

18. Zaporozhets D.Yu., Zaruba D.V., Kureichik V.V. Hybrid bionic algorithms for solving problems of parametric optimization, *World Applied Sciences Journal*, 2013, Vol. 23, pp. 1032-1036.
19. Bova V.V., Zammoev A.U., Dukkardt A.N. Evolyutsionnaya model' intellektual'nogo analiza raznorodnykh znaniy [An evolutionary model for intelligent analysis of heterogeneous knowledge], *Izvestiya KBNTs RAN [Izvestiya of Kabardino-Balkar scientific centre of the RAS]*, 2013, No. 4 (54), pp. 7-13.
20. Kuliev E.V., Lezhebokov A.A., Dukkardt A.N. Podkhod k issledovaniyu okrestnostey v roevykh algoritmakh dlya resheniya optimizatsionnykh zadach [Approach to research environs in swarms algorithm for solution of optimizing problems], *Izvestiya YuFU. Tekhnicheskie nauki [Izvestiya SFedU. Engineering Sciences]*, 2014, No. 7 (156), pp. 15-25.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Гатчин.

Бова Виктория Викторовна – Южный федеральный университет; e-mail: vvbova@yandex.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Bova Victoria Victorovna – Southern Federal University; e-mail: vvbova@yandex.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

УДК 002.53:004.89

Ю.А. Кравченко, В.В. Марков, А.А. Новиков

СЕМАНТИЧЕСКИЙ ПОИСК В SEMANTIC WEB*

Представлен механизм семантического поиска, основанного на сочетании методов распространения активации с традиционными поисковыми механизмами. Большинство пользователей привыкли выражать свои информационные потребности с точки зрения ключевых слов. В традиционных механизмах поиска, документ, как правило, извлекается, когда хотя бы одно из ключевых слов в строке запроса находится внутри концепта. В нашем подходе предполагается получение всех экземпляров концептов, которые связаны с введенным ключевым словом, даже если оно не встречается внутри самого концепта. Предлагаемый алгоритм может быть использован для онтологии, в которой все отношения между вершинами имеют как описание, основанное на определениях онтологии, так и весовой коэффициент, рассчитываемый методом отображения весовых коэффициентов. Алгоритм имеет в качестве отправной точки начальный набор концептов из онтологии, которые будут называться узлами или нодами. Исходный набор концептов получается в результате работы классических механизмов поиска. Все узлы имеют некоторое начальное значение активации. Алгоритм распространения активации используется для поиска понятий в онтологии на основе начального набора концептов с соответствующими начальными значениями активации. Алгоритм выполняется до тех пор, пока не будет достигнуто определенное состояние (например, заданный размер результирующего набора) или не осталось узлов, которые обрабатываются в приоритетной очереди.

Семантический поиск; онтология; Semantic Web; весовой коэффициент; алгоритм распространения активации.

Yu.A. Kravchenko, A.A. Novikov, V.V. Markov

SEMANTIC SEARCH IN SEMANTIC WEB

This article presents the mechanism of semantic search based on a combination of activation methods of dissemination of traditional search engines. Most people are accustomed to express their information needs in terms of keywords. Traditional search engines, the document is

* Работа выполнена при финансовой поддержке РФФИ (проект № 16-07-00703).

usually removed when at least one of the keywords in the query string is inside the concept. In our approach we are expected to obtain copies of all the concepts that are related to your keyword, even if it is not found within the concept. The proposed algorithm can be used for the ontology, in which all relations between the peaks have a description, based on the definitions of ontology, and a weighting factor, which is calculated by mapping the weighting factors. The algorithm has as a starting point an initial set of ontology concepts, which will be called nodes, or nodes. The initial set of concepts is the result of the work of classical search engines. All nodes have a initial value of activation. Spread activation algorithm is used to search for terms in the ontology based on the initial set of concepts with the corresponding initial values of activation. The algorithm runs as long until a certain condition (e.g., a predetermined size of the result set), or no more nodes are processed in a priority queue.

Semantic search; ontology; Semantic Web; weight mapping; spread activation algorithm.

Введение. В связи с растущим интересом к технологии Semantic Web можно предположить, что в скором времени станет доступен большой объем метаданных, описывающих ключевую информацию о ресурсах в сети Internet. Основной идеей является развитие поисковых процессов в гипермедиа приложениях с информацией, извлеченной из семантической модели предметной области. Для получения необходимых результатов в задачах семантического поиска все чаще используется сочетание методов распространения активации (spread activation) с традиционными поисковыми механизмами. Одной из основных проблем традиционных поисковых систем является то, что они, как правило, основаны на обработке ключевых слов [1]. Рассмотрим пример для ресурса научно-исследовательского учреждения, частично представленного на рис. 1.

Данный ресурс работает с людьми и публикациями в различных исследовательских областях. Здесь «Ключевое слово» не является концептом модели, а используется в диаграмме, чтобы отразить тот факт, что ключевое слово представлено внутри текста соответствующих экземпляров концептов. Например, ключевое слово «Web» определяется внутри концепта «Развитие Web приложений», так как оно находится внутри заголовка статьи. Ключевое слово «онтология» так же относится к данному концепту, т.к. оно появляется в описании статьи. Если пользователь будет искать узлы типа «профессор», то результат вполне может быть пустым, так как ключевое слово «Web» может не отображаться внутри текста описания (страницы) профессора. С другой стороны, анализируя семантику рассматриваемой предметной области, можно сделать вывод, что узел профессора, имеющего большое количество публикаций, связанных с этим ключевым словом, должен быть возвращен в качестве результата запроса. Таким образом, в рассмотренном примере узел профессора «Иванова» может быть представлен в результате запроса с ключевым словом «Web».

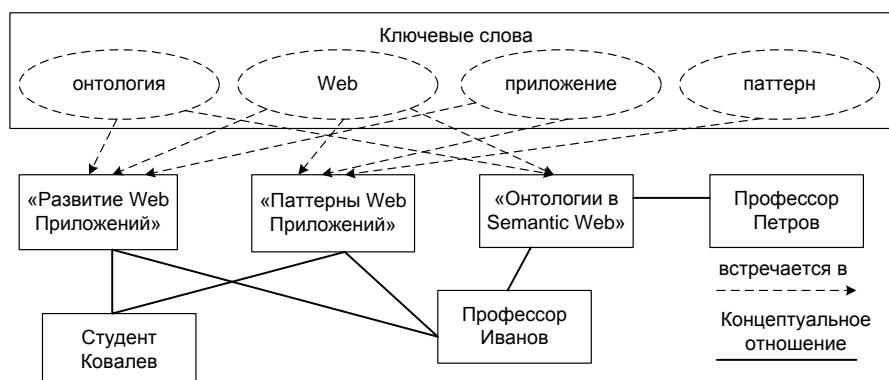


Рис. 1. Часть семантической модели

Описанный в примере механизм семантического поиска будет особенно эффективен, если пользователь ищет определенные экземпляры концептов, а не «произвольные» web-страницы, т.е. ключевые слова в запросе обозначают одно или несколько понятий. Другими словами, каждая страница является гипермедиа представлением узла в модели.

Некоторые существующие механизмы семантического поиска требуют от пользователя описание запроса с использованием других концептов, что накладывает на пользователя высокую семантическую нагрузку. Например, в большинстве семантических запросов, для нахождения информации о студентах, связанных с профессором «Ивановым», пользователь вынужден формировать запрос, состоящий из ключевых слов «Иванов», с указанием типа запрашиваемого понятия («Профессор»). Заранее заложенные типы понятий в ресурсе делают его более трудным для получения необходимой информации и описания запроса пользователем.

С учетом всех этих аспектов предлагается механизм семантического поиска, который позволяет пользователю выражать его потребности с точки зрения ключевых слов, но в то же время использует семантическую информацию и метаданные, относящиеся к предметной области, для получения результатов, которые невозможны при использовании традиционных механизмов поиска [2].

1. Существующие механизмы поиска. Большинство пользователей привыкли выражать свои информационные потребности с точки зрения ключевых слов. Механизм семантического поиска, который обрабатывает запрос как текстовую информацию и RDF аннотацию, представлен в статье [3]. Google также представляет собой очень интересный подход для поиска необходимой продукции. Это поисковая система, специализирующаяся на запросах, где пользователь выражает запрос с помощью ключевых слов, которые связаны с продуктом (бренд, название, модель и т.д.). Google пытается определить, какой продукт пользователь хочет найти, связывая ключевые слова в запросе с метаданными, которые описывают продукты в их базе знаний. Еще один интересный механизм семантического поиска SCORE [4]. Он использует методы автоматической классификации и извлечения информации, вместе с анализом метаданных и онтологии информации, позволяющие контекстный поиск сразу по нескольким ресурсам.

В традиционных механизмах поиска, документ, как правило, извлекается, когда хотя бы одно из ключевых слов в строке запроса находится внутри концепта. В нашем подходе предполагается получение всех экземпляров концептов, которые связаны с введенным ключевым словом, даже если оно не встречается внутри самого концепта. Отношения определяются с помощью алгоритма распространения активации, что дает возможность получить концепты, которые не содержат какое-либо из указанных в запросе слов [5].

Метод распространения активации является одним из часто используемых для обработки семантических сетей. Он успешно используется в различных областях, особенно в информационно-поисковых приложениях, т.к. он был разработан в области искусственного интеллекта (ИИ) в качестве основы для обработки семантических сетей и онтологий. Использование метода распространения активации поможет решить задачи обработки знаний в контексте Semantic Web.

Метод распространения активации работает в основном в качестве механизма поиска концептов. Учитывая первоначальный набор концептов и некоторых ограничений, во время прохождения через онтологию, достигаются другие узлы, которые тесно связаны с исходными концептами. Это очень мощный алгоритм по поиску семантической близости, где дан первоначальный набор понятий, в результате работы которого возвращается набор других понятий, тесно связанных с ними [6].

Метод распространения активации используется в семантических сетях (где каждое ребро в сети имеет описание, связанное с концептом) или в ассоциативных сетях (где каждое ребро имеет только весовой коэффициент, связанный с концептом).

Онтологии и их вершины несут гораздо больше информации, чем та, что указана в явном виде, т.к. есть много «скрытой» информации, которая кроется в связях между вершинами. В традиционных онтологиях, можно лишь указать на отсутствие или наличие связи между двумя экземплярами понятий. Но во многих случаях, необходимо выразить меру «прочности» в виде весового коэффициента, связанного с отношением между вершинами. Классический способ состоит в определении числового значения для каждого ребра в сети.

Основной идеей данной работы является извлечение знаний из онтологии, для получения весовых коэффициентов для каждого существующего экземпляра отношения в модели. Результатом является набор, где каждый экземпляр отношения имеет описание и весовой коэффициент. Похожая идея была предложена в работе [7], для описания нового подхода ранжирования результатов поиска в Semantic Web на основе онтологии. Для расчета весового коэффициента для каждого экземпляра отношения, используется метод отображения весовых коэффициентов (Weight Mapping), основанный на анализе ссылочной структуры базы знаний.

2. Алгоритм распространения активации. В статье предлагается гибридный алгоритм распространения активации, основанный на сочетании метода отображения весовых коэффициентов, описанного в статье, с традиционным методом распространения активации.

Отображение весовых коэффициентов. Необходимо дать числовую оценку для каждого экземпляра отношения в сети. Невозможно составить общую формулу для всех предметных областей, по которой будут рассчитываться лучшие весовые коэффициенты для каждого существующего экземпляра отношения в базе знаний. Например, в работе [7], расчет коэффициентов основывается на значимости отношений. Она пропорциональна специфике всех терминов, находящихся в отношении.

Предлагается использовать 3 различные меры оценки весового коэффициента связей между вершинами онтологии – кластерная, мера специфичности и комбинированная.

Кластерная оценка. Первая мера используется для установки степени сходства между двумя связанными экземплярами концептов. Используемая функция получена путем модификации функции из статьи [8] для концептов, которые связаны друг с другом. Приведенная ниже формула определяет сходство между концептами C_j и C_k .

$$W(C_j, C_k) = \frac{\sum_{i=1}^n n_{ijk}}{\sum_{i=1}^n n_{ij}}$$

Значение n_{ij} указывает на наличие связи между концептами C_j и C_k и равно 1, если концепты связаны между собой и 0 в противном случае. Значение n_{ijk} равно 1, если оба концепта C_j и C_k связаны с концептом C_i и 0 в противном случае. Таким образом, весовой коэффициент $W(C_j, C_k)$ определяет близость этих двух концептов в виде процентного соотношения. Кластерная мера аналогична мере близости, предложенной в работе [9], которая широко применяется в алгоритмах поиска ассоциативных правил. Идея этой оценки заключается в том, что концепты, которые имеют большое количество общих отношений с другими концептами, являются схожими. Еще одним важным моментом является то, что эта оценка сходства является асимметричной; Ограничения в использовании симметричных коэффициентов подобия, рассматривается в статье [10].

Оценка специфичности. Вторая мера схожа с мерой инвертированной частотной области, часто используемой в сфере извлечения информации. Мету специфичности полезно учитывать, когда пользователь хочет указать семантическую особенность или установить различия связи. Следующая формула используется для измерения меры специфичности:

$$W(C_j, C_k) = \frac{1}{\sqrt{n_k}}$$

Значение n_k равно числу экземпляров данного типа связей, имеющие узел k в качестве узла назначения. Таким образом, весовой коэффициент специфичности связи обратно пропорционален числу связей с концептом C_k . Если несколько концептов имеют такой же тип, что и C_j , и связаны с концептом C_k , то весовой коэффициент специфичности $W(C_j, C_k)$ будет высоким.

Комбинированная оценка. Третья мера является обобщенной оценкой, полученной как произведение двух предыдущих мер. Обе описанных ранее оценки могут быть использованы по отдельности в зависимости от выбранной семантики. Тем не менее, в общем случае комбинированная оценка, является предпочтительной. Первое слагаемое описывает схожесть двух концептов C_j и C_k . Второе указывает на специфичность связи между этими концептами. В сфере извлечения информации используют различные меры, в зависимости от предметной области и поставленных задач [11].

Вычисление значения весового коэффициента для экземпляра отношения может быть контекстно-зависимым. В зависимости от контекста, некоторые типы отношений могут быть важнее, чем другие. В данной статье все типы отношений имеют равный коэффициент важности. Для дальнейшего развития и получения более точных результатов необходимо предоставить пользователю возможность назначения коэффициентов важности для различных типов связей.

Гибридный алгоритм распространения активации. Предлагаемый алгоритм может быть использован для онтологии, в которой все отношения между вершинами имеют как описание, основанное на определениях онтологии, так и весовой коэффициент, рассчитываемый методом отображения весовых коэффициентов. Алгоритм распространения активации работает путем изучения концептов онтологии. Учитывая первоначальный набор концептов, алгоритм перемещается по связям в онтологии, получая в результате набор тесно связанных между собой концептов. Результирующий набор может содержать в себе концепты, которые явно не имеют отношений с концептами из начального набора.

Алгоритм имеет в качестве отправной точки начальный набор концептов из онтологии, которые будут называться узлами или нодами. Все узлы имеют некоторое начальное значение активации. Основная идея работы алгоритма заключается в том, что во время распространения активируются другие узлы, а в конце работы алгоритма получается набор узлов с вычисленным значением активации. Значение активации исходного набора узлов задается в качестве входного параметра алгоритма. Таким образом, можно установить различные значения активации исходных узлов в зависимости от их важности в контексте решаемой задачи. Если для узлов из начального набора не указано исходное значение активации, алгоритм устанавливает для каждого из этих узлов значение, равное единице. Все узлы, которые не находятся в исходном наборе имеют первоначальное значение активации, равное нулю [12, 13].

Узлы из начального набора располагаются в приоритетной очереди в порядке убывания значения активации. Узел с наибольшим значением активации извлекается из очереди для дальнейшей обработки. Если узел удовлетворяет всем описанным ограничениям, то активация распространяется на все узлы, связанные с исходным

$$I_j(t+1) = O_i(t) * w_{ij} * f_{ij} * (1 - \alpha),$$

где i – начальный узел, j – узел назначения, I – обозначает вход, а O – выход. Выходное значение i -го узла добавляется к входному значению узла j . Значение активации для каждого найденного узла вычисляется путем умножения выходного значения узла $O_i(t)$, весового коэффициента ребра w_{ij} , факторов f_{ij} и коэффициента $(1 - \alpha)$.

Функция $O_i(t)$ определяет выходное значение для узла i . Для нашего случая была выбрана линейная функция (значение на выходе равно входному значению). Значение w_{ij} равно весовому коэффициенту связи между узлами i и j . Значение f_{ij} соответствует удельному весу, связанному с диапазоном предметной области, в которой находятся узлы i и j . Значение α – процент активации, который теряется каждый раз, когда происходит выход за диапазон предметной области, выступающий в качестве коэффициента затухания. Т.е. более короткие связи имеют преимущество перед более длинными. Нулевое значение коэффициента α означает распространение активации без потерь. Значения всех рассмотренных выше параметров зависят от задачи и предметной области [14, 15].

Алгоритм выполняется до тех пор, пока не будет достигнуто определенное состояние (например, заданный размер результирующего набора) или не осталось узлов, которые обрабатываются в приоритетной очереди. Алгоритм проходит по ребру между узлами только 1 раз $O(E)$. Приоритетная очередь имеет начальную сложность $O(\log(V))$. Таким образом, общая сложность алгоритма распространения активации равна

$$O(|E| * \log(|V|)),$$

где E – число существующих отношений между концептами (ребер в графе), а V – число концептов (узлов в графе).

Главной проблемой алгоритмов распространения активации является прохождение по всем узлам онтологии и, как следствие, получения неверных результатов в конечном наборе концептов. Для решения данной проблемы необходимо вводить правила и ограничения. Они учитываются как до обработки узла из приоритетной очереди, так и после нее. Можно использовать следующие ограничения:

- ◆ ограничение по типу – когда активация не должна распространяться через узлы определенного типа;
- ◆ ограничение по количеству связей – когда активация не должна распространяться через узлы, которые связаны с большим количеством концептов, чем задано;
- ◆ ограничение расстояния – распространения активации должно быть завершено, если достигнуты узлы, находящиеся дальше, чем указанный порог.

3. Семантический поиск. Общая архитектура семантического поиска показана на рис. 2 [16]. Первые два шага совпадают с традиционным поиском. Пользователь выражает свой запрос с точки зрения ключевых слов, которые обрабатываются поисковой системой. Поисковая система имеет доступ ко всем существующим концептам в базе знаний, т.е. данным, которые связаны с метаданными, представленным в качестве узла онтологии. Для достижения поставленной цели каждому концепту из базы знаний сопоставляется узел в онтологии, являющийся конкатенацией всех свойств концепта.

Результатом традиционной поисковой системы является набор узлов, отсортированных по их сходству с ключевыми словами из запроса. Этот набор узлов подается в алгоритм распространения активации в качестве исходного набора узлов. Поисковая система определяет весовой коэффициент для каждого узла, который указывает относительную важность этого узла по отношению к запросу пользователя. Он используется в качестве начального значения активации

для узла. Таким образом, узлы, которые ранжированы поисковой системой будут иметь приоритет при работе алгоритма распространения, так как поиск начинается с узлов с максимальным значением активации [17].

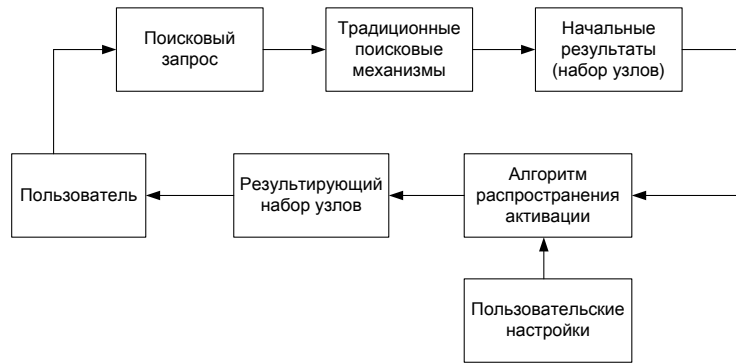


Рис. 2. Архитектура семантического поиска

Алгоритм распространения активации выполняется в соответствии с семантикой предметной области. Результатом работы алгоритма является набор узлов, при этом окончательный список узлов может не совпадать с результатами, полученными с использованием алгоритмов традиционного поиска. В зависимости от задачи, выполняемой пользователем, могут накладываться ограничения на некоторые связи между узлами онтологии. Кроме того, некоторые отношения могут быть более важными, чем другие. Все эти значения указываются при настройке алгоритма распространения активации инженером знаний, специалистом в конкретной предметной области. Значения весовых коэффициентов для всех типов отношений задаются с помощью конфигурационного XML файла [18].

Результирующий набор узлов упорядочивается по важности каждого узла для конкретного запроса пользователя. Например, исходя из рисунка 1, если запрос состоял из слов «онтология» и «web», то в результирующем наборе первым узлом будет идти «профессор Иванов», так как он имеет 3 публикации, связанные с ключевыми словами запроса, чем «профессор Петров», имеющий лишь 1 публикацию в этой области. Этот результат непосредственно связан с принципом работы алгоритма распространения активации.

Важным аспектом данного подхода является сохранение полного пути, пройденного алгоритмом, от начального до результирующего узла. Учитываются узлы с наибольшим весовым коэффициентом. Эта информация необходима специалисту для оценки и отладки поискового механизма в случае, если результирующий набор узлов не является удовлетворительным для конкретной задачи. Алгоритм распространения активации дает возможность «фильтровать» результаты поиска, например, пользователь может быть заинтересован в узлах, относящихся только к определенному типу (только узлы типа «Профессор» и т.д.).

4. Экспериментальные исследования. Первым приложением, выбранным для проведения тестирования, был web-сайт одной из кафедр (Dep) [19]. На данном web-сайте можно получить информацию об основных исследовательских направлениях, профессорах, проектах, студентах, лабораториях и публикациях. База знаний имеет около 2,630 узлов с 6,554 отношениями. Вторым приложением является сайт университета (Sfedu) [20], где так же представлена информация об институтах и кафедрах, профессорах и студентах, проектах и публикациях.

Для проведения экспериментальных исследований было выбрано два типа тестов. Первый заключается в поисковом запросе, который отражает лишь одну из тем, относящихся к области информационных технологий. Результатом будет являться динамически сгенерированная страница, связывающая сформированный запрос с множеством других узлов на сайте. Т.е. имея тему в качестве входных данных для запроса, в результате система предоставляет страницу с ключевыми ссылками на профессоров, студентов, публикации и т.д., связанные с данной темой. Второй заключается в использовании более чем одной темы в запросе. Идея заключается в том, чтобы сформировать первоначальный запрос, который будет носить весьма общий характер, чтобы проанализировать, как будет изменяться ранжирование узлов, при добавлении в запрос новой темы.

Экспериментальные исследования проводились в системе, разработанной на языке Java. В качестве традиционного поискового механизма использовался «движок» Lucene [21]. Система поддерживает весь процесс разработки семантического поиска, состоящего из следующих задач:

- ◆ импорт онтологии;
- ◆ настройка весовых коэффициентов;
- ◆ настройка алгоритма распространения активации;
- ◆ интеграция с рассматриваемыми приложениями.

Система имеет внутреннее представление онтологии и ее экземпляров, позволяющее осуществить преобразование импортируемой онтологии. Результаты экспериментов представлены в табл. 1.

Таблица 1

Краткое изложение полученных результатов

Приложение	Тип теста	Количество найденных узлов
Dep	1	20
Dep	2	13
Dep	2 (без вес. коэф.)	19
Sfedu	1	25
Sfedu	2	16
Sfedu	2 (без вес. коэф.)	25

Для сравнения предложенного гибридного подхода с традиционными поисковыми механизмами, был проведен тот же набор тестов с использованием семантической сети онтологии без функциональных возможностей отображения весовых коэффициентов. При этом результаты, полученные во время первого теста, полностью совпадали с результатами второго теста. Для достижения лучших результатов с использованием алгоритма семантического поиска необходимо предусмотреть множество нюансов, связанных с выбором наилучшего пути распространения активации. Несмотря на достижение хороших результатов, результаты, полученные с помощью гибридного подхода, были всегда выше, и требовали гораздо меньше усилий со стороны инженера знаний. Полученные результаты показывают, что алгоритм распространения активации с отображением весовых коэффициентов дает лучшие результаты и требует меньше времени для настройки, чем алгоритм семантического поиска.

Заключение. Предложен семантический поиск, основанный на традиционных поисковых механизмах в сочетании с информационным поиском на основе онтологий. Он позволяет определить неявные связи между концептами, описанными с помощью контекстной информации с семантически значимыми экземплярами онтологии, которые представлены в базе знаний.

Предложенный алгоритм распространения активации работает путем изучения концептов онтологии. Учитывая первоначальный набор концептов, алгоритм перемещается по связям в онтологии, получая в результате набор тесно связанных между собой концептов. Результирующий набор может содержать в себе концепты, которые явно не имеют отношений с концептами из начального набора. Значимость весовых коэффициентов, предложенных для использования в гибридном алгоритме распространения активации, выставляется лицом принимающим решение и зависит от конкретной предметной области и поставленных целей.

Главная проблема предложенного алгоритма распространения активации заключается в том, что не существует семантической интерпретации значения активации для всех узлов онтологии.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Курейчик В.М.* Особенности построения систем поддержки принятия решений // Известия ЮФУ. Технические науки. – 2012. – № 7 (132). – С. 92-98.
2. *Kravchenko Yu.A., Kureichik V.V.* Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8th IEEE International Conference “Application of Information and Communication Technologies – AICT 2014”. – 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.
3. *Davies J., Weeks R., and Krohn U.* QuizRDF: Search Technology for the Semantic Web // WWW2002 workshop on RDF & Semantic Web Applications, Proc. WWW2002, Hawaii, USA, 2008.
4. *Sheth A., Bertram C., Avant D., Hammond B., Kochut K., and Warke Y.* Managing Semantic Content for the Web // IEEE Internet Computing. – 2012. – No. 6 (4). – P. 80-87.
5. *Бова В.В.* Концептуальная модель представления знаний при построении интеллектуальных информационных систем // Известия ЮФУ. Технические науки. – 2014. – № 7 (156). – С. 109-117.
6. *Курейчик В.В., Родзин С.И.* О правилах представления решений в эволюционных алгоритмах // Известия ЮФУ. Технические науки. – 2010. – № 7 (108). – С. 13-21.
7. *Stojanovic N., Struder R., and Stojanovic L.* An Approach for the Ranking of Query Results in the Semantic Web. Proc. of ISWC '03 (Sanibel Island, FL, October 2003), SpringerVerlag, 2013. – P. 500-516.
8. *Chen H., and Ng T.* An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation); Symbolic Branch-and-Bound vs. Connectionist Hopfield Net Activation // Journal of the American Society for Information Science. – 2005. – No. 46 (5). – P. 348-369.
9. *Srikant R., and Agrawal R.* Mining generalized association rules // Proceedings of VLDB '95. – 2010. – P. 407-419.
10. *Peat H., and Willet P.* The limitations of term co-occurrence data from query expansion in document retrieval systems // Journal of the American Society for Information Science. – 2012. – No. 42 (5). – P. 378-383.
11. *Башмаков А.И., Башмаков И.А.* Интеллектуальные информационные технологии: учеб. пособие. – М.: МГТУ им. Н.Э. Баумана, 2005. – 304 с.
12. *Kureichik V.M., Rodzin S.I.* Evolutionary algorithms: genetic programming // Journal of Computer and Systems Sciences International. – 2002. – Vol. 41, No. 1. – P. 123-132.
13. *Анечков М.И., Бова В.В., Новиков А.А.* Эволюционный подход к созданию нейросетевой модели коллективного решения интеллектуальных задач // Известия Кабардино-Балкарского научного центра РАН. – 2015. – № 5 (67). – С. 24-30.
14. *Кулиев Э.В., Самойлов А.Н., Новиков А.А.* Когнитивная архитектура агентов мультиагентной системы // Информатизация и связь. – 2016. – № 2. – С. 116-120.
15. *Тузовский А.Ф., Чириков С.В., Ямпольский В.З.* Системы управления знаниями (методы и технологии) / под общ. ред. В.З. Ямпольского. – Томск: Изд-во НТЛ, 2005. – 260 с.
16. *Bova V.V., Kravchenko Y.A., Kureichik V.V.* Decision Support Systems for Knowledge Management // Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol. 3. – Springer International Publishing AG Switzerland. – P. 123-130.

17. *Кравченко Ю.А.* Синтез разнородных знаний на основе онтологий // Известия ЮФУ. Технические науки. – 2012. – № 11 (136). – С. 141-145.
18. *Кравченко Ю.А., Марков В.В.* Онтологический подход формирования информационных ресурсов на основе разнородных источников знаний // Известия ЮФУ. Технические науки. – 2013. – № 7 (144). – С. 116-120.
19. PUC-Rio Informatics Dept. [Электронный ресурс]. – URL: <http://www.inf.puc-rio.br> (дата обращения: 13.05.2016).
20. Южный Федеральный Университет [Электронный ресурс]. – URL: <http://sfedu.ru> (дата обращения: 13.05.2016).
21. *Курейчик В.М., Кажаров А.А.* Использование шаблонных решений в муравьиных алгоритмах // Известия ЮФУ. Технические науки. – 2013. – № 7 (144). – С. 11-17.

REFERENCES

1. *Kureychik V.M.* Osobennosti postroeniya sistem podderzhki prinyatiya resheniy [Features of decision making support system design], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 7 (132), pp. 92-98.
2. *Kravchenko Yu.A., Kureichik V.V.* Knowledge management based on multi-agent simulation in informational systems, *Conference proceedings. 8th IEEE International Conference "Application of Information and Communication Technologies – AICT 2014". 15-17 October 2014, Astana, Kazakhstan*, pp. 264-267.
3. *Davies J., Weeks R., and Krohn U.* QuizRDF: Search Technology for the Semantic Web, *WWW2002 workshop on RDF & Semantic Web Applications, Proc. WWW2002, Hawaii, USA, 2008*.
4. *Sheth A., Bertram C., Avant D., Hammond B., Kochut K., and Warke Y.* Managing Semantic Content for the Web, *IEEE Internet Computing*, 2012, No. 6 (4), pp. 80-87.
5. *Bova V.V.* Kontseptual'naya model' predstavleniya znaniy pri postroenii intellektual'nykh informatsionnykh sistem [Conceptual model of knowledge representation in the constructing intelligent information systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2014, No. 7 (156), pp. 109-117.
6. *Kureychik V.V., Rodzin S.I.* O pravilakh predstavleniya resheniy v evolyutsionnykh algoritmakh [On the rules for the submission decisions in evolutionary algorithm], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 7 (108), pp. 13-21.
7. *Stojanovic N., Struder R., and Stojanovic L.* An Approach for the Ranking of Query Results in the Semantic Web. Proc. of ISWC '03 (Sanibel Island, FL, October 2003), SpringerVerlag, 2013, pp. 500-516.
8. *Chen H., and Ng T.* An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation); Symbolic Branch-and-Bound vs. Connectionist Hopfield Net Activation, *Journal of the American Society for Information Science*, 2005, No. 46 (5), pp. 348-369.
9. *Srikant R., and Agrawal R.* Mining generalized association rules, *Proceedings of VLDB '95*, 2010, pp. 407-419.
10. *Peat H., and Willet P.* The limitations of term co-occurrence data from query expansion in document retrieval systems, *Journal of the American Society for Information Science*, 2012, No. 42 (5), pp. 378-383.
11. *Bashmakov A.I., Bashmakov I.A.* Intellektual'nye informatsionnye tekhnologii: ucheb. Posobie [Intelligent information technologies: a tutorial]. Moscow: MGTU im. N.E. Baumana, 2005, 304 p.
12. *Kureichik V.M., Rodzin S.I.* Evolutionary algorithms: genetic programming, *Journal of Computer and Systems Sciences International*, 2002, Vol. 41, No. 1, pp. 123-132.
13. *Anchekov M.I., Bova V.V., Novikov A.A.* Evolyutsionnyy podkhod k sozdaniyu neyrosetevoy modeli kollektivnogo resheniya intellektual'nykh zadach [An evolutionary approach to create neural network models of collective decision of intellectual tasks], *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [Izvestija Kabardino-Balkarskogo nauchnogo centra RAN], 2015, No. 5 (67), pp. 24-30.
14. *Kuliev E.V., Samoylov A.N., Novikov A.A.* Kognitivnaya arkhitektura agentov mul'tiagentnoy sistemy [Cognitive architecture of agents multi-agent systems], *Informatizatsiya i svyaz' [Informatization and communication]*, 2016, No. 2, pp. 116-120.

15. *Tuzovskiy A.F., Chirikov S.V., Yampol'skiy V.Z.* Sistemy upravleniya znaniyami (metody i tekhnologii) [The knowledge management system (methods and technology)], under ed. V.Z. Yampol'skogo. Tomsk: Izd-vo NTL, 2005, 260 p.
16. *Bova V.V., Kravchenko Y.A., Kureichik V.V.* Decision Support Systems for Knowledge Management, *Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol. 3.* Springer International Publishing AG Switzerland, pp. 123-130.
17. *Kravchenko Yu.A.* Sintez raznorodnykh znaniy na osnove ontologii [Synthesis of heterogeneous knowledge based on ontologies], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 11 (136), pp. 141-145.
18. *Kravchenko Yu.A., Markov V.V.* Ontologicheskii podkhod formirovaniya informatsionnykh resursov na osnove raznorodnykh istochnikov znaniy [Ontological approach formation of information resources based on knowledge disparate sources], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 116-120.
19. PUC-Rio Informatics Dept. [Electronic resource]. Available at: <http://www.inf.puc-rio.br> (accessed 13 May 2016).
20. Yuzhnyy Federal'nyy Universitet [Electronic resource]. Available at: <http://sfedu.ru> (13 May 2016).
21. *Kureychik V.M., Kazharov A.A.* Ispol'zovanie shablonnykh resheniy v murav'inykh algoritmakh [Template using for ant colony algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 11-17.

Статью рекомендовал к опубликованию д.т.н., профессор М.М. Ошхунов.

Кравченко Юрий Алексеевич – Южный федеральный университет; e-mail: yakravchenko@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Марков Владимир Васильевич – e-mail: v_v_mar@mail.ru; кафедра систем автоматизированного проектирования; доцент.

Новиков Антон Александрович – e-mail: anton.a.novikov@gmail.com; кафедра систем автоматизированного проектирования; аспирант.

Kravchenko Yury Alekseevich – Southern Federal University; e-mail: yakravchenko@sfedu.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Markov Vladimir Vasilyevich – e-mail: v_v_mar@mail.ru; the department of computer aided design; associate professor.

Novikov Anton Alexandrovich – e-mail: anton.a.novikov@gmail.com; the department of computer aided design; postgraduate.

УДК 002.53:004.89

Ю.А. Кравченко, И.О. Курсытс, Э.В. Кулиев

РАЗРАБОТКА ГЕНЕТИЧЕСКОГО АЛГОРИТМА РАСЧЕТА СЕМАНТИЧЕСКОЙ БЛИЗОСТИ В ЗАДАЧАХ УПРАВЛЕНИЯ ЗНАНИЯМИ*

Статья посвящена разработке нового подхода к оценке семантической близости для решения различного рода задач управления знаниями (knowledge management). В связи с постоянным ростом информационных потоков в различных сферах жизни задачи поиска новых способов хранения, представления, формализации, систематизации и обработки

* Исследование выполнено за счет гранта Российского научного фонда (проект № 14-11-00242) в Южном федеральном университете.