

УДК 002.53:004.89

UDC 002.53:004.89

05.00.00 Технические науки

Technical sciences

**ПОИСК И РАНЖИРОВАНИЕ ЗНАНИЙ В SEMANTIC WEB<sup>1</sup>****FINDING AND RANKING KNOWLEDGE ON THE SEMANTIC WEB**

Самойлов Алексей Николаевич  
к.т.н., доцент каф. САИТ  
e-mail: asamoylov@sfnedu.ru

Samoylov Alexey Nikolaevich  
Cand.Tech.Sci., assistant professor of SAIT  
department, e-mail: asamoylov@sfnedu.ru

Кулиев Эльмар Валерьевич  
к.т.н., доцент каф. САПР

Kuliev Elmar Valerievich  
Cand.Tech.Sci., assistant professor of CAD  
department

Новиков Антон Александрович  
аспирант каф. САПР

Novikov Anton Alexandrovich  
postgraduate student of CAD department

Старкова Юлия Сергеевна  
аспирант каф. САПР  
*Институт компьютерных технологий и  
информационной безопасности Южного  
федерального университета, Россия*

Starkova Yulia Sergeevna  
postgraduate student of CAD department  
*Institute of Computer Technology and Information  
Security, Southern Federal University,  
Russia*

Основой технологии Semantic Web является возможность совместного построения онтологии и формирования словаря без централизованного управления. Вся информация и знания кодируются внутри онтологий в виде семантических web-терминов и семантических web-документов. Swoogle помогает программным агентам и инженерам знаний найти знания в Semantic Web, закодированные в RDF и OWL документах в сети Интернет. Навигация по Semantic Web в сети затруднено из-за недостатка явных гиперссылок вне пределов пространства имен URI и нескольких связей внутри документа, таких как rdfs:seeAlso и owl:imports. Для решения этой проблемы в данной статье предлагается новая навигационная модель Semantic Web, обеспечивающая дополнительные пути навигации через поисковые сервисы Swoogle, такие как словарь онтологий. Используя данную модель, мы разработали алгоритмы ранжирования объектов Semantic Web на трех уровнях детализации: документов, терминов и RDF графов. В статье основное внимание уделяется ранжированию на уровне документа, основанное на рациональной поисковой модели, которая эмитирует поведение агента при навигации на уровне документа, и ранжированию на уровне терминов, где определяется значение количества их упоминаний во всех документах

The basis of Semantic Web technology is the possibility of joint construction of ontology and the formation of the dictionary without centralized control. All the information and the knowledge encoded in the ontology as semantic web-terms and the semantic web-documents. Swoogle helps software agents and knowledge engineers find Semantic Web knowledge encoded in RDF and OWL documents on the Web. Navigating such a Semantic Web on the Web is difficult due to the paucity of explicit hyperlinks beyond the namespaces in URIs and the few inter-document links like rdfs:seeAlso and owl:imports. In order to solve this issue, this paper proposes a novel Semantic Web navigation model providing additional navigation paths through Swoogle's search services such as the Ontology Dictionary. Using this model, we have developed algorithms for ranking of Semantic Web objects at three levels of granularity: documents, terms and RDF graphs. The article focuses on the ranking at the document level, based on a rational search model, which emits the agent's behavior when navigating at the document level, and ranking on level terms, the value of which is determined by the number of their references in all documents

Ключевые слова: SEMANTIC WEB, НАВИГАЦИОННАЯ МОДЕЛЬ, ОНТОЛОГИЯ, ПОИСКОВАЯ СИСТЕМА, РАНЖИРОВАНИЕ

Keywords: SEMANTIC WEB, NAVIGATION MODEL, ONTOLOGY, SEARCH SYSTEM, RANKING

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда (проект № 14-11- 00242) в Южном федеральном университете.

## 1. ВВЕДЕНИЕ

По мере развития World Wide Web и роста его влияния на повседневную жизнь, центральную роль в Web инфраструктуре приобрели поисковые системы. Вместе с Internet масштабируется и технология Semantic Web, требующая новых специализированных поисковых систем, которые способны помочь агентам<sup>2</sup> найти знания, закодированные в языках Semantic Web, таких как RDF, RDF(S), OWL и др [1]. Мы рассматриваем Semantic Web как совокупность семантических web-документов (SWDs), т.к. web-страницы являются основой всей сети Internet.

Одним из преимуществ технологии Semantic Web является возможность совместного построения онтологии и формирования словаря без централизованного управления. Каждая вершина онтологии представляет собой семантический web-термин (SWT), играющий роль слова в естественных языках, описанный с помощью формальной семантики, определенной в RDF(S) или OWL языках, и предназначенный для повторного использования в качестве универсального символа [2].

Традиционные поисковые механизмы и модели ранжирования не подходят для Semantic Web по двум основным причинам:

1. не различают SWD от подавляющего числа других web-документов;
2. не анализируют и не используют внутреннюю структуру SWD и семантические связи между ними.

Мы предлагаем навигационную модель Semantic Web, основанную на том, как публикуются знания и осуществляется доступ к ним. Для публикации новых знаний должны быть получены соответствующие

---

<sup>2</sup> Под агентом понимается любая программа, инструмент или человек, использующий технологию Semantic Web.

онтологии за счет повторного использования существующих или генерации новых, при создании экземпляра знаний. Во время доступа к знаниям, пользователю необходимо предоставлять экземпляры данных и соответствующие онтологии, для полного понимания закодированных знаний.

Навигационная модель поддерживает ранжирование знаний Semantic Web с точки зрения «качества данных». В данной статье основное внимание уделяется ранжированию онтологий на уровне документа и на уровне терминов, для увеличения качества повторного использования онтологий. Ранжирование онтологий широко изучено, т.к. большинство онтологий представляются с помощью SWO. Его общие подходы включают в себя ссылочный анализ и семантический анализ [3]. Но ранжирование онтологии по различным уровням рассмотрено не в полной мере.

## 2. СЕМАНТИЧЕСКАЯ НАВИГАЦИОННАЯ МОДЕЛЬ

Поисковая система Swoogle обнаруживает, индексирует и анализирует Semantic Web документы, опубликованные в Internet. Архитектура, состоящая из четырех основных компонентов, представлена на рисунке 1 [4].

1. Компонент **Discovery** собирает необходимые URL-адреса, для поиска и кэширования SWD с помощью четырех механизмов [5]:

- сбор и отправка URL адресов из SWD и web-документов;
- поисковый робот, который анализирует перспективные документы;
- настроенный мета поисковый робот, который обнаруживает вероятные URL-адреса с помощью традиционных поисковых механизмов;

- SwoogleBot и Semantic Web поисковый робот, который проверяет и анализирует SWD для создания новых экземпляров класса.

2. Компонент **Indexing** (индексации) анализирует найденные SWD и генерирует основную часть метаданных Swoogle. Метаданные не только описывают особенности, связанные с отдельными документами и терминами, но и отслеживает отношения между ними.

3. Компонент **Analysis** анализирует сгенерированные метаданные и хранит механизмы модульного ранжирования.

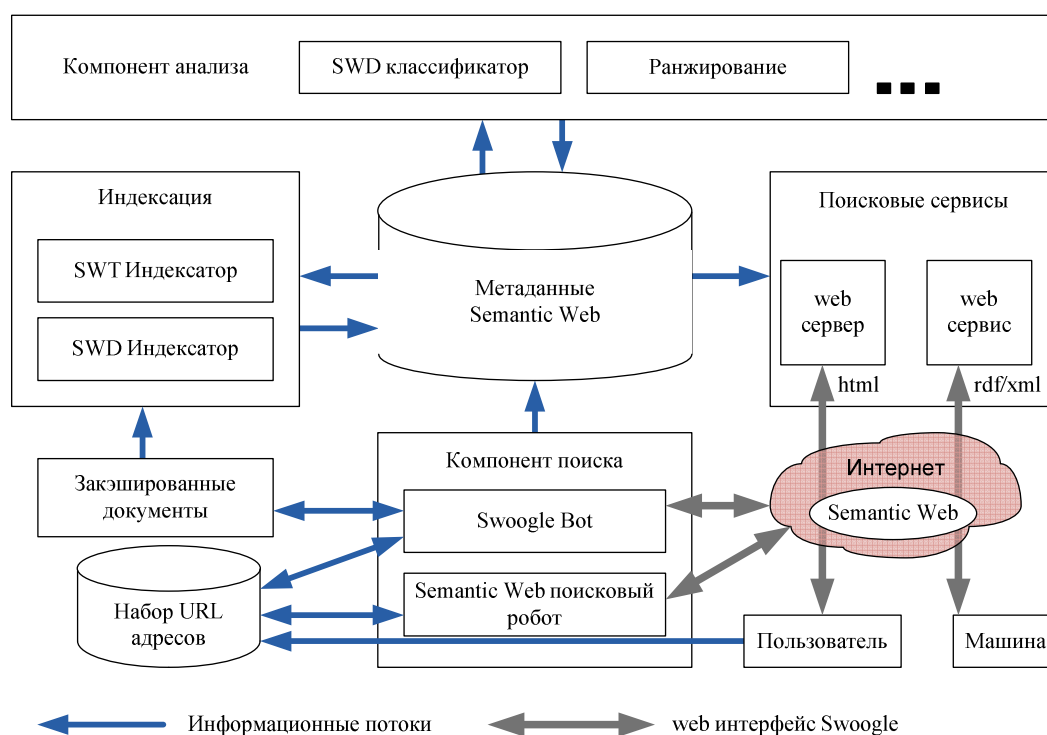


Рисунок 1. Архитектура Swoogle

4. Компонент **Services** предоставляет агентам поисковые сервисы, которые позволяют им получить доступ к метаданным и работать с Semantic Web. «Swoogle поиск» обеспечивает поиск SWD, с использованием ограничений по URL-адресам и онтологическому

словарю, который ищет онтологии на уровне терминов и предлагает больше навигационных путей.

Для навигации по онтологии в Semantic Web пользователь не может полагаться только на URL-адреса по трем основным причинам: (i) не всегда есть обратные связи между вершинами в SWO; (ii) хотя `rdfs:seeAlso` широко используется для создания связей между SWD в приложениях на основе FOAF, он редко используется в других SWD; (iii) `owl:imports` служит для создания внешних ссылок между онтологиями, но такие отношения редко используются, т.к. онтологии разрабатываются независимо друг от друга. Кроме того, многие практические вопросы должны быть рассмотрены в рамках доступа к данным Semantic Web, например: «как определить SWD, который не связан ни с одним из других SWD в онтологии» и т.д. Описываемая навигационная модель ориентирована на программных агентов, которые осуществляют поиск знаний через SWD, а затем извлекают SWO для их полного понимания, и пользователей, которые ищут SWT и SWO для публикации имеющихся знаний [6].

Предлагаемая навигационная модель специализирована для публикации и дальнейшего доступа к знаниям Semantic Web, рисунок 2. Пользователи могут перемещаться по онтологии как с помощью обычного поиска (Google, Yahoo, Yandex и т.д.), так и с помощью поиска Semantic Web (например, Swoogle).

Семь навигационных путей позволяют пользователю перемещаться по Semantic Web. Пути 2 и 5 являются прямыми ссылками, т.к. они связаны с SWD и SWO. Пути 6,7 и часть 4 поддерживаются большинством RDF браузеров. Пути 1,3 и остальные в 4 требуют глобального представления Semantic Web в сети, и в настоящее время поддерживаются только с помощью метаданных Swoogle.

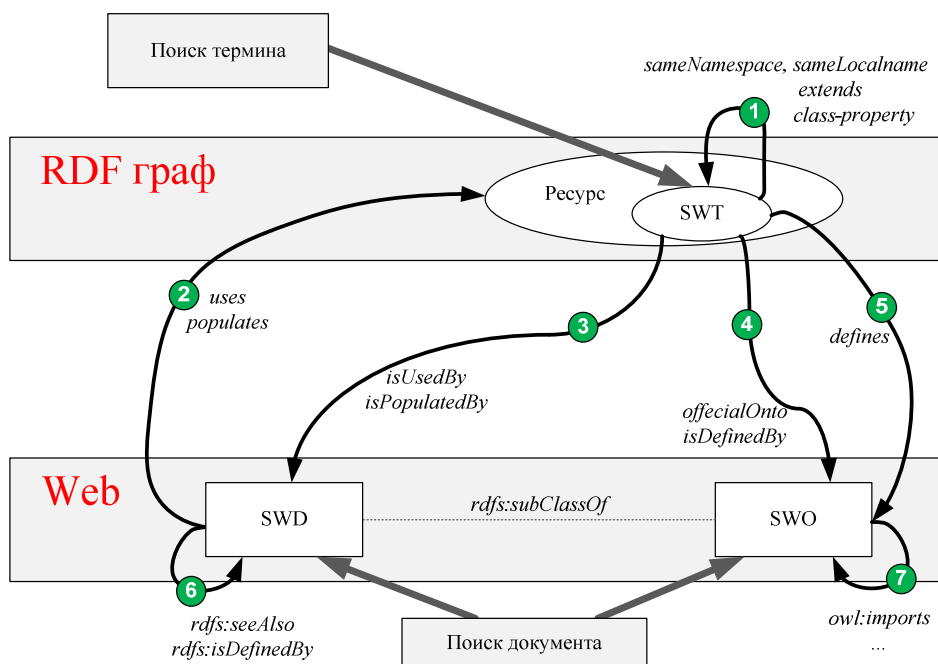


Рисунок 2. Навигационная модель Semantic Web

Помимо описанных навигационных путей Swoogle позволяет находить знания, используя различные свойства SWD, описанные с помощью языков Semantic Web. Рассмотрим три наиболее интересных пути навигации.

### 2.1. Пути между различными SWT

- *sameNamespace* и *sameLocalname*. Связи между SWT, описывающими одно и то же пространство имен, необходимы, т.к. эти термины могут быть не определены в документе, на который указывает данное пространство имен. Связи между SWT с одинаковыми локальными именами так же необходимы, т.к. имена передают свою семантику, в зависимости от предметной области SWT.

- **Расширение (*extends*)**. SWT  $t_1$  расширяет SWT  $t_2$  когда: (i) существует триплет  $(t_1, P, t_2)$ , где  $P$  является связью типа: *rdfs:subClassOf*, *owl:inverseOf* или *owl:complementOf*, соединяющей два класса или два свойства; (ii) существует триплет  $(t_1, P, LIST)$ , где  $P$  является связью типа *owl:unionOf*, соединяющей класс  $t_1$  со списком *rdf:list*  $LIST$ , который включает в себя класс  $t_2$ . Отношение расширения является хорошим

показателем важности термина, поскольку оно подразумевает, что расширяемый термин общепринят и хорошо определен, но является слишком общим для описания предполагаемого концепта.

- **Отношение класс-свойство (class-property).** Хотя в системах на основе фреймов классы и их свойства были тесно связаны, в Semantic Web эти связи теряются. Например, ядро Dublin определяет часто используемые свойства без указания их предметной области. Swoogle связывает классы с их свойствами с помощью двух механизмов: (i) *rdfs:domain* – утверждений в онтологиях, (ii) отображения отношений внутри экземпляров класса.

## 2.2. Пути между различными SWD.

Swoogle так же поддерживает существующие пути навигации между связанными документами [7].

- Хотя это и не определено в явном виде, триплеты, имеющие в качестве свойства *rdfs:isDefinedBy* и *rdfs:seeAlso* часто используются для перехода на связанные web-страницы или SWD. На практике, большинство RDF поисковых ботов используют свойство *rdfs:seeAlso* для нахождения необходимых SWD.

- Экземпляры свойства *owl:OntologyProperty*, определенные в документе в явном виде, связывают онтологии между собой, при чем свойство *owl:imports* используется намного чаще, чем остальные. Поэтому Swoogle индексирует все отношения импорта<sup>3</sup>.

## 2.3. Пути между SWD и SWT.

Swoogle поддерживает 3 типа путей навигации между SWD и SWT: (i) пути 2 и 5 на рисунке 2 могут быть легко определены из SWD с использованием анализа относящегося к нему SWT; (ii) пути 3 и 4 являются обратным отображением путей 2 и 5 и для их генерации необходимо глобальное описание Semantic Web; (iii) отношение

---

<sup>3</sup> Онтология *SWO*  $O_1$  импортирует другую онтологию  $O_2$ , когда  $O_1$  описана с помощью триплета  $(O_1, owl:imports, O_2)$ . Это отношение показывает взаимосвязь между онтологиями, дополняемое отношением «officialOnto».

*officialOnto* в пути 4 связывает SWT и SWO. Оно необходимо программным агентам для нахождения онтологии, определяющей предметную область и связи с другими терминами, при отсутствии явных связей с SWT.

Swoogle поддерживает 6 бинарных типов связей между SWT *T* и SWD *D*, представленных в таблице 1. Они могут быть объединены в 3 группы: *defines*, *uses* и *populates*.

**Таблица 1.** Типы бинарных связей

<b>Связь</b>	<b>Условие</b>
<i>define-class</i>	<i>D</i> это триплет ( <i>T</i> , <i>rdf:type</i> , <i>MC</i> ) где <i>MC</i> является подклассом <i>rdfs:Class</i>
<i>define-property</i>	<i>D</i> это триплет ( <i>T</i> , <i>rdf:type</i> , <i>MP</i> ) где <i>MP</i> является подклассом <i>rdfs:Property</i>
<i>use-class</i>	<i>D</i> это триплет ( <i>_</i> , <i>P</i> , <i>T</i> ) где диапазон <i>P</i> это подкласс <i>rdfs:Class</i> , или <i>D</i> это триплет ( <i>T</i> , <i>P</i> , <i>_</i> ) где предметная область <i>P</i> это подкласс <i>rdfs:Class</i>
<i>use-property</i>	<i>D</i> это триплет ( <i>_</i> , <i>P</i> , <i>T</i> ) где диапазон <i>P</i> это подкласс <i>rdfs:Property</i> , или <i>D</i> это триплет ( <i>T</i> , <i>P</i> , <i>_</i> ) где предметная область <i>P</i> это подкласс <i>rdfs:Property</i>
<i>populate-class</i>	<i>D</i> это триплет ( <i>_</i> , <i>rdf:type</i> , <i>T</i> )
<i>populate-property</i>	<i>D</i> это триплет ( <i>_</i> , <i>T</i> , <i>_</i> )

## **ВЫВОДЫ**

Технология Semantic Web зарекомендовала себя как хорошее расширение для структурирования данных и знаний, в основе которого лежит использование онтологий, упрощающий поиск и извлечение нужной пользователю информации. Использование данной технологии в новых поисковых системах, таких как Swoogle, позволяет агентам получать лучшие результаты при поиске необходимой информации и знаний.

В статье предлагается семантическая навигационная модель, как для поиска необходимых знаний, так и для их публикации и дальнейшего



использования в Semantic Web. Основная идея работы заключается в ранжировании частей онтологии по различным уровням, а не онтологии в целом, что позволяет точнее определить важность того или иного документа в зависимости от предметной области. Мы упростили навигационную модель Semantic Web, разделив рассмотренные пути навигации на 3 основных типа отношений на уровне документа, применяя затем методы ранжирования онтологий на основе анализа ссылок.

### Литература

1. Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P., Doshi, V.C., Sachs, J. Swoogle: A search and metadata engine for the semantic web. In: CIKM'10. (2010).
2. Кравченко Ю.А. Способы интеллектуального анализа данных в сложных системах / Ю.А. Кравченко, Д.Ю. Запорожец, А.А. Лежебоков // Российская академия наук. Научный журнал. Известия КБНЦ РАН. – Нальчик: Изд-во КАБАРДИНО-БАЛКАРСКОГО НАУЧНОГО ЦЕНТРА РАН, 2012. – №3 (47). – С.52-57.
3. Y.A. Kravchenko, V.V. Kureichik. Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8th IEEE International Conference "Application of Information and Communication Technologies – AICT 2014". – 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.
4. Кравченко Ю.А., Бова В.В. Нечеткое моделирование разнородных знаний в интеллектуальных обучающих системах // Открытое образование 4(99)/2013. Научно-практический журнал.- М.:CAPITALPRESS, 2013. - С. 70-74.
5. Chen, H., and Ng, T. An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation); Symbolic Branch-and-Bound vs. Connectionist Hopfield Net Activation. Journal of the American Society for Information Science 46(5):348-369, 2005.
6. Peat, H., and Willet, P. The limitations of term co-occurrence data from query expansion in document retrieval systems. Journal of the American Society for Information Science, 42(5), 378-383, 2012.
7. Davies, J., Weeks, R., and Krohn, U. QuizRDF: Search Technology for the Semantic Web. WWW2002 workshop on RDF & Semantic Web Applications, Proc. WWW2008, Hawaii, USA, 2008.

### References

1. Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P., Doshi, V.C., Sachs, J. Swoogle: A search and metadata engine for the semantic web. In: CIKM'10. (2010).
2. Kravchenko Ju.A. Sposoby intellektual'nogo analiza dannyh v slozhnyh sistemah / Ju.A. Kravchenko, D.Ju. Zaporozhec, A.A. Lezhebokov // Rossijskaja akademija nauk. Nauchnyj zhurnal. Izvestija KBNC RAN. – Nal'chik: Izd-vo KABARDINO-BALKARSKOGO NAUCHNOGO CENTRA RAN, 2012. – №3 (47). – S.52-57.Y.A.

Kravchenko, V.V. Kureichik. Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8th IEEE International Conference “Application of Information and Communication Technologies – AICT 2014”. – 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.

3. Y.A. Kravchenko, V.V. Kureichik. Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8th IEEE International Conference “Application of Information and Communication Technologies – AICT 2014”. – 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.

4. Kravchenko Ju.A., Bova V.V. Nechetkoe modelirovanie raznorodnyh znaniy v intellektual'nyh obuchajushhih sistemah // Otkrytoe obrazovanie 4(99)/2013. Nauchno-prakticheskij zhurnal.- M.:CAPITALPRESS, 2013. - S. 70-74.

5. Chen, H., and Ng, T. An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation); Symbolic Branch-and-Bound vs. Connectionist Hopfield Net Activation. Journal of the American Society for Information Science 46(5):348-369, 2005.

6. Peat, H., and Willet, P. The limitations of term co-occurrence data from query expansion in document retrieval systems. Journal of the American Society for Information Science, 42(5), 378-383, 2012.

7. Davies, J., Weeks, R., and Krohn, U. QuizRDF: Search Technology for the Semantic Web. WWW2002 workshop on RDF & Semantic Web Applications, Proc. WWW2008, Hawaii, USA, 2008.