

## Литература

1. Медведєва, О. М. Ціннісно-орієнтоване управління взаємодією в проектах: Методологічні основи [Текст] : автореф. дис. д-ра техн. наук : 25.02.2013 / О. М. Медведєва ; [Київський національний університет будівництва і архітектури] . – Київ, 2013 – 48 с.
2. Долан С. Управление на основе ценностей. Корпоративное руководство по выживанию, успешной жизнедеятельности и умению зарабатывать деньги в XXI веке [Текст] / С. Долан, С. Гарсия. – М. : Претекст, 2008. – 313 с.
3. Дасковский, В.С. Об оценке эффективности инвестиций [Текст] : В. С. Дасковский, В. И. Киселев // Экономист. – 2007. – №3. – С. 38–48.
4. A Guide to the Project Management Body of Knowledge [Текст] : An American National Standard / Ed. W.R. Duncan. – Third Edition. – PMI Standards Committee. ANSI / PMI 99 – 001–2004. – 388 p.
5. Performance Budgeting: OMB's Performance Rating Tool Presents Opportunities and Challenges for Evaluating Program Performance [Электронный ресурс] / D. C. Washington, The Government Accountability Office - Режим доступа : \www/ URL: http://www.gao.gov/cgi-bin/getrpt?GAO-04-550T/ – 11.03.2004 г. – Загл. с экрана.
6. A guidebook of Project & Program Management for Enterprise Innovation [Текст] Т. 1. : Project Management Association of Japan (PMAJ) / Representative Author : Prof. Shigenobu Ohara. – version 1.2. - PMCC/ENNA, Japan, 2008. 109 p.
7. Managing for Results: Measuring Program Results That Are Under Limited Federal Control [Электронный ресурс] / D. C. Washington, The Government Accountability Office - Режим доступа : \www/ URL: http://www.gao.gov/cgi-bin/getrpt?GAO/GGD-99-16/ – 11.12.1998 г. – Загл. с экрана.
8. Cleland, D. I. Strategic Management of Teams. [Текст] / D. I. Cleland. – New York. : John Wiley & Sons, Inc , 1996. – 292 p.
9. Forsberg, K. Visualizing Project Management [Текст] / K. Forsberg, H. Mooz, H. Cotterman. – New York. : John Wiley & Sons, Inc , 2000. – 364 p.
10. Kaplan, R. S. The Balanced Scorecard: Translating stratagem into Action [Текст] / R. S. Kaplan, D. P. Norton. – Boston.: Harvard Business School Press, 1996. – 364 p.

*У роботі запропонована методика розрахунку відстані між семантичними анотаціями профілів процесів. Вона заснована на розрахунку відстані між двома концептами онтології. Під час розрахунку враховуються як взаємне положення концептів в таксономії, так і схожість наборів їх властивостей. Запропонована методика знаходить своє застосування в задачах підбору компонентів за їх анотаціями та семантичним запитом користувача*

*Ключові слова: семантичний профіль, метрика семантичної відстані анотацій, OWL-S*

*В работе предложена методика расчета расстояния между семантическими аннотациями профилей процессов. Она основана на расчете расстояний между двумя концептами онтологии. При расчете учитываются как взаимное расположение концептов в общей иерархии, так и схожесть наборов их свойств. Описанная методика находит свое применение в задачах подбора компонентов по их аннотации и семантическому запросу пользователя*

*Ключевые слова: семантический профиль, метрика семантической близости аннотаций, OWL-S*

УДК 519.7:007.52

## МЕТРИКА БЛИЗОСТИ СЕМАНТИЧЕСКИХ АННОТАЦИЙ ДЛЯ ЗАДАЧИ СРАВНЕНИЯ ПРОФИЛЕЙ ПРОЦЕССОВ

Е. Л. Шевченко

Кандидат технических наук, доцент  
Кафедра программной инженерии  
Харьковский национальный университет  
радиоэлектроники  
пр. Ленина, 16, г. Харьков, Украина, 61166  
E-mail: olena.l.shevchenko@gmail.com

### 1. Введение

Одним из перспективных направлений в области формальных методов разработки ПО является задача

синтеза структурных схем программ, в частности на основе семантических вычислительных сетей [1]. В классическом варианте такие сети представляют отношения вычислимости между переменными и опера-

ционными блоками и позволяют по готовой базе компонент и набору входных данных получить множество потенциально вычисляемых выходных переменных и схемы их вычисления. При этом отношение вычислимости является единственным семантическим отношением в таких сетях. Поэтому операционные блоки в них могут различаться только набором входных и выходных параметров и нет возможности добавить в их описание дополнительную семантическую нагрузку, например, суть вычисляемого отношения, ограничения на применение операции, используемый в операции метод расчета и пр.

С другой стороны, как результат интеграции инициативы Semantic Web в технологии веб-сервисов, разработан язык OWL-S, который позволяет описывать семантику веб-сервиса и их возможности (т.н. профиль веб-сервиса) [2]. Такое описание профиля веб-сервиса является машинно-анализируемым и при разработке методов сравнения двух профилей становится возможным вычисление семантического расстояния между двумя сервисами, которые они описывают. Наглядный пример возможного подхода к динамической композиции веб-сервисов на основе их семантического описания приведен в [3].

Кроме уже упомянутой задачи синтеза структурных схем программ и динамической композиции веб-сервисов, данный аппарат находит свое применение в задачах поиска и рекомендации операционного блока (или веб-сервиса) по его описанию (например, в продвинутых help-системах).

## 2. Анализ предметной области и постановка задачи

На данном этапе разработки OWL-S профиль сервиса – это семантическое описание всех его входящих и выходящих параметров. Это достигается путем отображения каждого параметра на концепт, который определен в OWL-онтологии. Подробно особенности применения OWL 2.0 рассмотрены в [4]. Соответственно, последние работы, связанные со сравнением OWL-S профилей веб-сервисов [5], предлагают рассчитывать функцию семантической близости между двумя аннотациями как интеграцию расстояний, рассчитанных для концептов, прикрепленных к параметрам сравниваемых сервисов.

Такой подход обладает рядом существенных недостатков, например, он не учитывает сути выполняемого с параметрами действия. Так, при аннотации методов «добавить» и «удалить» элемент коллекции такие операции становятся неразличимыми.

В данной статье излагается расширение описанного выше подхода. В предыдущей работе автора была предложена усовершенствованная модель семантической аннотации на основе онтологии действий. Такая аннотация представляет собой граф. Узлы графа – это классы онтологии, описывающие выполняемое методом действие, его дополнительные характеристики (часть которых задается параметрами метода), а дуги – отношения между узлами в аннотации. В аннотации выделяют главный концепт, который соответствует основному действию, выполняемому методом.

Один из вариантов прототипа и аннотации метода приведен на рис. 1. Цель метода – по входящему пись-

му уметь определить, является ли оно ответом на одно из отправленных ранее с этого ящика писем, и если такое письмо существует, то вернуть его. Главное действие выражено узлом «Поиск».



Рис. 1. Пример простой аннотации метода

Для сравнения таких аннотаций необходимо учитывать не только узлы аннотации, но и дуги, задающие отношения между ними. В данной статье излагается один из возможных методов решения данной задачи.

## 3. Общий подход к вычислению расстояния между двумя семантическими аннотациями

Для определения степени семантической близости двух операционных компонент введем метрику расстояния между их аннотациями. В ней будет использоваться комбинация нескольких метрик семантического расстояния между двумя узлами (как классами так и свойствами) онтологии и основанная на них метрика расстояния между двумя графами аннотаций.

В данной статье предполагается, что все концепты, которые приходится сравнивать, находятся в одной онтологии, и, соответственно, в одной таксономии. В случае, если онтологии различны, до начала анализа они должны быть объединены.

Множество методов разработано для решения этой задачи, например [6,7].

### 3.1. Метрики близости узлов онтологии

В качестве меры близости двух узлов онтологии наиболее распространены два подхода [8]: а) на основе таксономий; б) на основе свойств.

*Метрика близости на основе таксономий.* Таксономическая составляющая онтологии формируется на основе отношения IS-A между концептами онтологии. *Метрика близости на основе таксономий*  $\sigma(c1, c2) \in [0..1]$  и вычисляется как

$$\sigma(c1, c2) = 1 - \delta(c1, c2), \quad (1)$$

где  $\delta(c1, c2)$  – взвешенное расстояние между узлами  $c1$  и  $c2$ . Для вычисления  $\delta(c1, c2)$  каждому узлу таксономии присваивается вес  $w(c) = \frac{1}{2^{l(c)}}$ , где  $l(c)$  – большая длина пути от корневого узла таксономии до  $c$ .

$$\delta(c1, c2) = [w(LCA(c1, c2)) - w(c1)] + [w(LCA(c1, c2)) - w(c2)] \quad (2)$$

здесь  $LCA(c1, c2)$  – наименьший (наиболее удаленный от корня) общий предок узлов  $c1$  и  $c2$ . Алгоритм нахождения  $LCA$  хорошо известен [9].

Данная мера расстояния имеет два свойства: 1) семантическая разница между узлами более высокого уровня более важна, чем между более удаленными от корня, 2) семантическая разница между узлами-братьями больше, чем между родителем и его ребенком.

Еще одним преимуществом метрики на основе таксономий является возможность применить её не только к классам онтологии, но и к свойствам, которые также упорядочиваются в виде самостоятельной иерархии.

Рассмотрим пример расчета расстояний между классами на основе таксономии действий (рис. 2)

$$\begin{aligned} \omega(\text{Action}) &= 0.25 \\ \omega(\text{UIAction}) &= 0.125 \\ \omega(\text{GUISystemAction}) &= \\ &= \omega(\text{TextModeSystemAction}) = 0.03125 \end{aligned}$$

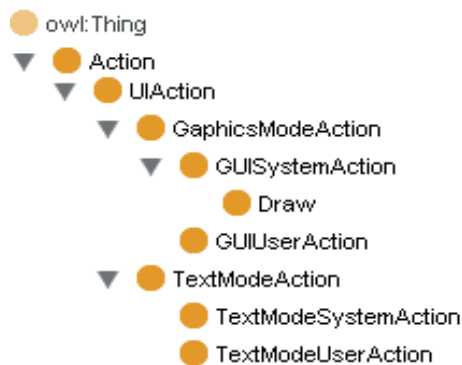


Рис. 2. Фрагмент таксономии действий

$$\begin{aligned} \delta(\text{UIAction}, \text{GUISystemAction}) &= \\ &= 0.125 - 0.03125 = 0,09375 \\ \sigma(\text{UIAction}, \text{GUISystemAction}) &= \\ &= 1 - 0.09375 = 0,90625 \\ \delta(\text{GUISystemAction}, \text{TextModeSystemAction}) &= \\ &= (0.125 - 0.03125) + (0.125 - 0.03125) = 0,1875 \\ \sigma(\text{GUISystemAction}, \text{TextModeSystemAction}) &= \\ &= 0.8125 \end{aligned}$$

*Метрика близости на основе свойств.* В онтологии узел характеризуется не только положением в иерархии, но и набором примитивных (data type) и объектных (object type) свойств. Фактически объектные свойства задают возможные отношения между классами онтологии. Чем больше общих свойств между двумя концептами и чем меньше различных, тем ближе они должны быть по расстоянию (3).

$$\delta(c1, c2) = \frac{2 \times |P(c1) \cap P(c2)|}{|P(c1) \cap P(c2)| + |P(c1) \cup P(c2)|} \quad (3)$$

В (3) P(c) – множество примитивных и объектных свойств класса c.

### 3.2. Метрика расстояния между двумя аннотациями

За основу разрабатываемой метрики взяты результаты работы [5].

Разница предлагаемых подходов заключается в том, что в [5] аннотация состоит из множества узлов онтологии и не содержит связей. В этом случае расстояние между аннотациями  $\text{Sim}(A1, A2) \in [0..1]$  вычисляется как среднее между расстояниями множеств входов и выходов аннотации П (5).  $\text{П}(C, C) \in [0..1]$  логично вычислять как нормированную сумму метрик близости наилучшего отображения между наборами узлов двух аннотаций (4).

$$\text{П}(C1, C2) = \frac{\text{MAX}_{c1 \in C1, c2 \in C2} \{ \text{П}(C1 - c1, C2 - c2) + \sigma(c1, c2) \}}{\text{MIN}(|C1|, |C2|)} \quad (4)$$

здесь C1 и C2 – множества узлов первой и второй аннотации соответственно.

$$\text{Sim}(A1, A2) = \frac{\text{П}(I1, I2) + \text{П}(O1, O2)}{2} \quad (5)$$

Для учета связей между узлами графа введем процедуру вычисления степени семантической близости аннотаций A1 и A2 методом обхода в глубину. В алгоритме (6):

Mark(c) – функция помечающая узел c как посещенный.

GetUnmarkedProps(P) – из множества дуг P выбирает только те, которые связаны хотя бы с одной непомеченной вершиной.

MakePair(c1, c2) – создает кортеж из двух узлов c1 и c2.

GetTarget(p) – из дуги p выбирает связанный с ней непомеченный узел.

$\text{MAX}_{p1 \in P1, p2 \in P2} \{ \text{MakePair}(\text{GetTarget}(p1), \text{GetTarget}(p2)) \}$  – находит наилучшее отображение между наборами связей P1 и P2 на основе метрики (1 и 2) для свойств и с учетом направления отношений (входящие/исходящие). SimSum(c1, c2, Sum, Count) – процедура вычисления степени семантической близости двух аннотаций, начиная с узлов c1 и c2.

$$\begin{aligned} \text{SimSum}(c1, c2, \text{Sum}, \text{Count}) \{ \\ \text{Mark}(c1), \text{Mark}(c2); \\ \text{Sum} += \sigma(c1, c2) \\ \text{Count}++; \\ P1 = \text{GetUnmarkedProps}(P(c1)), P2 = \text{GetUnmarkedProps}(P(c2)); \\ P3 = \text{MAX}_{p1 \in P1, p2 \in P2} \{ \text{MakePair}(\text{GetTarget}(p1), \text{GetTarget}(p2)) \} \\ \text{foreach}(p \text{ in } P3) \\ \quad \text{SimSum}(p.\text{first}, p.\text{second}, \text{Sum}, \text{Count}); \end{aligned} \quad (6)$$

Начинать вычисление следует с узла, представляющего главный глагол-действие аннотации.

Используя Sum и Count, полученные в SimSum, вычислить П(A1, A2) с учетом связей теперь возможно по формуле:

$$\text{П}(A1, A2) = \frac{\text{Sum}}{\text{Count}} \quad (7)$$

Недостатком (7) является отсутствие возможности учесть соотношение сопоставленных узлов A1 и A2 с оставшимися, которые могут задавать дополнительные ограничения, неучтенные одной из аннотированных операционных компонент. Для этого хорошо подходит формула (8), идея которой близка к (3):

$$\delta(A1,A2) = \frac{2 \times |\text{Nodes}(A1) \cap \text{Nodes}(A2)|}{|\text{Nodes}(A1) \cap \text{Nodes}(A2)| + |\text{Nodes}(A1) \cup \text{Nodes}(A2)|} \quad (8)$$

В (8)  $\text{Nodes}(A)$  – множество узлов аннотации A. При выполнении операций пересечения и объединения эквивалентными считаются пары унифицированных узлов из A1 и A2, в набор унифицированных узлов входят также узлы-OCL ограничений, для которых унификация без учета имен переменных дала положительный результат. Таким образом, окончательная формула для вычисления  $\Pi(A1,A2)$  будет иметь вид:

$$\Pi(A1,A2) = \varphi \times \frac{\text{Sum}}{\text{Count}} + (1 - \varphi) \times \delta(A1,A2) \quad (9)$$

Наиболее подходящее значение для коэффициента  $\varphi \in [0..1]$ , который задает вклад каждой из составляющих в  $\Pi(A1,A2)$ , может быть получено экспериментальным путем. Рассмотрим применение описанного выше метода (формулы 9) на примере нескольких аннотаций (рис. 3).

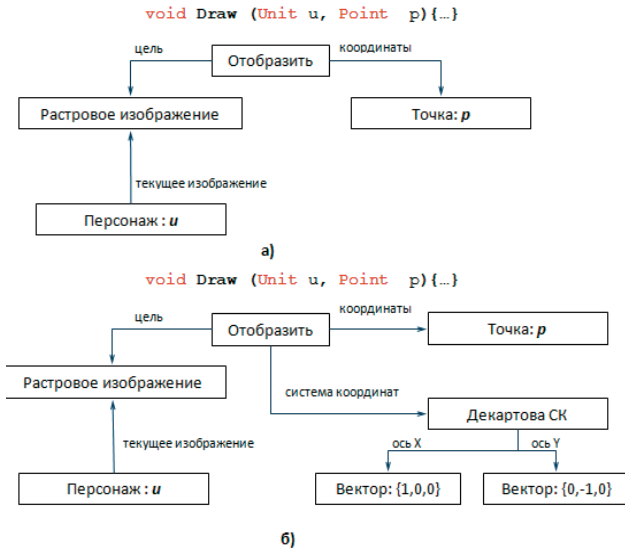


Рис. 3. Пример аннотаций разных методов:

- а) вывод текущего изображения персонажа u в точке p;
- б) вывод текущего изображения персонажа u в точке p в декартовой левосторонней системе координат

Положим, что  $\varphi = 0.6$ , тогда поскольку аннотация (б) отличается от (а) только тремя дополнительными узлами, то

$$\Pi(3.a,3.b) = 0.6 \times \frac{4}{4} + 0.4 \times \frac{2 \times 4}{4+7} = 0.6 + 0.291 = 0.891$$

В то же время аннотация с рис. 4 отличается от аннотации рис. 3а основным глаголом-действием, и соответственно, имеет несколько другой набор узлов и отношений. Объемы статьи не позволяют привести всю онтологию, подтверждающую правильность вы-

числения меры близости между различными узлами аннотации. Приведем уже готовые значения.

Сравнение начинается как всегда с пары основных узлов-действий:

$$\sigma(\text{ПрочитатьМножествоИзображений}, \text{Отобразить}) = 0,520$$

Наилучшее отображение отношений узла «Прочитать МножествоИзображений» (цель, источник, координаты) на отношения узла «Отобразить» (цель, координаты) формируется сравнением результатов вычисления метрики близости для всех элементов декартового произведения этих множеств. В данном случае будут сопоставлены отношения (цель, цель) и (координаты, координаты), хотя в данном случае отношение координаты и играет разную роль в каждом из действий.

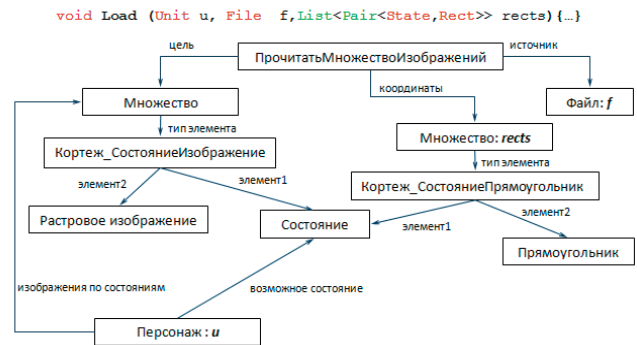


Рис. 4. Аннотация, пример 2

Соответственно, множество P3 из метода (6) будет составлено из пар (Растровое изображение, Множество) и (Множество:rects, Точка:p).

$$\sigma(\text{Множество}, \text{РастровоеИзображение}) = 0,594$$

$$\sigma(\text{Множество}, \text{Точка}) = 0,625$$

Узел «точка:p» больше связей не имеет, в то время как узел «Растровое изображение» имеет входящее отношение «текущее изображение» (рис.3а), сравнение которого с двумя отношениями входящим «изображения по состояниям» и исходящим «тип элемента» узла «Множество» (рис.4) дает единственную пару отношений («текущее изображение», «изображения по состояниям»), и, соответственно единственную пару узлов (Персонаж:u, Персонаж:u), метрика близости между которыми дает 1.

Итого, имеем унифицированные 4 узла и по (7) получим:

$$\Pi(3.a,4) = \frac{0.520 + 0.594 + 0.625 + 1}{4} = 0,68475$$

Неунифицированными остались 6 узлов из аннотации на рис. 4 По (9) при  $\varphi = 0.6$  получим:

$$\Pi(3.a,4) = 0.6 \times 0,68475 + 0.4 \times \frac{2 \times 4}{4+10} = 0.411 + 0.229 = 0.640$$

Это значительно меньше, чем рассчитанное ранее  $\Pi(3.a,3.б) = 0.891$ , чего и следовало ожидать.

#### 4. Выводы и перспективы дальнейших исследований

В данной статье усовершенствован метод расчета метрики близости между двумя семантическими аннотациями. Он позволяет сравнивать две аннотации, представленные в виде графов. Узлы такого графа – это классы некоторой онтологии предметной области, а дуги – отношения между ними.

Предложенный подход может быть использован для усовершенствования метода синтеза структурных схем программ на основе семантических вычислительных сетей. Усовершенствованный метод, кроме отношений вычислимости между узлами сети позволяет учитывать семантическую аннотацию

узлов, формализующую суть выполняемого узлом действия.

При интеграции графовых аннотаций в онтологию OWL-S (язык описания семантики веб-сервисов) представленная в статье метрика позволит точнее подбирать веб-сервисы по их профилю и запросу пользователя.

Направление дальнейших исследований может быть связано с усовершенствованием метрики близости семантических аннотаций путем расширения аннотации естественно-языковым описанием и применения статистических методов (наподобие tf-idf [10]) их сравнения для вычисления дополнительной компоненты формулы (9).

#### Литература

1. Новиков, Ф. А. Учебно-методическое пособие по дисциплине «Системы представления знаний» [Текст] / Ф.А. Новиков – СПб:ИТМО, 2007. – 119 с.
2. Martin, D. OWL-S: Semantic markup for web services. [Электронный ресурс] / World Wide Web Consortium (W3C). – Режим доступа / URL:www http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/ – 08.04.2013 – Заглавие с экрана.
3. Ermolayev, V. Towards a Framework for Agent-Enabled Semantic Web Service Composition [Текст] / V. Ermolayev, N. Keberle, S. Plaksin, O. Kononenko, V. Terziyan // International Journal of Web Services Research. – 2004. – №1(3). – P.63-87.
4. Плиско, Д. А. Расширение возможностей программного обеспечения при использовании баз знаний, основанных на OWL 2.0 [Текст] / Д. А. Плиско, А. Ю. Шевченко // Восточно-Европейский журнал передовых технологий. – 2010. – Т.4. – № 2 (46). – С. 32-36.
5. Ganjisaffar, Y. A Similarity Measure for OWL-S Annotated Web Services [Text] / Y. Ganjisaffar, H. Abolhassani, M. Neshati // Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence. – Hong Kong, 2006. – P. 621 – 624.
6. Stumme, G. Ontology merging for federated ontologies on the semantic web [Текст] / G. Stumme, A. Maedche // In Proceedings of the International Workshop for Foundations of Models for Information Integration (FMII-2001). – Sept. 2001. – P. 413-418.
7. Melnik, S. Similarity flooding: A versatile graph matching algorithm and its applications to schema matching [Текст] / S. Melnik, H. Garcia-Molina, E. Rahm // In Proceedings of the 18th International Conference on Data Engineering. – San Jose, CA: IEEE Computer Society, Feb. 2002. – 45 pp.
8. Zhong, J. Conceptual graph matching for semantic search [Текст] / J. Zhong, H. Zhu, J. Li, Y. Yu // Proceedings of the 2002 International Conference on Computational Science (ICCS2002). – Amsterdam, 2002. – P. 92-106.
9. Иванов, М. Наименьший общий предок. Нахождение за  $O(\sqrt{N})$  и  $O(\log N)$  с препроцессингом  $O(N)$  [Электронный ресурс] / Собрание алгоритмов e-maxx. – Режим доступа / URL:www http://e-maxx.ru/algo/lca – 08.04.2013 – Заглавие с экрана.
10. Федоровский, А.Н. Mail.ru на РОМИП-2005 [Текст] / А.Н. Федоровский, М. Ю. Костин // Сборник «Труды РОМИП'2005» Труды третьего российского семинара по оценке методов информационного поиска. Под ред. И. С. Некрестьянова. – Санкт-Петербург: НИИ Химии СПбГУ, 2005. – С. 106-124.