

WEB-СЕРВИС ДЛЯ ВИЗУАЛИЗАЦИИ АССОЦИАТИВНЫХ И СЕМАНТИЧЕСКИХ СЕТЕЙ

Кулаков Д.С., Кулаков А.С. (МГТУ им. Н.Э.Баумана)

Рассматриваются вопросы построения и визуализации семантических сетей и онтологий в рамках концепции Web 3.0 (Semantic web). В качестве основных технологий выбраны стандарты XML и средства разработки Adobe Flash.

Семантическая сеть — информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними. Объектами могут быть понятия, события, свойства, процессы. Таким образом, семантическая сеть является одним 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). 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.

Основной формой представления семантической сети является граф. Это наиболее удобно воспринимаемая человеком форма. Её недостатки проявляются, когда мы начинаем строить более сложные сети или пытаемся учесть особенности естественного языка.

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.

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), например, таковой является вышеупомянутая классификация биологических видов (с единственным отношением АКО). В неоднородных сетях количество типов отношений больше двух. Классические иллюстрации данной модели представления знаний представляют именно такие сети. Неоднородные сети представляют больший интерес для практических целей, но и большую сложность для исследования.

По арности, типичными являются сети с бинарными отношениями (связывающими ровно

два понятия). Бинарные отношения, действительно, они очень просты и удобно выглядят 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. При этом возникает сложность — как изобразить подобную связь на графе, чтобы не запутаться. Концептуальные графы снимают это затруднение, представляя каждое отношение в виде отдельного узла.

Помимо концептуальных графов существуют и другие модификации семантических сетей, это является ещё одной основой для классификации (по реализации).

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 сделаны программы обработки нужных отношений.

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. Название произошло от английского «IS A» (наиболее точный русский перевод, используемый в основном в научных кругах — «суть», например, «все зайцы суть млекопитающие»). 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, «Например».

Отношение между надмножеством и подмножеством называется АКО — «A Kind Of» («разновидность»). Элемент подмножества называется гипонимом, а надмножества — гиперонимом, 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.

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 части/целые объекты (отношение меронимии). Мероним — это объект, являющийся частью для другого. Двигатель — это мероним для автомобиля. Холоним — это объект, который включает в себя другое. Например, у дома есть крыша. Дом — холоним для крыши. Компьютер — холоним для монитора. Мероним и холоним — противоположные понятия.

Часто в семантических сетях требуется определить отношения синонимии и антонимии. Эти связи либо дублируются явно в самой сети, либо в алгоритмической составляющей. [2]

В общем случае под семантической сетью понимается выражение следующего вида:

$$S=(O=\{O^i, i=1, n\}, R=\{R^j, j=1, k\})$$

где

O – множество объектов конкретной предметной области;

$R^i | i=1, n$ – множество отношений между объектами;

i – тип отношений.[4]

Одним из наиболее интересных вопросов является использование семантических сетей. Попытка создания семантической сети на основе Всемирной паутины получила название семантической паутины. Эта концепция подразумевает использование языка RDF (языка разметки на основе XML) и призвана придать ссылкам некий смысл, понятный компьютерным системам. Это позволит превратить Интернет в распределённую базу знаний глобального масштаба. Концепция организации гипертекста напоминает однородную бинарную семантическую сеть, однако здесь есть существенное отличие:

1. Связь, осуществляемая гиперссылкой, не имеет семантики, т. е. 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.

2. 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.[2]

Разработанный web-сервис позволяет визуализировать ассоциативные и семантические сети практически любой структуры.

Цель данной работы – создание сервиса для визуализации сетей в виде графа с направленными связями вида A->B, работающего в web-браузере и принимающего входные параметры из адресной строки.

В ходе разработки сервиса были решены задачи построения ориентированного графа в среде Flash на основе XML документа, передачи параметров адресной строки Flash документу и преобразование их в XML документ.

Сервис работает в web-браузере и позволяет динамически добавлять вершины к графу путем ввода дополнительных данных в строку ввода адреса. Пример работы сервиса приведен ниже (см. рис. 1).

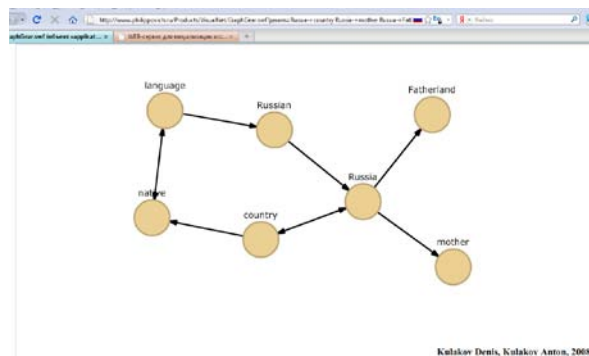


Рис.1 Пример работы Web-сервиса.

У пользователя имеется возможность создавать вершины графа, соединенные между собой ребрами различной направленности. Для этого достаточно ввести в строку адреса необходимые параметры в виде пар ассоциативно связанных с собой понятий. В предыдущем примере данная строка выглядит следующим образом:

<http://www.philippovich.ru/Products/VisualNet/GraphGear.swf?param=Russia->country Russia->mother Russia->Fatherland country->Russia country->native native->language language->native language->Russian Russian->Russia>

Пользователь также может динамически добавлять новые вершины и рёбра к графу, а также изменять местоположение вершин на экране.[1]

Для разработки Web-сервиса использовался пакет Adobe Flash - мультимедийная платформа, используемая главным образом для создания

векторной анимации и интерактивных приложений, а также для интеграции видеороликов в веб-страницы. За прототип сервиса был взят [сервис визуального словарика](http://www.creativesynthesis.net/blog/2007/08/19/graph-gear-an-opensource-platform-for-graph-visualization-now-available/) и opensource ресурс для визуализации графов (<http://www.creativesynthesis.net/blog/2007/08/19/graph-gear-an-opensource-platform-for-graph-visualization-now-available/>).[2]

Наиболее интересной программной частью является разбор параметра, введенного пользователем и формирование XML документа. Ниже представлен программный код, реализующий парсинг адресной строки и компоновку XML документа:

```

var my_array:Array = new Array();
var obj_array:Array = new
Array();
var obj_array_init:Array = new
Array();

//var param:String = new
String();
var my_string:String = new
String();
var firststep:Array = new
Array();
var secondstep:Array = new
Array();
//my_array = [[1,2]];
//my_array.push([[3],[4]]);
//trace(my_array[0][1]);
Result = '';

this.onLoad = function(){
//param='a->b b->c';
my_string = param;
firststep = my_string.split('
');
for (i = 0; i <
firststep.length; i++)
{
d =
firststep[i].lastIndexOf('->');
my_array_a =
firststep[i].slice(0,d);
my_array_b =
firststep[i].slice(d+2,firststep[i].
length);
//trace('a = ' +
my_array_a);
//trace('b = ' +
my_array_b);

my_array.push([my_array_a,my_a
rray_b]);
Result = Result +
my_array_a + "->" + my_array_b +
"\n";
}
_global.str = '';

```

```

_global.str = '<?xml
version="1.0"?>';
_global.str += '<graph
animation="true" action="drag"
title="Simple Graph Demo"
bgcolor="ffffff" linecolor="000000"
viewmode="display" width="725"
height="400" type="directed"
segmentlength="105">';
/*obj_array.slice(0,obj_array.
length);
obj_array_init.slice(0,obj_arr
ay_init.length);*/

for (i = 0; i <
my_array.length; i++)
{

obj_array_init.push(my_array[i
][0]);

obj_array_init.push(my_array[i
][1]);
}
trace(obj_array);
//-----
Удаление повторов-----
-----
/*for (k = 0; k <
obj_array.length; k++ )
{
for (l = k+1; l <
obj_array.length; l++ )
{

if
(obj_array[k].to_global.string()==ob
j_array[l].to_global.string())
obj_array.splice(l,1);
}
}
trace(my_array); */
for (k = 0; k <
obj_array_init.length; k++ )
{
for (l = k+1; l <
obj_array_init.length; l++ )
{

if
(obj_array_init[k].toString()==obj_a
rray_init[l].toString())
obj_array_init[l] = 0;
}
}
for (k = 0; k <
obj_array_init.length; k++ ) if
(obj_array_init[k].toString()!=0)
obj_array.push(obj_array_init[k]);

```

```

//-----
Объекты-----

    for (i = 0; i <
obj_array.length; i++) _global.str
+= '<node type="CircleTextNode"
id="n' + (i+1) + '" text="' +
obj_array[i] + '" scale="110"
color="ECD093"
textcolor="000000"/>'; //89BD0F
ECD093
    //trace(find_number(my_array[0
][0]));

    //trace(my_array);
//-----
Связи-----

    for (i = 0; i <
my_array.length; i++)
    {
        n1 =
find_number(my_array[i][0]);
        n2 =
find_number(my_array[i][1]);
        _global.str += '<edge
sourceNode="n' + n1 + '"
targetNode="n' + n2 + '" label=""
textcolor="555555"/>';
    }
    trace(_global.str);
    //z =
find_number(my_array[1][0]);
    //trace(_global.str);
    //trace(obj_array);
    //var gl = new GraphLoader();
    gl.processLiveXML(_global.str)
;

    /*delete obj_array;
delete obj_array_init;*/
    while(obj_array_init.length)ob
j_array_init.pop();
    while(obj_array.length)obj_arr
ay.pop();
    }

function find_number (my_array){
    n = 0;
    for (p = 0; p <
obj_array.length; p++)
        if (my_array.toString()
== obj_array[p].toString()) n = p+1;

```

```

return n;
}

```

Таким образом, разработан web-сервис для визуализации ассоциативных и семантических сетей. В дальнейшем возможна доработка сервиса, расширение его функциональности.[3]

Одной из наиболее важных задач в развитии сервиса является визуализация силы связи между понятиями. Для удобства восприятия стрелочки, обозначающие связи между понятиями должны располагаться так, чтобы попарные расстояния между ними отражали силу связи. Алгоритм оптимального отображения сети на плоскость реализует один из методов решения известной в математике задачи многомерного шкалирования. Он позволяет по заданной матрице попарных расстояний между объектами (весов связей) построить их размещение в пространстве заданной размерности (в данном случае - на плоскости).[5]

Другой не менее важной задачей является предоставление пользователю возможности устанавливать типы связей или отношений между понятиями. А также стоит задача наглядного отображения отношений между объектами в виде, например, подписей к ребрам, соединяющим вершины или в виде различных стилей для дуг, соответствующих разным типам отношений.

Данный web-сервис создан технически таким образом, что делает возможным не только самостоятельное использование продукта, но и позволяет интегрировать данный сервис в различные интернет проекты, например, основанные на концепции семантической паутины.

Источники, использованные в работе

1. <http://www.artlebedev.ru/tools/technogrette/etc/flash-xml/>
2. <http://ru.wikipedia.org/>
3. <http://www.creativesynthesis.net/blog/2007/08/19/graph-gear-an-opensource-platform-for-graph-visualization-now-available/>
4. <http://ito.edu.ru/2008/Kursk/IV/IV-0-4.html>
5. http://www.rco.ru/technology.asp?ob_no=1920