

ANALYSIS OF SERVICE CALLS FOR CONSTRUCTION OF THE SEMANTIC NETWORK OF SERVICES

Roman K. Fedorov⁽¹⁾, Alexander S. Shumilov⁽¹⁾, Mikhail L. Voskoboynikov⁽¹⁾

⁽¹⁾Matrosov Institute for System Dynamics and Control Theory SB RAS, Irkutsk, Russia

The paper proposes a method for obtaining the missing information in the construction of a semantic network, based on the analysis of statistical data about the service use by users. The advantage of the method is that the semantic connection between the two services is confirmed by a workable example demonstrating the transfer of data from one service to another. In addition, together with the semantic network, we obtain the connectivity of the parameters by the data, the possible values of the parameters, the frequency characteristics of users' use of services and their compositions, the time of the services execution, etc.

Key words: SOA, WPS, geospatial data processing, Semantic network, service composition.

АНАЛИЗ ПРИМЕНЕНИЯ СЕРВИСОВ ДЛЯ ПОСТРОЕНИЯ СЕМАНТИЧЕСКОЙ СЕТИ СЕРВИСОВ

Федоров Р.К.⁽¹⁾, Шумилов А.С.⁽¹⁾, Воскобойников М. Л.⁽¹⁾

⁽¹⁾ Институт динамики систем и теории управления имени В.М. Матросова СО РАН,
г. Иркутск

В работе предлагается метод для получения недостающей информации при построении семантической сети, основанный на анализе статистических данных о применении сервисов пользователями. Преимуществом данного метода является то, что семантическая связь между двумя сервисами подтверждается работоспособным примером, демонстрирующим передачу данных от одного сервиса другому. Кроме того вместе с семантической сетью получаем связанность параметров по данным, возможные значения параметров, частотные характеристики использования пользователями сервисов и их композиций, время выполнения сервисов и т.д.

Ключевые слова: SOA, WPS, обработка геопространственных данных, семантическая сеть, композиция сервисов.

Введение. В современном мире активно развивается сервис-ориентированный подход (SOA) [1] и растет количество Web сервисов, выполняющих удаленную обработку данных. Например, сервисы обработки данных дистанционного зондирования земли, видеоданных, пространственных данных и т.д. Создаются сервисы хранения и предоставления различных данных. В частности, облачные хранилища данных, такие как интернет-сервисы: Dropbox, OneDrive, Google Drive, iCloud, Яндекс.Диск, Облако Mail.Ru и т.д. Необходимо отметить, что часто сервисы выполняют аналогичные функции. Например, для поиска кратчайшего маршрута между двумя точками существует как минимум три различных сервиса, предоставляемых OSM, Google, Yandex. Создаются каталоги сервисов, содержащих

метаинформацию и позволяющих выполнять поиск необходимых сервисов. Разработаны и применяются стандарт WPS [9], набор спецификаций REST [3] и т.д., унифицирующие вызов сервисов, передачу и получение данных.

Для решения сложной задачи может потребоваться композиция Web сервисов [7]. Формирование композиции сервисов комбинаторно сложный процесс, в котором нужно определить решаемую каждым сервисом задачу и согласовать по входным и выходным данным их сочетания. Если в настольных средах разработки приложений программист имеет узкий набор методов, представляемых средой разработки или собственными библиотеками программ, то в случае с Web сервисами программист вынужден просмотреть значительное количество их сочетаний. Одним из способов, упрощающих разработку композиции сервисов, является формирование семантической сети сервисов [7]. Семантическая сеть сервисов определяет возможные связи сервисов в соответствии с передаваемыми данными, т.е. дуги этой сети задают возможность элементарной композиции – цепочки выполнения из двух сервисов, когда результаты работы одного сервиса могут быть применены в качестве входных данных другого сервиса. Одним из способов формирования семантической сети является использование метаинформации [3,8] и онтологий [1]. В частности осуществляется построение семантической сети на основе WSDL. Следует отметить, что автоматическое построение семантической сети является сложной задачей по следующим причинам:

- сервисы имеют недостаточно метаинформации, либо она вообще отсутствует или рассчитана на человека;

- параметры некоторых сервисов могут быть связаны со многими другими сервисами. Например, сервисы алгебры растров могут использоваться во многих композициях сервисов для анализа пространственных данных. Построение семантической связи из трех последовательных сервисов, где промежуточным сервисом являются сервисы алгебры изображений, требует проработанной онтологии предметной области. Ее построение достаточно сложный процесс, в котором очень сложно рассмотреть все возможные сущности и связи;

- требуется согласованность метаинформации сервисов и онтологии предметной области.

Основная идея метода. Для получения недостающей информации при построении семантической сети предлагается метод, основанный на анализе статистических данных о применении сервисов пользователями. В рамках геопортала ИДСТУ СО РАН [6] формируется распределенная вычислительная среда, которая применяется преимущественно для задач обработки пространственных данных. Разработаны вычислительные сервисы, реализующие стандарт WPS, и сервисы хранения данных, каталог WPS сервисов и системы выполнения сервисов и их композиций. Реализовано взаимодействие между вычислительными сервисами и сервисами хранения данных, что позволяет упростить использование сервисов пользователем. В системе выполнения сервисов и их композиций ведется сбор данных о произведенных вызовах сервисов. Данные включают название сервиса и его адрес, значения входных и выходных параметров, время выполнения сервисов, успешность выполнения, ошибки выполнения и т.д. Для каждого сервиса известно его расположение, список выходных и входных параметров, типы данных. Введем ниже следующие обозначения, необходимые для описания работы метода.

$s = \langle name, I, O \rangle$ – сервис, где *name* – это имя сервиса, *I* – множество входных параметров, *O* – множество выходных параметров сервиса. Далее будем обозначать $s.I$ и $s.O$ параметры, принадлежащие определенному сервису. Пользовательский вызов сервиса имеет одинаковую структуру: исходные параметры, известные пользователю, и параметры, которые должны быть получены.

$t = \langle V_I, V_O \rangle$ – вызов сервиса *s* со значениями V_I, V_O параметров $s.I$ и $s.O$.

Log – это множество успешных вызовов сервисов, совершенных пользователями. Причиной неудачного вызова сервиса могут быть некорректные значения параметров. Поэтому статистические данные фильтруются и оставляются только успешные вызовы.

Построение семантической сети сводится к задаче определения наличия семантической связи между параметрами сервисов (Semantic causal link) $s_i.O$ и $s_j.I$.

Множество параметров можно разделить по значениям на передаваемые непосредственно в теле вызова и на передаваемые в виде ссылки файлы. В теле вызова сервиса и его результата могут передаваться и получатся строковые и числовые параметры, по значениям которых довольно сложно определить наличие семантической связи между параметрами сервисов. Параметры, имеющие одинаковое значение, могут иметь разный смысл. По параметрам файлам можно точно определить, что файл был передан от одного сервиса другому. Каждый файл однозначно идентифицируется URL, который используется для его скачивания WPS сервисом.

В методе для анализа вызовов используется следующий алгоритм построения семантической сети сервисов.

```
Input: Log
Output: list – семантические связи между сервисов
for each  $t_i$  in Log:
    for each  $t_j$  in Log:
        for each  $v_m$  in  $t_i.VO$ :
            for each  $v_n$  in  $t_j.VI$ :
                If (typeof  $v_m$  = file and typeof  $v_n$  = file and  $v_m$ 
                =  $v_n$ ) then
                    add ( $t_i, t_j$ ) into list
                end if
            end for
        end for
    end for
end for
```

Данный алгоритм производит поиск сочетания вызовов сервисов, в которых файл, являющийся результатом работы одного сервиса, передается в качестве параметра другому сервису. Предложенный метод построения семантической сети апробирован на небольшом фрагменте данных о применении сервисов пользователями. Источники данных представлены в виде сервисов, у которых присутствуют только исходящие дуги. На рисунке 1 представлена часть полученной сети. Несвязанные сервисы (одиночные вершины) не отображены. Необходимо отметить, что при таком подходе семантика всех параметров сервисов не определяется. Среди положительных моментов укажем, что вместе с семантической сетью получаем связанность параметров по данным, возможные значения

параметров, частотные характеристики использования пользователями сервисов и их композиций, время выполнения сервисов и т.д.

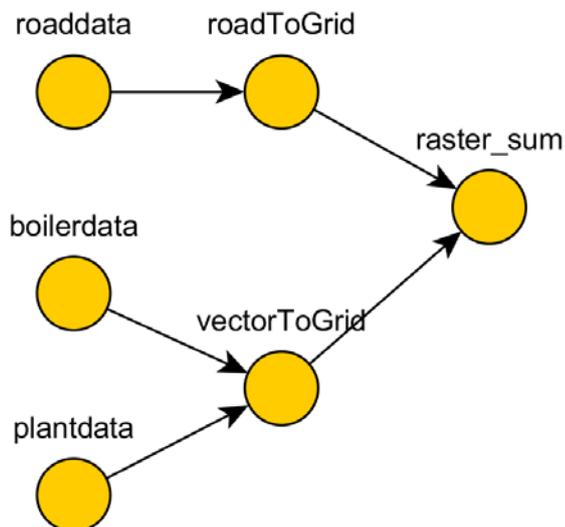


Рис. 1. Полученная семантическая сеть

Заключение. Представленный в статье метод позволяет сформировать более полную семантическую сеть на основе статистических данных применения сервисов. Применение семантической сети значительно упрощает разработку композиции сервисов за счет автоматического поиска и построения цепочек сервисов по данным. Преимуществом данного метода является то, что семантическая связь между двумя сервисами подтверждается работоспособным примером, демонстрирующим передачу данных от одного сервиса другому. Кроме того вместе с семантической сетью получаем связанность параметров по данным, возможные значения параметров, частотные характеристики использования пользователями сервисов и их композиций, время выполнения сервисов и т.д. Анализ полученных данных является перспективным в следующих направлениях:

- автоматическое формирование композиции сервисов для типичных операций пользователя;
- определение альтернативных композиций сервисов;
- распознавание семантики параметров сервисов;
- подсказка разработчику возможных значений параметров;
- и т.д.

Апробация метода в области обработки пространственных данных обусловлена тем, что часто используются базовые пространственные данные, семантика которых заранее известна, формируются реляционные данные с пространственными атрибутами, предоставляемые сервисам. Следовательно, в этой области значительная часть семантической сети в той или иной мере определена, что подчеркивает перспективность исследования.

Работа выполнена при финансовой поддержке РФФИ (грант № 16-07-00411а, 16-07-00554-а, 17-47-380007-р-а, 18-07-00758-а), Президиума РАН проект №27, Интеграционных программ СО РАН проект №184, ИНЦ СО РАН.

Результаты получены при использовании сетевой инфраструктуры ЦКП «Интегрированная информационно-вычислительная сеть Иркутского научно-образовательного комплекса» (<http://net.icc.ru>).

ЛИТЕРАТУРА

- [1] Grimm, S. Ontologies and the Semantic Web / S. Grimm, A. Abecker, J. Volker, R. Studer // Handbook of semantic web technologies: foundations and technologies, VOLS 1 and 2. — 2011. — P. 507-579.
- [2] Hoffmann, J. Web Service Composition / J. Hoffmann, I. Weber // Encyclopedia of Social Network Analysis and Mining. — Springer-Verlag, 2014.
- [3] Kopecky, J. SAWSDL: Semantic annotations for WSDL and XML schema / J. Kopecky, T. Vitvar, C. Bournez, J. Farrell // IEEE INTERNET COMPUTING. — 2007. — Vol. 11, № 6. — P. 60–67.
- [4] Pautasso, C. RESTful Web service composition with BPEL for REST / C. Pautasso // Data knowledge. — 2009. — Vol. 68, № 9. — P. 851–866.
- [5] Welke, R. Service-oriented architecture maturity / R. Welke, R. Hirschheim, A. Schwarz // Computer. — 2011. — Vol. 44, № 2. — P. 61–67.
- [6] Фёдоров Р. К., Бычков И. В., Шумилов А. С., Ружников Г. М. Система планирования и выполнения композиций веб-сервисов в гетерогенной динамической среде // Вычислительные технологии. — 2016. — Т. 21, № 6. — С. 18 – 35.
- [7] Lecue, F. A formal model for semantic Web service composition / F. Lecue, A. Leger // The Semantic Web – ISWC. — 2006. — Vol. 4273 — P. 385-398
- [8] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne et al., “Owl-s: Semantic markup for web services,” W3C member submission, vol. 22, pp. 2007–04, 2004.
- [9] P. Schut, P. OpenGIS ® Web Processing Service / P. Schut // Open Geospatial Consortium. — 2007. — № 6. — P. 1–3.