

УДК 004.7

2237.24

Чубин И. М.

ИСПОЛЬЗОВАНИЕ СИСТЕМЫ ИМЕН РЕЙЗЕРА ДЛЯ ПОИСКА СУЩНОСТЕЙ В РАСПРЕДЕЛЕННОЙ СИСТЕМЕ

Национальный технический университет Украины "КПИ"

В работе показано, как распределенная система имен Рейзера может использоваться для поиска сущностей по ключевым словам в распределенной системе. Описан механизм построения имени, которое может использоваться для поиска сущностей по ключевым словам, в том случае, если пространство имен может быть представлено в виде двухъярусного графа.

Введение

В последнее время большую актуальность получила задача управления сущностями в больших распределенных системах и одна из ее важных подзадач – поиск сущностей. Один из подходов к решению этой задачи базируется на использовании семантики Рейзера – специального способа организации и использования пространств имен, упрощающего поиск информации. Вопросы использования семантики Рейзера в централизованных системах досконально исследованы Рейзером в его работах [1].

Семантика Рейзера может использоваться для организации распределенной системы имен: в работе [2] предложена структура такой распределенной системы имен, а в работе [3] предложен алгоритм разрешения имен в этой системе. Основываясь на этой структуре можно построить распределенную поисковую систему, обладающую высоким уровнем масштабируемости по Ньюману [4].

Постановка задачи

Существует распределенная система с множеством объектов, характеризующихся ключевыми словами. Необходимо показать, что система имен Рейзера может использоваться в качестве основы распределенной поисковой системы, т.е. что с ее помощью можно решать задачу поиска объекта системы по ключевым словам, и система сохранит при этом свои свойства распределенности и масштабируемости.

Поиск объекта в распределенной системе

Распределенная система DS представляет собой множество взаимодействующих сущностей E_i :

$$DS = \{E_i\}$$

Сущность характеризуется именем и идентификатором:

$$E = (Id, Name)$$

Идентификатор Id однозначно указывает сущность внутри системы, однако он непригоден для наименования сущности людьми и внешними системами. Для этого используется имя $Name$.

В том случае, если работа распределенной системы предполагает, что необходимо выполнять поиск объектов, сущность характеризуется еще и набором ключевых слов:

$$E = (Id, Name, Keys)$$

Ключевые слова $Keys$ — неупорядоченное множество слов, характеризующих сущность:

$$Keys = \{key_i\}.$$

Имя $Name$, идентификатор Id и ключевые слова key_i представляют собой битовые или символьные строки.

Система имен позволяет получить информацию о сущности посредством имен. Она предлагает механизм, позволяющий по имени сущности определить (в частности) ее идентификатор:

$$Id = NS(Name).$$

Пространством имен с возможностью поиска (searchable namespace) позволяет определить имя (и идентификатор) сущности по множеству ключевых слов, описывающих ее:

$$Id = SNS(Keys)$$

или

$$Name = SNS(Keys).$$

Поиск сущности в распределенной системе — это процесс, в результате которого по множеству ключевых слов, характеризующих сущность распределенной системы, определяется ее идентификатор или имя.

Поиск объекта в пространстве имен Рейзера

Описанная выше задача поиска, то есть нахождение сущностей по набору характеризующих их ключевых слов, может решаться системой имен Рейзера самостоятельно без использования дополнительных поисковых машин.

Поисковые машины могут существовать, но в этом случае они будут выполнять скорее функции интерфейса между поисковой системой и программами или человеком. Кроме того, поисковые машины на базе примитивов системы имен Рейзера могут реализовывать более сложные методы поиска. Здесь рассматривается непосредственное обращение к системе имен, когда поиск выполняется средствами только системы имен.

Рассмотрим частный случай пространства имен Рейзера, когда вся информация о сущностях может быть представлена двухъярусным графом (то есть графом с диаметром 2). На первом ярусе графа располагаются ключевые слова, а на втором — сущности, которые они характеризуют. Тогда, имя любой сущности можно записать в форме $/[k_1 k_2 k_3 \dots k_n]$ где $k_1, k_2,$

$\dots k_n$ — ключевые слова, характеризующие сущность. При разрешении имени будут возвращаться идентификаторы сущностей, соответствующих запросу, т.е. имена сущностей, которые *характеризуются* всеми словами из заданного множества $\{k_1, k_2, \dots, k_n\}$. Этому соответствуют все сущности, в число ключевых слов которых входят искомые слова.

Например, пусть существует множество сущностей $E = \{e_1, e_2, \dots, e_5\}$, характеризующихся словами из множества $K = \{k_1, k_2, \dots, k_6\}$ следующим образом:

$$K_{e_1} = \{k_1, k_2\}$$

$$K_{e_2} = \{k_1, k_3, k_4\}$$

$$K_{e_3} = \{k_3, k_5\}$$

$$K_{e_4} = \{k_3, k_4, k_5\}$$

$$K_{e_5} = \{k_5, k_6\}$$

Взаимосвязь сущностей и ключевых слов может быть представлена графом пространства имен Рейзера, изображенным на рисунке 1.

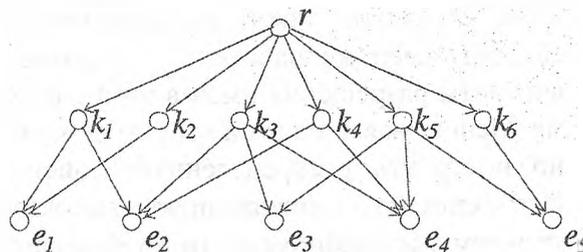


Рис. 1. Пример представления в виде графа пространства имен Рейзера взаимосвязи сущностей и ключевых слов

Пространство имен Рейзера может быть распределенным. Предположим, например, что изображенное на рисунке 1 пространство имен разбито на три зоны: RZ_1 , RZ_2 и RZ_3 (см. рисунок 2). Эти зоны обслуживаются серверами s_1 , s_2 и s_3 соответственно.

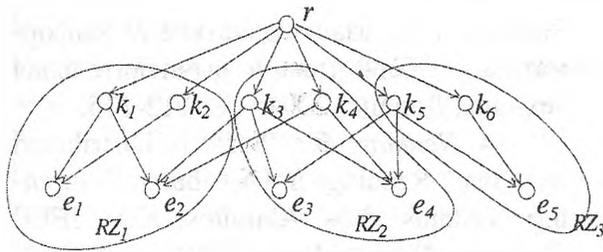


Рис. 2. Пример представления в виде графа распределенного пространства имен Рейзера взаимосвязи сущностей и ключевых слов с разбиением по зонам

Обращение к системе имен выполняется клиентом c , который направляет запрос на разрешение агенту системы имен a , который уже, в свою очередь, выполняет разрешение запроса обращаясь к нужным серверам имен $s2, s3$, в соответствии с алгоритмом, описанным в [3]. Процесс поиска сущности по ключевым словам k_3, k_4, k_5 показан на рисунке 3. Видно, что в результате поиска будет найдена одна сущность e_4 , соответствующая заданным ключевым словам.

Таким образом, система имен Рейзера пригодна для поиска по ключевым словам. В том случае, если пространство имен представлено двухъярусным графом, где ключевые слова занимают первый ярус графа, а искомые сущности второй, поисковый запрос должен формироваться на основе ключевых слов запроса и иметь вид: $[k_1 k_2 k_3 \dots k_n]$. Здесь k_1, k_2, k_3 и т.д. – ключевые слова, характеризующие искомую сущность.

Система сохраняет свою масштабируемость. Как показано в [5], основные характеристики системы определяются величиной коэффициента межзональности, который зависит от количества зон в распределенном пространстве имен (n) и вероятности попадания слов в одну зону p следующим образом:

$$k_1(p) = \frac{2(n-1)p(1-p) + (n-2)(1-p)^2}{(n-1)p^2 + (1-p)^2}$$

Рис. 3 отражает зависимость коэффициента межзональности от количества зон и вероятности попадания слов в одну зону. Здесь еще лучше видно, что на больших количествах зон, характер зависимости коэффициента межзональности от вероятности попадания слов в одну зону практически не меняется.

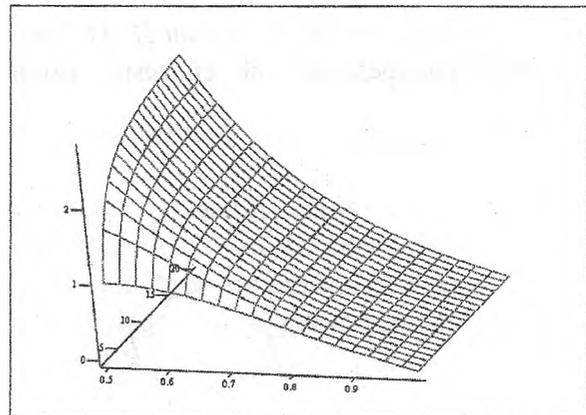


Рис. 3. Зависимость коэффициента межзональности k_1 от количества серверов имен и вероятности принадлежности слов одной зоне (рабочий диапазон вероятности, малое количество серверов имен)

Выводы

Распределенное пространство имен Рейзера может использоваться для поиска сущностей по ключевым словам непосредственно, без применения централизованных поисковых машин. Отказ от централизованных элементов системы повышает уровень ее масштабируемости.

В данной работе рассмотрен поиск в пространстве имен представленном двухъярусным графом. Система имен Рейзера может выполнять поиск в многоярусных пространствах имен, т.е. пространствах имен представляемых многоярусными графами с диаметром больше 2. Разработка методов поиска в многоярусных пространствах имен позволит строить качественно новые поисковые системы, обладающие улучшенными способностями поиска в структурированной информации.

Литература

1. Reiser, H. Future Vision, www.namesys.com

2. Кулаков Ю. А., Чубин И. М.: Разрешение имени в распределенном пространстве имен Рейзера // Параллельные вычисления и задачи управления. Сборник научных трудов, -М.: 2004, - С. 54-57.

3. Кулаков Ю. А., Чубин И. М. Элементы распределенной системы имен

Рейзера и их взаимодействие // Информатика, кибернетика и вычислительная техника, Донецк, 2004. - С. 112-115.

4. Neuman, B.: "Scale in Distributed Systems." Readings in Distributed Computing Systems. Los Alamitos, CA: IEEE Computer Society Press, 1994. - pp. 463-489.

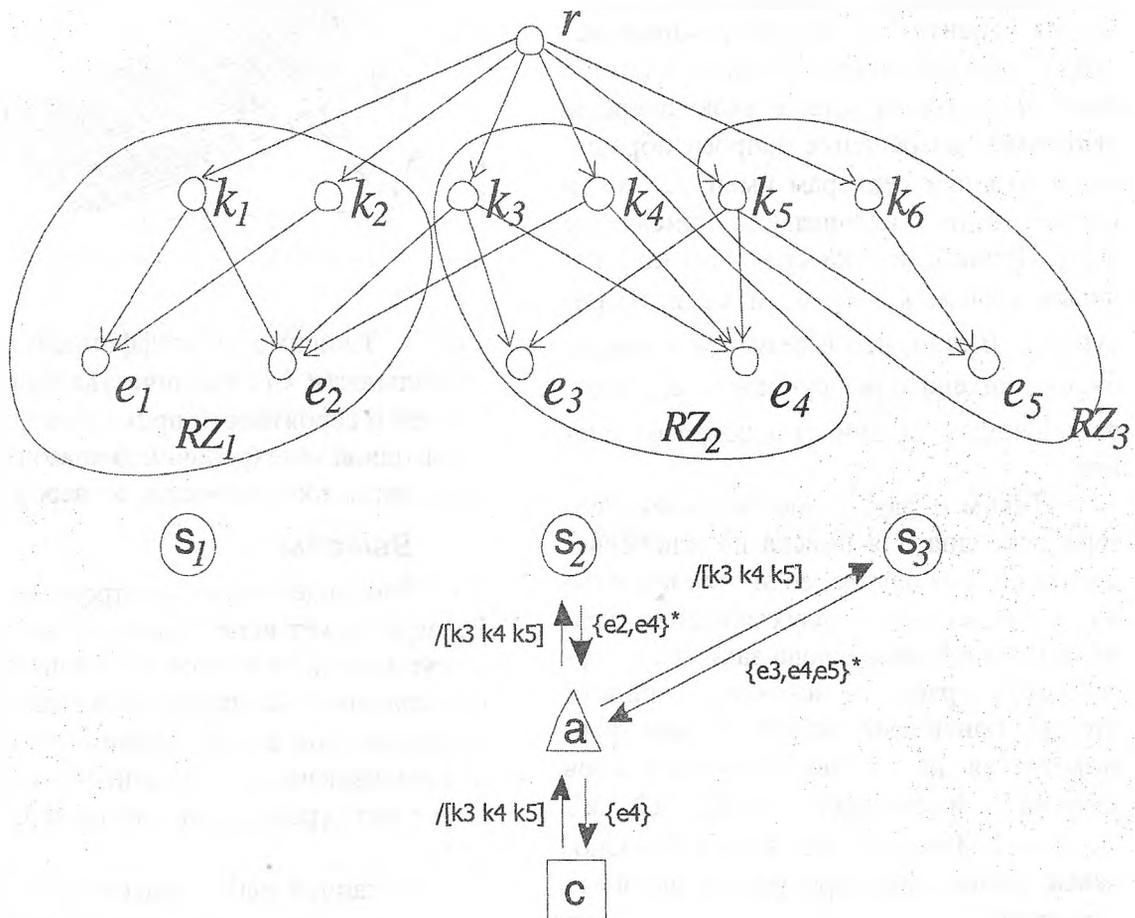


Рис. 3. Взаимодействие элементов распределенной системы имен при поиске