей предметної області.
Ключові слова: онтологічна модель, концепт, пертинентний зв'язок, інтелектуальний аналіз

Чалая Л.Э., Чижевский А.В., Волощук Е.Б. Метод поиска пертинентных связей между концептами проектируемых онтологий. В статье предлагается метод определения наиболее пертинентных связей между концептами формируемых онтологических моделей. Вычислительная схема метода, основанная на модификации алгоритма Гинзбурга, позволяет повысить качество автоматически создаваемых онтологий. Метод может эффективно использоваться для задач семантического поиска в системах интеллектуального анализа электронных текстов и формирования онтологических моделей предметной области.
Ключевые слова: онтологическая модель, концепт, пертинентная связь, интеллектуальный анализ

Chala L.E., Chyzhevskyi A.V., Voloshchuk O.B. Method of search of pertinent connections between concepts of the designed ontologies. In the article the method of determination most of pertinent connections between concepts of the designed ontological models is proposed. The calculus procedure of method, based on modification of Ginsburg's algorithm, allows improving quality automatically created ontologies. A method can be effectively used for the tasks of semantic search in the intellectual analysis systems of e-texts and forming of ontological models of subject domain.
Keywords: ontological model, concept, pertinent connection, intellectual analysis

Постановка проблемы. Основной задачей современных систем поиска и предварительной обработки web-документов является оперативное предоставление пользователям сети Интернет необходимой информации. При этом результаты поиска не всегда оказываются удовлетворительными, так как поисковые Интернет-сервисы могут выдавать по запросам пользователей большое количество условно релевантных web-данных, которые далеко не всегда удовлетворяют истинным интересам пользователей. Кроме того, такие результаты могут быть существенно зашумлены нерелевантными ссылками. Все это приводит к снижению эффективности получения пользователями необходимой значимой информации из сети Интернет, ресурсы которой постоянно растут. В связи с этим особенно актуальными становятся автоматические методы работы с большими объемами информации. В последнее время, в частности, получили широкое распространение исследования в области автоматического синтеза онтологических моделей, позволяющих повысить эффективность систем семантического поиска по запросам пользователей (в корпусе текстов, электронных библиотеках, в сети Интернет) [1]. Актуальными также являются задачи использования онтологии как основы для спецификации и разработки программного обеспечения, поддержки общего доступа к информации, поиска информации, взаимодействия при объединении информации, создании порталов знаний, разработке пользовательского интерфейса программных систем, редакторов информации и интеллектуальных систем [2]. Качество формируемых онтологий, используемых для создания поисковых систем, во многом определяется полнотой учета в онтологической модели наиболее значимых концептов для корпуса анализируемых текстов с учетом их тематической специфики (под концептами будем в дальнейшем понимать наиболее значимые слова и словосочетания в анализируемом тексте, которые могут быть учтены в онтологической модели). В связи с этим целесообразно решить задачу формирования множества концептов будущей онтологии с учетом связей между ними. В работах [3, 4] уже были рассмотрены решения автоматического построения онтологий, в частности методы нахождения концептов для онтологии и связей между ними. Данное исследование ставит перед собой целью усовершенствование и дополнение алгоритмов и методов автоматического синтеза онтологических моделей.

Методы нахождения концептов при автоматическом синтезе онтологий и нахождения шаблонных связей между ними (типа «часть-целое» и «отношение») рассматриваются, в частности, в работе [5]. Однако результаты экспериментального исследования этих методов показали, что при поиске слов и словосочетаний, которые могут использоваться в качестве концептов, сформированное

множество концептов-претендентов не всегда соответствует такому же множеству, составленному экспертом предметной области. Это приводит к тому, что некоторые важные понятия предметной области могут не попасть в автоматически создаваемую онтологию. Кроме того, в этих методах отсутствует процедура общего ранжирования по значимости списка всех концептов-претендентов, а осуществляется лишь раздельное ранжирование слов и словосочетаний, входящих в этот список.

В частности, возникают следующие проблемы:

– не всегда удается правильно найти связи между концептами;

– не всегда удается выделить концепты, имеющие связь с наибольшим количеством других концептов;

– найденные связи между концептами будущей онтологии не всегда актуальны для конкретной предметной области. При этом не только повышается используемый объем памяти и увеличивается время на создание онтологии и обработку запросов к ней, но и избыточным становится объем онтологии, что снижает оперативность дальнейшего ее применения.

В данной статье рассматривается возможность частичного устранения перечисленных трудностей на основе комбинированного применения и модификации существующих методов определения пертинентных связей между концептами формируемых онтологических моделей.

Целью данной статьи является модификация и программная реализация методов автоматического поиска актуальных связей между концептами проектируемой онтологии для заданной предметной области.

Установление связей между концептами проектируемой онтологии. Выделим три основных подхода для решения задачи установления связей между концептами проектируемой онтологии:

- поиск слова-претендента на связь в онтологии и последующий подбор концептов, для которых актуальна эта связь (метод 1);

- определение для рассматриваемого концепта списка вероятных слов-претендентов на использование в качестве связи для этого концепта и последующий подбор концепта для установления связи (метод 2);

- нахождение в онтологии двух концептов, которые необходимо связать, и последующий подбор связи для данных концептов (метод 3).

Достоинства первого подхода (метод 1):

– поиск в тексте слов-связей и концептов осуществляется раздельно. Это означает, что концепт и связь не обязательно должны составлять в тексте словосочетание при поиске данной связки в тексте программы автоматического синтеза онтологии;

– возможность варьировать количество учитываемых связок «концепт-связь-концепт» с помощью настраиваемых коэффициентов (уменьшать в случае нахождения большого количества ненужной информации и увеличивать в случае недостаточного количества связей в онтологии).

Недостатки первого подхода:

– в общем множестве найденных связей между концептами присутствуют несущественные или несуществующие связи;

– некоторые важные концепты предметной области не имеют связей сформированного множества с другими концептами проектируемой онтологии.

Устранению отмеченных недостатков способствует комбинированное применение второго и третьего подходов (метод 2 и метод 3).

Предлагаемый ниже метод поиска связей для онтологии, основанный на таком комбинированном подходе, назовем методом главного концепта.

Метод главного концепта. Предлагаемый метод предполагает необходимость вычисления вероятности применения слова в качестве пертинентной связки для рассматриваемого концепта.

Рассмотрим вначале некоторые свойства слов, которые в формируемой онтологии будут применяться в качестве слов-отношений, связывающих концепты в онтологии.

Задача автоматического определения таких связок является далеко не тривиальной. Рассмотрим пример определения слов связок для концептов следующего текстового фрагмента:

1) «на основе алгоритма Гинзбурга был разработан метод выделения ключевых слов»;

2) «разработанный алгоритм синтезирует функциональную модель»;

3) «осуществляется определение для концепта онтологии необходимого списка связей».

В этом примере словами-связками между понятиями являются соответственно слова «разработан», «синтезирует», «определение». Здесь в качестве слов-связок могут применяться как слова

специфичные для рассматриваемой предметной области, так и достаточно общие, которые могут присутствовать в любом тексте. Можно отметить, что слово-связка вероятнее всего будет находиться в тексте между понятиями, которые оно связывает. Вследствие этого целесообразно определить степень специфичности претендента на слово-связку в контексте понятия, которое будет связывать данное слово-связка. Для решения этой задачи предлагается использовать алгоритм Гинзбурга [6]. В соответствии с этим алгоритмом, если слово-связка входит в контекст леммы-понятия в рамках рассматриваемого текста, то считают, что оно специфично в контексте данного понятия. Введем понятие тройки элементов, используемых для реализации процедуры предварительного отбора наиболее пертинентных связок для проектируемой онтологии. К элементам такой тройки отнесем: слово, обозначающее связь между двумя концептами (L_1) и собственно два концепта (W_1 и W_2), каждое из которых может быть представлено одним словом либо словосочетанием. Таким образом, тройку можно представить в виде: «слово№1, связь, слово№2»:

$$W_1 \leftrightarrow L_1 \leftrightarrow W_2 . \quad (1)$$

Отметим, что если концепт представлен словосочетанием, то в тройку вносится главное слово словосочетания.

Выделим четыре возможных варианта представления любой тройки в зависимости от уровня специфичности слова-связки по отношению к понятиям:

$$W_1 \xleftarrow{F_{(w_1,L_1)}\uparrow} L_1 \xleftarrow{F_{(w_2,L_1)}\uparrow} W_2 ; \quad (2)$$

$$W_1 \xleftarrow{F_{(w_1,L_1)}\uparrow} L_1 \xleftarrow{F_{(w_2,L_1)}\downarrow} W_2 ; \quad (3)$$

$$W_1 \xleftarrow{F_{(w_1,L_1)}\downarrow} L_1 \xleftarrow{F_{(w_2,L_1)}\uparrow} W_2 ; \quad (4)$$

$$W_1 \xleftarrow{F_{(w_1,L_1)}\downarrow} L_1 \xleftarrow{F_{(w_2,L_1)}\downarrow} W_2 , \quad (5)$$

где символ $F_{(w_i,L_1)}\uparrow$ означает, что слово-связка L_1 специфична для слова W_i, а $F_{(w_i,L_1)}\uparrow$ означает, что слово-связка L_1 не специфична для слова W_i.

На основе статистического анализа текстов рассматриваемой предметной области могут быть определены коэффициенты вероятности принадлежности определенной тройки к одному из вариантов ее представления: (2), (3), (4) или (5). Например, для корпуса текстов из электронной библиотеки методических указаний Харьковского национального университета радиоэлектроники по технической тематике было определено количество троек, принадлежащих к одному из четырех типов, и рассчитаны (как среднее арифметическое по всей выборке текстов) соответствующие вероятности: $p_1 = 0,25$; $p_2 = 0,15$; $p_3 = 0,1$; $p_4 = 0,5$ [3].

На основании полученных значений p_1, p_2, p_3, p_4 определим вероятности выбора слова в качестве связки в зависимости от его положения в предложении по отношению к концептам. Рассмотрим варианты положения слова-связки в предложении относительно концептов, которые оно связывает. Назовем «нормальным порядком» расположения концептов в тексте, если первый концепт располагается в тексте раньше, чем второй, и «обратным порядком», если второй концепт располагается в тексте раньше, чем первый. В зависимости от положения слова-связки в предложении относительно концептов рассматриваемую тройку можно отнести к одной из трех возможных групп (рис.1). Первая группа содержит тройки, в которых слово-связка находится в предложении между первым и вторым концептами, а концепты расположены в нормальном порядке. Вторая группа содержит тройки, в которых слово-связка находится в предложении между первым и вторым концептами, а концепты расположены в обратном порядке. К этой группе отнесем также тройки, в которых слово-связка расположено в тексте раньше, чем концепты. Третья группа содержит тройки, в которых слово-связка расположено в тексте после концептов.

На основе значений p_1, p_2, p_3, p_4 можно определить коэффициенты k, соответствующие вероятности автоматического выбора определенной тройки в качестве актуальной для проектируемой онтологии. Коэффициенты k представляют собой отношение числа троек определенного типа к общему числу троек, актуальных для онтологии заданной предметной области. Для рассмотренного выше примера получены следующие значения этого коэффициента в зависимости от варианта (группы) положения слова-связки в предложении относительно концептов: группа (1) – $k = 0,7$; группа (2) –

$k = 0,2$; группа (3) – $k = 0,1$.

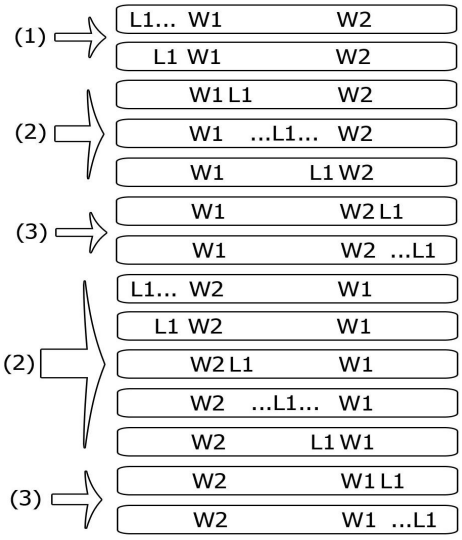

Рис.1. Варианты положения слова-связки L_1 в тексте относительно концептов W_1 и W_2.

При принятии решения о занесении той или иной тройки в проектирующуюся онтологию, кроме расположения элементов тройки необходимо учитывать наличие слов между ними и их количество. Очевидно, что целесообразнее вносить в онтологию тройки, элементы которой следуют непосредственно друг за другом, чем тройки, между концептами и связкой которой находятся фрагменты предложения.

Назовем расстоянием между элементами тройки количество слов, которые находятся в предложении между двумя любыми элементами тройки. Обозначим через N расстояние в предложении между двумя концептами W_1 и W_2 рассматриваемой тройки.

Тогда вероятность актуальности рассматриваемой тройки в зависимости от положения ее элементов в предложении можно определить следующим образом:

$$P_{place} = \frac{k * \left(\frac{|m-n|}{\min(n,m)+2} + 1 \right)}{n+m+1}, \quad (6)$$

где n – расстояние от L_1 до W_1, $n = N$, если между L_1 и W_1 находится W_2; m – расстояние от L_1 до W_2, $m = N$, если между L_1 и W_2 находится W_1.

В соответствии с (6), чем больше расстояние между словом-связкой и концептами в тройке, тем меньше вероятность ее актуальности для проектируемой онтологии. Также необходимо отметить, что приведенная формула учитывает приоритет троек, у которых расстояние слова-связки хотя бы с одним из концептов является намного меньше среднего значения такого расстояния для всей совокупности рассматриваемых концептов.

Предлагаемый метод главного концепта имеет ряд преимуществ. В частности, его можно применять для поиска связей в онтологии, в которой до этого не была определена ни одна связь. Дополнительным преимуществом метода является возможность задавать здесь список концептов, для которых необходимо найти связь (например, всех концептов, имеющихся в онтологии в данный момент). Это позволяет определить максимально возможное количество актуальных (пертинентных) связей для проектируемой онтологии. Следует отметить, что целесообразно искать слово-связку для выбранного концепта/понятия только в тех предложениях, где встречается собственно сам этот концепт.

Алгоритм установления связей в онтологии по методу главного концепта можно представить набором следующих действий:

- выбор концепта/понятия и нормализация его до одного слова (W_1), для которого следует

сформировать тройку в проектируемой онтологии;

- определение множества слов $M\uparrow(W_1)$, входящих в контекстное множество данного концепта W_1 (из множества всех слов в предложениях, где присутствует данный концепт с понятием $M(W_1)$), а также множества слов $M\downarrow(W_1)$, не входящих в контекстное множество данного концепта (по алгоритму Гинзбурга [6]);

- определение наиболее вероятного типа связи ($F_{(w_i,L_1)}\uparrow$ или $F_{(w_i,L_1)}\downarrow$) между концептом W_1 и предполагаемым словом-связкой L_1;

- определение множества $M(L_i)$, состоящего из претендентов на слова-связки, удовлетворяющих установленному типу связи ($M(L_1)$ принимается как $M\uparrow(W_1)$ или как $M\downarrow(W_1)$);

- определение для каждого L_i (из множества $M(L_i)$) множества слов $M\uparrow(W_2)$, входящих в контекстное множество данного слова L_i (из множества всех слов, входящих в одно предложение с данным словом L_i и данным словом W_1, во всех предложениях, где присутствуют L_i и W_1), а также множества слов, $M\downarrow(W_2)$ не входящих в контекстное множество данного концепта (по алгоритму Гинзбурга [6]);

- определение множеств $M_i(W_2)$, состоящих из претендентов на концепт, связываемый с концептом W_1 при помощи слова-связки L_i, удовлетворяющих установленному типу связи (для каждого L_i из множества $M(L_i)$);

- определение наиболее вероятного типа связи ($F_{(w_2,L_1)}\uparrow$ или $F_{(w_2,L_1)}\downarrow$) между будущим словом-связкой L_1 и концептом W_2 (для каждого L_i из множества $M(L_i)$;

- включение в онтологию наиболее вероятной связки из множества $M(T_i)$ возможных вариантов троек.

Наиболее вероятный тип связи $F_{(w_i,L_1)}$ определяется по следующим зависимостям:

$$F_{(w_i,L_1)}=\begin{cases}F_{(w_q,L_1)}\uparrow, if\ P_{L_1}>P_{L_2},\\ F_{(w_q,L_1)}\downarrow, if\ P_{L_1}<P_{L_2}\end{cases}, \tag{7}$$

$$P_{L_1}=\frac{\left|M\uparrow(W_1)\right|}{N_{all1}}*(p_1+p_2), \tag{8}$$

$$P_{L_2}=\frac{\left|M\downarrow(W_1)\right|}{N_{all1}}*(p_3+p_4), \tag{9}$$

где N_{all1} – мощность объединения множеств $M\uparrow(W_1)$ и $M\downarrow(W_1)$, определяемая количеством всех слов в предложениях с W_1.

В (8,9) вероятности p_i суммируются, т.к. типы троек, соответствующие вероятностям p_1 и p_2, удовлетворяют условию типа связи $F_{(w_i,L_1)}\uparrow$, а p_3 и p_4 – условию типа связи $F_{(w_2,L_i)}\downarrow$.

Наиболее вероятный тип связи $F_{(w_2,L_i)}$ определяется по следующим зависимостям:

$$F_{(w_2,L_i)}=\begin{cases}F_{(w_2,L_i)}\uparrow, if\ P_{L_3}>P_{L_4},\\ F_{(w_2,L_i)}\downarrow, if\ P_{L_3}<P_{L_4}\end{cases}, \tag{10}$$

$$P_{L_3} = \begin{cases} \frac{\left|M \uparrow (W_2)\right|}{N_{all2}} * p_1 * (N_{all2} - N_{L_i}), if\ F_{(w_1,L_1)} = F_{(w_1,L_1)} \uparrow, \\ \frac{\left|M \uparrow (W_2)\right|}{N_{all2}} * p_2 * (N_{all2} - N_{L_i}), if\ F_{(w_1,L_1)} = F_{(w_1,L_1)} \downarrow \end{cases}, \quad (11)$$

$$P_{L_4} = \begin{cases} \frac{\left|M \downarrow (W_2)\right|}{N_{all2}} * p_3 * (N_{all2} - N_{L_i}), if\ F_{(w_1,L_1)} = F_{(w_1,L_1)} \uparrow, \\ \frac{\left|M \downarrow (W_2)\right|}{N_{all2}} * p_4 * (N_{all2} - N_{L_i}), if\ F_{(w_1,L_1)} = F_{(w_1,L_1)} \downarrow \end{cases}, \quad (12)$$

где N_{all2} – мощность объединения множеств $M \uparrow (W_2)$ и $M \downarrow (W_2)$; N_{L_i} – количество повторений L_i.

В формулах (11) и (12) вероятности P_{L3} и P_{L4} умножаются на разность общего количества слов в предложениях, где присутствуют элементы рассматриваемой тройки и вхождения в эти предложения слова-связки L_i. Таким образом, чем чаще в предложении встречается определенное слово, тем меньше вероятность его выбора в качестве слова-связки, т.к. маловероятно, что одно и тоже слово будет выполнять в предложении и роль слова-связки для двух понятий и просто встречаться в предложении в каких-либо других контекстах.

На заключительном этапе алгоритма определяется множество $M(T_i)$ – множество троек, для которых определены типы связей $F_{(w_i,L_1)}$ и $F_{(w_2,L_1)}$ ((2), (3), (4) или (5) соответственно). При этом предлагается ранжировать элементы из данного множества в соответствии со значениями вероятностей их выбора в качестве троек, актуальных для проектируемой онтологии. Вероятность выбора тройки T_i в качестве актуальной для онтологии P_{T_i} рассчитывается по следующей зависимости:

$$P_{T_i} = \begin{cases} p_{place} * f_{(w_1,L_1)} * f_{(w_2,L_1)} * f_{(w_1,w_2)}, if\ T_i \subset (2), \\ \frac{p_{place} * f_{(w_1,L_1)} * f_{(w_1,w_2)}}{f_{(w_2,L_1)}}, if\ T_i \subset (3), \\ \frac{p_{place} * f_{(w_2,L_1)} * f_{(w_1,w_2)}}{f_{(w_1,L_1)}}, if\ T_i \subset (4), \\ \frac{p_{place} * f_{(w_1,w_2)}}{f_{(w_1,L_1)} * f_{(w_2,L_1)}}, if\ T_i \subset (5) \end{cases}, \quad (13)$$

где p_{place} – вероятность, рассчитываемая по формуле (6); $f_{(w_1,L_1)}$ – сила связи, рассчитанная для W_1 и L_1 из тройки T_i по алгоритму, описанному в [3]; $f_{(w_2,L_1)}$ – сила связи, рассчитанная для W_2 и L_1 из тройки T_i по алгоритму, описанному в [3]; $f_{(w_1,w_2)}$ – сила связи, рассчитанная для W_1 и W_2 из тройки T_i по алгоритму, описанному в [3]; (2, 3, 4, 5) – один из четырех определенных типов тройки T_i.

Эта формула позволяет учесть (в зависимости от типа рассматриваемой тройки), насколько сила семантической связи между элементами тройки влияет на вероятность ее выбора в качестве актуальной для проектирующейся онтологии. Если тип связи между двумя элементами определен как $F_{(w_i,L_1)} \uparrow$, то с возрастанием силы связи между элементами тройки W_i и L_1 возрастает вероятность актуальности данной тройки для онтологии. Если же тип связи между двумя элементами определен как $F_{(w_i,L_1)} \downarrow$, то чем больше сила связи между элементами тройки W_i и L_1, тем меньше вероятность актуальности

данной тройки для онтологии. Также следует отметить, что чем больше сила связи между концептами тройки, тем тройка актуальнее для проектирующейся онтологии. Кроме того, при оценивании связей в онтологии по методу главного концепта можно найти или одну тройку, для которой вероятность актуальности для онтологии наиболее высока, или же найти множество троек, для которых вероятность P_{T_i} выше либо равна $\min P_{T_i}$, и считать, что все тройки, входящие в это множество актуальны для проектирующейся онтологии:

$$\min P_{T_i} = K_1 * \max(P_{T_i}), \tag{14}$$

где K_1 – настраиваемый коэффициент, позволяющий исключить из рассмотрения тройки, заведомо неактуальные для проектируемой онтологии.

Оценка эффективности разработанного метода. По предложенному методу был разработан программный модуль ««Concept-Ont-M», который может эффективно использоваться для задач семантического поиска в системах анализа электронных текстов и автоматического создания онтологий. Проведенные экспериментальные исследования показали, что метод главного концепта в целом работает гораздо эффективнее, чем методы, основанные на поиске отдельных связей. Оценка эффективности проводилась по двум параметрам: *R* – точность поиска связей (отношение правильно найденных связей к общему количеству найденных связей); *P* – полнота поиска связей (отношение правильно найденных связей к общему количеству связей, выявленных экспертом). Результаты экспериментальных исследований (для корпуса текстов из электронной библиотеки методических указаний Харьковского национального университета радиоэлектроники по технической тематике): для метода поиска отдельных связей и метода главного концепта значения *R* составляют 57% и 78% соответственно; значения *P* – 78% и 82% соответственно.

Выводы и перспективы дальнейших исследований. Проведенные исследования позволяют сделать вывод, что важным этапом автоматического построения онтологий является формирование пертинентных связей между концептами. Модификация и программная реализация метода нахождения таких связей с учетом расположения элементов «концепт-связка» в тексте позволили повысить возможности автоматического создания онтологий. В частности, предложенный метод можно применять для поиска связей в онтологии, в которой до этого не была определена ни одна связь. При этом особое внимание следует уделить задачам ранжирования однословных/многословных концептов и выявления связей типа «отношения» между ними. Научная новизна предложенного метода состоит в возможности определения степени специфичности претендента на слово-связку в контексте леммы-понятия в рамках анализируемого текста. При проведении дальнейших исследований целесообразно усовершенствовать предложенный метод, дополнив его анализом более сложных типов связок в онтологической модели.

1. Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic Web (Ч. 3) / В.Ф. Хорошевский// Искусственный интеллект и принятие решений.-2011.- № 2.-С. 15–36.
2. Ландэ Д.В. Интернетика: Навигация в сложных сетях: модели и алгоритмы / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов – М.: Либроком, 2009. – 264 с.
3. Чалая Л.Э., Формирование множества связных концептов для автоматического синтеза онтологий [Текст] / Л.Э. Чалая, А.В. Чижевский// International Journal "Information Theories and Applications". – Vol. 21, Number 3. – 2014. – P. 203 – 212.
4. Зябрев И.Н., Пожарков О.В., Пожаркова И.Н. Использование спектральных характеристик лексем для улучшения поисковых алгоритмов.// Труды РОМИП 2010. –Казань: Казанский ун-т. – С. 40-48.
5. Воронина И.Е. Алгоритмы определения семантической близости по их окружению в тексте / И.Е. Воронина, А.А. Кретов, И.В. Попова. // Вестник ВГУ: системный анализ и информационные технологии.– 2011.-№ 2. – С. 15–36.
6. Гинзбург Е. Л. Идиоглоссы: проблемы выявления и изучения контекста. / Е. Л. Гинзбург // Семантика языковых единиц: Доклады VI Международной конференции. Т. I., М., 1998. – С. 26–28.