

УДК 519.711.3

ПОВЫШЕНИЕ РЕЛЕВАНТНОСТИ ПРИ ПОИСКЕ ВЕБ-СЕРВИСОВ В МУЛЬТИОНТОЛОГИЧЕСКОЙ СРЕДЕ

Л.С. Глоба¹, М.В. Ковальский²

Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина

¹lgloba@its.kpi.ua, ²mkovalskyi@luxoft.com

Аннотация

Кратко рассмотрены существующие подходы к решению задачи поиска веб-сервисов. Выявлены проблемы низкой релевантности, интероперабельности, существенного влияния человеческого фактора в процессе поиска веб-сервисов, которые решаются добавлением семантической информации к их описанию. Проанализированы существующие модели аннотированных онтологий веб-сервисов, а также предложена их модификация, позволяющая получать более релевантный ответ на запрос пользователя. Предложена концепция промежуточного абстрактного объекта - «шаблона сервиса», в который преобразовываются пользовательские запросы с одной стороны, а с другой — онтологические модели независимо от языка их описания. С целью повышения релевантности ответа предлагается также использовать входные и выходные параметры операций веб-сервисов, сравнивая их как синтаксически, так и семантически. Согласно полученным результатам, релевантность ответа на поисковый запрос увеличилась, а вероятность возникновения «ложного обнаружения» уменьшилась.

Ключевые слова: веб-сервисы, онтология, релевантность.

Введение

Веб-сервисы предоставляют единообразные программные интерфейсы, позволяющие их использовать конечными пользователями при необходимости получения информационных услуг в среде Интернет [1].

В текущий момент стоит задача нахождения такой модели построения распределенных систем на базе веб-сервисов, которая бы позволяла автоматизированно обнаруживать веб-сервисы согласно заданным критериям с наибольшей точностью (релевантностью) [2].

Релевантность поисковых запросов¹ является серьезной проблемой в набирающем всё большие объёмы данных Интернете. Например, реестр UDDI (Universal Description, Discovery and Integration) позволяет производить поиск веб-сервиса исключительно с помощью ключевых слов и предопределенных категорий [3]. Таким образом, окончательный выбор веб-сервиса производится пользователем.

В ряде работ предложено множество решений в этом направлении, в частности — семантические веб-сервисы, в основе которых лежит дополнение существующих WSDL-описаний (Web Services Description Language) посредством аннотирования онтологическими моделями, разрабатываемыми независимо от описаний веб-сервисов [4]. Однако они характеризуются низкой релевантностью в процессе поиска веб-сервисов, что осложняет их использование конечными пользователями при необходимости получения услуг в среде Интернет.

¹ **Релевантность** (лат. *relevo* — поднимать, облегчать) - семантическое соответствие поискового запроса и поискового образа документа. В более общем смысле, близко к понятию «адекватность», при этом оценивается не только степень соответствия, но и степень практической применимости результата поиска. *Прим.ред.*

1 Анализ существующих подходов

Де-факто стандартом обнаружения веб-сервисов является реестр UDDI, который предоставляет категоризированный поиск, а также поиск по ключевым словам. Множество исследователей считает, что такой подход не является масштабируемым, поскольку требует полной репликации всех опубликованных в реестре веб-сервисов [4].

Для решения этой проблемы были предложены варианты архитектуры «peer-to-peer» [5], которые включают в себя решения распределенного хранения каталога веб-сервисов с использованием полной или частичной репликации на множество серверов.

В [6] рассмотрена возможность применения таксономий процессов, описывающих не сущности, а действия, применимые к этим сущностям. Такой метод сопоставления использует семантические отношения между процессами для сравнения пользовательских запросов и моделей, описывающих веб-сервисы.

Работа [7] использует принцип подстановки Барбары Лисков² относительно концептов онтологической модели. Любой концепт может быть заменен своим дочерним концептом, если последний предоставляет больший или равный интерфейс, который имеет избыточность в некотором контексте. Таким образом, более детализированные онтологические модели можно адаптировать для сравнения с менее детализированными. Недостатком этого метода является то, что предусматривается аннотирование веб-сервисов онтологиями исключительно средствами DAML (DARPA Agent Markup Language), который не поддерживает описания входных и выходных данных. Анализ полученных описаний возможен только по описаниям и названиям веб-сервисов, что является естественным ограничением DAML. Поэтому значительно возрастает возможность ложного обнаружения.

Предложенный в [8] метод состоит в преобразовании веб-сервисов в вид RDF-графов (Resource Description Framework). Далее полученные графы сравниваются, при этом проводится декомпозиция узлов, представляющих сложные типы, к более простым типам.

Также был представлен способ сопоставления онтологических моделей средствами специально разработанного языка LARKS (Language for Advertisement and Request for Knowledge Sharing) [9]. Ядро системы, описанной в указанной работе, состоит из пяти так называемых «фильтров» - критериев поиска. Такая концепция «фильтров» основывается не на сравнении отдельных компонентов, а на аппроксимации онтологической модели с целью получения множества ключевых слов, которые определяют содержание модели. Далее это множество ключевых слов сопоставляется с пользовательским запросом. LARKS разработан с учётом особенностей мультионтологической среды, но при этом считается, что любая онтологическая модель, подлежащая сравнению, использует некоторый единый глоссарий или генерализированную онтологическую модель.

В [10] была дана классификация различий, которые возможны между онтологическими моделями в связи с использованием различных языков моделирования и намерений, с которыми некоторая модель создавалась. Это же исследование показывает, что применение статистических и лингвистических алгоритмов для сравнения свойств концептов представляется обоснованным.

Существуют также решения из области машинного обучения, например, LSD (Learning Source Descriptions) [10], которые основаны на вероятностном распределении экземпляров объектов внутри разнообразных онтологических моделей.

² Принцип подстановки Барбары Лисков (*Liskov Substitution Principle, LSP*) в объектно-ориентированном программировании является специфичным определением *подмнога*, предложенным Барбарой Лисков в 1987 году. Этот принцип служит критерием оценки качества принимаемых решений при построении иерархий наследования и фактически означает, что поведение наследуемых классов не должно противоречить поведению, заданному базовым классом. *Примеч. ред.*

На практике, несмотря на значительное количество исследований в данном направлении, всё же остаётся существенное влияние человеческого фактора в процессе поиска веб-сервисов, а также низкая релевантность результатов поиска, что вынуждает конечного пользователя выполнять ручной подбор необходимого ему сервиса.

Метод поиска аннотированных онтологическими моделями веб-сервисов, предложенный в [4], основан на оценке синтаксического подобия названий и описаний концептов, представляющих сами веб-сервисы. Концепты, описывающие искомые веб-сервисы, и ключевые слова из пользовательского запроса сравниваются с помощью синтаксического подобия, использующего различные алгоритмы сравнения строк (нормализация текста, стемминг³, алгоритмы с использованием N-грамм). Этот метод показывает хорошие результаты для моноонтологической среды, в которой используется единая онтология. Основываясь на результатах проведённого исследования утверждается, что релевантность ответа на пользовательский запрос значительно уменьшается в мультионтологической среде [4].

При этом для мультионтологической среды, использующей различные онтологические модели, разработанные разрозненно, метод, описанный в работе [4], не подходит. Анализ рассмотренных подходов позволяет сделать вывод о недостаточной эффективности вышеописанных методов. Однако, метод поиска аннотированных онтологическими моделями веб-сервисов, изложенный в [4], показывает наибольшую эффективность в случае поиска веб-сервисов в моноонтологической среде, поэтому его доработка с целью использования в мультионтологической среде является наиболее целесообразной.

Для использования метода поиска аннотированных онтологическими моделями веб-сервисов в мультионтологической среде предложены следующие модификации.

Во-первых, не использовать конкретный язык описания, в данном случае — DAML. Несмотря на то, что используемый алгоритм не зависит от конкретной реализации онтологической модели, он зависит от языка описания. С целью независимости от языка описания онтологических моделей предлагается подход с использованием программного промежуточного абстрактного объекта, который является отображением онтологической модели и называется «шаблоном сервиса». В такие объекты преобразовываются пользовательские запросы с одной стороны, а с другой — онтологические модели независимо от языка их описания.

Во-вторых, в работе [4] оценка подобия веб-сервисов производилась исключительно с помощью названий и описаний концептов, описывающих веб-сервис. С целью повышения релевантности ответа предлагается использовать также входные и выходные параметры операций, сравнивая их как синтаксически, так и семантически, например, учитывая кардинальность множества принимаемых параметром значений.

2 Метод оценки подобия веб-сервиса искомому

Подход к поиску веб-сервиса в мультионтологической среде сводится к трём последовательным шагам:

- 1) создание шаблона сервиса на основе пользовательского запроса;
- 2) сравнение шаблона сервиса с множеством веб-сервисов, которые были идентифицированы как вероятные кандидаты;
- 3) возврат веб-сервисов, удовлетворяющих минимально допустимой оценке подобия, пользователю в виде упорядоченного списка.

Семантический шаблон сервиса описывает пользовательский запрос. Он позволяет задавать набор требуемых пользователю операций, их свойств, входных и выходных данных. У

³ Stemming (англ.) - морфологический поиск, т.е. поиск слова во всех его морфологических формах. *Прим.ред.*

шаблона сервиса нет конкретной реализации, поскольку его следует рассматривать как промежуточную абстракцию – веб-сервис-посредник (proxy web service).

Более формально шаблон сервиса (ST, Service Template) можно определить следующим образом:

$$ST = \langle N_{ST}, D_{ST}, OP_{ST} \langle N_{OP}, D_{OP}, O_{OP}, I_{OP} \rangle \rangle,$$

где N_{ST} – название веб-сервиса, D_{ST} – текстовое описание веб-сервиса, OP_{ST} – множество операций веб-сервиса.

Каждая из операций веб-сервиса, в свою очередь, определяется с помощью N_{OP} – названия операции, D_{OP} – текстового описания операции, O_{OP} и I_{OP} – выходные и входные данные операции.

Шаблон сервиса сравнивается с множеством сервис-кандидатов (CS, Candidate Service) – шаблонов сервиса, полученных при анализе веб-сервисов из предопределённого множества, подлежащего перебору. Сравнение состоит в оценке синтаксического и семантического подобия:

$$\Theta(ST, CS) = [w_H \cdot H(ST, CS) + w_\Phi \cdot \Phi(ST, CS)] / [w_H + w_\Phi],$$

где $\Theta(ST, CS)$ – суммарная оценка, $H(ST, CS)$ – синтаксическое подобие, $\Phi(ST, CS)$ – семантическое (функциональное) подобие, w_i представляет весовой коэффициент соответствующей оценки подобия, предназначенный для более гибкого управления критериями сравнения.

Синтаксическое подобие вычисляется аналогично базовому методу [4] с использованием стемминга, т.е. процесса нахождения основы слова, представляющего описание или название, и дальнейшее вычисление расстояния Хемминга – числа позиций, в которых символы двух слов различны.

Алгоритм, используемый для нахождения основы слова, заключается в использовании таблицы правил для преобразования входного слова в нормализованную форму. Таблица правил определяет префиксы и суффиксы, которые необходимо удалять из слов или заменять на другие префиксы или суффиксы. Такой подход оправдан, поскольку показывает хорошее быстродействие и, в то же время, достаточно прост. Например, алгоритм, использующий таблицу, определяющую правило полного соответствия слова и его основы, является менее производительным, хотя и обеспечивает большую эффективность для языков, имеющих сложные правила образования префиксов или суффиксов [11].

Семантическое подобие вычисляется с использованием множества критериев, которые учитывают оценку подобия двух веб-сервисов с использованием атрибутов и отношений концептов – подобие по входным и выходным данным операций веб-сервисов, концептуальное подобие и его составляющая – подобие по свойствам концептов.

Концептуальное подобие веб-сервисов $C(ST, CS)$ определяется как:

$$C(ST, CS) = [w_H \cdot H(ST, CS) + w_P \cdot P(ST, CS)] / [w_H + w_P],$$

где $C(ST, CS)$ – концептуальное подобие, $H(ST, CS)$ – синтаксическое подобие, $P(ST, CS)$ – подобие по свойствам концептов, w_i – весовые коэффициенты соответствующих оценок подобия.

Концептуальное подобие используется также при оценке подобия по входным и выходным данным $IO(OP_{ST}, OP_{CS})$:

$$IO(OP_{ST}, OP_{CS}) = (I(OP_{ST}, OP_{CS}) \cdot O(OP_{ST}, OP_{CS}))^{1/2},$$

где $IO(OP_{ST}, OP_{CS})$ – подобие по входным и выходным данным операций, $I(OP_{ST}, OP_{CS})$ – подобие по входным данным операций, $O(OP_{ST}, OP_{CS})$ – подобие по выходным данным операций, OP_{ST} – множество операций шаблона сервиса, OP_{CS} – множество операций сервиса-кандидата.

Оценка подобия по входным данным определяется как:

$$I(OP_{ST}, OP_{CS}) = \max[I(OP_{ST} - OPi_{ST}, OP_{CS} - OPi_{CS}) + C(OPi_{ST}, OPi_{CS})],$$

где $I(OP_{ST}, OP_{CS})$ – подобие по входным данным, OP_{ST} и OP_{CS} – множество операций шаблона сервиса и сервиса-кандидата соответственно, OPi_{ST} и OPi_{CS} – элементы множества операций шаблона сервиса и сервиса-кандидата, $C(OPi_{ST}, OPi_{CS})$ – концептуальное подобие элементов множества.

Оценка подобия по выходным данным операций производится аналогичным образом.

На рисунке 1 схематически представлена иерархия используемых критериев.

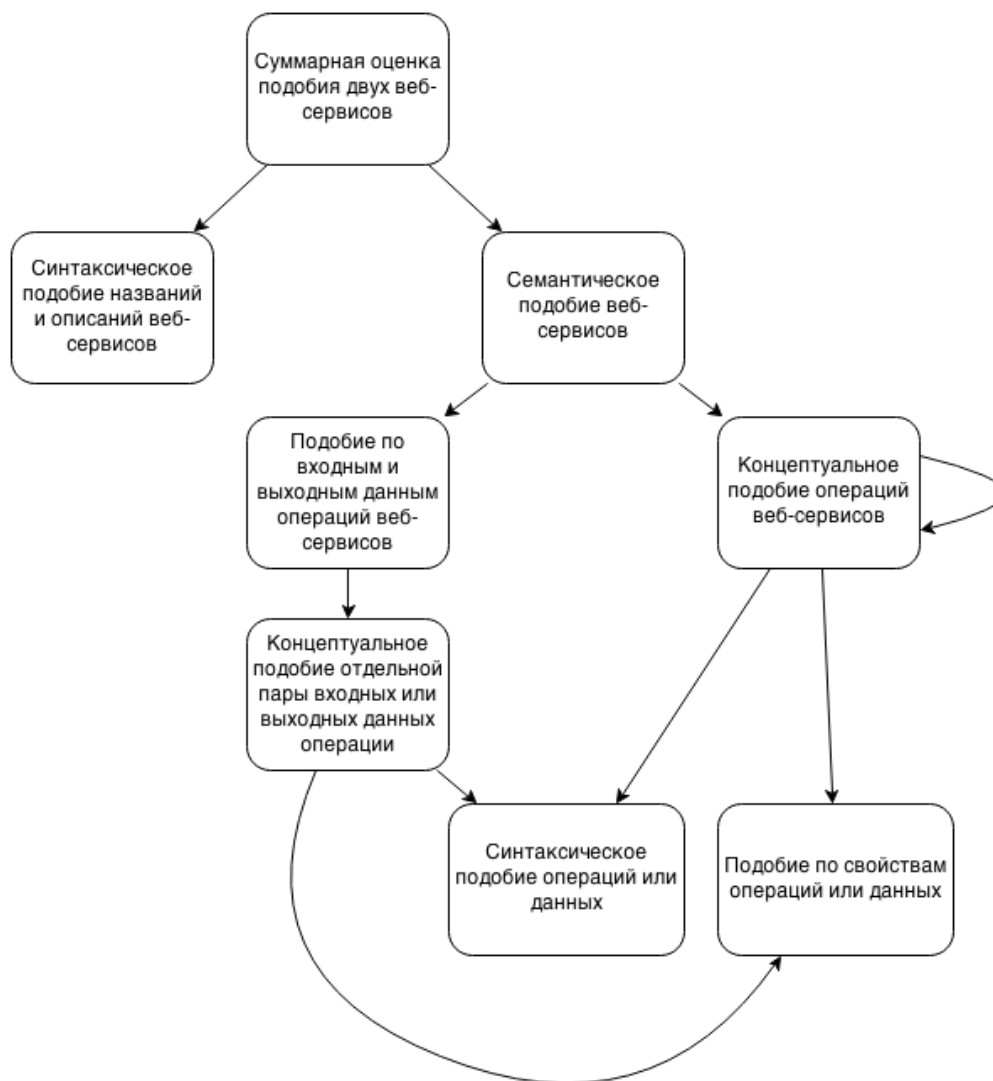


Рисунок 1 – Схематическое представление иерархии используемых критериев

Следует отметить, что концептуальное подобие операций веб-сервисов определено как задача динамического программирования, т.е. рекурсивная функция, поскольку включает в себя перебор всех возможных комбинаций операций с целью нахождения такого их отображения, которое даёт максимальную оценку подобия по этому критерию.

Наибольшее влияние на оценку семантического подобия имеет её составляющая – подобие по свойствам концептов, которыми описаны операции веб-сервиса.

Вычисление оценки подобия по свойствам концептов заключается в поочередном сравнении пар свойств двух выбранных для сравнения концептов. Во время сравнения этих свойств формируются их отображения, показывающие эквивалентность свойств между концептами. При этом следует обеспечить максимально возможную среднюю оценку подобия отображаемых свойств:

$$P(C_{ST}, C_{CS}) = \max (P(CP_{ST} - CPi_{ST}, CP_{CS} - CPi_{CS}) + p(CPi_{ST}, CPi_{CS})),$$

где $P(C_{ST}, C_{CS})$ – оценка подобия по свойствам, CP_{ST} и CP_{CS} – множество свойств концепта, описывающего операцию шаблона сервиса и сервиса-кандидата соответственно, CPi_{ST} и CPi_{CS} – отдельные свойства концепта, описывающего операцию шаблона сервиса и сервиса-кандидата соответственно. Подобие отдельных свойств $p(CPi_{ST}, CPi_{CS})$ описывается следующим образом:

$$p = k \cdot (H \cdot K \cdot \chi)^{1/3} - 0.05 \cdot \|\text{свойства, не имеющие отображения}\|,$$

где H – синтаксическое подобие, K – подобие по кардинальности, χ – подобие по ограничениям, применяемым к типу.

Множитель k является константой, значение которой составляет 1 в случае, если оба сравниваемых свойства являются инверсными функциональными свойствами, и 0,8 – в противном случае. Инверсными функциональными называют такие свойства, которые имеют значения, уникальные для каждого экземпляра объекта. Иными словами, это свойства концепта, которые определяют его идентичность в пределах предметной области.

Подобие по ограничениям вычисляется с учётом типа данных, представляющего свойство. Свойство концепта может быть представлено с помощью примитивного типа или другого концепта. Исходя из этого, возможны три случая:

- 1) Оба сравниваемых свойства представлены примитивным типом данных. Оценка подобия по ограничениям сводится к вычислению количества информации, теряемой при преобразовании из одного типа в другой. На практике используются эвристические оценки в пределах от 0 (полная потеря информации) до 1 (полное сохранение информации).
- 2) Оба сравниваемых свойства являются онтологическими концептами. Задача оценки подобия по свойствам двух концептов сводится к рекурсивному вычислению оценки и является задачей динамического программирования.
- 3) Одно из свойств представлено примитивным типом, а другое – концептом. В этом случае оценка равна нулю, поскольку типы несовместимы. Такая оценка может быть нежелательной, поскольку данный случай может возникнуть в связи с различным уровнем детализации двух сравниваемых концептов, тогда нужно использовать некоторые оценки по умолчанию или же сравнивать такие свойства с помощью синтаксического подобия.

Подобие по кардинальности также играет значительную роль при сравнении двух свойств. Оценка подобия по кардинальности предусматривает подсчёт кардинальности множества принимаемых свойством значений. Кардинальность множества принимаемых свойством значений может быть как конечной малой величиной (в случае с типом перечисления), так и условно конечной величиной (например, вещественные числа, количество возможных значений которых ограничено только конкретной реализацией окружения).

Оценка подобия по кардинальности также является эмпирической и равняется:

- 1) единице, если кардинальности множества принимаемых значений двух свойств равны;
- 2) единице, если оба свойства являются инверсными функциональными;
- 3) 0,9, если кардинальность свойства, представляющего шаблон сервиса, меньше;
- 4) 0,7, если кардинальность свойства, представляющего шаблон сервиса, больше.

В случае, если количество свойств концепта сервиса-кандидата не меньше количества свойств концепта шаблона сервиса, представляется возможным отображение «один к одно-

му» каждого свойства шаблона. В противном случае, когда отображение имеет вид «один ко многим», следует вводить штраф, уменьшающий оценку подобия по свойствам пропорционально количеству свойств, имеющих несколько отображений. С помощью моделирования значение штрафа было установлено на уровне 0,05.

В конечном итоге, в результате сравнения каждый сервис-кандидат получает нормированную в пределах [0, 1] оценку подобия. Считаем, что оценка со значением 1 соответствует наилучшему совпадению в выборке. Затем веб-сервисы, удовлетворяющие минимально допустимой оценке, возвращаются пользователю в виде упорядоченного списка.

3 Полученные результаты

Метод был проверен на наборе веб-сервисов для оценки повышения релевантности ответа на пользовательский запрос. Веб-сервисы были получены из индекса компании xIgnite и аннотированы двумя разными онтологиями. Первая была получена в результате анализа предметной области фондовых бирж при помощи материалов NASDAQ, находящихся в открытом доступе в сети Интернет. Вторая онтология была создана на основе понятий, используемых компанией xIgnite при описании собственных веб-сервисов.

Такой выбор используемых данных обусловлен тем, что они в полной мере удовлетворяют необходимым требованиям.

Во-первых, рассматриваются четыре сценария, которые могут возникнуть при сравнении двух онтологических концептов в моноонтологической (в пределах одной онтологической модели), и четыре сценария относительно мультионтологической среды (в пределах нескольких онтологических моделей). Таковыми являются сравнение (в моноонтологической среде под эквивалентным концептом подразумевается идентичный):

- 1) эквивалентных концептов;
- 2) концепта и его подконцепта (дочернего концепта);
- 3) концепта и его надконцепта (родительского концепта);
- 4) несвязанных концептов.

Набор используемых для анализа веб-сервисов включает в себя 14 различных веб-сервисов, 13 из которых содержат по две операции и один – одну операцию.

Моделирование производится в два этапа. Сначала шаблон сервиса для пользовательского запроса описывается заранее определенным (базовым) концептом из онтологической модели. Затем, для каждого из восьми вышеописанных сценариев, сервис-кандидат описывается одним из концептов, удовлетворяющих условиям сценария; а потом производится оценка.

Второй этап заключается в аннотировании операций, принадлежащих веб-сервисам из набора, и их свойств; входящих и исходящих данных при помощи ранее выбранного базового концепта. Далее производится оценка по базовому и расширенному методам.

На рисунке 2 изображены результаты первого этапа моделирования. Сценарии 1-4 соответствуют моноонтологической среде, а сценарии 5-8 – мультионтологической.

Из полученных результатов следует, что суммарная оценка при использовании синтаксического и семантического подобия ниже оценки только по синтаксическому подобию для сценариев 3, 4, 7, 8. Следовательно, учёт семантического подобия снижает суммарную оценку для несвязанных концептов и концептов, которые неспособны в полной мере покрыть свойства базового концепта. Это, в свою очередь, понижает вероятность так называемого «ложного обнаружения».

Суммарная оценка для сценария 5 тоже уменьшилась, что также связано с неполным покрытием свойств базового концепта, что является следствием различий между уровнем детализации онтологических моделей.

Анализируя сценарии 1, 2, 6, можно сделать вывод о том, что суммарная оценка повысилась для концептов, удовлетворяющих свойствам базового концепта (в данном случае – идентичные и дочерние по отношению к базовому концепту). Таким образом, при поиске веб-сервиса предпочтение будет отдано веб-сервисам, которые входят в эту категорию.

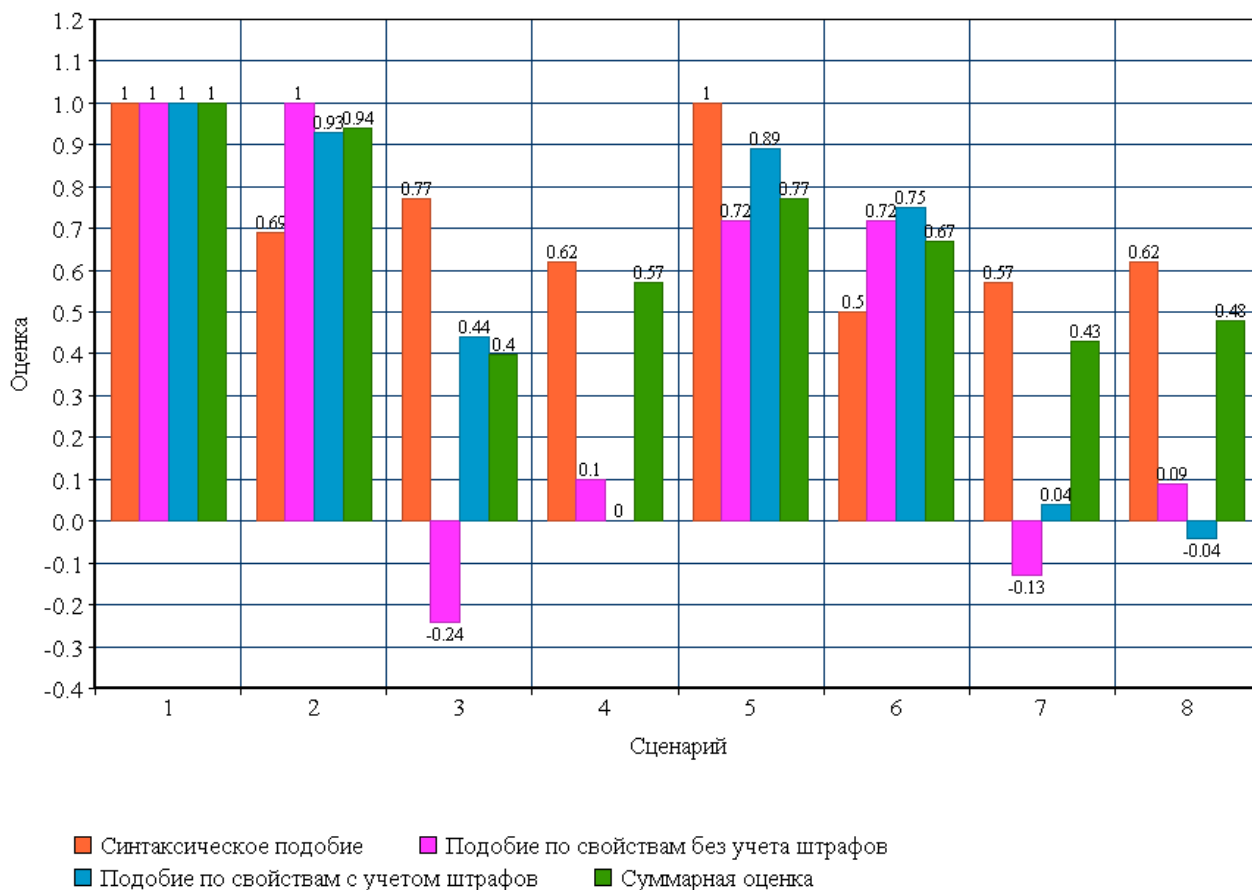


Рисунок 2 – Подобие двух онтологических концептов согласно видам оценок

На рисунке 3 показаны результаты второго этапа моделирования. Операции веб-сервисов 1-7 имеют большее количество свойств, чем может предоставить базовый концепт, при этом веб-сервис 7 предоставляет только одну операцию, в то время как шаблон-сервиса, описывающий пользовательский запрос, содержит две операции.

Согласно полученным результатам, оценка метода, учитывающего синтаксическое и семантическое подобие, оказалась ниже оценки только по синтаксическому подобию для веб-сервисов 1-7. Таким образом, веб-сервисы, которые не могут полностью покрыть требования шаблона сервиса, отсеиваются.

Задавая при поиске минимально допустимую оценку подобию значением 0,7, получаем, что при использовании предложенного метода получим пять веб-сервисов (9, 10, 12, 13 и 14), а при использовании метода, учитывающего только синтаксическое подобие, - семь (4, 6, 10-14).

Следовательно, оценка семантического подобию помогает уменьшить вероятность возникновения «ложного обнаружения» (в данном случае – веб-сервисов 4, 6, 11)⁴.

⁴ В приведенном примере авторы сочли возможным не отмечать повышение вероятности «ложного обнаружения», которая проявилась у веб-сервиса 9 (см. рисунок 3). *Прим.ред.*

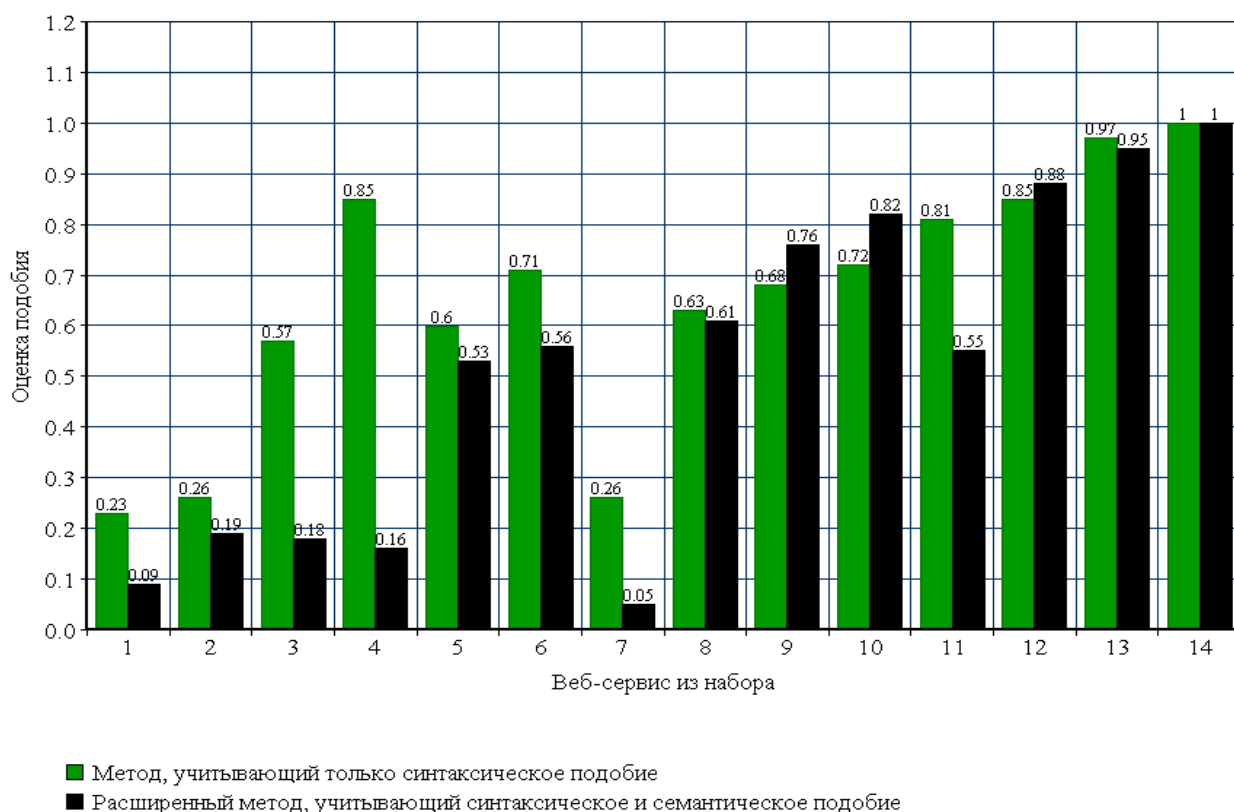


Рисунок 3 – Сравнение оценки подобия веб-сервисов из набора и пользовательского запроса

Заключение

Обнаружение веб-сервисов является исключительно семантической проблемой. Текущий стандарт веб-сервисов на основе UDDI и WSDL не использует семантическую информацию, поэтому релевантность обнаруженных пользователями веб-сервисов остается невысокой.

Предложено усовершенствование метода поиска веб-сервиса в мультионтологической среде, суть которого заключается в дополнительном использовании программного промежуточного абстрактного объекта – «шаблона сервиса», а также в использовании для сравнения сервисов дополнительно типов входных и выходных параметров операций веб-сервиса с целью увеличения релевантности ответа на запрос пользователя.

Дальнейшие исследования направлены на расширение предложенного метода за счёт поддержки описаний ошибок в терминах WSDL, которые могут возникнуть во время обработки запроса, а также требований к QoS (Quality of Service). Необходимо определить оптимальные оценки вместо эвристических для подобия по свойствам концепта с помощью экспертных систем и проверить предлагаемый метод на большем количестве реальных веб-сервисов и их онтологических описаний.

Также метод может быть расширен за счёт поддержки композиции веб-сервисов во время их обнаружения, что позволит формировать цепочку их вызовов. Это необходимо в ситуациях, при которых требуемый веб-сервис не был найден в реестре, но, комбинируя другие существующие, можно получить искомый. Предлагается проводить такую композицию с помощью онтологических описаний входных и выходных параметров операций каждого веб-сервиса в цепочке.

СПИСОК ИСТОЧНИКОВ

- [1] *Fensel, D.* Semantic web services / D. Fensel, F.M. Facca, E. Simperl, I. Toma. - Springer, 2011 - 357 p.
- [2] *Pedrinaci, C.* Semantic web services / C. Pedrinaci, J. Domingue, A.P. Sheth // In Handbook of semantic web technologies. - Springer Berlin Heidelberg, 2011 - pp. 977-1035.
- [3] *Toch, E.* Humans, semantic services and similarity: A user study of semantic Web services matching and composition / E. Toch, I. Reinhartz-Berger, D. Dori // Web Semantics: Science, Services and Agents on the World Wide Web, 9(1), 2011 - pp. 16-28.
- [4] *Sheth, A.* Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics / A. Sheth // In Interoperating Geographic Information Systems. - Kluwer Academic Publishers, 1998 - pp. 5-30.
- [5] *Schmidt, C.* A Peer-to-Peer Approach to Web Service Discovery / C. Schmidt, M. Parashar // World Wide Web, Internet and Web Information Systems. - Kluwer Academic Publishers, Vol. 7, No. 2, 2004 – pp. 211–229.
- [6] *Klein, M.* Searching for Services on the Semantic Web using Process Ontologies / M. Klein, A. Bernstein // First Semantic Web Working Symposium (SWWS-1), Stanford, CA USA, 2001.
- [7] *Gonzales-Castillo, J.* Description logics for matchmaking of services / J. Gonzales-Castillo, D. Trastour, C. Bartolini // Proc. of Workshop on Application of Description Logics, Vol. 44. ADL-2001.
- [8] *Trastour, D.* A Semantic Web Approach to Service Description for Matchmaking of Services / D. Trastour, C. Bartolini, J. Gonzalez-Castillo // Proc. 1st Semantic Web Working Symposium, CA, 2001 - 15 p.
- [9] *Sycara, K.* Dynamic Service Matchmaking among Agents in Open Information Environments / K. Sycara, J. Lu, M. Klusch, S. Widom // Journal ACM SIGMOD Record, Special Issue on Semantic Interoperability in Global Information Systems, 28(1). 1999 - pp. 47-53.
- [10] *Magnini, B.* Linguistic Based Matching of Local Ontologies / B. Magnini, L. Serafini, M. Speranza // In Working notes of MeaN-02 (Workshop held in conjunction with AAI-2002), Edmonton, Alberta, Canada, July 28 - August 1, 2002 - pp. 42-50.
- [11] *Porter, M.F.* An Algorithm for Suffix Stripping / M.F. Porter // The journal Program, Vol. 14, No. 3, July 1980 - pp 130-137; Program: electronic library and information systems, Vol. 40 Iss: 3, pp. 211–218.

INCREASING WEB SERVICES DISCOVERY RELEVANCY IN THE MULTI-ONTOLOGICAL ENVIRONMENT

L.S. Globa¹, M.V. Kovalskyi²

National Technical University of Ukraine «Kiev Polytechnic Institute», Kiev, Ukraine

¹lgloba@its.kpi.ua, ²mkovalskyi@luxoft.com

Abstract

The existing approaches to the solution of web services discovery issues are reviewed briefly in this paper. The problems of low relevancy, interoperability, high amount of the human factor existing in the process of web services discovery are found and can be solved by addition of semantic data to web service description. The existing models of ontology annotated web services are analyzed and their modification is proposed to allow achievement of higher discovery relevancy. The concept of the intermediate abstract object called «service template», in which end user requests and ontological models are converted, is introduced. It is also proposed to use inputs and outputs of web services operations to match them by syntactic and semantic criteria. As shown in the paper, web service discovery response relevancy has increased, and probability of false matching has decreased.

Key words: *web services, ontology, relevancy.*

References

- [1] *Fensel, D.* Semantic web services / D. Fensel, F.M. Facca, E. Simperl, I. Toma. - Springer, 2011 - 357 p.
- [2] *Pedrinaci, C.* Semantic web services / C. Pedrinaci, J. Domingue, A.P. Sheth // In Handbook of semantic web technologies. - Springer Berlin Heidelberg, 2011 - pp. 977-1035.

- [3] **Toch, E.** Humans, semantic services and similarity: A user study of semantic Web services matching and composition / E. Toch, I. Reinhartz-Berger, D. Dori // Web Semantics: Science, Services and Agents on the World Wide Web, 9(1), 2011 - pp. 16-28.
- [4] **Sheth, A.** Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics / A. Sheth // In Interoperating Geographic Information Systems. - Kluwer Academic Publishers, 1998 - pp. 5-30.
- [5] **Schmidt, C.** A Peer-to-Peer Approach to Web Service Discovery / C. Schmidt, M. Parashar // World Wide Web, Internet and Web Information Systems. - Kluwer Academic Publishers, Vol. 7, No. 2, 2004 – pp. 211–229.
- [6] **Klein, M.** Searching for Services on the Semantic Web using Process Ontologies / M. Klein, A. Bernstein // First Semantic Web Working Symposium (SWWS-1), Stanford, CA USA, 2001.
- [7] **Gonzales-Castillo, J.** Description logics for matchmaking of services / J. Gonzales-Castillo, D. Trastour, C. Bartolini // Proc. of Workshop on Application of Description Logics, Vol. 44. ADL-2001.
- [8] **Trastour, D.** A Semantic Web Approach to Service Description for Matchmaking of Services / D. Trastour, C. Bartolini, J. Gonzalez-Castillo // Proc. 1st Semantic Web Working Symposium, CA, 2001 - 15 p.
- [9] **Sycara, K.** Dynamic Service Matchmaking among Agents in Open Information Environments / K. Sycara, J. Lu, M. Klusch, S. Widom // Journal ACM SIGMOD Record, Special Issue on Semantic Interoperability in Global Information Systems, 28(1). 1999 - pp. 47-53.
- [10] **Magnini, B.** Linguistic Based Matching of Local Ontologies / B. Magnini, L. Serafini, M. Speranza // In Working notes of MeaN-02 (Workshop held in conjunction with AAAI-2002), Edmonton, Alberta, Canada, July 28 - August 1, 2002 - pp. 42-50.
- [11] **Porter, M.F.** An Algorithm for Suffix Stripping / M.F. Porter // The journal Program, Vol. 14, No. 3, July 1980 - pp 130-137; Program: electronic library and information systems, Vol. 40 Iss: 3, pp. 211–218.

Сведения об авторах



Глоба Лариса Сергеевна, 1954 г. рождения. Окончила Харьковский авиационный институт в 1978 г., к.т.н. (1984), д.т.н. (1996), проф. с 2001 г., с 2012 г. – академик Международной академии информационных технологий (Минск, Белоруссия). Заведует кафедрой информационно-телекоммуникационных систем Национального технического университета Украины «Киевский политехнический институт». Сфера научных интересов: разработка интеллектуальных корпоративных систем производства и систем управления, распределенные системы, технологии Web-сервисов. В списке научных трудов около 400 работ, среди которых учебники и учебные пособия.

Larysa Sergiivna Globa (b. 1954) graduated from Kharkov Aviation Institute in 1978, Ph.D., from 1984, Doctor of Technical Sciences from 1996, Prof. from 2001, academician of the International Academy of Information Technology (Minsk, Belorussia) from 2012. The head of the chair of Information and Telecommunication Systems of National Technical University of Ukraine "Kiev Polytechnic Institute". Research interests: development of intelligent enterprise production and management systems, distributed systems, Web-services technology. She is co-author of about 400 scientific works, a significant number of works issued abroad.



Ковальский Михаил Витальевич, 1994 г. рождения. Окончил Национальный технический университет Украины «Киевский политехнический институт» в 2014 г. Сфера научных интересов – распределенные в сети Интернет системы и облачные вычисления.

Mikhail Vitalievich Kovalskyi (b.1994) graduated from the National Technical University of Ukraine "Kiev Polytechnic Institute" in 2014. Works in the area of the distributed Internet-systems and cloud computing.