



Факты и правила в семантической сети

Рассмотренные выше отношения, записанные в виде субъект — объект, представляют собой неизменные знания, т. е. факты. Знание всех известных фактов о каждом объекте может потребовать довольно много времени. В качестве примера уместно привести реальные отношения. Для любых двух родственников есть название отношения между ними: дядя-племянник, свекровь-зять и т. п. Таким образом, для семьи из $n = 10$ человек число отношений будет равно n^2 ($n^2 = 90$). При этом часть отношений являются первичными (супруга и родитель-ребенок), остальные отношения вытекают из первичных. Если информацию о том, как вторичные отношения делятся на основе первичных, записать в виде правил, то для каждого объекта можно заносить в базу знаний только первичные факты. Семейных отношений это означает сокращение, не менее, чем, в $n/3$. Если считать, что каждый член семьи является чьим-то ребенком и супругом, а также чьим-то супругом, и не более того.

Один из стандартов языка представления правил является SWRL — Semantic Web Rule Language (<http://www.w3.org/Submission/SWRL/>). Этот язык является расширением XML, и конструкции на нем назначены исключительно для машинной интерпретации. Многие правила обычно предоставляют вариант “human readable” или в формате, подобном правилам на Прологе, для их создания и отладки. Ниже приведен фрагмент правила определения отношения «родитель» для чтения человеком и исходный текст данного правила на языке SWRL.

```

isParent (?x1, ?x2) and hasBrother (?x2, Px3) => hasIncl
(?x1, Px3)

hasBrother"/>

```

ганавливая правила для объектов семантической сети, мы можем кнуться с проблемой открытого или закрытого мира (Open or Closed d Assumption). Допущение об открытом мире предполагает, что о не располагает полной информацией об окружающем мире, овательно, выводы должны делаться исключительно на основании что известно. Допущение о закрытом мире предполагает, что вся ормация известна наблюдателю. В качестве примера можно ести родственное отношение типа «мачеха». Пусть в базе знание тся следующие факты:

дрей является родителем Егора.

ния является супругой Андрея.

оответствии с допущением закрытого мира, Юлия является мачехой а, поскольку в базе нет сведений о том, что она его мать. В открытом : Юлию можно считать мачехой Егора только в том случае, если стно, что его матерью является не Юлия, а другая женщина.

именение правил крайне полезно в тех случаях, когда в базе знаний ржится неполная информация. Пусть, например, в предыдущем ере указано, Андрей — это человек, но нет сведений, что Юлия — век. Тогда все правила наподобие

и X человек, то X имеет фамилию

смогут быть применены к Юлии. Если же создать правило

и X человек И X супруг YТОУ человек

можно будет установить факт, что Юлия тоже человек. Помимо жительного эффекта от использования правил имеется и статок: комбинаторная сложность, которая по мере увеличения ма базы знаний довольно быстро вырастает до космических итабов. Так, например, если в достаточно маленькой базе знаний тся 100 фактов и 10 правил по три факта в каждом, то общее чество попыток применить правила к фактам может достигать знения $10 * 100 * 100 * 100 = 10^7$, поскольку в каждое правило едовательно будут подставляться все возможные факты. Очевидно, такая «наивная реализация» поиска в семантической сети изнеспособна. Как в любой задаче поиска, здесь нужно решать

лему сокращения комбинаторной сложности. В качестве примера решения обработки правил можно привести алгоритм Rete (Рити) (https://en.wikipedia.org/wiki/Rete_algorithm), основной смысл которого заключается в том, что строится дерево, каждый узел которого отвечает части условий правил и хранит список фактов, удовлетворяющих этим условиям. Поскольку в ходе применения правил постоянно возникают новые факты, они прогоняются по сети, и списки фактов при вершинах обновляются. Узким местом алгоритма Rete является большой требуемый объем памяти, поскольку одни и те же факты многократно дублируются в списках при вершинах графа.

В качестве одного из альтернативных путей решения можно запускать возможные правила для каждого документа один раз и сохранять результаты в виде фактов. Для базы родственных связей это будет означать, что в документ вначале заносятся только первичные связи (родители-дети и супруги), затем из них вычисляются все производные отношения (внучатные племянники и т. д.), которые на основе этих данных затем пополняют базу знаний. После этого все факты могут извлекаться одинаково быстро. Такой подход является очевидностью вывода на основе прецедентов (Case based reasoning), и можно считать аналогом навыков в человеческом интеллекте. На деле, мы почти всегда действуем по аналогиям, например, в живой речи. Если бы мы при построении каждой фразы применяли грамматику языка, то скорость речи не превышала бы нескольких предложений в час. Это особенно заметно, когда мы переводим русский язык на иностранный язык. Если языковая конструкция нам знакома (мы ранее неоднократно использовали ее), то перевод идет в быстром темпе. Если мы создаем предложение в первый раз, то процесс перевода занимает десятки и сотни раз.