

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.

информации из гетерогенных источников представляются актуальными на сегодняшний день. Основная проблема в области поиска знаний заключается в постоянно растущей сложности идентификации и использования ключевой информации. Для ее решения предложена модификация ранее разработанного фильтра знаний, функционирующего на основе построения семантического дерева таксономии понятий, как систематизации сложноорганизованных областей действительности и знания, имеющих иерархическое строение, с целью определения и упорядочивания терминов и их синонимов, с дальнейшим преобразованием пользовательского запроса в наиболее эффективную форму. В мета-модель фильтра знаний добавлен блок оценки семантической близости, на основе которого будет производиться выдача наиболее подходящих с точки зрения семантики результатов. Проведен анализ поставленной задачи: дано определение семантической близости, рассмотрены различные подходы для ее расчета. Предложена графовая модель компонентов триплетов онтологий, приведена формула вычисления семантической близости. Для повышения эффективности подхода разработан генетический алгоритм расчета семантической близости в соответствии с графовой моделью. Предложен ряд модифицированных генетических операторов кроссинговера и мутации. Проведены вычислительные эксперименты на тестовых примерах (бенчмарках). Вычислена временная сложность разработанного алгоритма. Проведенные серии тестов и экспериментов показали эффективность применения данного подхода.

Управление знаниями; онтологии; мета-модели; семантическая близость; графовая модель; генетический алгоритм.

Yu.A. Kravchenko, I.O. Kursitys, E.V. Kuliev

THE DEVELOPMENT OF GENETIC ALGORITHM FOR SEMANTIC SIMILARITY ESTIMATION IN TERMS OF KNOWLEDGE MANAGEMENT PROBLEMS

This article is devoted to the development of a new approach for semantic similarity estimation to solve different knowledge management problems. Due to information flows constantly growing in various life spheres the problems of searching new ways of storing, representation, formalization, systematization and processing of information from heterogeneous sources are relevant today. The main problem in the field of knowledge search is the complexity of identification and usage of key information, which is increasing constantly. To solve this problem we propose to modify previously developed knowledge filter running on the basis of the semantic concepts taxonomy tree as an systematization of complex areas of the reality and hierarchical knowledge in order to define and arrange terms and its synonyms with a further query transformation into the most effective form. The knowledge filter meta-model is supplemented by a semantic similarity estimation block, that allows us to obtain the most appropriate results in the context of semantics. We analyzed the assigned problem, gave a definition of the term 'semantic similarity', and observed different ways of its estimation. To solve the problem we proposed the graph model containing components of ontology triplets. The semantic similarity formula is presented in this paper. To increase the efficiency we developed the genetic algorithm for semantic similarity estimation in accordance with the graph model. A set of genetic operators of crossover and mutation is proposed for genetic algorithm work. Experiments carried out on benchmarks show the efficiency of developed approach.

Knowledge management; ontologies; meta-model; semantic similarity; graph model; genetic algorithm.

Введение. В современном мире процесс развития общества характеризуется постоянно возрастающей ролью информационных технологий в науке, производстве и управлении. Повсеместное распространение информационных технологий и доступа к сети Интернет привело к тому, что за последние годы существенно увеличился объем генерируемой, передаваемой и обрабатываемой информации. В связи с этим актуальными на сегодняшний день являются задачи поиска новых способов хранения информации, ее представления, формализации и систематиза-

ции, а также автоматической обработки. Таким образом, растет интерес к всеобъемлющим базам знаний, которые возможно использовать для различных практических задач. Огромный интерес вызывают системы, способные без участия человека извлечь какие-либо сведения из текста. Как результат, на фоне вновь возникающих потребностей развиваются новые технологии, призванные решить заявленные проблемы.

В настоящее время проблемой поиска знаний занимаются специалисты в области управления личными и корпоративными знаниями (Knowledge Management), так как это позволит сократить временные и трудовые затраты на выполнение задач и принятие решений, как на рабочем месте, так и в повседневной жизни [1, 2]. Особенно остро стоит проблема поиска знаний в сети Интернет, т.к. при постоянно растущем количестве информации, извлечь действительно нужную и важную информацию становится все труднее [3].

Основная проблема в области поиска знаний заключается в постоянно растущей сложности идентифицирования и использования ключевой информации. Одним из способов решения данной проблемы является усовершенствование подходов семантического моделирования в направлении интерпретации и использования в качестве априорных данных профилей поиска других пользователей, преследующих схожие цели [4, 15–16].

1. Фильтр знаний. Для решения проблем, возникающих в сфере управления знаниями, в работе [4] было предложено использовать фильтр предварительной обработки поисковых запросов – фильтр знаний. Важной проблемой, препятствующей повышению эффективности запросов поиска информации в неоднородных распределенных источниках знаний, является семантическое несоответствие представления исследователя (пользователя) о составлении запроса и форме его выражения и того, как запрос должен быть выражен на основе конкретного интерфейса используемой информационной системы. Фильтр знаний функционирует на основе построения семантического дерева таксономии понятий, как систематизации сложноорганизованных областей действительности и знания, имеющих иерархическое строение, с целью определения и упорядочивания терминов и их синонимов, с дальнейшим преобразованием пользовательского запроса в наиболее эффективную форму.

Оценка качества модифицированного запроса проводится на основе многокритериальной модели поддержки принятия решений, которая определяет «вес» запроса по синтаксическим, семантическим, категориальным и приоритетным критериям.

Основными достоинствами предложенного в [4] фильтра знаний являются:

- 1) поддержка семантического поиска релевантных знаний на основе онтологических моделей;
- 2) использование информационных энциклопедических справочных систем различной функциональности для усовершенствования формы поискового запроса;
- 3) повышение эффективности запроса пользователя на основе использования репозитория прецедентов.

Кроме того, в работе [4] разработана мета-модель фильтра знаний, позволяющая определить семантику сквозного потока операций, управляющего всем поисковым процессом, включая спецификацию запроса, переформулировку запроса, разбиение запроса, выбор веб-сервиса, выбор источника знаний, результаты ранжирования и представление рекомендаций (рис. 1).

Для повышения эффективности работы представленного фильтра знаний предлагается ввести дополнительный блок расчета семантической близости между объектами поиска, который будет использоваться агентом ранжирования.

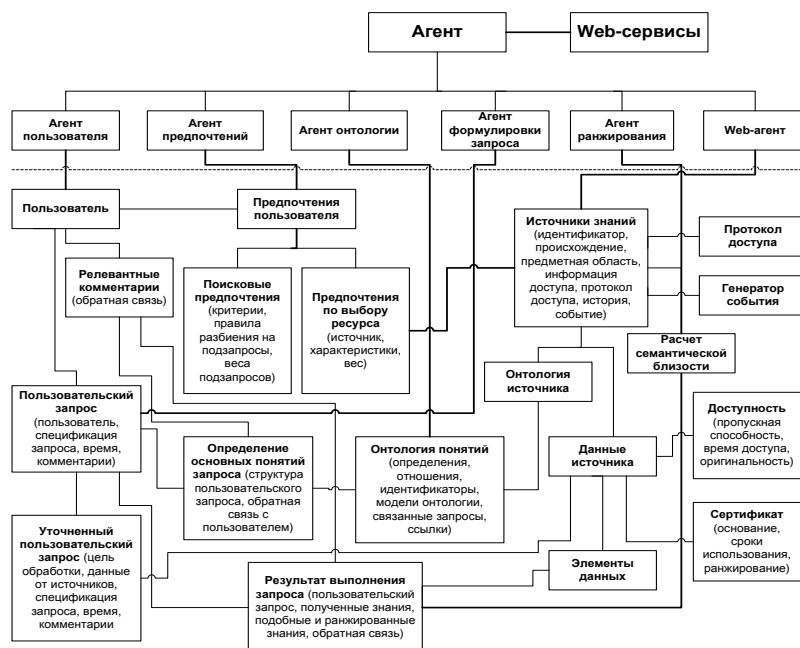


Рис. 1. Мета-модель фильтра знаний

Согласно данной модели, агент пользователя отвечает за взаимодействие и применение пользовательских настроек, которыми управляет Агент предпочтений. Предпочтения включают в себя: вес термина, обозначающий его относительную значимость; важность результатов запроса; предпочтения различных источников знаний и т.д. Агент Предпочтений может также изучить предпочтение пользователя на основе опыта и обратной связи, связанной с предыдущими запросами.

Агент онтологии имеет доступ к образам моделей исследуемых предметных областей, определенных с помощью открытого стандарта конструкции онтологии – веб-языка Онтологии (OWL). Энциклопедические справочные системы позволяют агенту онтологии использовать обобщающие тематические термины и синонимы.

Первоначальный запрос пользователя, попадая Агенту формулировки запроса, впоследствии совершенствуется Агентом онтологии на основе семантической обработки, обеспеченной доступными моделями онтологий предметных областей и энциклопедическими справочными системами. Переформулированный запрос разбивается Агентом формулировки запроса на подзапросы, предназначенные для надлежащих источников знаний. Это поддерживает семантическое посредничество терминологии, используемой в онтологии модели предметной области и справочных системах с используемыми локальными источниками. Кроме того, перевод запроса необходим, чтобы получить данные из предназначенных неоднородных источников знаний.

Агент ранжирования несет ответственность за сбор результатов подзапросов от различных источников знаний, оценивая их по семантическим критериям согласно требованиям Агента предпочтений. Предпочтения определяются на основе веса авторитетности используемого ресурса (от 0 до 10) и веса термина, включенного в запрос.

Основная роль Web-агента – получение переформулированного запроса, после обработки Агентом онтологии и разбиения на подзапросы Агентом формулировки запроса. Web-агент отправляет подзапросы и адаптирует неоднородные распределенные источники знаний с учетом аспектов: пользовательской настройки

ресурсов; авторитетности и репутации ресурса; соглашения об условиях использования; оценки размера ответов на подзапрос; меры качества обслуживания сетевого трафика и динамической рабочей нагрузки ресурса [5, 20].

Каждый агент в представленной мета-модели управляет соответствующими классами объектов, обрабатывает спецификации и веб-сервисы. Например, пользователь определяет Пользовательские предпочтения, которые могут быть специализированы для Поисковых Предпочтений и Предпочтений по выбору ресурса. Пользователь излагает Пользовательский запрос, у которого есть несколько Понятий запроса, которые поочередно взаимодействуют с Онтологией понятий. Агент Онтологии управляет и Пользовательским запросом и Онтологией понятий, которые предоставляются Онтологией Источника. Онтология источника и Данные источника – специализации Источника. Источником управляет Web-агент [4, 17].

Пользовательский запрос содержит несколько уточненных запросов, каждый из которых изложен к нескольким источникам данных, которые обеспечивают один или несколько элементов данных в качестве результата запроса. На основе возвращенного результата запроса пользователь может использовать обратную связь относительно уместности результата и других комментариев.

Результаты поиска влияют на развитие метаданных, связанных с предпочтениями пользователя, формулировкой запроса, использованием источника данных и ранжированием результата. Мета-модель фильтра знаний может быть реализована в виде реляционной базы данных [4].

2. Семантическая близость. Под семантикой текста обычно понимается его «смысл», который автор текста хотел передать посредством символов, однако для компьютерной системы смысл текста строго зависит от контекста, где он определяется и обрабатывается. В области семантического веба лучшим средством представления семантики является онтология.

Онтологические модели обладают следующими преимуществами:

- ◆ Обеспечение простоты развития системы;
- ◆ Возможность получения в результате запроса не «сырых» данных, а логического вывода на основе хранимых данных;
- ◆ Легкость моделирования сложных связей и отношений, по сравнению с базами данных;
- ◆ Возможность использования согласованной (разделяемой всеми) терминологии с точно определенной семантикой, позволяющего интегрировать и использовать данные (информацию) из разных источников;
- ◆ Возможность динамического изменения данных [6, 19].
- ◆ Под онтологической моделью (онтологией) O понимается знаковая система $\langle C, P, I, L, T \rangle$, где:

C – множество элементов, которые называются понятиями;

P – множество элементов, называемые свойствами (двуместными предикатами);

I – множество экземпляров понятий;

L – множество текстовых меток или значений понятий и свойств;

T – частичный порядок на множестве C и P .

Оценкой близости между документом и запросом является числовое значение, которое выражает степень сходства между ними; оценка близости называется оценкой семантической близости, если и только если она определена на основе семантики документов и запросов.

В работе [7] сделан обзор методов вычисления мер семантической близости термов внутри онтологий. Выделяют следующие типы мер семантической близости:

1) таксономические – на основе иерархических (родовидовых, таксономических) связей;

2) реляционные – на основе неиерархических (ассоциативных, проблемно-специфических, «горизонтальных») связей между терминами онтологии и атрибутивные.

Методики на основе иерархических отношений подразделяются на:

1) основанные на определении кратчайшего пути (количество ребер или вершин) между вершинами;

2) основанные на определении глубины таксономического дерева;

3) учитывающие глубину наименьшей общей родовой вершины – ближайшего общего родителя (least common subsumer – LCS);

4) основанные на понятии общей специфичности двух вершин.

При этом вводятся некоторые ограничения, например, ограничивают конфигурацию пути: длину пути и количество перегибов.

Для расчета меры семантической близости на основе неиерархических (ассоциативных связей) используют сравнение с третьим понятием плюс рекурсивное уточнение [7].

Атрибутивную меру близости рассчитывают по значениям общих атрибутов двух понятий [8].

Интерес представляют так называемые гибридные методы, используемые для учета всех характеристик двух термов при расчете меры семантической близости и содержащие оценку близости экземпляров, состоящую из трех частей – таксономической, реляционной и атрибутивной [8]:

$$S(i_1, i_2) = t \cdot S^t(i_1, i_2) + p \cdot S^p(i_1, i_2) + a \cdot S^a(i_1, i_2), \quad (1)$$

где i_1, i_2 – объекты, между которыми требуется найти семантическую близости; t, p, a – коэффициенты, определяющие «вес» меры каждого типа в общей мере.

Недостатком большинства мер, основанных на онтологических структурах, является симметричность. Экспертные оценки показывают, что мера близости не всегда симметрична как для иерархических отношений, так и для ассоциативных. Другой актуальной задачей является определение семантической близости онтологических термов, не связанных иерархически, однако, имеющих определенную проблемно-специфическую («горизонтальную», ассоциативную) связь.

Задача определения семантической близости является особо актуальной в узкоспециализированных предметных областях, так как при выделении специализированной области знания из более общей, многие понятия оказываются расположенными слишком далеко от своего родового понятия, которое является чересчур общим и принадлежит более общей предметной области, в которую «вложена» рассматриваемая. Таким образом, в специализированных предметных областях оказывается значительное количество терминов, не имеющих родовидовых связей, либо таксономия для них имеет всего один уровень (предок-потомок).

Для устранения указанных недостатков в работе [9] было предложено использовать графовую модель расчета семантической близости.

С помощью набора предикатов P онтологии могут описываться различные отношения между понятиями и экземплярами. Эти отношения задаются с использованием простых утверждений (триплетов) $\langle s, p, o \rangle$, где s и o – это субъект и объект высказывания, а $p \in P$ – это предикат онтологии O [9].

Будем считать, что любому свойству $p \in P$ может быть задан весовой коэффициент (семантический вес) $rv \in [0, 1]$, задающий смысловую близость между субъектом и объектом утверждения (значение коэффициента 1 означает, субъект и объект считаются сходными по семантике, значение коэффициента 0 – не похожими). Задание значений коэффициентов rv для предикатов выполняется специалистами-разработчиками в соответствии с их пониманием онтологии и потребностями решаемых задач.

Построим неориентированный граф G всех имеющихся субъектов и объектов триплетов, имеющихся в базе знаний, в соответствии со следующими правилами:

- ◆ используются только те триплеты, у которых значения весовых коэффициентов предикатов не равны нулю ($pv \neq 0$);
- ◆ вершинами графа являются субъекты и объекты триплетов, а ребра графа, соединяющие субъекты с объектами имеют веса, равные значению pv предиката того триплета, с использованием которого они были сформированы;
- ◆ инверсное отношение (на основе предиката owl:inverseOf) между предикатами $p1$ ($pv1$) и $p2$ ($pv2$) добавляет в граф два ребра с весами $pv1$ и $pv2$;
- ◆ симметричное отношение добавляет в граф два ребра с равными весами, например, <owl:sameAs> добавляет два ребра со значениями $pv = 1.0$.

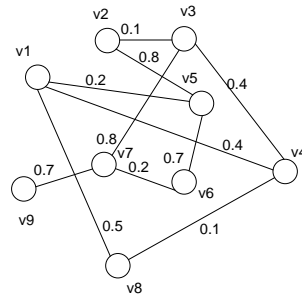


Рис. 2. Абстрактная графовая модель компонентов триплетов

Обозначим через $Sim(\alpha, \beta)$ семантическую близость между элементами α и β , где $\alpha, \beta \in C \cup I \cup P \cup T$. С учетом использования указанной графовой модели семантическая близость будет определяться как

$$Sim(\alpha, \beta) = \max_{i=1 \rightarrow k} (Sim_{PATH_i}(\alpha, \beta)), \quad (2)$$

где k – число возможных путей графа GO от вершины α до вершины β .

Под путём $PATH(\alpha, \beta)$ между двумя вершинами α и β графа GO понимается набор рёбер (предикатов) ведущих от вершины α до вершины β , с учётом их направленности.

Значение семантической близости между элементами α и β по направлению пути i $Sim(\alpha, \beta)$ определяется по следующей формуле:

$$Sim_{PATH_i}(\alpha, \beta) = \prod_{j=1}^{h_i} pv_{i,j} \quad (3)$$

где h_i – число семантических отношений между элементами α и β на пути i ; $pv_{i,j}$ – значение веса ребра на основе j -го семантического предиката на пути i .

Тогда формула определения семантической близости между элементами онтологии α и β будет иметь вид

$$Sim(\alpha, \beta) = \max_{i=1 \rightarrow k} (Sim_{PATH_i}(\alpha, \beta)) = \max_{i=1 \rightarrow k} \left(\prod_{j=1}^{h_i} pv_{i,j} \right). \quad (4)$$

Величина $Sim(\alpha, \beta)$ удовлетворяет следующим свойствам:

- 1) $Sim(\alpha, \beta) \in [0, 1]$;
- 2) $Sim(\alpha, \beta) = 0$ при отсутствии пути от α к β ;
- 3) $Sim(\alpha, \alpha) = Sim(\beta, \beta) = 1$.

В исключительном случае $Sim(\alpha, \beta)$ может равняться 1, при условии существовании инверсного отношения между элементами α, β .

Для вычисления семантической близости между вершинами α и β графа G используется функция $Sim(\alpha, \beta)$, которая возвращает максимальное значение близости по возможным путям между вершинами α и β . Функция $Sim(\alpha, \beta)$ вызывает на выполнение функцию $PATH(\alpha, \beta)$, которая в свою очередь вызывает функцию $PATH(\alpha, \beta, Path)$.

Функция $PATH(\alpha, \beta, Path)$ является рекурсивной и вызывается для каждого не пройденного ребра, исходящего из вершины α , при этом параметр α и список ребер по пути $Path$ изменяются.

Работа функции $PATH(\alpha, \beta, Path)$ прекращается при следующих условиях:

(1) По новому пути $Path$ не получается максимального значения близости ($<maxWeight$), при этом условие ($maxWeight > 0$) означает существование пути между исходной вершиной α и конечной вершиной β .

(2) Вершина α была пройдена ($\alpha \in PassedVertex$).

(3) Вершины α и β совпали ($\alpha = \beta$).

Так как количество компонентов онтологий может быть непредсказуемо большим, можно говорить о том, что задача оценки семантической близости между ними относится к классу NP-полных. Это означает, что для них не существуют точные эффективные, отличные от полного перебора, алгоритмы, а полный перебор не даст решения за полиномиальное время.

На сегодняшний день конкурентоспособность при решении многих NP-трудных задач доказали генетические алгоритмы (ГА). Их эффективность особенно отмечена в тех задачах, где математические модели настолько сложны, что применение стандартных методов оптимизации, например, ветвей и границ, динамического или линейного программирования, крайне затруднено [10, 14].

3. Генетический алгоритм расчета семантической близости. Для повышения эффективности подхода, описанного в [9] предлагается использовать генетический алгоритм расчета семантической близости.

Преимущества ГА перед другими оптимизационными и поисковыми процедурами заключаются в следующем:

1) в ГА поиск оптимизации задачи осуществляется не с одним решением, а с множеством решений. Это позволяет находить новые решения на основе старых решений, которые являются лучшими среди множества. Таким образом, свойства оптимальных решений развиваются;

2) решение задачи является закодированной структурой, а не совокупностью параметров. Благодаря этому увеличивается скорость процесса обработки данных, т.е. быстродействие оптимизационного поиска;

3) для оценки приспособленности той или иной хромосомы в дополнение к подсчету ее целевой функции задаются правила выживания в исследуемой популяции. Это развивает разнообразие популяции, что способствует повышению вероятности нахождения правильного решения;

4) при создании популяции, скрещивании или изменении хромосом используются вероятностные правила, которые вносят в направленность генетического поиска элементы случайности, что позволяет повысить шансы преодоления локальных оптимумов [11, 13, 18].

Помимо вышеперечисленных качеств, генетические алгоритмы обладают следующими свойствами:

1) достаточно широкая область применения алгоритмов;

2) возможность комбинирования генетических алгоритмов с другими методами, в том числе и неэволюционными алгоритмами;

3) эффективность алгоритма в поиске решений в пространстве большой размерности;

4) отсутствие ограничений на вид целевой функции;

5) наглядность схемы построения генетических алгоритмов [12].

Разработанный генетический алгоритм работает с графовой моделью субъектов и объектов триплетов, описанной в разделе 2. Задача оценки семантической близости состоит в нахождении максимального произведения весовых коэффициентов на пути, ведущем от вершины α к вершине β .

Хромосома представляет собой последовательность вершин, представляющая собой путь из вершины α в вершину β . Хромосомы генерируются случайным образом. Приспособленность хромосом определяет целевая функция.

Подсчет целевой функции производится в соответствии с заданными весовыми коэффициентами предикатов по принципу выявления такой последовательности вершин, при которой произведение коэффициентов будет максимальным. Пусть pv_1, pv_2, \dots, pv_n – весовые коэффициенты предикатов, заданные специалистами при разработке онтологии. Тогда ЦФ будет сводиться к максимизации

$$\prod_{i=1}^n pv_i \rightarrow \max. \quad (5)$$

Для решения задачи оценки семантической близости с помощью генетического алгоритма использовались следующие правила [11, 13–14]:

1) С заданной вероятностью реализуется оператор кроссинговера над генотипами наиболее приспособленных хромосом. Далее с вероятностью 0,5 выбирается один из потомков $P_i(t)$ и сохраняется как член новой популяции.

2) К $P_i(t)$ применяется оператор инверсии и оператор мутации с заданными вероятностями. Полученный генотип сохраняется как $P_k(t)$.

Для работы генетического алгоритма был применен ряд генетических операторов. Это двухточечный, модифицированный трехточечный и модифицированный жадный операторы кроссинговера; одноточечный оператор мутации, модифицированный оператор мутации на основе дихотомии, модифицированный оператор мутации с равновероятным распределением и оператор инверсии.

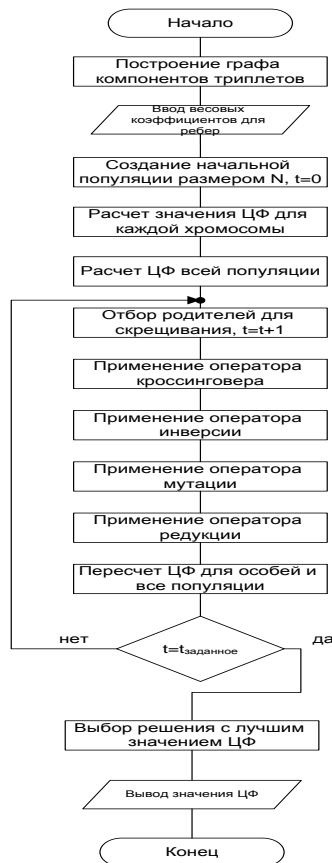


Рис. 3. Генетический алгоритм расчета семантической близости

Условие останова генетического алгоритма – достижение заданного количества популяций.

4. Экспериментальные исследования. Экспериментальные исследования, проведенные с разным количеством вершин графа, показали, что временная сложность алгоритма полиномиальная $O(n^2)$. График временной сложности алгоритма показан на рис. 4.

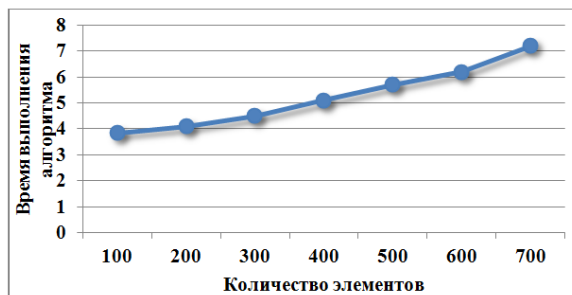


Рис. 4. График временной сложности алгоритма

В табл. 1 показаны результаты вычислительных экспериментов, проведенных с разным количеством вершин и ребер графа. В экспериментах участвовали разработанный генетический алгоритм оценки семантической близости и жадный алгоритм.

Таблица 1

Результаты вычислительных экспериментов

Граф компонентов триплетов G(X,U)		ЦФ ГА	ЦФ ЖА
Число вершин	Число ребер		
100	70	0,00003985	0,000028
200	90	0,000005893	0,0000042
300	110	0,00000065	0,00000045
400	130	0,0000000123	0,000000011
500	150	0,000000000485	0,00000000026
600	170	0,0000000000458	0,00000000000250
700	190	0,0000000000008523	0,0000000000000569

Как показывает табл. 1, генетический алгоритм в среднем является на 29 % более эффективным, чем жадный алгоритм.

Заключение. В XXI веке развитие общества неразрывно связано с развитием информационных технологий. За последние годы существенно увеличился объем генерируемой, передаваемой и обрабатываемой информации. В связи с этим актуальными на сегодняшний день являются задачи поиска новых способов хранения информации, ее представления, формализации и систематизации, а также автоматической обработки. Распространенным методом решения указанных проблем является использование семантики обрабатываемых понятий.

В исследованиях связанных с моделированием естественной языковой семантики особую актуальность приобретают разработки количественных методов измерения семантической информации в языке и в тексте. Одной из важных задач в этой области является расчет семантической близости (или инверсной величины – семантического расстояния) между единицами лексикографической системы (словаря, тезауруса, онтологии).

В статье проанализированы распространенные способы оценки семантической близости на основании иерархических и неиерархических связей. Предложена графовая модель компонентов триплетов онтологий, приведена формула для расчета семантической близости. Разработан генетический алгоритм, работающий с указанной графовой моделью, предложен ряд генетических операторов. Проведенные экспериментальные исследования показали эффективность такого подхода.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Dorsey P.* Personal knowledge management [e-resource]. – URL: http://www.360doc.com/content/05/1228/22/2563_51065.shtml (дата обращения 16.02.2016).
2. *Martin J.* Personal Knowledge Management. The Basis of Corporate and Institutional Knowledge Management // *Managing Knowledge: Case Studies in Innovation*. – Alberta: University of Alberta, faculty of Extension, 2000. – Vol. 6.
3. *Ландэ Д.В.* Поиск знаний в Internet. – М.: Диалектика, 2005. – 271 с.
4. *Bova V.V., Kravchenko Y.A., Kureichik V.V.* Models for Supporting of Problem-Oriented Knowledge Search and Processing // *Proceedings of the First International Scientific Conference “Intelligent Information Technologies for Industry” (ITI’16)*. – Vol. 1. – P. 287-295
5. *Kerschberg L., Jeong H., Kim W.* Emergent Semantic in Knowledge Sifter: An Evolutionary Search Agent based on Semantic Web Services. In: Spaccapietra, S., Aberer, K., Cudre-Mauroux, P. (eds.) // *Journal on Data Semantic VI. LNCS*. – 2006. – Vol. 4090. – P. 187-209.
6. *Гончар А.Д.* Сравнительный анализ баз данных и баз знаний (онтологий) применимо к моделированию сложных процессов // *Современные научные исследования и инновации*. – 2014. – № 5. <http://web.snauka.ru/issues/2014/05/34325> (дата обращения: 27.03.2016).
7. *Крюков К.В., Панкова Л.А., Пронина В.А.* Меры семантической близости в онтологии // *Проблемы управления*. – 2010. – № 5. – С. 2-14.
8. *Maedche A., Zacharias V.* Clustering Ontology-Based Metadata in the Semantic Web // *Proceedings PKDD-2002, LNAI 2431*. – 2002. – P. 348-360.
9. *Ле Хаой, Тузовский А.Ф.* Разработка семантических электронных библиотек на основе онтологических моделей // *Труды XV Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL 2013* (Ярославль, Россия, 14.10–17.10.2013). – Ярославль: Ярославского государственного университета им. П.Г. Демидова, 2013.
10. *Курейчик В.М., Кажаров А.А.* Использование шаблонных решений в муравьиных алгоритмах // *Известия ЮФУ. Технические науки*. – 2013. – № 7 (144). – С. 11-17.
11. *Гладков Л.А., Курейчик В.М., Курейчик В.В.* Генетические алгоритмы. – М.: Физмат-лит, 2006. – 320 с.
12. *Курейчик В.М.* Особенности построения систем поддержки принятия решений // *Известия ЮФУ. Технические науки*. – 2012. – № 7 (132). – С. 92-98.
13. *Курейчик В.В., Родзин С.И.* О правилах представления решений в эволюционных алгоритмах // *Известия ЮФУ. Технические науки*. – 2010. – № 7 (108). – С. 13-21.
14. *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.
15. *Bova V.V., Kravchenko Y.A., Kureichik V.V.* Development of Distributed Information Systems: Ontological Approach // *Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015)*. Vol. 3. – Springer International Publishing AG Switzerland, 2015. – P. 113-122.
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, 2015. – P. 123-130.
17. *Kravchenko Y.A., and Kureichik V.V.* Knowledge management based on multi-agent simulation in informational systems // *8th IEEE International Conference “Application of Information and Communication Technologies – AICT 2014”*. – 2014. – P. 264-267.
18. *Запорожец Д.Ю., Курейчик В.В.* Гибридный алгоритм решения задач транспортного типа // *Известия ЮФУ. Технические науки*. – 2013. – № 7 (144). – С. 80-85.

19. Fishwick P.A., Miller J.A. Ontologies for Modeling and Simulation: Issues and Approaches // In Proceedings of, Winter Simulation Conference. – 2004. – P. 259-264.
20. Тузовский А.Ф., Чуриков С.В., Ямпольский В.З. Системы управления знаниями (методы и технологии) / под общ. ред. В.З. Ямпольского. – Томск: Изд-во НТЛ, 2005. – 260 с.

REFERENCES

1. Dorsey P. Personal knowledge management [e-resource]. Available at: http://www.360doc.com/content/05/1228/22/2563_51065.shtml (accessed 16 February 2016).
2. Martin J. Personal Knowledge Management. The Basis of Corporate and Institutional Knowledge Management, *Managing Knowledge: Case Studies in Innovation*. Alberta: University of Alberta, faculty of Extension, 2000, Vol. 6.
3. Lande D.V. Poisk znaniy v Internet [Search knowledge in the Internet]. Moscow: Dialektika, 2005, 271 p.
4. Bova V.V., Kravchenko Y.A., Kursityis I.O. Models for Supporting of Problem-Oriented Knowledge Search and Processing, *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'16)*, Vol. 1, pp. 287-295
5. Kerschberg L., Jeong H., Kim W. Emergent Semantic in Knowledge Sifter: An Evolutionary Search Agent based on Semantic Web Services. In: Spaccapietra, S., Aberer, K., Cudre-Mauroux, P. (eds.), *Journal on Data Semantic VI. LNCS*, 2006, Vol. 4090, pp. 187-209.
6. Gonchar A.D. Sravnitel'nyy analiz baz dannykh i baz znaniy (ontologiy) primenimo k modelirovaniyu slozhnykh protsessov [Comparative analysis of databases and knowledge bases (ontologies) applicable to the modeling of complex processes] *Sovremennye nauchnye issledovaniya i innovatsii* [Modern scientific researches and innovations], 2014, No. 5. Available at: <http://web.snauka.ru/issues/2014/05/34325> (accessed 27 March 2016).
7. Kryukov K.V., Pankova L.A., Pronina V.A. Mery semanticheskoy blizosti v ontologii [Measures of semantic closeness in the ontology], *Problemy upravleniya* [Problems of Management], 2010, No. 5, pp. 2-14.
8. Maedche A., Zacharias V. Clustering Ontology-Based Metadata in the Semantic Web, *Proceedings PKDD-2002, LNAI 2431*, 2002, pp. 348-360.
9. Le Khoay, Tuzovskiy A.F. Razrabotka semanticheskikh elektronnykh bibliotek na osnove ontologicheskikh modeley [The development of semantic digital libraries on the basis of ontological models], *Trudy XV Vserossiyskoy nauchnoy konferentsii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kolleksii» - RCDL 2013 (Yaroslavl', Rossiya, 14.10–17.10.2013)* [Proceedings of the XV all-Russian scientific conference "Electron-related libraries: advanced methods and technologies, digital collections" - RCDL 2013 (Yaroslavl, Russia 14.10–17.10.2013)]. Yaroslavl': Yaroslavskogo gosudarstvenno-go universiteta im. P.G. Demidova, 2013.
10. 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.
11. Gladkova L.A., Kureychik V.M., Kureychik V.V. Geneticheskie algoritmy [Genetic algorithms]. Moscow: Fizmatlit, 2006, 320 p.
12. 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.
13. 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.
14. 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.
15. Bova V.V., Kravchenko Y.A., Kureichik V.V. Development of Distributed Information Systems: Ontological Approach, *Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015)*. Vol. 3. Springer International Publishing AG Switzerland, 2015, pp. 113-122.
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, 2015, pp. 123-130.

17. *Kravchenko Y.A., and Kureichik V.V.* Knowledge management based on multi-agent simulation in informational systems, *8th IEEE International Conference "Application of Information and Communication Technologies – AICT 2014"*, 2014, pp. 264-267.
18. *Zaporozhets D.Yu., Kureychik V.V.* Gibridnyy algoritm resheniya zadach transportnogo tipa [Hybrid algorithm solving transport type problems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 80-85.
19. *Fishwick P.A., Miller J.A.* Ontologies for Modeling and Simulation: Issues and Approaches, In *Proceedings of, Winter Simulation Conference*, 2004, pp. 259-264.
20. *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.

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.

Кулиев Эльмар Валерьевич – e-mail: ekuliev@sfedu.ru; кафедра систем автоматизированного проектирования; ассистент.

Курситыс Илона Олеговна – e-mail: i.kursitys@mail.ru; тел.: +79518421327; кафедра систем автоматизированного проектирования; аспирант.

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

Kuliev Elmar Valerievich – e-mail: ekuliev@sfedu.ru; the department of computer aided design; assistant.

Kursitys Iona Olegovna – e-mail: i.kursitys@mail.ru; phone: +79518421327; the department of computer aided design; post-graduate student.