

Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет информационных технологий и управления
Кафедра интеллектуальных информационных технологий

РЕФЕРАТ

по дисциплине «Обработка изображений в интеллектуальных системах»
на тему

СЕМАНТИЧЕСКАЯ ПАУТИНА

Студент гр. 121701
Проверил

Р. В. Липский
С. А. Самодумкин

Минск 2023

СОДЕРЖАНИЕ

1	Семантическая паутина	2
1.1	Онтологии	2
1.2	Семантические сети	3
1.2.1	Структура семантических сетей	5
1.2.2	Классификация семантических сетей	5
1.2.3	Семантические отношения	6
1.3	Универсальный идентификатор ресурса – URI	7
1.3.1	Структура URI	8
1.3.2	Недостатки URI	9
1.4	Архитектура семантической паутины	10
1.5	SPARQL	14
1.6	Проекты, реализованные при помощи семантической паутины	16
1.6.1	Дублинское ядро	16
1.6.2	RSS 0.90/1.0	16
1.6.3	FOAF	17
1.6.4	DBPedia	17
1.7	Web 3.0	17
1.8	Критика семантической паутины	18
	Список использованных источников	19

1 СЕМАНТИЧЕСКАЯ ПАУТИНА

В сентябре 1998 года Тим Бернерс-Ли, создатель Всемирной паутины, впервые представил термин «Семантическая паутина» как следующий этап развития сети. Он также предложил использовать термин «гигантский глобальный граф» (GGG) в качестве синонима для описания семантической паутины, аналогично существующему термину WWW (World Wide Web). Концепция семантической паутины была принята и продвигается консорциумом Всемирной паутины.

Семантическая паутина, также известная как Semantic Web, представляет собой глобальную сеть, доступную для всех, которая создается путем установления стандартов представления информации в формате, удобном для автоматической обработки компьютерами.

Сравнивая с обычной Всемирной паутиной, основанной на HTML-страницах, семантическая паутина отличается тем, что информация представлена в виде машинно-читаемых элементов, называемых узлами семантической сети, которые опираются на онтологии. Это позволяет программам-клиентам получать и обрабатывать информацию из Интернета в виде утверждений вида «предмет - вид взаимосвязи - другой предмет» и делать логические выводы на их основе. Семантическая паутина функционирует параллельно с обычной Всемирной паутиной, используя протокол HTTP и идентификаторы ресурсов URI.

1.1 Онтологии

Онтология – это попытка всеобъемлющей и подробной формализации некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из структуры данных, содержащей все релевантные классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области.

Онтологии используются в процессе программирования как форма представления знаний о реальном мире или его части. Основные сферы применения — моделирование бизнес-процессов, семантическая паутина, искусственный интеллект.

Современные онтологии строятся по большей части одинаково, независимо от языка написания. Обычно они состоят из экземпляров, понятий, атрибутов и отношений:

– Экземпляры (англ. instances) – это объекты, основные нижнеуровневые компоненты онтологии; могут представлять собой как физические объекты (люди, дома, планеты), так и абстрактные (числа, слова). Строго говоря, онтология может обойтись и без конкретных объектов, однако, одной из главных целей онтологии является классификация таких объектов,

поэтому они также включаются.

– Понятия (англ. concepts) или классы (англ. classes) – абстрактные группы, коллекции или наборы объектов. Они могут включать в себя экземпляры, другие классы либо же сочетания и того, и другого. Пример: Понятие «люди», вложенное понятие «человек». Чем является «человек» — вложенным понятием, или экземпляром (индивидом) — зависит от онтологии. Понятие «индивиды», экземпляр «индивид». Классы онтологии составляют таксономию — иерархию понятий по отношению вложения.

– Объекты в онтологии могут иметь атрибуты. Каждый атрибут имеет по крайней мере имя и значение и используется для хранения информации, которая специфична для объекта и привязана к нему. Например, объект Автомобиль-модели-А имеет такие атрибуты, как: название, число дверей, двигатель и т.д. Значение атрибута может быть сложным типом данных.

– Важная роль атрибутов заключается в том, чтобы определять отношения (зависимости) между объектами онтологии. Обычно отношением является атрибут, значением которого является другой объект. Предположим, что в онтологии автомобилей присутствует два объекта — автомобиль Автомобиль-модели-А и Автомобиль-модели-Б. Пусть Автомобиль-модели-Б это модель-наследник Автомобиль-модели-А, тогда отношение между Автомобиль-модели-А и Автомобиль-модели-Б определим как атрибут «isSuccessorOf» со значением «Автомобиль-модели-А» для объекта Автомобиль-модели-Б (следует заметить, что в языках описания онтологий существуют предопределенные отношения наследования).

1.2 Семантические сети

Семантическая сеть — информационная модель предметной области, имеет вид ориентированного графа. Вершины графа соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними. Объектами могут быть: понятия, события, свойства, процессы. Таким образом, семантическая сеть — это один из способов представления знаний.

В названии соединены термины из двух наук: семантика в языкознании изучает смысл единиц языка, а сеть в математике представляет собой разновидность графа — набора вершин, соединённых дугами (рёбрами), которым присвоено некоторое число. В семантической сети роль вершин выполняют понятия базы знаний, а дуги (причем направленные) задают отношения между ними. Таким образом, семантическая сеть отражает семантику предметной области в виде понятий и отношений.

Неправильно приравнивать друг другу понятия «Семантическая сеть» (англ. Semantic Network) и «Семантическая паутина» (англ. Semantic Web).

Идея систематизации на основе каких-либо семантических отношений

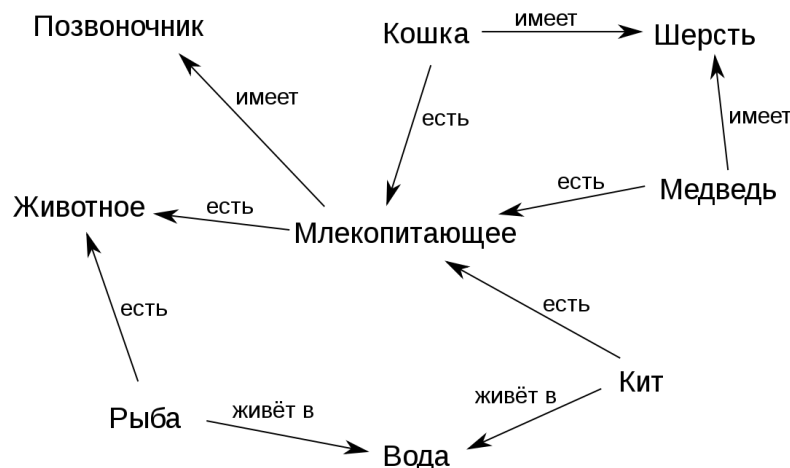


Рисунок 1.1 – Пример семантической паутины

предлагалась ещё учёными ранней науки. Примером этого может служить биологическая классификация Карла Линнея 1735 г. Если рассматривать её как семантическую сеть, то в данной классификации используется отношение подмножества, современное АКО (от англ. «A Kind Of», «разновидность»).

Прародителями современных семантических сетей можно считать экзистенциальные графы, предложенные Чарльзом Пирсом в 1909 г. Они использовались для представления логических высказываний в виде особых диаграмм. Пирс назвал этот способ «логикой будущего».

Важным направлением в исследовании сетей стали работы немецкого психолога Отто Зельца 1913 и 1922 гг. В них для организации структур понятий и ассоциаций, а также изучения методов наследования свойств он использовал графы и семантические отношения. Исследователи М. Андерсон (1973), Д. Норман (1975) и другие использовали эти работы для моделирования человеческой памяти и интеллектуальных свойств.

Компьютерные семантические сети были детально разработаны Ричардом Риченсом (англ.) в 1956 году в рамках проекта Кембриджского центра изучения языка по машинному переводу. Процесс машинного перевода подразделяется на 2 части: перевод исходного текста в промежуточную форму представления, а затем эта промежуточная форма транслируется на нужный язык. Такой промежуточной формой как раз и были семантические сети. В 1961 г. появилась работа Мастерман (англ.), в которой она, в частности, определяла базовый словарь для 15000 понятий. Эти исследования были продолжены Робертом Симмонсом (1966), Й. Уилксом (1972) и другими учёными.

Труды по семантическим сетям часто ссылаются на работу американ-

ского психолога Росса Квиллиана (Quillian) о «семантической памяти».

1.2.1 Структура семантических сетей

Математика позволяет описать большинство явлений в окружающем мире в виде логических высказываний. Семантические сети возникли как попытка визуализации математических формул. Основным представлением для семантической сети является граф. Однако не стоит забывать, что за графическим изображением непременно стоит строгая математическая запись и что обе эти формы отображают одно и то же.

Графическое представление

Основной формой представления семантической сети является граф. Понятия семантической сети записываются в овалах или прямоугольниках и соединяются стрелками с подписями — дугами (см. рис.). Это наиболее удобно воспринимаемая человеком форма. [источник не указан 4351 день] Её недостатки проявляются, когда мы начинаем строить более сложные сети или пытаемся учесть особенности естественного языка. Схемы семантических сетей, на которых указаны направления навигационных отношений, называют картами знаний, а их совокупность, позволяющая охватить большие участки семантической сети, атласом знания.

Математическая запись

В математике граф представляется множеством вершин V и множеством отношений между ними E . Используя аппарат математической логики, приходим к выводу, что каждая вершина соответствует элементу предметного множества, а дуга — предикату.

Лингвистическая запись

В лингвистике отношения фиксируются в словарях и в тезаурусах. В словарях в определениях через род и видовое отличие родовое понятие занимает определённое место. В тезаурусах в статье каждого термина могут быть указаны все возможные его связи с другими родственными по теме терминами. От таких тезаурусов необходимо отличать тезаурусы информационно-поисковые с перечнями ключевых слов в статьях, которые предназначены для работы дескрипторных поисковых систем.

1.2.2 Классификация семантических сетей

Для всех семантических сетей справедливо разделение по арности и количеству типов отношений.

По количеству типов отношений, сети могут быть однородными и неоднородными. Однородные сети обладают только одним типом отношений (стрелок), таковой является вышеупомянутая классификация биологических видов (с единственным отношением АКО).

В неоднородных сетях количество типов отношений больше одного. Классические иллюстрации данной модели представления знаний представляют именно такие сети. Неоднородные сети представляют больший интерес для практических целей, но и большую сложность для исследования. Неоднородные сети можно представлять как переплетение древовидных многослойных структур. Примером такой сети может быть Семантическая сеть Википедии.

Можно классифицировать семантические сети по арности: типичными являются сети с бинарными отношениями (связывающими ровно два понятия). Бинарные отношения очень просты и удобно изображаются на графе в виде стрелки между двух концептов. Кроме того, они играют исключительную роль в математике. На практике, однако, могут понадобиться отношения, связывающие более двух объектов — N-арные.

Классификация по размеру:

- Для решения конкретных задач, например, тех, которые решают системы искусственного интеллекта.
- Семантическая сеть отраслевого масштаба должна служить базой для создания конкретных систем, не претендуя на всеобщее значение.
- Глобальная семантическая сеть. Теоретически такая сеть должна существовать, поскольку всё в мире взаимосвязано. Возможно, когда-нибудь такой сетью станет Всемирная паутина.

1.2.3 Семантические отношения

Количество типов отношений в семантической сети определяется её создателем, исходя из конкретных целей. В реальном мире их число стремится к бесконечности. Каждое отношение является, по сути, предикатом, простым или составным. Скорость работы с базой знаний зависит от того, насколько эффективно реализованы программы обработки нужных отношений.

Иерархические отношения:

– отношения между объектом и множеством, обозначающим, что объект принадлежит этому множеству, называется отношением классификации (ISA). Говорят, что множество (класс) классифицирует свои экземпляры.[3] (пример: «Шарик является собакой» = Шарик является объектом типа собака). Иногда это отношение именуют также MemberOf, InstanceOf или подобным образом. Связь ISA предполагает, что свойства объекта наследуются от множества.

– Отношение между надмножеством и подмножеством (называется АКО — «A Kind Of», «разновидность»). (Пример: «собака является животным» = тип с именем собака является подтипом типа животные). Элемент подмножества называется гипонимом (собака), а надмножества — гиперо-

нимом (животное), а само отношение называется отношением гипонимии. Альтернативные названия — «SubsetOf» и «Подмножество». Это отношение определяет, что каждый элемент первого множества входит и во второе (выполняется ISA для каждого элемента), а также логическую связь между самими подмножествами: что первое не больше второго и свойства первого множества наследуются от второго. Отношение АКО (Род-Вид) часто используется для навигации в информационном пространстве.

— объект, как правило, состоит из нескольких частей, или элементов. Например, компьютер состоит из системного блока, монитора, клавиатуры, мыши и т. д. Важным отношением является HasPart, описывающее связь частей и целого — отношение меронимии.

Вспомогательные: в семантических сетях часто используются также следующие отношения:

- функциональные связи (определяемые обычно глаголами «производит», «влияет»...);
- количественные (больше, меньше, равно...);
- пространственные (далеко от, близко от, за, под, над...);
- временные (раньше, позже, в течение...);
- атрибутивные (иметь свойство, иметь значение);
- логические (И, ИЛИ, НЕ);
- лингвистические.

Этот список может сколь угодно продолжаться: в реальном мире количество отношений огромно.

1.3 Универсальный идентификатор ресурса – URI

URI — символьная строка, позволяющая идентифицировать какой-либо ресурс: документ, изображение, файл, службу, ящик электронной почты и т. д. Прежде всего, речь идёт о ресурсах сети Интернет и Всемирной паутины. URI предоставляет простой и расширяемый способ идентификации ресурсов. Расширяемость URI означает, что уже существуют несколько схем идентификации внутри URI, и ещё больше будет создано в будущем.

URI является либо указателем ресурса URL, либо именем ресурса URN, либо одновременно обоими.

URL — это URI, который, помимо идентификации ресурса, предоставляет ещё и информацию о местонахождении этого ресурса. А URN — это URI, который только идентифицирует ресурс в определённом пространстве имён (и, соответственно, в определённом контексте), но не указывает его местонахождение. Например, URN urn:ISBN:0-395-36341-1 — это URI, который указывает на ресурс (книгу) 0-395-36341-1 в пространстве имён ISBN, но, в отличие от URL, URN не указывает на местонахождение этого ресурса: в нём не сказано, в каком магазине её можно купить или на каком

сайте скачать. Впрочем, в последнее время появилась тенденция говорить просто URI о любой строке-идентификаторе, без дальнейших уточнений. Так что, возможно, термины URL и URN скоро уйдут в прошлое.

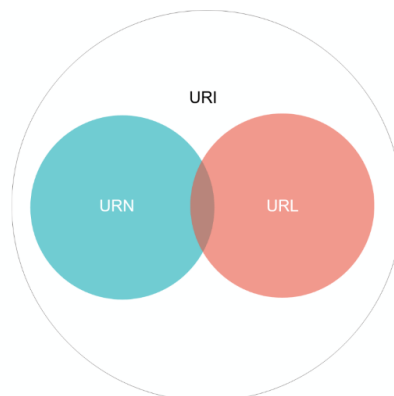


Рисунок 1.2 – Соответствие URL и URI

Поскольку URI не всегда указывает на то, как получить ресурс, в отличие от URL, а только идентифицирует его, это даёт возможность описывать с помощью RDF (Resource Description Framework) ресурсы, которые не могут быть получены через Интернет (например, личность, автомобиль, город и проч.).

1.3.1 Структура URI

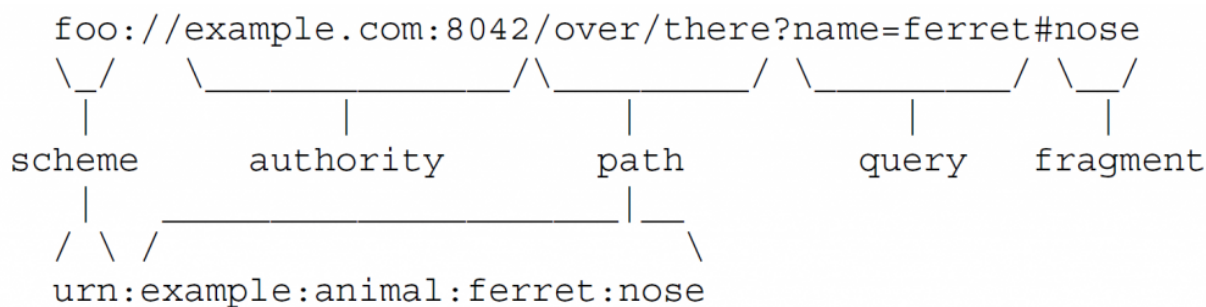


Рисунок 1.3 – Структура URI и URL

- схема – схема обращения к ресурсу (часто указывает на сетевой протокол), например http, ftp, file, ldap, mailto, urn

- иерархическая часть – содержит данные, обычно организованные в иерархической форме, которые, совместно с данными в неиерархическом компоненте запрос, служат для идентификации ресурса в пределах видимости URI-схемы. Обычно иер-часть содержит путь к ресурсу (и, возможно, перед ним, адрес сервера, на котором тот располагается) или идентификатор ресурса (в случае URN).

– запрос – содержит данные, организованные в неиерархической форме, которые, вместе с данными в иерархическом компоненте пути, служит, чтобы идентифицировать ресурс в рамках схемы URI.

– фрагмент – позволяет косвенно идентифицировать вторичный ресурс посредством ссылки на первичный и указанием дополнительной информации. Вторичный идентифицируемый ресурс может быть некоторой частью или подмножеством первичного, некоторым его представлением или другим ресурсом, определённым или описанным таким ресурсом.

этот необязательный компонент URI описан выше. фрагмент (тоже необязательный компонент) RFC 3986:

позволяет косвенно идентифицировать вторичный ресурс посредством ссылки на первичный и указанием дополнительной информации. Вторичный идентифицируемый ресурс может быть некоторой частью или подмножеством первичного, некоторым его представлением или другим ресурсом, определённым или описанным таким ресурсом.

1.3.2 Недостатки URI

URL стал фундаментальным нововведением в Интернете, поэтому принципы URI документально закреплялись так, чтобы обеспечить полную совместимость с URL. Отсюда появился и большой недостаток URI, пришедший как наследство от URL. В URI, как и в URL, можно использовать только ограниченный набор латинских символов и знаков препинания (даже меньший, нежели в ASCII). Иными словами, если мы захотим использовать в URI символы кириллицы, или иероглифы, или, скажем, специфические символы французского языка, то нам придётся кодировать URI. Например, строка вида: `https://example.com/Кириллица` кодируется в URL как: `https://example.com/%D0%9A%D0%B8%D1%80%D0%B8%D0%BB...`

Поскольку такому преобразованию подвергаются буквы всех алфавитов, кроме используемой в английском языке латиницы, то URI со словами на других языках (даже европейских) утрачивают способность восприниматься людьми. А это входит в грубое противоречие с принципом интернационализма, провозглашаемого всеми ведущими организациями Интернета, включая W3C и ISOC. Эту проблему призван решить стандарт IRI (англ. Internationalized Resource Identifier) — международных идентификаторов ресурсов, в которых можно было бы без проблем использовать символы Юникода, и которые не ущемляли бы права других языков. Хотя заранее сложно сказать, смогут ли когда-либо идентификаторы IRI заменить URI, имеющие столь широкое употребление.

Ещё одной интересной вариацией URI является расширяемый идентификатор ресурса XRI (англ. Extensible Resource Identifier), разработанный организацией OASIS. Этот формат стремится создавать идентификаторы,

которые были бы совершенно независимы от контекста, то есть не зависели бы ни от протокола, ни от домена, ни от пути, ни от приложения, ни от платформы — были бы совершенно независимыми.

Также и сам создатель URI, Тим Бернерс-Ли, говорил, что система доменных имён, лежащая в основе URL, — плохое решение, навязывающее ресурсам иерархическую архитектуру, мало подходящую для гипертекстового веба.

1.4 Архитектура семантической паутины

Семантическая паутина включает в себя семейство стандартов для языков описания, которые составляют ее техническую основу. Это семейство стандартов включает XML, XML Schema, RDF, RDF Schema, OWL и другие. Каждый из этих языков описания представляет собой уровень абстракции, который постепенно повышается от более низкого к более высокому уровню, обеспечивая возможность более сложного и точного описания информации в семантической паутине.

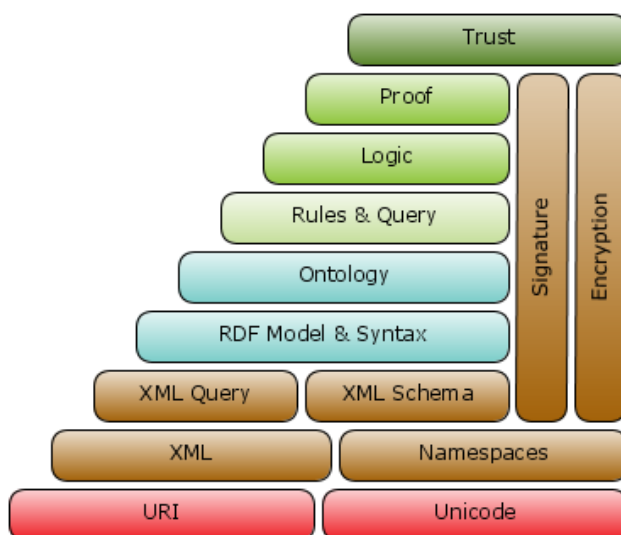


Рисунок 1.4 – Архитектура семантической паутины

– XML предоставляет синтаксис для определения структуры документа, подлежащего машинной обработке. Синтаксис XML не несёт семантической нагрузки. XML (аббр. от англ. eXtensible Markup Language) — «расширяемый язык разметки». Рекомендован Консорциумом Всемирной паутины (W3C). XML разрабатывался как язык с простым формальным синтаксисом, удобный для создания и обработки документов как программами, так и человеком, с акцентом на использование в Интернете. Язык называется расширяемым, поскольку он не фиксирует разметку, используемую в документах: разработчик волен создать разметку в соответствии с потребностями конкретной области, будучи ограниченным

лишь синтаксическими правилами языка.

Пример XML-документа:

```
<?xml version="1.0"?>
<greeting>Hello, world!</greeting>
```

– XML Schema определяет ограничения на структуру XML-документа. Стандартный синтаксический анализатор языка XML в состоянии проверить произвольный XML-документ на соответствие его структуры так называемой схеме документа, описанной в XML Schema.

Пример схемы XML Schema:

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="country">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="country_name" type="xs:string"/>
        <xs:element name="population" type="xs:decimal"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Пример документа, соответствующего этой схеме:

```
<?xml version="1.0" encoding="utf-8"?>
<country>
  <country_name>France</country_name>
  <population>59.7</population>
</country>
```

– RDF представляет собой простой способ описания экземплярных данных в формате субъект-отношение-объект (рис. 1.5), в котором в качестве любого элемента этой тройки используются только идентификаторы ресурсов (за исключением объекта, которому разрешено быть литералом). Существует стандартизованное отображение этих троек на XML-документы предопределённой структуры (то есть консорциумом W3 определена схема XML-документов, содержащих RDF-описания), а также на другие форматы представления (например, в нотацию N3).

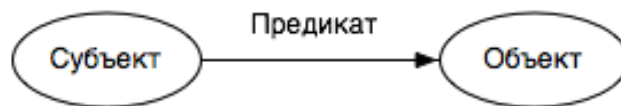


Рисунок 1.5 – Пример RDF-тройки

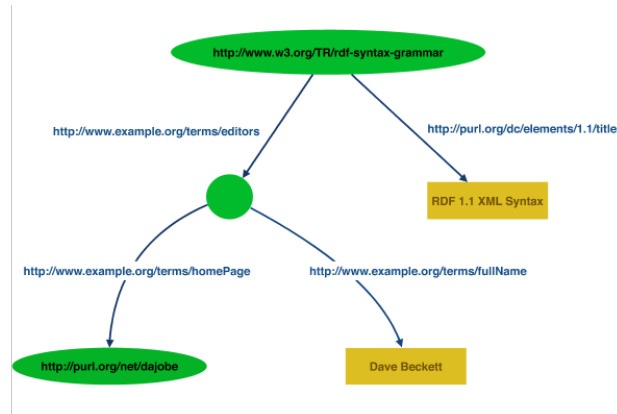


Рисунок 1.6 – Пример RDF-текста

RDF/XML — это заданный консорциумом W3C синтаксис выражения (т.н. сериализации) графа RDF в виде документа XML. Согласно определению W3C, «RDF/XML — это нормативный синтаксис записи RDF».

Пример RDF/XML-документа:

```
<?xml version="1.0"?>

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:si="https://www.w3schools.com/rdf/">

  <rdf:Description rdf:about="https://www.w3schools.com">
    <si:title>W3Schools</si:title>
    <si:author>Jan Egil Refsnes</si:author>
  </rdf:Description>

</rdf:RDF>
```

В данном RDF документе говорится о веб-сайте W3Schools. Документ содержит информацию о некоторых аспектах этого сайта, представленных в виде RDF триплетов.

В документе определены два пространства имен: "rdf" (<http://www.w3.org/1999/02/22-rdf-syntax-ns>) и "si" (<https://www.w3schools.com/rdf/>).

Внутри `<rdf:RDF>` элемента определено `<rdf:Description>`, который описывает ресурс с идентификатором `"https://www.w3schools.com"`. Затем внутри `<rdf:Description>` указаны два свойства: `<si:title>` с значением `"W3Schools представляющее заголовок (название) сайта"`. `<si:author>` с значением `"Jan Egil Refsnes представляющее автора сайта"`.

RDF/JSON представляет набор RDF-троек в виде вложенных структур данных. Каждый уникальный субъект в наборе троек представлен в виде ключа в JSON-объекте (также известном как ассоциативный массив, словарь или хэш-таблица). Значение каждого ключа - это объект, ключи которого являются URI свойств, связанных с каждым субъектом. Значение каждого ключа свойства - это массив объектов, представляющих значение каждого свойства.

Пример RDF/JSON-документа:

```
{
  "http://example.org/about" :
  {
    "http://purl.org/dc/elements/1.1/title": [
      {
        "type" : "literal" ,
        "value" : "Anna's Homepage."
      } ]
  }
}
```

В данном RDF/JSON-документе содержится описание ресурса с идентификатором `"http://example.org/about"`. Описание включает свойство `"http://purl.org/dc/elements/1.1/title"` которое имеет значение `"Anna's Homepage."`.

- RDF Schema описывает набор атрибутов (здесь их точнее назвать отношениями), таких, как `rdfs:Class`, для определения новых типов RDF-данных. Языком поддерживается также отношение наследования типов `rdfs:subClassOf`.

- OWL расширяет возможности по описанию новых типов (в частности, добавлением перечислений), а также позволяет описывать новые типы данных RDF Schema в терминах уже существующих (например, определять тип, являющийся пересечением или объединением двух существующих).

- Микроданные (HTML microdata) — международный стандарт семантической разметки HTML-страниц, с помощью атрибутов, описывающих смысл информации, содержащейся в тех или иных HTML-элементах. Такие

атрибуты делают контент страниц машиночитаемым, то есть позволяют в автоматическом режиме находить и извлекать нужные данные.

Форматы описания метаданных в Семантической паутине предполагают проведение логического вывода на этих метаданных, и разрабатывались с оглядкой на существующие математические формализмы в этой области. Формализм, лежащий в основе формата, даёт возможность делать заключения о свойствах программ, обрабатывающих данные в этом формате.

Особенно сильно это относится к языку OWL. Базовым формализмом для него являются дескрипционные логики, а сам язык разбит на три вложенных подмножества (в порядке вложенности): OWL Lite, OWL DL и OWL Full. Доказано, что логический вывод на метаданных с выразительностью OWL Lite выполняется за полиномиальное время (другими словами, задача вывода принадлежит к классу P). OWL DL описывает максимальное обладающее разрешимостью подмножество дескрипционных логик, но некоторые запросы по таким данным могут требовать экспоненциального времени выполнения. OWL Full реализует все существующие конструкторы дескрипционных логик за счёт отказа от обязательной разрешимости запросов.

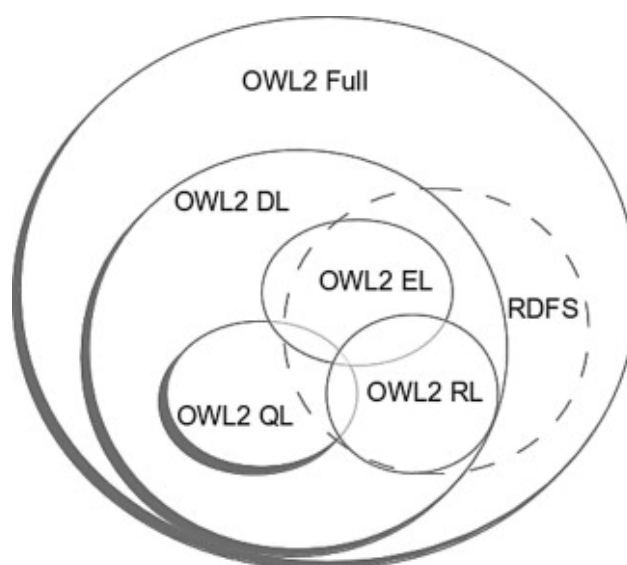


Рисунок 1.7 – Диаграмма Эйлера для семейства языков OWL

Простая структура предикатов языка RDF, в свою очередь, позволяет использовать при его обработке опыт из теорий логических баз данных, логики предикатов и т. д.

1.5 SPARQL

SPARQL (рекурсивный акроним от англ. SPARQL Protocol and RDF Query Language) — язык запросов к данным, представленным по модели RDF, а также протокол для передачи этих запросов и ответов на них.

SPARQL является рекомендацией консорциума W3C и одной из технологий семантической паутины. Предоставление SPARQL-точек доступа (англ. SPARQL-endpoint) является рекомендованной практикой при публикации данных во всемирной паутине

Как и в SQL, SPARQL декларирует не только чтение, но и манипуляцию данными. Стандарт SPARQL Update описывает команды INSERT и DELETE для добавления и удаления триплетов в хранилище. Однако, обе эти команды оперируют той или иной выборкой, так что основа SPARQL — это команда SELECT.

Общая схема SPARQL-запроса SELECT выглядит так:

PREFIX

- # префиксные объявления - служат для указания сокращений URI,
- # используемых в запросе.

FROM ...

- # источники запроса - определяют какие RDF-графы запрашиваются.

SELECT ...

- # состав результата - определяет возвращаемые элементы данных.

WHERE {...}

- # шаблон запроса - определяет, что запрашивать из набора данных.

ORDER BY ...

- # модификаторы запроса - ограничивают, упорядочивают,
- # преобразуют результаты запроса.

SPARQL позволяет пользователям писать глобально однозначные запросы. Например, следующий запрос возвращает имена и адреса электронной почты каждого человека в мире:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?name ?email
```

```
WHERE {
```

```
  ?person a foaf:Person.
```

```
  ?person foaf:name ?name.
```

```
  ?person foaf:mbox ?email.
```

```
}
```

Приведённые параметры используются для описания человека, включённого в FOAF. Это иллюстрирует видение Семантической паутины как единой огромной базы данных. Каждый идентификатор в SPARQL, URI, глобально однозначен, в отличие от «email» или «e-mail», обычно используемых в SQL.

Язык SPARQL определяет четыре различных варианта запросов для различных целей:

- SELECT запрос – извлекает необработанные значения из точки доступа SPARQL и возвращает результаты в формате таблицы.
- CONSTRUCT запрос – извлекает информацию из точки доступа SPARQL в формате RDF и преобразовывает результаты к определённой форме.
- ASK запрос – формирует запрос типа Истина/Ложь.
- DESCRIBE запрос – получает описание RDF-ресурса. Реализация поведения DESCRIBE-запросов определяется разработчиком точки доступа SPARQL.

Каждая из этих форм запроса содержит блок WHERE для указания ограничений, хотя в случае запроса DESCRIBE этот блок не является обязательным.

1.6 Проекты, реализованные при помощи семантической паутины

1.6.1 Дублинское ядро

Одним из первых серьёзных и популярных проектов, основанных на принципах Семантической паутины, стал проект «Дублинское ядро» (англ. Dublin Core), реализуемый инициативной организацией Dublin Core Metadata Initiative (DCMI). Это открытый проект, цель которого — разработать стандарты метаданных, которые были бы независимы от платформ и подходили бы для широкого спектра задач. Конкретнее, DCMI занимается разработкой словарей метаданных общего назначения, стандартизирующих описания ресурсов в формате RDF.

1.6.2 RSS 0.90/1.0

Версии 0.90 и 1.0 формата RSS основаны на RDF. Информация в нём представляется как и в RDF, тройками субъект-отношение-объект. Необходимо отметить, что несмотря на то, что ему присущи многие недостатки Семантической паутины (например, дублирование информации), этот простейший формат быстро стал чрезвычайно популярным за счёт узкой категоризации подмножества используемых метаданных. Отличие RSS от RDF состоит в том, что субъектом тройки всегда является сайт-источник RSS-файла, а в качестве отношений используются самые очевидные свойства документов, имеющие отношение к часто обновляющимся источникам информации: дата написания, автор, постоянная ссылка, и т. д. Другими словами, RSS — узкоспециализированное подмножество RDF.

Заметим, что формат RSS версии 2.0, хотя и не является форматом, основанным на RDF, позволяет внедрение произвольного XML-содержимого, находящегося в собственных пространствах имён XML. Это позволяет

использовать RDF-описания также и в нём (используя пространство имён `rdf`).

1.6.3 FOAF

Проект «Friend of a Friend» («Друг друга») позволяет описывать отношение знакомства с помощью RDF. Любой его участник может идентифицировать себя уникальным образом с помощью URI (например, `mailto`-адресом электронной почты, адресом блога, и т. п.), создать свой профиль, используя предопределённые для FOAF отношения на языке RDF, и перечислить идентификаторы людей, которых этот участник знает. Это описание может обрабатываться автоматически; на его основе можно строить сети доверия, анализировать структуру социальных групп, и т. д.

1.6.4 DBpedia

DBpedia — проект, направленный на извлечение структурированной информации из данных, созданных в рамках проекта Wikipedia. DBpedia позволяет пользователям запрашивать информацию, основанную на отношениях и свойствах ресурсов Википедии, в том числе ссылки на соответствующие базы данных. Начат группой добровольцев из Свободного университета Берлина и Лейпцигского университета, в сотрудничестве с OpenLink Software, и впервые был опубликован в 2007 году. Проект DBpedia использует Resource Description Framework (RDF) для представления извлеченной информации. По состоянию на апрель 2010, базы данных DBpedia состоят из более чем 1 млрд единиц информации, из которых 257 млн были взяты из английской версии Википедии и 766 млн извлечены из версий на других языках.

1.7 Web 3.0

Web 3.0 — концепция, ориентированная на развитие интернет-технологий, которая подразумевает обеспечение понимания информации в Web в первую очередь человеком, а Web 3.0 обеспечит взаимодействие и понимание данных в Web компьютерными системами.

Одной из трактовок термина Web 3.0 является соотнесение его с семантической паутиной. Главная мысль этой концепции базируется на внедрении метаязыка, описывающего содержание сайтов для организации автоматического обмена между серверами. Описательные механизмы семантической паутины действительно разработаны (RDF, DAML, OIL, OWL), однако на этапе обработки и вывода информации появляется ряд проблем:

- необходимость дополнительных затрат для создания семантической версии каждого сайта, что делает технологию значительно менее доступной;

- отсутствие гарантии адекватного описания веб-мастерами собственных ресурсов;
- невозможность принятия единого формата описания свойств ресурсов в условиях существующей конкуренции из-за корпоративно-рекламной политики создателей ресурса и наличия широкого поля для манипуляций описательными механизмами.

1.8 Критика семантической паутины

В 2006 году журнал «IEEE Intelligent Systems» опубликовал новую статью Тима Бернерса-Ли «Semantic Web Revisited» («Семантическая паутина: пересмотр»), в которой автор называет описанный подход к организации информации в вебе «простой идеей, до сих пор в большой степени нереализованной», несмотря на все преимущества, предоставляемые Семантической паутиной в случае её внедрения.

На сегодняшний день нет общедоступных средств просмотра и прямого использования информации, предоставляемой сайтами в Семантическую паутину. Редкие образцы разрозненны, а программы-клиенты не выходят за уровень локальных исследовательских проектов отдельных энтузиастов.

Комментаторы указывают на различные причины, которые препятствуют активному развитию Семантической паутины, начиная с человеческого фактора (люди склонны избегать работы по поддержке документов с метаданными, открытыми остаются проблемы истинности метаданных, и т. д.), и заканчивая косвенным указанием Аристотеля на отсутствие очевидного способа деления мира на различимые концепты. Это ставит под сомнение возможность существования онтологии верхнего уровня, критической для Семантической паутины.

Необходимость описания метаданных так или иначе приводит к дублированию информации. Каждый документ должен быть создан в двух экземплярах: размеченным для чтения людьми, а также в машинно-ориентированном формате. Этот недостаток Семантической паутины был главным толчком к созданию так называемых микроформатов и языка RDF. Последний является вариантом языка RDF и отличается от него тем, что не определяет собственного синтаксиса, а предназначен для внедрения в XML-атрибуты XHTML-страниц. Кроме того, в самих стандартах HTML появляются семантические теги.

Лёгкость машинной обработки позволяет посреднику вроде поисковой системы выбирать, какую информацию предоставлять пользователю. Это делает возможным предоставление только необходимой части информации. В результате пользователь может получать информацию, не посещая сайт-источник, а посредник может выводить информацию, не давая ссылки на источник (или давая её при условии, что посредник узнает о переходе). Та-

ким образом уменьшается посещаемость сайта-источника, и увеличивается посещаемость посредника; а от посещаемости зависит действие Интернет-рекламы, обеспечивающей основное финансирование многих сайтов. Цитирование частей новостей веб-агрегаторами было предметом судебных разбирательств с разными исходами. Для сайтов, публикующих контент для привлечения читателей к размещаемой рядом рекламе, предоставление контента в машинно-читаемом виде не выглядит выгодным, однако для сайтов, чей контент сам по себе является информацией о рекламируемом объекте — например, для сайтов производителей товаров и услуг — предоставление информации о рекламируемых объектах в машинно-читаемом виде выгодно, так как способно облегчать её распространение и использование. Это облегчает автоматическую цензуру. Предпочтение поисковыми системами размеченных страниц делает неразмеченные частью (относительно) Глубокой паутины. Облегчается нахождение персональных данных и их связей с другой информацией (авторы, местонахождения, социальные связи).