

УДК 621.4

Л.С. Сорока¹, В.В. Косенко²¹Харьковский национальный университет им. В.Н. Каразина, Харьков²ГП "Харьковский НИИ технологии машиностроения", Харьков

МОДЕЛИРОВАНИЕ БАЗОВОГО ДЕРЕВА РЕШЕНИЙ ДЛЯ РАСПРЕДЕЛЕННОГО ПОИСКА В ИНФОКОММУНИКАЦИОННЫХ СЕТЯХ

В статье предлагается математическая модель базового дерева решений для распределенного поиска в инфокоммуникационных сетях. Определены интенционалы и экстенционалы семантической сети на стратифицированном расширении дерева решений, что позволяет перейти к синтезу семантической сети с целью конструирования сетевыми интеллектуальными агентами механизмов вывода в соответствии с требуемыми атрибутами и заданными отношениями.

Ключевые слова: дерево решений, семантическая сеть, интенционал, экстенционал, инфокоммуникационная сеть, интеллектуальный агент.

Введение

Постановка проблемы. Одной из основных тенденций развития современных информационно-коммуникационных сетей (ИКС) становится расширение доступности информационных ресурсов сетей для отдельных абонентов. В связи с этим отдельные абоненты ИКС становятся все более активными потребителями их вычислительных ресурсов и участниками создания информационных материалов, непосредственно владеющими техникой доступа к информационным ресурсам ИКС.

С другой стороны, повышение активности отдельных абонентов ИКС обусловлено распределенностью информационно-вычислительных ресурсов современных ИКС, в связи с чем в распределенных системах резко возрастает роль коммуникаций как на уровне прикладных задач, так и на уровне технических систем.

Бурное развитие технических средств информационно-коммуникационных сетей и повышение активности доступа к информационно-вычислительным ресурсам отдельных абонентов повысили интерес к проблеме эффективного доступа к информационным ресурсам сети [1 – 6]. Особенно остро при этом в настоящее время стоит проблема целенаправленного информационного поиска (Information Retrieval, IR), под которым понимается комплексная деятельность по сбору, организации, поиску, извлечению и распространению информации при помощи компьютерных технологий [7].

Для многих подходов к организации полнотекстового поиска [8, 9] оказывается достаточно инвертированного файла или подобной структуры данных. Но при извлечении релевантных данному запросу документов необходимо определить способ обработки запроса и вычисления значения релевантности для каждой пары "запрос-документ". Однако, при этом

можно отметить ряд очевидных недостатков, затрудняющих поиск релевантных текстов, например:

– избыточность: в индексе используются синонимы, выражающие одни и те же понятия;

– слова текста считаются независимыми друг от друга, что не соответствует свойствам связного текста;

– многозначность слов: поскольку многозначные слова могут рассматриваться как дизъюнкция двух или более понятий, выражающих различные значения многозначного слова, то маловероятно, что все элементы этой дизъюнкции интересуют пользователя.

Этих недостатков лишено так называемое концептуальное индексирование, т.е. такое индексирование, когда текст индексируется не по словам, а по понятиям, которые обсуждаются в данном тексте. При такой технологии все синонимы сведены к одному и тому же понятию; многозначные слова отнесены к разным понятиям; связи между понятиями и соответствующими словами описаны и могут быть использованы при анализе текста.

Анализ литературы. Для того чтобы попытаться реализовать схему автоматического концептуального индексирования и концептуального поиска, необходимо иметь ресурс, описывающий систему понятий данной предметной области. Такой ресурс содержится в семантической сети (СС), суть которой состоит в автоматизации "интеллектуальных" задач обработки значения (в семантическом смысле) тех или иных ресурсов, имеющихся в сети [10]. Обработкой и обменом информацией должны заниматься интеллектуальные агенты (ИА), размещенные в сети. На сегодня разработан целый ряд стандартов и рекомендаций, реализовано множество проектов.

Но, несмотря на отдельные успехи, до сих пор нельзя сказать, что идея СС реализована на практике [11].

Для описания семантики в сети определена спецификация RDF (Resource Description Framework) [12]. RDF предоставляет простой, но мощный язык описания ресурсов, основанный на триплетях (triple-based) "Субъект-Предикат-Объект" и спецификации URI. В Концептуально RDF дает минимальный уровень для представления знаний в Сети и опирается на ранние стандарты, лежащие в основе Web. Начата работа над форматом обмена правилами – RIF (Rule Interchange Format). Его назначение – соединить в одном стандарте несколько формализмов для описания правил (по которым может осуществляться нетривиальный логический вывод): логику клауз Хорна, логики высших порядков, продукционные модели и т.п.

Однако во всех рассмотренных выше стандартах и подходах к построению СС отсутствует четкий формализм выявления функциональных и логических зависимостей на этапе формирования промежуточных последовательностей вывода в соответствии с требуемыми атрибутами и заданными отношениями, формируемыми в процессе расширенной обработки при концептуальном поиске [13]. Особенно остро данная проблема стоит при организации распределенного поиска. Один из возможных подходов к реализации такого формализма при распределенном поиске предполагает сочетание аппарата построения деревьев решений и семантических сетей, расширенных стратифицированной структурой.

Целью данной статьи является описание модели базового дерева решений для семантической сети, формирующей механизмы вывода в соответствии с требуемыми атрибутами и заданными отношениями.

Результаты исследований

Деревья решений дают возможность устанавливать логические закономерности между информационными сегментами и представляют удобный аппарат, на базе которого можно построить семантическую сеть, позволяющую хранить и обрабатывать выводы, полученные при использовании деревьев решений.

Для моделирования воспользуемся логикой предикатов первого порядка, позволяющей:

- производить вычисления как в узлах дерева, так и на его ребрах, а на основе результатов вычислений принимать решения;
- применять операции разбиения для выбора отдельных фрагментов;
- уточнять решения при дальнейшем раскрытии верхних вершин выводов;
- расширить по вертикали и горизонтали многоуровневую модель.

В качестве графа моделируемого дерева решений будем рассматривать дерево G_X с корнем:

$$G_X = (\bar{X}, R), \quad (1)$$

где $\bar{X} = (X^0, \bar{X}^1, \dots, \bar{X}^{m-1})$ – кортеж, состоящий из множества узлов различных рангов; X^0 – корень дерева; $\bar{X}^i = (X_1^i, X_2^i, \dots, X_{\ell_i}^i)$, $0 \leq i \leq m-1$ – множество узлов i -го ранга; $R = \{r_{jv}^i\}$; $0 \leq i \leq m-2$; $1 \leq j \leq \ell_i$; $1 \leq v \leq \ell_{i+1}$ – множество дуг графа, представляющих собой направленные связи между узлами; i указывает ранг вершины j , из которого выходит связь; v – номер вершины $(i+1)$ -го ранга, в которую входит связь.

На первом этапе формализации построим граф базового дерева решений $G_X^{(B)}$, узлы которого представляют подструктуру ССС, способную выполнить автономную подзадачу поиска (частичное решение или частичный вывод). Для этого определим область вывода $D = \{d_k \mid k = \overline{1, m}\}$, где d_k – k -ая переменная вывода с доменом D_k и область решений $X = \{x_\xi \mid \xi = \overline{1, n}\}$, где x_ξ – ξ -ая переменная решения с доменом X_ξ . При построении дерева решений будем рассматривать переменные двух типов: "решение" и "вывод". Если $\Pi_\zeta(x_\xi)$, $\zeta = \overline{1, r}$ – j -ое отношение, определенное на домене переменной x_ξ , то приход по дереву решений в эту вершину означает определение цепочки

$$\Pi_\zeta(x_\xi) \in \Pi_\zeta^* = \{\Pi_1(x_\xi), \dots, \Pi_r(x_\xi)\}. \quad (2)$$

Пусть $\Pi(d_k)$ – отношение, определенное на домене переменной d_k , а $\sigma_\ell(\Pi(d_k))$, $\ell = \overline{1, r}$ – отношение, определенное на множестве значений, удовлетворяющих отношению Π . Определим, что эксцидентность будет нулевой только для узлов вывода, т.е. только на этом множестве можно ввести операцию раскрытия вершины \mathfrak{Z} :

– родительский узел – "решение", дочерний узел – "вывод":

$$\mathfrak{Z}(x_\xi) \rightarrow \Pi(d_k); \quad (3)$$

– родительский узел – "решение", дочерний узел – "решение":

$$\mathfrak{Z}(x_\xi) \rightarrow (\Pi_j(x_\xi) = x_\xi); \quad (4)$$

– родительский узел – "вывод", дочерний узел – "вывод":

$$\mathfrak{Z}(\Pi(d_k)) \rightarrow \sigma_\ell(\Pi(d_k)); \quad (5)$$

– родительский цзел – "вывод", дочерний узел – "решение":

$$\mathfrak{Z}(\Pi(d_k)) \rightarrow x_\xi \rightarrow \dots \quad (6)$$

После операции раскрытия вершин над множествами переменных вывода и решений строится базовое дерево решений $\mathfrak{S} = (\mathfrak{Z}(D), \mathfrak{Z}(X))$ с графом $G_X^{(B)}$, на котором можно выделить такие ветви:

- факт \aleph_p (путь от корневого узла дерева до конечной вершины);
- элементарный факт $v_p \subset \aleph_p$ (конкретный элемент пути решения \aleph_p с операцией \Im);
- частичный факт \aleph_r (путь от узла дерева до узла вывода, не являющегося конечным).

Следующий этап – наполнение базового дерева семантической информацией и организация его взаимодействия с сетевыми интеллектуальными агентами.

Для работы с ветвями на дереве решений $\aleph = (\Im(D), \Im(X))$, расширенным стратифицированной структурой W , определим следующие операции:

- разбиение дерева $\aleph(\aleph)$; операция заключается в выделении частичного дерева \aleph_λ по уровням стратификации W или по «вертикальным» признакам;
- объединение λ ($\lambda > 1$) частичных деревьев с дополнением $\wp(\aleph_{\lambda_1}, \dots, \aleph_{\lambda_\lambda})$;

– уточняющее объединение $\wp(\aleph_\lambda)$, при котором конечной вершине вывода частичного дерева сопоставляется другое частичное дерево, корневая вершина которого является выбранной вершиной вывода (рис. 9).

Обозначим множество атрибутов рассматриваемого факта, полученное на основании структуры W , как $A = (A_1, \dots, A_n)$, а множество отношений, связанных с данным фактом как $R = (R_1, \dots, R_\psi)$, причем интенционал отношения R_v ($v \in \overline{1, \psi}$) определяется следующим образом:

$$\text{Int}(R_v) = \left\{ A^{(v)} \subset A; \text{dom}(A^{(v)}) \right\}, \quad (14)$$

где $A \supset A^{(v)}$ – подмножество атрибутов рассматриваемого отношения; $\text{dom}(A^{(v)})$ – соответствующие данным атрибутам домены, объединение которых является базовым множеством для передачи информации соответствующим интеллектуальным агентам.

В принятых обозначениях экстенционал отношения R_v ($v \in \overline{1, \psi}$) определяется как

$$\text{Ext}(R_v) = \left\{ F_\theta^{(v)}(R_v) \right\}, \quad (15)$$

где $F_\theta^{(v)}(R_v)$ – расширенный θ -й элементарный факт рассматриваемого отношения, задаваемый совокупностью пар атрибутов в соответствии со структурой W , т.е. являющийся конкретизацией данного отношения.

Отметим, что в графической интерпретации расширенный элементарный факт является подграфом семантической сети, имеющим звездообразную структуру, причем его корень принадлежит страте более высокого порядка, чем подчиненные узлы.

Выводы

Таким образом, предложена модель базового дерева решений для семантической сети, формирующей механизмы вывода в соответствии с требуемыми атрибутами и заданными отношениями, которая позволяет при распределенном поиске в инфокоммуникационной сети рассматривать их в тесной взаимосвязи, в результате чего появляется возможность формализации процесса поиска и конструирования сетевыми интеллектуальными агентами соответствующих механизмов вывода.

Направление дальнейших исследований связано с формализацией математической модели связи базового дерева решений, семантической сети и соответствующих интеллектуальных агентов.

Список литературы

1. Кучук Г.А. Управление ресурсами инфотелекоммуникаций / Г.А. Кучук, Р.П. Гахов, А.А. Паинев. – М.: Физматлит, 2006. – 218 с.
2. Кучук Г.А. Побудова черги при самоподібному трафіку / Г.А. Кучук // Системи обробки інформації. – Х.: ХВУ. – 2004. – Вип. 6. – С. 134 – 137
3. Кучук Г.А. Метод дослідження фрактального мережного трафіка / Г.А. Кучук // Системи обробки інформації. – Х.: ХУПС, 2005. – Вип. 5(45). – С. 74 – 84.
4. Кучук Г.А. Метод визначення моментів зміни характеристик трафікового процесу / Г.А. Кучук // Системи озброєння і військова техніка. – 2008. – № 4(16). – С. 151-155.
5. Кучук Г.А. Модель процесу буферизації пакетів в апаратних засобах мультиплексування / Г.А. Кучук // Системи обробки інформації. – Х.: ХУПС, 2006. – Вип. 2(51). – С. 65-70.
6. Кучук Г.А. Конвергенція різномірного трафіка в інформаційно-телекомунікаційних мережах / Г.А. Кучук, Ю.В. Паржин, І.І.Сидоренко // Системи обробки інформації. – Х.: ХУПС, 2008. – Вип. 6 (73). – С. 75 – 77.
7. Алгоритмы интеллектуального анализа данных. – [Электрон. ресурс]. – Режим доступа: <http://msdn2.microsoft.com/ru-ru/library/ms175595.aspx>.
8. Voorhees E.M. Natural Language Processing and Information Retrieval / E.M. Voorhees [Электрон. ресурс]. – Режим доступа: <http://citeseer.ist.psu.edu/voorhees99.natural.html>.
9. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа / Д.В. Ландэ. – М.: Вильямс, 2005. – 272 с.
10. Бондарев В.Н. Искусственный интеллект / В.Н. Бондарев, Ф.Г. Аде. – Севастополь: СевНТУ, 2002. – 615 с.
11. Fridman A. Knowledge Integrating in Situative Modelling System for Nature-Technical Complexes / A.Fridman, A. Oleynik, O. Fridman // Proc. European Simulation and Modelling Conference (ESMc2004). – Paris, France, 2004. – P. 49-58.
12. Тауберер Д. Краткое введение в RDF / Д. Тауберер; пер. с англ. А. Скробов [Электрон. ресурс]. – Режим доступа: <http://xmlhack.ru/texts/D6-RDF>.
13. W3C GRDDL Specification [Электрон. ресурс]. – Режим доступа: <http://www.w3.org/RDF/YR/GRDDL>.

Поступила в редколлегию 30.09.2009

Рецензент: д-р техн. наук, проф. И.В. Рубан, Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков.

**МОДЕЛЮВАННЯ БАЗОВОГО ДЕРЕВА РІШЕНЬ ДЛЯ РОЗПОДІЛЕНОГО ПОШУКУ
В ІНФОКОМУНІКАЦІЙНИХ МЕРЕЖАХ**

Л.С. Сорока, В.В. Косенко

У статті пропонується математична модель базового дерева рішень для розподіленого пошуку в інфокомунікаційних мережах. визначені інтенціонали і екстенціонали семантичної мережі на стратифікованому розширенні дерева рішень, що дозволяє перейти до синтезу семантичної мережі з метою конструювання мережевими інтелектуальними агентами механізмів виводу відповідно до необхідних атрибутів і заданих відносин.

Ключові слова: *дерево рішень, семантична мережа, інтенціонал, екстенціонал, інфокомунікаційна мережа, інтелектуальний агент.*

**MODELLING OF BASE TREE OF DECISIONS
FOR THE DISTRIBUTED SEARCH IN INFOCOMMUNICATION NETWORKS**

L.S. Soroka, V.M. Prikhod'ko

In the article the mathematical model of base tree of decisions is offered for the up-diffused search in infocommunication networks. Intensionals and extensionals of semantic network is certain on the stratified expansion tree of decisions, that allows to pass to the synthesis of semantic network with the purpose of constructing of mechanisms conclusion network intellectual agents in accordance with the required attributes and set relations.

Keywords: *tree of decisions, semantic network, intensional, extensional, infocommunication network, intellectual agent.*