

Программные системы и вычислительные методы

Правильная ссылка на статью:

Лукичев Р.В. Эволюция технологий семантического веба: проблемы и перспективы // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71719 EDN: JJXDYW URL:

https://nbpublish.com/library_read_article.php?id=71719

Эволюция технологий семантического веба: проблемы и перспективы

Лукичев Руслан Владимирович

ORCID: 0000-0002-2293-2410

кандидат философских наук

магистр; факультет программной инженерии и компьютерной техники; Университет ИТМО
Директор по развитию; ООО "Первая Фестивальная Компания"

198332, Россия, г. Санкт-Петербург, Ленинский пр-т, 100к3, кв. 113

✉ ruslanlukichev@gmail.com



[Статья из рубрики "Базы данных"](#)

DOI:

10.7256/2454-0714.2024.3.71719

EDN:

JJXDYW

Дата направления статьи в редакцию:

16-09-2024

Аннотация: Статья посвящена рассмотрению ключевых технологий семантического веба, анализу их особенностей, проблемных сторон и точек роста, что представляется особенно актуальным в контексте вопросов импортозамещения и повышения информационной безопасности нашей страны. Особое внимание уделено RDF-графам, которые основаны на онтолого-ориентированном подходе, а также языку OWL как основному инструменту организации машиночитаемых структур данных со сложными взаимосвязями между сущностями, иерархией классов и свойств. В отдельном блоке проанализированы ограничения, связанные с вопросами безопасности семантических баз данных, необходимости их упрощения, стандартизации и развития профильного программного обеспечения, отвечающего критериям юзабилити. Кроме того, обозначены перспективы дальнейшего совершенствования указанных технологий в контексте Интернета вещей и искусственного интеллекта. В исследовании применена комплексная методологическая база, которая подразумевает использование преимущественно

общенаучных методов, в частности, системного и аналитического: в статье обобщены и проанализированы актуальные наработки, связанные с технологиями семантического веба, что позволило выявить целый ряд проблем, нуждающихся в решении. В частности, имеющиеся на сегодняшний день инструменты зачастую имеют высокий порог входа, отличаются избыточно сложным, невыразительным интерфейсом без функций дополняющих подсказок и визуализации запросов. Кроме того, языки семантического веба нуждаются в стандартизации и введении общего протокола с целью упрощения процесса работы с разноформатными данными, агрегированными из разных источников. Другими важными вопросами являются обеспечение достоверности и актуальности информации, ее целостности и конфиденциальности, а также контекстуальной обусловленности логических выводов и соответствия запросам пользователя. В числе ключевых перспектив названо создание интеллектуальной автономной среды, устройства в которой смогут свободно обмениваться данными и взаимодействовать друг с другом на семантическом уровне с целью оказания высококачественных персонализированных услуг. Положения статьи могут быть взяты за основу при разработке отечественных систем структуризации и описания данных, доступных для машинной обработки, а также профильных лекционных курсов в высших учебных заведениях.

Ключевые слова:

семантический веб, онтологии, графовые базы данных, модели данных, семантическая сеть вещей, RDF, RDFS, OWL, SPARQL, XML

Актуальность темы и методология исследования

Семантический веб – это концепция, предложенная ровно четверть века назад сэром Тимом Бернерсом Ли, создателем Всемирной паутины. Еще в преддверии эпохи Web 2.0 он пророчески заявлял о необходимости научить компьютеры самостоятельно описывать те или иные явления, затем делать выводы и, наконец, аргументированно рассуждать [\[1, с. 184\]](#). На страницах книги «Плетя паутину: истоки и будущее Всемирной паутины» основатель и бессменный глава Консорциума W3C рассуждал о необходимости становления семантического веба как информационной сети, данные в которой были бы представлены в машиночитаемом формате или конвертированы в таковой, – с целью их прямой или косвенной обработки компьютерами, постоянно взаимодействующими между собой [\[1, с. 177\]](#).

Семантический веб все еще находится в процессе становления и включает в себя целый ряд технологий, в том числе RDF, OWL, SPARQL, обзору и анализу которых посвящена настоящая статья. Поставленная исследовательская задача представляется весьма актуальной.

Во-первых, более глубокое понимание природы подобных семантических инструментов позволяет совершенствовать механизмы обработки больших данных в самых разных предметных областях, включая медицину, экономику, военно-промышленный комплекс, науку, образование и многое другое, а также улучшать процесс подбора и интерпретации запрашиваемой пользователем информации в поисковых системах. Все это способствует высвобождению человеческих и временных ресурсов, отводимых на работу с данными, и автоматизирует обмен информацией между различными устройствами, тем самым претворяя в жизнь концепцию Интернета вещей.

Во-вторых, большинство статей отечественных авторов за последние несколько лет, посвященных семантическому вебу, раскрывают лишь некоторые его аспекты, не затрагивая проблемы и перспективы их развития в комплексе, либо предлагают изучение вопросов использования подобных технологий в том или ином отдельно взятом виде профессиональной деятельности. Например, от онлайн-образования [2], сферы финансов и инвестирования [3] до археологических изысканий [4] или геоаналитики улично-дорожных сетей [5]. Впрочем, широта и разнохарактерность применения семантического инструментария и связанного с ним онтолого-ориентированного подхода лишь подчеркивает необходимость его углубленного и всестороннего рассмотрения.

В-третьих, выбранная тема получает особое значение в контексте проблемы импортозамещения. Изучение технологий семантического веба поможет создать российские аналоги зарубежных веб-сервисов, обеспечивая независимость от иностранных компаний и повышая информационную безопасность нашей страны. Основные выводы и положения настоящего исследования могут быть взяты за теоретическую основу при разработке отечественных систем структуризации и описания данных, доступных для машинной обработки, а также профильных лекционных курсов в высших учебных заведениях.

Что касается методологии исследования, то в статье применена комплексная методологическая база, которая подразумевает использование преимущественно общенаучных методов, в частности, системного и аналитического, с целью систематизации и обобщения релевантных теоретических и научно-практических изысканий как отечественных, так и зарубежных авторов. Также в настоящей работе немаловажную роль играет индуктивный метод: поочередное рассмотрение ключевых особенностей ряда основополагающих семантических технологий способствует формированию более целостного взгляда на проблему их текущих ограничений и перспектив дальнейшего развития.

Обзор основных семантических технологий

Отправной точкой в развитии семантического веба можно по праву назвать создание языка XML (eXtensible Markup Language) в 1996 году – расширяемого языка разметки, предназначенного для создания, хранения и передачи структурированной информации в машиночитаемом формате. XML-документ предполагает обязательное наличие так называемого пролога, с указанием версии языка, и корневого элемента, который в свою очередь содержит произвольное количество вложенных элементов, состоящих из открывающих и закрывающих тегов с соответствующими им значениями и атрибутами, наименование которых определяет пользователь. При всех достоинствах указанного языка разметки, в том числе его независимости от платформы, расширяемости и ориентированности на быстрый и надежный обмен информацией между программами и устройствами, он имеет существенный недостаток, а именно способен описывать лишь структуру данных, а не заложенный в них смысл.

По этой причине основным компонентом семантического веба стала технология, получившая название RDF (Resource Description Framework). Она не зависит от предметной области и поддерживает графовую модель данных, предполагающую наличие триплета из субъекта (сущности или ресурса), объекта и связи между ними, называемой предикатом. Такого рода триплет позволяет машине формировать логические утверждения из предоставленной в ее распоряжение информации.

Если модель RDF описывает субъектно-объектные отношения, то взаимосвязи между сущностями оказались представлены в его расширении, – примитивном языке онтологий RDF Schema. Он обеспечивает возможность задать семантику конкретной области знаний, определив словарь терминов посредством организации иерархии классов и свойств.

Понятие онтологии – одно из ключевых в концепции семантического веба. Оно подразумевает иерархически структурированный, формализованный набор знаний о той или иной предметной области, включающий в себя описание характерных для неё классов, подклассов и их экземпляров, а также свойств и зависимостей между ними. Для этих целей рабочие группы W3C консорциума разработали отдельный, более выразительный и эффективный, чем RDFS, язык, – OWL (Ontology Web Language), обладающий широким набором инструментов для формирования сложных онтологий и основанный на дескрипционных логиках. Это «семейство логик, созданное специально для представления терминологически знаний, которое имеет собственную богатую историю и особенности, хорошо известные в соответствующем профессиональном обществе» [\[6, с. 88\]](#).

На сегодняшний день остается актуальной вторая версия языка OWL, которая призвана расширить его первоначальный стандарт и предоставить дополнительные возможности для моделирования сложных отношений между данными, например, такие как эквивалентность, равенство и логические комбинации классов, их непересекаемость и непротиворечивость, специальные характеристики свойств и поддержка логического вывода. Кстати, существует несколько диалектов OWL, в том числе OWL2 DL (Description Logics), обеспечивающий вычислительную эффективность и результативную поддержку логического вывода на основе дескрипционных логик, но при этом не имеющий полной совместимости с RDF; и OWL2 Full – напротив, с максимальной выразительной мощностью и абсолютной структурной и семантической совместимостью с RDF, однако без гарантий эффективной поддержки логического вывода [\[6, с. 94\]](#).

Также нельзя не упомянуть о еще одном важном инструменте: по аналогии с языком SQL в семантических базах данных существует свой язык запросов для эффективного взаимодействия с RDF-графами, – это SPARQL (SPARQL Protocol and RDF Query Language). Он имеет SQL-подобный синтаксис и поддерживает различные виды запросов, включая выборку данных в хранилищах триплетов, создание новых записей, их модификацию и удаление. SPARQL использует шаблоны для формирования запросов, а также позволяет их объединять и разделять, что способствует более гибкой настройке поиска и обработки данных и созданию сложных запросов.

Семантический веб: ограничения и точки роста

Рассмотренные выше технологии обеспечивают эффективную работу с данными в семантическом вебе, делая их более информативными, связанными и доступными для машинной обработки. Однако, несмотря на значительный прогресс в этой области, у них существует ряд проблемных моментов, требующих решения, равно как и многообещающие перспективы дальнейшего совершенствования и развития.

Одна из, возможно, не столь очевидных проблем, препятствующих широкому распространению семантических технологий, – это их сложность и высокий порог входа, а также нехватка интуитивно понятных инструментов для работы. Существующие стандарты семантической сети следует упростить, сделав их более доступными и привлекательными для начинающих разработчиков [\[7, с. 14\]](#). А интерфейсы должны быть

простыми и удобными в использовании как для специалиста, так и для рядового пользователя, и отвечать критериям так называемого юзабилити [8, с. 14587]. Некоторые отечественные авторы предлагают в этой связи собственные наработки – например, редактор запросов на языке SPARQL. Существующие решения не отличаются выразительностью интерфейса и не обладают функциями дополняющих подсказок и визуализации запросов [9, с. 87], что и было предложено специалистами из Пермского государственного национального исследовательского университета в собственной версии такого редактора.

Кроме того, применяемые в семантическом вебе языки имеют множество диалектов, что существенно затрудняет их изучение и стандартизацию. Очевидно, необходимо введение единого общего протокола, поскольку агрегация данных из различных источников остается сложной задачей из-за большого разнообразия их форматов и структур.

Еще одна проблема – достоверность и актуальность данных: быстро меняющаяся информация требует постоянного обновления и синхронизации с первоисточниками, что представляется весьма трудоемким процессом, производительность которого также нуждается в оптимизации. Не меньшую роль играет и релевантность агрегированной информации поставленной пользователем задаче, а также контекстуальная обусловленность логических выводов. Как отмечают зарубежные исследователи, стандарты семантического веба действительно облегчают обмен данными и их интеграцию, но полностью раскрывают свой потенциал лишь благодаря возможности адаптации данных под различные цели [10, с. 3389].

Наконец, одной из ключевых проблем для любого типа данных, включая семантический веб, является их безопасность, в том числе целостность и конфиденциальность. Например, остается до сих пор нерешенным вопрос предотвращения атак, основанных на инъекциях вредоносного кода в запрос на языке SPARQL, а имеющиеся предложения по защите семантических сетей не представляются эффективными [11, с. 41]. Возможное решение состоит в создании специальных криптографических средств, разработкой которых уже занимаются, в частности, представители Южного федерального университета.

Говоря о перспективах, отметим, что основными на сегодняшний день направлениями для дальнейшей прикладной интеграции семантического веба являются такие области, как большие данные, машинное обучение, искусственный интеллект и Интернет вещей. Концепции и технологии семантической сети активно интегрируются в сети связанных объектов для решения проблем совместимости, ограничивающих развитие Интернета вещей, что в конечном счете привело к появлению нового термина, известного как семантическая сеть вещей (Semantic Web of Things – SWoT) [12, с. 265]. Автоматическое извлечение и агрегирование информации и возможность создания интеллектуальных агентов, способных предоставлять персонализированные услуги, адаптированные к потребностям пользователя, а также интероперабельность с широким спектром устройств открывает путь к созданию интеллектуальной автономной среды, устройства и приборы в которой смогут свободно обмениваться данными и взаимодействовать друг с другом на семантическом уровне.

Заключение

Семантический веб – ключевой элемент Интернета будущего, где знания окажутся равнодоступными и для людей, и для машин – с учетом их контекста и смысловой

составляющей. Используемые для этого технологии, в том числе XML, RDF, RDFS, OWL, SPARQL, обеспечивают возможности персонализированного поиска, агрегирования сведений из разных источников, ускоренной и эффективной обработки больших объемов данных, обмена информацией между различными устройствами и находят применение в самых разных областях, таких как образование, наука, здравоохранение, транспорт и финансы. Решение существующих проблем семантического веба и реализация его перспективных направлений развития позволят достичь нового уровня взаимодействия с данными, что открывает широкие возможности для последующих технологических инноваций.

Библиография

1. Berners-Lee, T. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. – NY: HarperCollins Publishers, 2000. – 246 pp.
2. Шполянская И. Ю., Середкина Т. А. Технологии Semantic Web в организации поддержки онлайн обучения. // Системный анализ в проектировании и управлении. – Вып. XXIV, № 3, 2020. – С. 343-350. doi:10.18720/SPBPU/2/id20-231
3. Иващук Ю. С., Орлянская Н. П., Тешев В. А. Разработка базы знаний инвестиционной деятельности на основе онтологического моделирования. // Вестник Адыгейского государственного университета. – Серия 5: Экономика, № 2 (320), 2023. – С. 90-98. doi:10.53598/2410-3683-2023-2-320-90-98
4. Петров И. Д., Терехова Ю. В. Расширение логической модели предметной области археологических исследований с использованием онтологий. // Успехи в химии и химической технологии. – Вып. 34, № 6 (229), 2020. – С. 133-135.
5. Смирнов А. В., Тесля Н. Н. Онтолого-ориентированная геоаналитика для определения мест концентрации ДТП на участках улично-дорожной сети. // Труды Кольского научного центра РАН. Серия: Технические науки. – Вып. 14, № 7, 2023. – С. 79-85. doi:10.37614/2949-1215.2023.14.7.008
6. Антониоу, Г. Семантический веб / Г.Антониоу, П. Грос, Ф. ван Хармелен, Р. Хоекстра; пер. с англ. Т. Шульга. – М.: ДМК Пресс, 2016. – 240 с.
7. Hogan, A. The Semantic Web: Two decades on. // Semantic Web. – Vol. 11, pp. 169-185, 2020. – doi:10.3233/SW-190387, 2020.
8. Hassan, B. Towards Semantic Web: Challenges and Needs. // International Journal Of Engineering And Computer Science. – Vol. 4, issue 10, Oct 2015. – pp. 14585-14588. doi:10.48550/arXiv.2105.02708
9. Турова И. А., Постановов И. С. Разработка интеллектуального редактора SPARQL-запросов. // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – Вып. 19, № 4, 2021. – С. 85-95.
10. Jat, A. Semantic web technologies: challenges and applications. // Journal of Critical Reviews. – Vol. 7, Issue 17, 2020. – pp. 3388-3390. doi:10.31838/jcr.07.17.417.
11. Чудинов П. Ю., Бабенко Л. К., Рогозов Ю. И. Анализ проблем защиты информации в семантических сетях. // Известия Южного федерального университета. Технические науки. – № 5 (229), 2022. – С. 37-47.
12. F. Z. Amara, M. Hemam, M. Djeddar and M. Maimor. Semantic Web and Internet of Things: Challenges, Applications and Perspectives. // Journal of ICT Standardization. – Vol. 10. No. 2, 2022. – pp. 261-291. doi: 10.13052/jicts2245-800X.1029

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Рецензируемая статья посвящена изучению эволюции технологий семантического веба, в ней рассматриваются проблемы и перспективы развития таких технологий.

Методология исследования базируется на использовании общенаучных методов, в частности, системного подхода и аналитического метода, которые применяются с целью систематизации и обобщения релевантных теоретических и научно-практических изысканий отечественных и зарубежных авторов, а также индуктивного метода для поочередного рассмотрения особенностей различных семантических технологий.

Актуальность работы авторы справедливо связывают с необходимостью познания природы применяемых семантических инструментов, комплексного представления проблем и перспектив развития технологий семантического веба и потребностью в создании российских аналогов зарубежных веб-сервисов.

Научная новизна рецензируемого исследования состоит в выявлении и систематизации проблемы и перспектив эволюции технологий семантического веба.

Структурно в статье выделены следующие разделы: Актуальность темы и методология исследования, Обзор основных семантических технологий, Семантический веб: ограничения и точки роста, Заключение и Библиография.

В публикации приведен обзор следующих технологий: XML (eXtensible Markup Language), RDF (Resource Description Framework), OWL (Ontology Web Language), OWL2 DL (Description Logics), OWL2 Full, SPARQL (SPARQL Protocol and RDF Query Language). Среди проблем, препятствующих широкому распространению семантических технологий, названы такие как: их сложность и высокий порог входа, нехватка интуитивно понятных инструментов для работы; наличие множества диалектов, затрудняющих их изучение и стандартизацию; требование постоянного обновления и синхронизации с первоисточниками в условиях быстро меняющейся информации. В Заключении авторы приходят к выводу о том, что технологии XML, RDF, RDFS, OWL, SPARQL обеспечивают возможности персонализированного поиска, агрегирования сведений из разных источников, ускоренной и эффективной обработки больших объемов данных, обмена информацией между различными устройствами и находят применение в самых разных областях, таких как образование, наука, здравоохранение, транспорт и финансы. Естественно, публикации не содержит решений всех проблем технологий семантического веба, но проведенная работа по их обобщению и систематизации представляется необходимой. Говоря о перспективах развития рассматриваемых технологий, авторы отмечают основные направления дальнейшей прикладной интеграции семантического веба: большие данные, машинное обучение, искусственный интеллект и Интернет вещей. Библиографический список включает 21 источник – публикации отечественных и зарубежных ученых на русском и английском языках по рассматриваемой теме, на которые в тексте имеются адресные ссылки, подтверждающие наличие апелляции к оппонентам.

Рецензируемый материал соответствует направлению журнала «Программные системы и вычислительные методы», отражает результаты проведенной авторами работы, содержит элементы научной новизны и практической значимости, может вызвать интерес у читателей, рекомендуется к опубликованию.