

МЕХАНИЗМЫ ПОДОБИЯ КОНЦЕПТУАЛЬНЫХ ЗАВИСИМОСТЕЙ ДЛЯ СЕМАНТИЧЕСКОЙ ФИЛЬТРАЦИИ ЭКЗЕМПЛЯРНЫХ ЗАПРОСОВ

Исмаилова Лариса Юсифовна

К.т.н., в.н.с., e-mail: lyu.ismailova@gmail.com

Косиков Сергей Владимирович

С.н.с., e-mail: kosikov.s.v@gmail.com

Беляцкая Полина Васильевна

Стажер, e-mail: polina@awg.ru

Никулин Илья Алексеевич

Стажер, e-mail: ilia.nikulin@gmail.com

Институт актуального образования «ЮрИнфоР-МГУ»,
119435, г. Москва, ул. Малая Пироговская 5

Аннотация. В работе рассматривается задача повышения качества информационного поиска в сети Интернет. Рассматривается представление информации в виде информационного графа и механизм поиска на основе экземплярного запроса. Предлагается использование семантических критериев подобия графов, которое даёт возможность организовать семантическую фильтрацию решений. Представлены варианты определения критериев подобия, даны определения и представлены выражения для некоторых классов критериев. Показано, что подгонка фреймов может использоваться для установления критериев семантического подобия.

Ключевые слова: информационный поиск, информационный граф, экземплярный запрос, критерий подобия.

Введение. Развитие информационных технологий в целом делает актуальной задачу поиска информации, относящейся к решению конкретной проблемы. Особенно актуальной эта задача становится в связи с развитием Интернет-технологий. Уже в силу объёма поиск релевантной информации представляет собой весьма сложную задачу. Дополнительно осложняет положение большая разнородность информации.

Информация представлена в разных форматах, которые могут быть (и фактически являются) несовместимыми между собой. Несовместимость может проявляться на принципиальном уровне (например, текстовая информация и аудиоинформация, кодирующая музыкальные произведения, принципиально различны между собой). Другим типом несовместимости является несовместимость на уровне представления. Так, до недавнего времени практически каждый текстовый процессор использовал собственный формат представления текстовых данных. Все эти факторы осложняют задачу поиска информации и делают актуальной задачу развития средств поиска данных.

Одним из подходов к решению задачи информационного поиска является изучение структур данных, обеспечивающих достаточно универсальное представление информации. Одной из таких структур, обеспечивающей универсальное и вместе с тем достаточно легко воспринимаемое человеком представление информации, являются графы. В общем случае

(ориентированный) граф задаётся множеством вершин E и множеством дуг V , являющимися подмножеством $E \times E$, где \times обозначает декартово произведение соответствующих множеств. Информационные графы представляют собой базовые графы, которые моделируют различные типы информации посредством вершин и дуг.

Графы могут быть использованы для представления информации различного типа. В частности, на графах основан подход к представлению знаний при помощи семантических сетей [1]. Семантические сети ориентированы на модель предметной области, в которой выделяются сущности и их связи. С использованием графа вершины представляют сущности, а дуги – связи между ними.

При усложнении модели предметной области усложняется и её графовое представление. Например, возможно использование дуг, помеченных различным образом. Возможно также вложение графов в целом в качестве вершин в другие графы, что придаёт всей структуре гиперграфовый характер. В случае рассмотрения моделей общего характера связь двух элементов структуры может рассматриваться как концептуальная зависимость [2].

Структурный характер графа открывает возможность подходов к определению механизма поиска ответа на запрос в графовой структуре. Поиск ответа на запрос предполагает определение в графе, представляющем массив данных, известных информационной системе, подграфов, соответствующих запросу. Соответствие при этом может пониматься различным образом.

Один из возможных способов понимания состоит в том, что запрос задаёт пример структуры графа, который должен быть получен в качестве ответа. Этот подход получил название экземплярных запросов [3]. Экземплярные запросы особенно полезны в случаях, когда пользователь знает один единственный элемент среди тех, которые, как ожидается, будут в требуемом наборе результатов, а на систему возлагается вывод из него остальных элементов ожидаемого набора. Другими словами, пользовательский запрос даёт только пример представляющих интерес элементов и взаимосвязей, которые, как ожидается, будут возвращены с помощью поисковой системы/механизма.

Извлечение ответа на экземплярный запрос является многошаговой процедурой, привлекающей решения из области семантического моделирования и информационных графов. Запрос, полученный информационной системой, обрабатывается – подготавливается к поиску подграфов, подобных заданному. Затем производится собственно поиск, выявляющий подобные структуры. При необходимости производится улучшение графа запроса за счёт добавления вершин и дуг, соединённых с ранее полученными результатами. Затем результаты ранжируются для получения наиболее релевантных из них.

Важнейшим этапом оказывается выявление подобных структур, требующее их корректного определения. От способа определения подобных структур зависит степень соответствия найденных подграфов запросу пользователя. Наличие вариативных критериев подобия даёт возможность осуществлять “семантическую фильтрацию” получаемой в результате извлечения ответа на запрос информации: от нахождения образцов-экземпляров запроса, через семантическое подобие, к ранжированию экземплярных результатов. Фильтрация предполагает установление и поддержание окрестностей подобия графов, которые могут быть использованы для семантической параметризации качества поиска.

1. Подходы к задаче повышения качества информационного поиска. Задача информационного поиска неоднократно изучалась, в том числе с точки зрения возможности

определения структур, обеспечивающих получение информации, соответствующей запросу пользователя. Разрабатываются техники, обеспечивающие получение ответов на менее структурированные и менее специализированные запросы [4]. Среди таких техник интерес представляют ослабление запросов (query relaxation) [5], семантические расширения [6], ответы на запросы со статистическим управлением (statistic-driven) [7], анализ протоколов [8, 9] и другие.

С формальной точки зрения запрос на информационный поиск может рассматриваться как выражение некоторого логического языка (обычно – варианта языка исчисления предикатов первого порядка). Системы и техники, обеспечивающие расширение состава данных, вовлекаемых в поиск ответа, фактически выполняют преобразование формулы, описывающей исходный запрос пользователя, в другую формулу, которая, как предполагается, лучше соответствует информационной потребности пользователя.

Большинство существующих техник нацелены на то, чтобы вовлечь в обработку не только информацию, явным образом указанную в запросе, но и представляющую интерес для конкретного пользователя информацию, связанную с явно указанной. В различных техниках эта задача решается различным образом. Такие техники особенно полезны в тех случаях, когда пользователь фактически не может корректно описать свои информационные потребности на языке запроса.

Техника ослабления запроса основана на замене условий запроса, представляющихся чрезмерно жёсткими (сильно сужающими пространство поиска) на менее жёсткие условия, обеспечивающие большее пространство поиска. Ослабление может проводиться различными путями, в том числе удалением из запроса конъюнктивных членов или расширением области изменения переменных в запросе. Существенно при этом, что вариант процедуры ослабления выбирается, исходя из логической структуры запроса, и не принимает во внимание семантику данных, к которым выполняется запрос.

Семантические расширения предполагают анализ терминов-констант, участвующих в запросе, и определение для них некоторых других констант, семантически связанных с исходными. После этого в запросе может быть осуществлена замена констант с более узкими семантическими границами на константы с более широкими – например, замена менее общего термина на более общий. Успех применяемой техники в значительной степени зависит от мощности семантической компоненты системы, т.е. от степени поддержки формальной техники определения семантических объёмов подвыражений запроса. Не исключена также возможность семантических ошибок.

Ответы на запросы со статистическим управлением пытаются опираться на статистические связи запросов типа «с поиском термина А часто производится поиск термина В». При этом термины могут быть связаны отношениями синонимии, близости или смежности и т. п. Возможен также учёт области, к которой отнесён термин, для уточнения его отношений с другими терминами; при этом, например, уточнение междисциплинарных терминов может выполняться с учётом дисциплины. В целом функционирование такой системы тем более успешно, чем более стандартным оказывается запрос пользователя. Кроме того, статистические взаимосвязи не позволяют выявлять основания, по которым они устанавливаются. Поэтому оказывается невозможным учитывать семантику систематическим образом.

Анализ протоколов основан на возможности получения информации о предыдущем взаимодействии пользователя с той же или другими поисковыми системами. Набор терминов и других параметров прошлых запросов может использоваться для уточнения семантических характеристик текущего запроса, создавая необходимый семантический контекст для интерпретации. Однако эта техника также зависима от степени стереотипности запроса.

Большинство предлагаемых решений опирается на предположение о том, что метки дуг графа, на котором устанавливается подобие, несут информацию о семантике предметной области. Это предположение, справедливое, например, в случае представления графа средствами RDF, в общем случае не выполняется. Оно существенно ограничивает класс информационных графов, к которому могут быть применены методы установления семантических окрестностей.

В отличие от рассматриваемых техник, экземплярные запросы предполагают выявление структуры семантического графа, лежащего в основании запроса, что обеспечивает возможность систематического учёта семантики. Критерии подобия семантических графов при этом оказываются независимыми от конкретной (зачастую достаточно случайной) структуры запроса пользователя.

Необходимо подчеркнуть отличие принятой в работе техники экземплярных запросов от QBE (query by example). В QBE запрос также представлен как пример, однако он используется при взаимодействии с механизмом поиска для обеспечения способа ввода информации, более удобного для пользователя (user-friendly). При использовании принятой техники запрос описывает образец элемента из требуемого множества, однако элементы множества могут иметь семантические характеристики, отличающиеся от указанных в запросе, поскольку семантическое подобие может быть установлено (в том числе) на основе характеристик, не упомянутых в запросе явно.

2. Задача определения критериев подобия. Анализ подходов к задаче установления семантического подобия графовых структур показывает одновременно большое разнообразие используемых критериев подобия, и вместе с тем отсутствие общей основы для систематического описания их семантики. Указанное обстоятельство обуславливает актуальность исследования задачи определения критериев подобия в более общей постановке. В частности, представляет интерес возможность учёта следующих факторов при обеспечении поиска в семантической сети и составлении критериев подобия:

- возможность связывания семантической информации как с вершинами, так и с дугами семантической сети;
- использование в ходе определения подобия графов метаинформации, заданной в виде специализированных меток дуг;
- возможность учёта концептуальных иерархий предметной области;
- возможность работы с составными графами, полученными в результате применения концептуальных операций;
- возможность рассмотрения графов, содержащих в качестве вершин подграфы, также имеющие вид информационных графов.

Критерии подобия неразрывно связаны с механизмами их установления. Возможно определение критериев подобия, учитывающих только структуру базового графа, лежащего

в основании рассматриваемого информационного графа. Возможен также учёт меток, приписанных вершинам и/или дугам графа.

Метки графа могут учитываться различным образом. Способ их учёта связан с семантическим уровнем, приписываемым меткам в конкретной модели информационных графов. Метка может отражать непосредственную семантику конкретной связи сущностей в предметной области, а может соответствовать целому классу связей с материально или структурно сходной семантикой. В последнем случае можно говорить о семантическом типе метки. Соответственно установление подобия может проводиться с точностью до семантического типа метки.

Особый интерес при установлении подобия вызывает возможность перехода от семантических сетей к сетям фреймов. Сети фреймов также могут быть охарактеризованы как информационные графы, но их структура значительно богаче. В частности, на фреймах может быть установлено отношение, определяющее включение классов сущностей – ISA-иерархия. Кроме того, дуги фреймов могут быть связаны с кванторными конструкциями, имеющими специфическую семантику. Адекватным семантическим механизмом, определяющим подобие фреймов, является подгонка (*fitting*), что вызывает интерес к изучению возможности использования подгонки при означивании экземплярных запросов.

3. Критерии подобия. Информационный поиск на основе экземплярных запросов в общем виде выполняется в два этапа. На первом этапе из запроса Q извлекается семантическая структура, которой должны соответствовать элементы, включаемые в ответ. Мы будем считать эту структуру информационным графом G . На втором шаге в информационном графе, представляющем содержание системы, выделяются подграфы, подобные G . Таким образом, ответ системы на запрос может быть представлен в виде:

$$\{G' \mid G = Sem(Q) \ \& \ Similar(G, G')\}, \quad (1)$$

где Sem – функция выделения семантической структуры, $Similar$ – функция семантического подобия графов.

Задача определения семантической структуры запроса (представленной в виде графа) может решаться методами концептуального моделирования [10, 11, 12]. Она находится вне рамок настоящей работы. Мы сосредоточимся на задаче определения критериев подобия, которую будем решать в рамках аппликативной парадигмы. При её использовании критерий подобия задаётся функцией, устанавливающей подобие информационных графов. Таким образом:

$$Similar(G, G') = \text{существует } f: G = f(G'). \quad (2)$$

Процедура установления подобия в этом случае сводится к построению необходимой функции заданного класса. Класс функций, задающих подобие, далее называется критерием подобия. Например, критерий подобия, основанный на изоморфизме графов, приводит к необходимости построения конкретных функций, задающих изоморфизм заданных графов. Методы задания семантической структуры исследовались отдельно [11, 12].

3.1. Критерии подобия, основанные на базовом графе. Базовый граф задается в виде $G = (E, V)$, где E – множество вершин графа, V – множество дуг графа, являющееся подмножеством $E \times E$. Как правило, в качестве ответа рассматривается подграф базового графа. Подграф определяется, как граф $G' = (E', V')$, где E' является подмножеством E , а V' является подмножеством V . Условие того, что G' является графом, гарантирует, что V'

является подмножеством $E' \times E'$. Если при этом дуга $v = (e_1, e_2)$ входит в V' тогда и только тогда, когда v входит в V , то подграф называется полным.

Переходим собственно к определению критериев подобия.

А. Изоморфизм. Изоморфизмом графов G и G' называется функция $f: E \rightarrow E'$, такая, что:

f взаимно однозначна;

$v = (e_1, e_2)$ входит в V тогда и только тогда, когда $v' = (f(e_1), f(e_2))$ входит в V' .

В качестве критерия подобия изоморфизм обычно устанавливается между графом запроса и подграфом информационного графа, в котором осуществляется информационный поиск.

В. Стягивание графа. Определим на вершинах графа следующее отношение $R(n)$, где n – натуральное число.

(e, e') входит в $R(n)$ тогда и только тогда, когда в графе существует путь $(e = e_1, e_2, \dots, e_k = e')$ длиной не более n .

Теперь в качестве критерия подобия можно использовать функцию $f: E \rightarrow E'$, такую, что:

f взаимно однозначна;

$v = (e_1, e_2)$ входит в V тогда и только тогда, когда $(f(e_1), f(e_2))$ входит в $R(n)$.

Легко показать, что отношение $R(n)$ следует из отношения $R(m)$ при $m > n$. Поэтому отношения $R(n)$ образуют частичный (на самом деле линейный) порядок.

Отношение $R(n)$ зависит от числа n , которое может быть выбрано, исходя из решаемой задачи информационного поиска. Таким образом, рассматриваемый критерий подобия обеспечивает параметризацию задачи информационного поиска.

Рассмотренные выше критерии работают с базовым графом, лежащим в основе информационного графа, поэтому их круг применения ограничен. Как правило, их применение оправдано в случаях, когда семантика дуг является достаточно однородной.

3.2. Критерии подобия, учитывающие метки дуг. Для обеспечения учёта меток дуг дополним определение графа. Будем рассматривать множество M меток дуг и определим граф как $G = (E, V, F)$, где E и V интерпретируются аналогично первоначальному определению, а F – функция вида $V \rightarrow M$ (функция пометки дуг).

А. Изоморфизм. Определение изоморфизма графов может быть дополнено следующим пунктом:

$$F(e_1, e_2) = F'(f(e_1), f(e_2)), \quad (3)$$

где F и F' – функции пометки дуг для графов G и G' соответственно. На основе изоморфизма графов получается критерий подобия, аналогичный предыдущему случаю.

Приведённое определение требует сохранения всех меток. На практике, однако, метки зачастую соответствуют разноплановым семантическим механизмам, в которых специфика предметной области отражена в различной степени. Поэтому более интересным является сохранение части меток.

Для формализации соответствующего критерия будем рассматривать условие ρ , заданное на множестве M . Определение изоморфизма может далее быть модифицировано следующим образом:

$$F(e_1, e_2) = F'(f(e_1), f(e_2)), \text{ если } \rho(F(e_1, e_2)). \quad (4)$$

Отметим, что в этом случае условие p также задаёт механизм параметризации.

В. Отображение меток. Более общим типом критерия подобия является критерий, позволяющий задавать отображение меток. Для этого необходимо задать функцию $g: M \rightarrow M$. Теперь требование, связывающее определения меток в сопоставляемых графах, может быть записано следующим образом:

$$g(F(e_1, e_2)) = F'(f(e_1), f(e_2)), \text{ если } p(F(e_1, e_2)). \quad (5)$$

Важным частным случаем оказывается возможность установления подобия графов с точностью до заданного отношения эквивалентности. Одним из способов задания отношения эквивалентности на множестве является отображение элементов множества в классы эквивалентности. В этом случае в качестве функции g должна быть выбрана функция, дающая по элементу множества M канонического представителя соответствующего класса.

С. Условие общего вида. Наиболее общий вид критерия, учитывающего метки дуг информационного графа, может быть получен следующим образом. Рассмотрим отношение Φ , являющееся подмножеством $M \times M'$. Выдвинем следующее требование к функции f , задающей определение критерия:

$$\Phi(F(e_1, e_2), F'(f(e_1), f(e_2))). \quad (6)$$

Указанное требование задаёт наиболее общий тип параметризации. Однако в силу его общего характера средства задания отношения Φ оказываются достаточно сложными, поэтому на практике удобно использовать сформулированные выше более частные критерии.

3.3. Критерии подобия, учитывающие метки вершин. Для обеспечения учёта меток вершин дополним определение графа. Будем рассматривать множество C меток дуг и определим граф как $G = (E, V, F, S)$, где E , V и F интерпретируются аналогично ранее данному определению, а S – функция вида $E \rightarrow C$ (функция пометки вершин). Рассмотрим соответствующие критерии подобия.

А. Изоморфизм. Определение изоморфизма графов может быть дополнено следующим пунктом:

$$S(e) = S'(f(e)), \quad (7)$$

где S и S' – функции пометки вершин для графов G и G' соответственно.

В. Условия общего вида. К меткам вершин информационного графа может быть применена та же техника уточнения критериев подобия, которая в предыдущем пункте была применена к меткам дуг. Соответствующие определения параллельны определениям предыдущего пункта и могут быть опущены.

3.4. Критерии подобия для сетей фреймов. Рассмотрим вариант определения критериев подобия для сетей фреймов. Будем использовать упрощённую модель сети фреймов. Модель определим следующим образом:

$$MF = (C, A, V, P, R, ISA), \quad (8)$$

где C – множество концептов;

A – множество констант;

V – множество переменных;

P – множество предикатов;

R – множество ролей;

ISA – отношение частичного порядка на $C \times C$.

Будем считать, что переменным и константам приписаны типы – концепты из множества C , что будем обозначать через $a : c, x : c$, где a, x, c – элементы множеств A, V, C соответственно.

Будем рассматривать простые фреймы следующего вида:

$$F = [p(r_1 : o_1 : c_1; r_2 : o_2 : c_2; \dots, r_n : o_n : c_n)], \quad (9)$$

где p – элемент множества P , r_i – элемент множества R , o_i – элемент множества A или множества V , c_i – элемент множества C . Фреймы рассматриваемого вида могут быть многими способами представлены в графическом виде, и, следовательно, могут рассматриваться как образующие информационный граф.

Определим критерий подобия, согласованный с базовой процедурой подгонки фреймов. Рассмотрим отношение Fit , определяемое следующим образом.

- (1) $o : c \text{ Fit } o : c$, если o – элемент множества A или V ;
- (2) $x : c \text{ Fit } a : c$, если x – элемент множества V и o – элемент множества A ;
- (3) $o : c \text{ Fit } o : c'$, если o – элемент множества A или V , c – элемент множества C и $c' \text{ ISA } c$;
- (4) $o : c \text{ Fit } o' : c'$, если $o : c \text{ Fit } o'' : c''$ и $o'' : c'' \text{ Fit } o' : c'$.

Указанное отношение задаёт базовые способы подгонки констант и переменных.

Определим теперь отношение Fit на фреймах:

$$[p(r_1 : o_1 : c_1; r_2 : o_2 : c_2; \dots, r_n : o_n : c_n)] \text{ Fit } [p(r_1 : o'_1 : c'_1; r_2 : o'_2 : c'_2; \dots, r_n : o'_n : c'_2)], \quad (10)$$

если для всех i $o_i : c_i \text{ Fit } o'_i : c'_i$.

Указанное отношение задаёт критерий подгонки фреймов, согласованный с базовыми способами подгонки. Рассмотрим возможности расширения критерия на фреймы, снабжённые кванторной приставкой. Будем рассматривать кванторы следующих видов:

$$[n] \quad < n] \quad [n > \quad \text{all}.$$

Будем рассматривать фреймы вида QF , где Q – кванторная приставка

$$Q = q_1 x_1 q_2 x_2 \dots q_n x_n, \quad (11)$$

где q_i – квантор, x_i – переменная. Определим функции hB и lB из множества Q в множество $N+$, где $N+ = N \cup \{inf\}$ и N – множество натуральных чисел. Будем считать, что для всех n из N выполнено $n < inf$.

$$\begin{aligned} hB [n] &= n & lB [n] &= n \\ hB < n] &= 0 & lB < n] &= n \\ hB [n >= n & & lB [n] &= inf \\ hB \text{ all} &= 0 & lB \text{ all} &= inf. \end{aligned}$$

Теперь мы можем определить критерий подобия фреймов с учётом кванторной приставки. Пусть

$$\begin{aligned} Fq &= Q F & Q &= q_1 x_1 q_2 x_2 \dots q_n x_n, \text{ и} \\ Fq' &= Q' F' & Q &= q'_1 x_1 q'_2 x_2 \dots q'_n x_n. \end{aligned}$$

Будем считать, что

$$Fq \text{ Fit } Fq', \text{ если } F \text{ Fit } F' \text{ и для всех } i \text{ имеем } hB q_i < hB q'_i \text{ и } lB q_i > lB q'_i.$$

Указанные способы определения критериев подобия на фреймах не являются исчерпывающими. Возможности определения критериев, более точно согласующихся со структурой рассматриваемых фреймов, а также расширения рассматриваемого подхода на другие типы фреймов являются перспективным направлением исследований.

Заключение. В работе предлагается решение задачи повышения качества информационного поиска в сети Интернет, рассматриваемой как частный случай информационного графа. Предлагаемый подход основан на использовании графа экземплярного запроса, формирование ответа на который предполагает использование механизмов подобия графов. Использование семантических критериев подобия даёт возможность организовать семантическую фильтрацию решений, основанную на рассмотрении семантических окрестностей запроса.

Представлены варианты подхода к определению критериев подобия, учитывающие как структуру базового графа, так и связанные с ним метки, причём может быть учтена как формальная идентичность меток, так и сопоставленные им семантические свойства. Даны формальные определения и представлены выражения для некоторых классов критериев.

Особый интерес вызывает возможность применения методов семантического подобия к сетям фреймов, рассматриваемых как информационные графы с дополнительной структурой. Показано, что подгонка фреймов может использоваться для установления критериев семантического подобия, в том числе в случае использования кванторных операций.

Работа поддержана грантами РФФИ 16-07-00912, 17-07-00893.

СПИСОК ЛИТЕРАТУРЫ

1. Kosikov S.V., Wolfengagen V.E., and Ismailova L.Yu. The Presentation of Evolutionary Concepts. Springer International Publishing. Cham. 2018. Pp. 113 – 125.
2. Ismailova L.Yu., Wolfengagen V.E., Kosikov S.V. Basic Constructions of the Computational Model of Support for Access Operations to the Semantic Network. *Procedia Computer Science*. 2018. Volume 123, Pp. 183 – 188. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.030>. (<http://www.sciencedirect.com/science/article/pii/S1877050918300310>).
3. D. Mottin, M. Lissandrini, Y. Velegrakis, and T. Palpanas. Exemplar queries: Give me an example of what you need. *PVLDB*. 2014. №7(5).
4. S. Bergamaschi, F. Guerra, S. Rota, and Y. Velegrakis. A hidden markov model approach to keyword-based search over relational databases. *ER*. 2011.
5. C. Mishra and N. Koudas. Interactive query refinement. *EDBT*. 2009.
6. S. Bergamaschi, E. Domnori, F. Guerra, R. Trillo Lado, and Y. Velegrakis. Keyword search over relational databases: a metadata approach. *SIGMOD*. 2011.
7. S. Gauch and J. B. Smith. Search improvement via automatic query reformulation. *TOIS*. 1991. №9(3). Pp. 249 – 280.
8. I. Bordino, G. De Francisci Morales, I. Weber, and F. Bonchi. From machu picchu to rafting the urubamba river: anticipating information needs via the entity-query graph. In *WSDM*. 2013.
9. D. Mottin, T. Palpanas, and Y. Velegrakis. Entity Ranking Using Click-Log Information // *IDA Journal*. 2013. №17(5).
10. Wolfengagen V.E., Ismailova L.Yu., Kosikov S.V., Nikulin I.A., Parfenova I.A., Kcholodov V.A. Means for Ensuring Compatibility of Heterogeneous Data Models in an Interactive Visualization Environment // *Procedia Computer Science*. 2018. Volume 123. Pp. 195 – 202. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.088>.

(<http://www.sciencedirect.com/science/article/pii/S1877050918300899>)

11. Wolfengagen V.E., Ismailova L.Yu., Kosikov S.V. The Typing System to Provide Compositional Thinking About Data Flows // *Procedia Computer Science*. 2018. Volume 123. Pp. 246 – 251. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.038>.

(<http://www.sciencedirect.com/science/article/pii/S1877050918300395>)

12. Wolfengagen V.E., Ismailova L.Yu., Kosikov S.V. Model of Conversion of Data Objects for Defining the Object-Relation Mapping // *Procedia Computer Science*. 2018. Volume 123. Pp. 541 – 546. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.082>.

(<http://www.sciencedirect.com/science/article/pii/S1877050918300838>)

UDK 004

**MEANS OF THE SIMILARITY OF CONCEPTUAL DEPENDENCIES
FOR SEMANTIC FILTERING OF EXEMPLAR QUERIES**

Larisa Yu. Ismailova

PhD, Leading Researcher, e-mail: lyu.ismailova@gmail.com

Sergej V. Kosikov

Senior Researcher, e-mail: kosikov.s.v@gmail.com

Polina V. Belyatskaya

Trainee, e-mail: polina@avg.ru

Ilya A. Nikulin

Trainee, e-mail: ilia.nikulin@gmail.com

Institute for contemporary education “JurInfoR-MGU” 5, Malaja Pirogovskaja, 119435,
Moscow, Russia

Abstract. The problem of improvement of quality of information search on the Internet is considered. Submission of information in the form of the information graph and a search engine on the basis of exemplar queries is considered. Use of semantic criteria of similarity of graphs which gives the possibility to organize a semantic filtration of decisions is offered. Options of determination of criteria of similarity are presented, definitions are given and expressions for some classes of criteria are presented. It is shown that fitting of frames can be used for establishment of criteria of semantic similarity.

Keywords: information search, information graph, exemplar queries, criterion of similarity.

References

1. Kosikov S.V., Wolfengagen V.E., and Ismailova L.Yu. The Presentation of Evolutionary Concepts. Springer International Publishing. Cham. 2018. Pp. 113 – 125.
2. Ismailova L.Yu., Wolfengagen V.E., Kosikov S.V. Basic Constructions of the Computational Model of Support for Access Operations to the Semantic Network // *Procedia Computer Science*. 2018. Volume 123, Pp. 183 – 188. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.030>.

- (<http://www.sciencedirect.com/science/article/pii/S1877050918300310>).
3. D. Mottin, M. Lissandrini, Y. Velegrakis, and T. Palpanas. Exemplar queries: Give me an example of what you need. PVLDB. 2014. №7(5).
 4. S. Bergamaschi, F. Guerra, S. Rota, and Y. Velegrakis. A hidden markov model approach to keyword-based search over relational databases. ER. 2011.
 5. C. Mishra and N. Koudas. Interactive query refinement. EDBT. 2009.
 6. S. Bergamaschi, E. Domnori, F. Guerra, R. Trillo Lado, and Y. Velegrakis. Keyword search over relational databases: a metadata approach. SIGMOD. 2011.
 7. S. Gauch and J. B. Smith. Search improvement via automatic query reformulation. TOIS. 1991. №9(3). Pp. 249 – 280.
 8. I. Bordino, G. De Francisci Morales, I. Weber, and F. Bonchi. From machu picchu to rafting the urubamba river: anticipating information needs via the entity-query graph. In WSDM. 2013.
 9. D. Mottin, T. Palpanas, and Y. Velegrakis. Entity Ranking Using Click-Log Information // IDA Journal. 2013. №17(5).
 10. Wolfengagen V.E., Ismailova L.Yu., Kosikov S.V., Nikulin I.A., Parfenova I.A., Kholodov V.A. Means for Ensuring Compatibility of Heterogeneous Data Models in an Interactive Visualization Environment // Procedia Computer Science. 2018. Volume 123. Pp. 195 – 202. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.088>. (<http://www.sciencedirect.com/science/article/pii/S1877050918300899>)
 11. Wolfengagen V.E., Ismailova L.Yu., Kosikov S.V. The Typing System to Provide Compositional Thinking About Data Flows // Procedia Computer Science. 2018. Volume 123. Pp. 246 – 251. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.038>. (<http://www.sciencedirect.com/science/article/pii/S1877050918300395>)
 12. Wolfengagen V.E., Ismailova L.Yu., Kosikov S.V. Model of Conversion of Data Objects for Defining the Object-Relation Mapping // Procedia Computer Science. 2018. Volume 123. Pp. 541 – 546. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.082>. (<http://www.sciencedirect.com/science/article/pii/S1877050918300838>)