

НЕКОТОРЫЕ РЕКОМЕНДАЦИИ ПО РЕГИСТРАЦИИ ДОМЕННЫХ ИМЕН В СЕТИ INTERNET

к.т.н. А.А. Феклистов, к.т.н. О.Я. Лазарева, А.В. Низовцев

Предлагаются научно - методические рекомендации по определению доменных имен при регистрации их в международной компьютерной сети Internet на основе использования оригинальных алгоритмов лингвистического анализа текстов в заданной предметной области.

Одним из первых видов работ при размещении информации в международной компьютерной сети Internet является определение доменного имени. Важность своевременной регистрации значимого для организации доменного имени приобретает особую актуальность в условиях киберсквоттинга (cybersquatting), то есть покупки (захвата) физическими или юридическими лицами доменных имен, созвучных, например, названий торговых марок, с целью последующей их перепродажи. Кроме того, возможно использование доменного имени для недобросовестной конкуренции, а также хулиганства. Существующая сегодня новая индустрия захвата и перепродажи доменных имен привела к тому, что в доменной зоне “.com” зарегистрировано около 6.000.000 доменных имен (порядка 97 % словарных статей словаря английского языка). Самой дорогой покупкой считается доменное имя <<business.com>>, которое за \$7.500.000 недавно приобрела компания eCompanies. Анализ структуры и содержания некоторых известных доменных имен выявил некоторые закономерности их формирования:

- доменные имена содержательно выражают наиболее емкие и употребительные понятия для данной предметной области;
- доменные имена могут часто являться названиями торговых марок, организаций (компаний), а также их аббревиатурами и слоганами;
- доменное имя может содержать от 3 до 5 структурных компонентов;
- с синтаксической точки зрения доменное имя представляется, как правило, номинативным словосочетанием. В частном случае, это может быть отдельное слово с номинативным значением.

Для «раннего» обнаружения и, следовательно, своевременной регистрации значимых для данной организации доменных имен предлагается выполнить следующие мероприятия.

1. Формирование множества информационных ресурсов по заданной предметной области, например, в виде каталогов, в которых

хранятся локальные копии Web-сайтов партнеров (конкурентов), коммерческие предложения, размещаемые ими в сети, электронные сообщения, а также другая естественно-языковая текстовая информация, представленная в сети.

2. Преобразование файлов информационных ресурсов в ASCII либо Unicode формат. Последний формат особенно актуален для обработки файлов на таких языках как, например, китайский, японский, корейский. В операционной системе Linux данная операция может быть выполнена, например, внешней программой форматирования `lynx`:

$$\$ascii = 'lynx -dump $filename'$$

(для получения файлов в ASCII формате).

3. Последовательное выделение дескрипторов в отдельных файлах, в информационных ресурсах и, в целом, в предметной области. Для автоматизации процесса формирования списка дескрипторов может быть использована специальная методика выделения дескрипторов из текста на основе контекстного и морфологического анализа, использующая словарь квазиокончаний [1]. Сущность методики сводится к выявлению в тексте, рассматриваемом как последовательность лексем, словосочетаний заданной синтаксической структуры. В качестве основных структур предлагается выбирать два типа синтаксических структур: адъективные именные словосочетания и атрибутивные именные словосочетания. Формально (с использованием нотаций БНФ) они могут быть описаны следующим образом:

$$S1 := N | (A + S1); S2 := S1 + [S2'] \dots,$$

где $N := \langle \text{существительное} \rangle | \langle \text{аббревиатура} \rangle | \langle \text{наименование} \rangle | \langle \text{обозначение} \rangle;$

$A := \langle \text{прилагательное} \rangle | \langle \text{причастие} \rangle.$

Элемент $S2'$ обозначает компонент, выраженный генетивной формой (только для флективных языков, например русский, украинский, немецкий). Для других языков $S2'$ совпадает с $S2$. С учетом рекурсивности описанных конструкций каждый из вложенных компонентов может рассматриваться как самостоятельный элемент, имеющий аналогичную структуру. Таким образом, каждое из выделенных словосочетаний «порождает» некоторое множество производных дескрипторов.

Необходимость такой процедуры порождения производных дескрипторов объясняется тем, что потенциально некоторый дескриптор может являться компонентом нескольких различных «порождающих» дескрипторов.

Объём выборки анализируемых текстов определяется пользователем и зависит от наличия у него соответствующих информационных материалов. Достоверность полученных результатов объективно зависит от объема используемых для анализа текстов. Экспериментальные исследо-

вания роста словаря дескрипторов в зависимости от объема выборки для различных предметных областей показали, что насыщение словаря (когда прирост словаря становится меньше 1 %) происходит при объеме текста порядка 600.000 словоупотреблений.

4. Экспертная оценка полученного списка дескрипторов, которая включает в себя удаление из списка общезначимых слов, предлогов, междометий и т. п., а также выделение экспертами в данной предметной области дескрипторов, которые наиболее ёмко и полно характеризуют сущность данной организации.

5. Проверка возможности регистрации доменного имени в базах данных регистраторов доменных имен. Осуществляется, как правило, на основе сервиса WHOIS. Не рекомендуется осуществлять проверку на серверах, которые не являются официальными регистраторами, так как указание адреса в строке ввода может привести к получению потенциальными киберсквоттерами информации о возможном доменном имени. Для наблюдения за системой доменных имен правительством США назначена организация ICANN (Internet Corporation for Assigned Names and Numbers).

6. Регистрация доменного имени (если оно свободно) либо организация работ по передаче регистратора доменного имени (если оно занято). В настоящее время около 80 % рынка регистрации доменных имен принадлежат организациям Network Solutions [1] и Register.com, Inc. [2]. На данном этапе выполняется оформление юридических документов, подтверждающих право собственности нового владельца доменного имени, расчет стоимости выполняемых работ и, непосредственно, проведение мероприятий по регистрации.

7. Настройка параметров DNS-серверов местных провайдеров услуг Internet для размещения и предоставления доступа к информации.

Представленные рекомендации описывают основные мероприятия, которые необходимо выполнить при регистрации доменного имени и могут рассматриваться как основа для проведения дополнительных научно-методических исследований.

ЛИТЕРАТУРА

1. <http://www.networksolutions.com> .
2. <http://www.register.com> .

Поступила 23.05.2002

ФЕКЛИСТОВ Андрей Александрович, канд. техн. наук, нач. НИЛ научного центра при ХВУ. В 1993 году окончил Харьковское высшее военное командно-инженерное училище им. Крылова Н. И. Область научных интересов – искусственный интеллект.

ЛАЗАРЕВА Ольга Ярославна, ст. научный сотрудник научного центра при ХВУ. В 1981 году окончила ХГУ. Область научных интересов – математическая лингвистика.

НИЗОВЦЕВ Андрей Валерьевич, нач. смены ИВЦ ХВУ. В 1986 году окончил Харьковское ВВКИУ. Область научных интересов – ассоциативный поиск неявной информации.