Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

Российский экономический университет имени Г.В. Плеханова

Факультет: **Математической экономики и информатики**

Кафедра**Информатики**

**Направление: 230700** «**Прикладная информатика»**

**Магистерская программа: Информационная бизнес-аналитика**

**О Т Ч Е Т**

**по производственной и научно-исследовательской практикам**

Выполнил студент гр. 429

Шамсиев Артур Айдарович

###### \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

###### Проверили:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(должность, ФИО руководителя от предприятия)*

*\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*(оценка) (подпись)*

*\_\_\_\_\_\_\_\_\_\_\_\_\_*

МП *(дата)*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*(должность, ФИО руководителя от кафедры)*

*\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*(оценка) (подпись)*

*\_\_\_\_\_\_\_\_\_\_\_\_\_*

*(дата)*

**Москва**

**2015**

# Введение

В основе исследований в области искусственного интеллекта лежит подход, связанный со знаниями. Опора на знания — базовая парадигма искусственного интеллекта и экспертных систем. Как и многие фундаментальные научные категории (например, алгоритм, интеллект, деятельность и т. д.), понятие «знание» относится к интуитивно определяемым. В Большой Советской Энциклопедии дается следующее его толкование: «Знание — проверенный практикой результат познания действительности, верное ее отражение в сознании человека. Знания бывают житейскими, донаучными, художественными, научными (теоретическими и эмпирическими)». Знания о некоторой предметной области представляют собой совокупность сведений об объектах этой предметной области, их существенных свойствах и связывающих их отношениях, процессах, протекающих в данной предметной области, а также методах анализа, возникающих в ней ситуаций и способах разрешения, ассоциируемых с ними проблем.

В «Словаре русского языка» Ожегова знание определяется как «постижение действительности сознанием» и «совокупность сведений, познаний в какой-нибудь области». Интерпретация знаний как «совокупности сведений, образующих целостное описание, соответствующее некоторому уровню осведомленности об описываемом вопросе, предмете, проблеме и т. д.» дана в толковом словаре искусственного интеллекта. Семантические сети же являются моделью представления знаний.

Семантическая сеть – это [информационная модель](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) [предметной области](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D0%B4%D0%BC%D0%B5%D1%82%D0%BD%D0%B0%D1%8F_%D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%8C) (domain model), имеющая вид ориентированного [графа](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29), вершины которого соответствуют сущностям (объектам) предметной области, а дуги (рёбра) задают отношения между ними и называются предикатами. Объектами могут быть понятия, события, свойства, процессы, ресурсы, документы. Таким образом, семантическая сеть является одним из способов [представления знаний](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D0%B4%D1%81%D1%82%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B7%D0%BD%D0%B0%D0%BD%D0%B8%D0%B9). В названии соединены термины из двух наук: [семантика](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D0%BA%D0%B0) в языкознании изучает смысл единиц языка, а сеть в математике представляет собой разновидность [графа](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29) — набора вершин, соединённых ребрами которым присвоено некоторое число. В семантической сети роль вершин выполняют [понятия](https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%BD%D1%8F%D1%82%D0%B8%D0%B5) базы знаний, а дуги (причем направленные) задают [отношения](https://ru.wikipedia.org/wiki/%D0%9E%D1%82%D0%BD%D0%BE%D1%88%D0%B5%D0%BD%D0%B8%D0%B5) между ними. Таким образом, семантическая сеть отражает семантику предметной области в виде понятий и отношений.

Неправильно приравнивать друг другу понятия «Семантическая сеть» (Semantic Network) и «[Семантическая паутина](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BF%D0%B0%D1%83%D1%82%D0%B8%D0%BD%D0%B0)» (Semantic Web). Семантическая паутина является частным случаем семантической сети, а именно семантической сетью, вершины графа которой представляют собой веб ресурсы в сети интернет и описанные на языке RDF или OWL.

**История развития идеи**

Идея систематизации знаний на основе каких-либо семантических отношений предлагалась учеными еще до начала научно технической революции. Примером этого может служить [биологическая классификация](https://ru.wikipedia.org/wiki/%D0%91%D0%B8%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0) [Карла Линнея](https://ru.wikipedia.org/wiki/%D0%9B%D0%B8%D0%BD%D0%BD%D0%B5%D0%B9,_%D0%9A%D0%B0%D1%80%D0%BB) [1735 года.](https://ru.wikipedia.org/wiki/1735) Если рассматривать её как семантическую сеть, то в данной классификации используется отношение подмножества, современное [AKO](https://ru.wikipedia.org/w/index.php?title=AKO&action=edit&redlink=1) (от англ. «A Kind Of», «разновидность, экземпляр»). Такое отношение еще называют «IsA»

Прародителями современных семантических сетей можно считать [экзистенциальные графы](https://ru.wikipedia.org/w/index.php?title=%D0%AD%D0%BA%D0%B7%D0%B8%D1%81%D1%82%D0%B5%D0%BD%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%B3%D1%80%D0%B0%D1%84&action=edit&redlink=1), предложенные [Чарльзом Пирсом](https://ru.wikipedia.org/wiki/%D0%9F%D0%B8%D1%80%D1%81,_%D0%A7%D0%B0%D1%80%D0%BB%D1%8C%D0%B7_%D0%A1%D0%B0%D0%BD%D0%B4%D0%B5%D1%80%D1%81) в [1909 г](https://ru.wikipedia.org/wiki/1909). Они использовались для представления [логических](https://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B8%D0%BA%D0%B0) 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 [Дж. Андерсон](https://ru.wikipedia.org/w/index.php?title=%D0%90%D0%BD%D0%B4%D0%B5%D1%80%D1%81%D0%BE%D0%BD,_%D0%94%D0%B6%D0%BE%D0%BD_%D0%A0%D0%BE%D0%B1%D0%B5%D1%80%D1%82&action=edit&redlink=1) (1973), [Д. Норман](https://ru.wikipedia.org/wiki/%D0%9D%D0%BE%D1%80%D0%BC%D0%B0%D0%BD,_%D0%94%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%B4) (1975) и другие использовали эти работы для моделирования человеческой памяти и интеллектуальных свойств.

Компьютерные семантические сети были детально разработаны Ричардом Риченсом в [1956 году](https://ru.wikipedia.org/wiki/1956_%D0%B3%D0%BE%D0%B4) в рамках проекта Кембриджского центра изучения языка по [машинному переводу](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4). Процесс машинного перевода подразделяется на две 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](https://en.wikipedia.org/wiki/Margaret_Masterman), в которой она, в частности, определяла базовый словарь для 15000 понятий. Эти исследования были продолжены Робертом Симмонсом (1966), [Й. Уилксом](https://ru.wikipedia.org/wiki/%D0%A3%D0%B8%D0%BB%D0%BA%D1%81,_%D0%99%D0%BE%D1%80%D0%B8%D0%BA) (1972) и другими учёными.

Труды по семантическим сетям часто ссылаются на работу американского психолога Росса Квиллиана о «семантической памяти». В этой работе описываются так называемые ассоциативные карты памяти (Mind Map), которые имею структуру подобную семантической сети.

**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](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29). 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

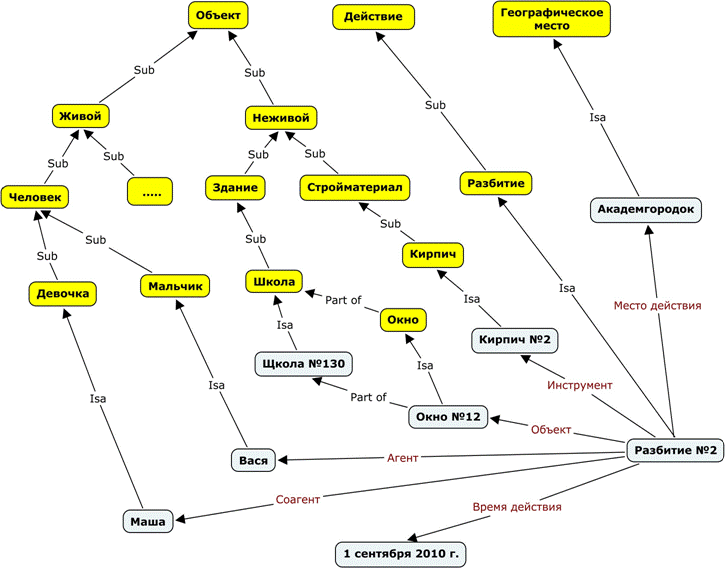
Основной формой представления семантической сети является граф. Понятия семантической сети записываются в овалах или прямоугольниках и соединяются стрелками с подписями — дугами. Это наиболее удобно воспринимаемая человеком форма. Её недостатки проявляются, когда мы начинаем строить более сложные сети или пытаемся учесть особенности естественного языка. Даже для небольших сетей размер графического изображения получается слишком большим и практически нечитаемым. Схемы семантических сетей, на которых указаны направления навигационных отношений, называют картами знаний, а их совокупность, позволяющая охватить большие участки семантической сети, атласом знания. 

Рисунок 1. Пример графического представления семантического графа

###### Математическая запись

В математике [граф](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29) 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](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D0%B4%D0%B8%D0%BA%D0%B0%D1%82).

###### Лингвистическая запись

В лингвистике отношения фиксируются в словарях и в [тезаурусах](https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%B7%D0%B0%D1%83%D1%80%D1%83%D1%81). 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.

**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

Наиболее часто возникает потребность в описании отношений между элементами, множествами и частями объектов. Отношение между объектом и множеством, обозначающим, что объект принадлежит этому множеству, называется отношением классификации (ISA). Говорят, что множество (класс) классифицирует свои экземпляры. (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. Обратное к ISA отношение используется для обозначения примеров, поэтому так и называется — «Example», или по-русски «Пример». Иерархические отношения образуют [древовидную структуру](https://ru.wikipedia.org/wiki/%D0%94%D1%80%D0%B5%D0%B2%D0%BE%D0%B2%D0%B8%D0%B4%D0%BD%D0%B0%D1%8F_%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D0%B0).

Отношение между надмножеством и подмножеством (называется AKO — «A Kind Of», «разновидность»). (Пример: «собака является животным» = тип с именем собака является подтипом типа животные). Элемент подмножества называется [гипонимом](https://ru.wikipedia.org/wiki/%D0%93%D0%B8%D0%BF%D0%BE%D0%BD%D0%B8%D0%BC) (собака), а надмножества — [гиперонимом](https://ru.wikipedia.org/wiki/%D0%93%D0%B8%D0%BF%D0%B5%D1%80%D0%BE%D0%BD%D0%B8%D0%BC) (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 вторым. 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. В этом случае свойства первого множества не наследуются вторым. [Мероним и холоним](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%80%D0%BE%D0%BD%D0%B8%D0%BC_%D0%B8_%D1%85%D0%BE%D0%BB%D0%BE%D0%BD%D0%B8%D0%BC) — противоположные понятия:

Мероним — объект, являющийся частью для другого. (Двигатель — мероним автомобиля.)

Холоним — объект, который включает в себя другое. (Например, у дома есть крыша. Дом — холоним крыши. Компьютер — холоним монитора.)

Часто в семантических сетях требуется определить отношения [синонимии](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BD%D0%BE%D0%BD%D0%B8%D0%BC) и [антонимии](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D1%82%D0%BE%D0%BD%D0%B8%D0%BC). Эти связи либо дублируются явно в самой сети, либо определяются алгоритмической составляющей.

##### 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. Например, между понятиями может использоваться отношение «совершенно разные вещи» или подобное: Не\_имеют\_отношения\_друг\_к\_другу(Солнце, Кухонный\_чайник).

**Семантическая паутина**

Концепция организации [гипертекста](https://ru.wikipedia.org/wiki/%D0%93%D0%B8%D0%BF%D0%B5%D1%80%D1%82%D0%B5%D0%BA%D1%81%D1%82) напоминает однородную бинарную семантическую сеть, однако здесь есть существенное отличие:

Связь, осуществляемая гиперссылкой, не имеет семантики, то есть не описывает смысла этой связи. Назначение семантической сети состоит в том, чтобы описать взаимосвязи объектов, а не дополнительную информацию по предметной области. Человек может разобраться, зачем нужна та или иная гиперссылка, но компьютеру эта связь не понятна.

Страницы, связываемые гиперссылками, являются документами, описывающими, как правило, проблемную ситуацию в целом. В семантической сети вершины (то, что связывают отношения) представляют собой понятия или объекты реального мира.

Попытка создания семантической сети на основе [Всемирной паутины](https://ru.wikipedia.org/wiki/%D0%92%D1%81%D0%B5%D0%BC%D0%B8%D1%80%D0%BD%D0%B0%D1%8F_%D0%BF%D0%B0%D1%83%D1%82%D0%B8%D0%BD%D0%B0) получила название семантической паутины. Эта концепция подразумевает использование языка [RDF](https://ru.wikipedia.org/wiki/RDF) (языка описания структуры на основе [XML](https://ru.wikipedia.org/wiki/XML)) и призвана придать ссылкам и UNI адресам некий смысл, понятный [компьютерным](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BC%D0%BF%D1%8C%D1%8E%D1%82%D0%B5%D1%80) системам. Это позволит превратить [Интернет](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82) в распределённую [базу знаний](https://ru.wikipedia.org/wiki/%D0%91%D0%B0%D0%B7%D0%B0_%D0%B7%D0%BD%D0%B0%D0%BD%D0%B8%D0%B9) глобального масштаба.

Семанти́ческая паути́на (Semantic Web) — это общедоступная глобальная [семантическая сеть](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C), формируемая на базе [Всемирной паутины](https://ru.wikipedia.org/wiki/%D0%92%D1%81%D0%B5%D0%BC%D0%B8%D1%80%D0%BD%D0%B0%D1%8F_%D0%BF%D0%B0%D1%83%D1%82%D0%B8%D0%BD%D0%B0) путём стандартизации представления [информации](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D1%8F) существующий в сети в виде, пригодном для машинной обработки.

В обычной Всемирной паутине, основанной на [HTML](https://ru.wikipedia.org/wiki/HTML)-страницах, информация заложена в тексте страниц и предназначена для чтения и понимания человеком. Семантическая паутина состоит из машинно-читаемых элементов — узлов [семантической сети](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C), с опорой на [онтологии](https://ru.wikipedia.org/wiki/%D0%9E%D0%BD%D1%82%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F_%28%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29). Благодаря этому, программы-[клиенты](https://ru.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B8%D0%B5%D0%BD%D1%82_%28%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29) получают возможность непосредственно получать из интернета утверждения вида «предмет — вид взаимосвязи — другой предмет» и вычислять по ним [логические заключения](https://ru.wikipedia.org/wiki/%D0%92%D1%8B%D0%B2%D0%BE%D0%B4_%28%D1%80%D0%B0%D1%81%D1%81%D1%83%D0%B6%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5%29). Семантическая паутина работает параллельно с обычной Всемирной паутиной и на её основе, используя протокол [HTTP](https://ru.wikipedia.org/wiki/HTTP) и идентификаторы ресурсов [URI](https://ru.wikipedia.org/wiki/URI).

Название «Семантическая паутина» было впервые введено [сэром Тимом Бернерсом-Ли](https://ru.wikipedia.org/wiki/%D0%91%D0%B5%D1%80%D0%BD%D0%B5%D1%80%D1%81-%D0%9B%D0%B8,_%D0%A2%D0%B8%D0%BC) (изобретателем Всемирной паутины) еще в сентябре 1998 года, и называется им «следующим шагом в развитии Всемирной паутины». Позже в своём блоге он предложил в качестве синонима термин «гигантский глобальный граф» ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) giant global graph, GGG, по аналогии с WWW). Концепция семантической паутины была принята и в данный момент активно продвигается [консорциумом Всемирной паутины](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BD%D1%81%D0%BE%D1%80%D1%86%D0%B8%D1%83%D0%BC_%D0%92%D1%81%D0%B5%D0%BC%D0%B8%D1%80%D0%BD%D0%BE%D0%B9_%D0%BF%D0%B0%D1%83%D1%82%D0%B8%D0%BD%D1%8B).

Семантическая паутина — это надстройка над существующей Всемирной паутиной, придуманная для того, чтобы сделать размещаемую в Интернете информацию пригодной для машинной обработки. Доступная в сети информация удобна для прочтения человеком. Семантическая паутина создана для того, чтобы сделать информацию пригодной для автоматического анализа, синтеза выводов и преобразования, как самих данных, так и сделанных на их основе заключений в различные представления, полезные на практике.

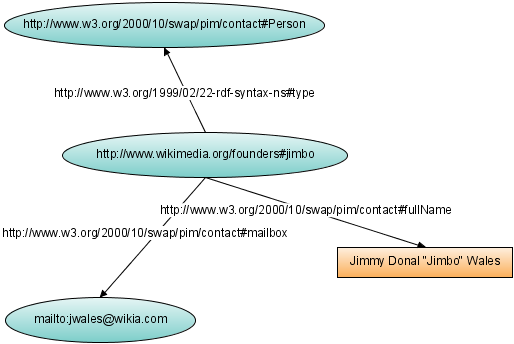


Рисунок 2. RDF граф

[Граф](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29) визитной карточки основателя [Википедии](https://ru.wikipedia.org/wiki/%D0%92%D0%B8%D0%BA%D0%B8%D0%BF%D0%B5%D0%B4%D0%B8%D1%8F) в формате [RDF](https://ru.wikipedia.org/wiki/RDF). Элементы этого графа — как узлы, так и дуги (кроме литерала, изображённого в оранжевом прямоугольнике) — являются [URI](https://ru.wikipedia.org/wiki/URI).

Машинная обработка возможна благодаря двум характеристикам семантической паутины:

* наличию [URI](https://ru.wikipedia.org/wiki/URI);
* использованию [семантических сетей](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C) и [онтологий](https://ru.wikipedia.org/wiki/%D0%9E%D0%BD%D1%82%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F_%28%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29).

[URI](https://ru.wikipedia.org/wiki/URI) — унифицированный идентификатор ресурса или адрес, используемый для указания [ссылок](https://ru.wikipedia.org/wiki/%D0%93%D0%B8%D0%BF%D0%B5%D1%80%D1%81%D1%81%D1%8B%D0%BB%D0%BA%D0%B0) на какой-либо объект (например, [веб-документ](https://ru.wikipedia.org/wiki/%D0%92%D0%B5%D0%B1-%D1%81%D1%82%D1%80%D0%B0%D0%BD%D0%B8%D1%86%D0%B0), [файл](https://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%B9%D0%BB) или ящик [электронной почты](https://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BF%D0%BE%D1%87%D1%82%D0%B0)). URI используются для именования объектов. Каждый объект глобальной семантической сети имеет уникальный URI. URI однозначно называет некоторый объект. Отдельные URI создают не только для страниц, но и для объектов реального мира (людей, городов, художественных произведений и так далее), и даже для абстрактных понятий (например, «имя», «должность», «цвет»). Благодаря уникальности URI одни и те же предметы можно называть одинаково в разных местах семантической паутины. Используя URI, можно собирать информацию о одном предмете из разных мест. Рекомендуется включать в адрес URI название одного из протоколов Всемирной паутины ([HTTP](https://ru.wikipedia.org/wiki/HTTP) или [HTTPS](https://ru.wikipedia.org/wiki/HTTPS)). То есть адрес URI рекомендуют начинать с «http://» или «https://»). Такой адрес можно одновременно использовать как адрес [URI](https://ru.wikipedia.org/wiki/URI) и как адрес веб-страницы ([URL](https://ru.wikipedia.org/wiki/URL)). На веб-страницах, адреса URL которых совпадают с URI, W3C рекомендует размещать описание предмета. Описание желательно предоставлять в двух форматах:

* в формате, удобном для чтения человеком;
* в формате, удобном для чтения машиной.

###### Использование [семантических сетей](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C) и [онтологий](https://ru.wikipedia.org/wiki/%D0%9E%D0%BD%D1%82%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F_%28%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29).

Данные во Всемирной паутине, как правило, представлены в виде текста, записанного на [естественных языках](https://ru.wikipedia.org/wiki/%D0%95%D1%81%D1%82%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA). Такие тексты предназначены для восприятия человеком, но машина может понять их смысл, используя один из методов [обработки естественного языка](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%BA%D0%B0_%D0%B5%D1%81%D1%82%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%BE%D0%B3%D0%BE_%D1%8F%D0%B7%D1%8B%D0%BA%D0%B0). Методы выполняют [частотный](https://ru.wikipedia.org/wiki/%D0%A7%D0%B0%D1%81%D1%82%D0%BE%D1%82%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7) анализ и/или [лексический](https://ru.wikipedia.org/wiki/%D0%9B%D0%B5%D0%BA%D1%81%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7) анализ текста.

В качестве формата, удобного для чтения машиной, [W3C](https://ru.wikipedia.org/wiki/W3C) предлагает использовать [язык](https://ru.wikipedia.org/wiki/%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) [RDF](https://ru.wikipedia.org/wiki/RDF). Язык RDF позволяет описывать структуру семантической сети в виде [графа](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29). Каждому узлу и каждой дуге графа можно назначить отдельный URI. Утверждения, записанные на языке RDF, можно интерпретировать с помощью онтологий. Для создания и описания онтологий рекомендуют использовать [языки](https://ru.wikipedia.org/wiki/%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) [RDF Schema](https://en.wikipedia.org/wiki/RDF_Schema) и [OWL](https://ru.wikipedia.org/wiki/Web_Ontology_Language). Онтологии создаются для получения из данных [логических заключений](https://ru.wikipedia.org/wiki/%D0%92%D1%8B%D0%B2%D0%BE%D0%B4_%28%D1%80%D0%B0%D1%81%D1%81%D1%83%D0%B6%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5%29). В основе онтологий лежат математические формализмы, определяемые дискрипционной логикой.

**Архитектура семантической паутины**

Техническую часть Семантической паутины составляет семейство стандартов на языки описания, включающее [XML](https://ru.wikipedia.org/wiki/XML), [XML Schema](https://ru.wikipedia.org/wiki/XML_Schema), [RDF](https://ru.wikipedia.org/wiki/RDF), [RDF Schema](https://ru.wikipedia.org/wiki/RDF_Schema), [OWL](https://ru.wikipedia.org/wiki/Web_Ontology_Language), а также некоторые другие. Располагая их в порядке повышения уровня абстракции, реализуемого тем или иным языком, получаем:

* [**XML**](https://ru.wikipedia.org/wiki/XML) предоставляет синтаксис для определения структуры документа, подлежащего машинной обработке. Синтаксис XML не несёт семантической нагрузки.
* [**XML Schema**](https://ru.wikipedia.org/wiki/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.
* [**RDF**](https://ru.wikipedia.org/wiki/RDF) представляет собой простой способ описания экземплярных данных в формате субъект-отношение-объект, в котором в качестве любого элемента этой тройки используются только идентификаторы ресурсов. 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](https://ru.wikipedia.org/wiki/%D0%9D%D0%BE%D1%82%D0%B0%D1%86%D0%B8%D1%8F_3)).
* [**RDF Schema**](https://ru.wikipedia.org/wiki/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**](https://ru.wikipedia.org/wiki/Web_Ontology_Language) расширяет возможности по описанию новых типов (в частности, добавлением [перечислений](https://ru.wikipedia.org/wiki/%D0%9F%D0%B5%D1%80%D0%B5%D1%87%D0%B8%D1%81%D0%BB%D0%B8%D0%BC%D1%8B%D0%B9_%D1%82%D0%B8%D0%BF)), а также позволяет описывать новые типы данных RDF Schema в терминах уже существующих (например, определять тип, являющийся [пересечением](https://ru.wikipedia.org/wiki/%D0%9F%D0%B5%D1%80%D0%B5%D1%81%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D0%BC%D0%BD%D0%BE%D0%B6%D0%B5%D1%81%D1%82%D0%B2) или [объединением](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%8A%D0%B5%D0%B4%D0%B8%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5_%D0%BC%D0%BD%D0%BE%D0%B6%D0%B5%D1%81%D1%82%D0%B2) двух существующих).
* [**Микроданные (HTML microdata)**](https://ru.wikipedia.org/wiki/%D0%9C%D0%B8%D0%BA%D1%80%D0%BE%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5)— это 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.

**Семантический рабочий стол**

Семантический рабочий стол — это устройство, на котором каждый хранит всю свою цифровую информацию — документы, мультимедиа и сообщения. Они интерпретируются как ресурсы семантической паутины, каждый из которых имеет уникальный идентификатор ([URI](https://ru.wikipedia.org/wiki/URI)), все данные доступны, и их можно запросить как [RDF](https://ru.wikipedia.org/wiki/Resource_Description_Framework)-граф. Ресурсы из сети могут сохраняться, а авторским контентом можно делиться с другими. Онтологии позволяют пользователю выражать персональные интеллектуальные модели и формировать семантическую связку, соединяющую информацию и системы. Приложения действуют указанным образом, и хранение, считывание, взаимодействие происходят через онтологии и протоколы семантической паутины. Семантический рабочий стол — существенное дополнение к пользовательской памяти.

Представление семантического рабочего стола может рассматриваться как ответ на осознаваемые проблемы существующих пользовательских интерфейсов. Не обладая качественными [метаданными](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%B0%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5), компьютеры не способны без труда узнавать множество зачастую необходимых атрибутов файлов. Например, предположим, что кто-то скачивает документ, созданный конкретным автором, по конкретной теме; хотя наверняка из документа будет ясна его тема, автор, источник и, возможно, информация об [авторском праве](https://ru.wikipedia.org/wiki/%D0%90%D0%B2%D1%82%D0%BE%D1%80%D1%81%D0%BA%D0%BE%D0%B5_%D0%BF%D1%80%D0%B0%D0%B2%D0%BE), компьютеру может быть не так просто получить эту информацию и обрабатывать её в различных приложениях — таких, как файловые менеджеры, локальные поисковики файлов и т. д. Это означает, что компьютер не сможет осуществлять поиск, фильтрацию или иные действия над информацией с максимальной эффективностью. Это основная проблема, которой занимается [семантическая паутина](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D0%BC%D0%B0%D0%BD%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BF%D0%B0%D1%83%D1%82%D0%B8%D0%BD%D0%B0).

Также существует проблема зависимости различных файлов друг от друга. Например, в UNIX-подобных операционных системах [электронные письма](https://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BF%D0%BE%D1%87%D1%82%D0%B0) хранятся отдельно от файлов. Ни одна из проблем не имеет отношения к задачам, заметкам или запланированным действиям, которые могут храниться в программе-календаре. Контактные данные могут храниться в другой программе. Так или иначе, все эти формы представления информации могут одновременно быть актуальными и необходимыми для какой-либо конкретной задачи.

В связи с этим, пользователь часто будет получать — с помощью [браузера](https://ru.wikipedia.org/wiki/%D0%91%D1%80%D0%B0%D1%83%D0%B7%D0%B5%D1%80) или иной программы — доступ к большому количеству информации из интернета, которая отделена от информации, локально хранимой на компьютере.

Семантический рабочий стол — это попытка решить часть или все эти проблемы разом путём расширения возможностей операционной системы, делая возможным управление всеми данными с помощью технологий семантической паутины. Улучшенные пользовательские интерфейсы (или плагины к существующим приложениям), основанные на такой интеграции данных, могут давать пользователю целостностное представление о хранимых знаниях. Некоторые операционные системы, такие как [BeOS](https://ru.wikipedia.org/wiki/BeOS), имеют файловые системы, схожие по функциональности с реляционными базами данных, которые хранят метаданные о документе прямо в файловой системе, что является существенным шагом на пути к семантическому рабочему столу.

Существуют различные интерпретации идеи семантического рабочего стола. В наиболее узком смысле она может интерпретироваться как добавление механизмов привязывания считываемых компьютером [метаданных](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%B0%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5) к файлам. В крайних случаях она может трактоваться как полная замена существующим пользовательским интерфейсам, унифицирующая все формы данных и обеспечивающая единый последовательный интерфейс. В зависимости от решаемой проблемы, есть много степеней градации между двумя перечисленными вариантами.

Рисунок 3. Структура Semantic Desktop

**Заключение**

Хоть технологии Semantic Web разрабатывались для развертывания семантической сети в глобальной сети Интернет, они достаточно универсальный для построения семантических графов, распределенных по электронным сетям любого уровня. Так, с помощью языка описания семантических документов RDF и идентификаторов URI можно выстроить семантическую сеть на уровне предприятия или даже группы предприятий (например, для стандартизации финансовых показателей в конгломерате компаний). Более того, в семантическую сеть можно даже превратить ресурсы одного персонального компьютера с помощью определенных приложений, таких как персональный информационный менеджер или даже специализированных операционных систем, построенных на идее семантически соединенных файлов и их метаданных в компьютере (BeOS).

Практической реализацией этой идеи является так называемый Семантический рабочий стол – компьютерный пользовательский интерфейс, который дает возможность управления данными на персональном компьютере таким образом, что упрощается автоматическая обработка этих данных и взаимодействие между различными приложениями. Не обладая качественными [метаданными](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%B0%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5), компьютеры не способны без труда узнавать множество зачастую необходимых атрибутов файлов. Однако если позволить компьютеру управлять данными с помощью технологий семантической паутины, возможна смысловая обработка всей информации в системе, что превращает персональный компьютер в персональную базу знаний.

Семантические графы и их электронные представления хоть и находятся на заре своего развития, имеют поистине безграничный потенциал. Их не зря называют будущим Интернета и баз знаний.