УДК 681.5

В.В. Онищенко

Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков

СПОСОБ СБОРА РЕЛЕВАНТНЫХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СТРУКТУРНО-ЛИНГВИСТИЧЕСКОЙ КЛАССИФИКАЦИИ И ИДЕНТИФИКАЦИИ

В статье рассматривается способ сбора релевантной поисковому запросу графической информации в компьютерных системах и сетях (КСС) с использованием метода структурно-лингвистической классификации и идентификации изображений, отличающийся от известных способов сбора тем, что, наряду с применением традиционных поисковых запросов по ключевым словам, предлагается использовать сортировку списка полученных ссылок на основе реализации процедур распознавания графических объектов, содержащихся в найденных документах.

Ключевые слова: *семантическая сеть распознавания, признаки классификации и идентификации, поиск и сбор изображений, поисковый запрос, релевантность.*

Введение

По мере роста объема информационных ресурсов совершенствуются поисковые средства и приемы, позволяющие найти необходимую информацию. Поиск информации (в частности изображений) и ее последующий сбор является достаточно сложным процессом, имеющим свои определенные подходы.

Создатель всемирной паутины Тим Бернерс Ли предсказал функционирование Интернета как семантической сети, которая позволит анализировать, понимать запросы пользователей и выдавать абсолютно релевантные данные [1].

Таким образом, можно сделать вывод о том, что разработка новых методов и способов сбора релевантной графической информации (изображений) в структурированных, слабоструктурированных и неструктурированных базах данных на основе концепции семантической обработки информации является актуальной научно-технической задачей.

Сбор релевантной графической информации должен быть основан на применении новых подходов в ее индексировании, которые не связаны с заданием текстовых индексов.

Альтернативой индексированию по ключевым словам может служить применение высокоэффективных процедур непосредственного распознавания – классификации и идентификации изображений, найденных традиционными методами. Кроме того, если удастся построить некоторое структурно-лингвистическое описание изображений, более адекватное чем описание ключевыми словами, то это даст возможность создать новую систему индексации изображений, позволяющую осуществлять эффективный сбор релевантных изображений. Подобная индексация может проводиться распределенными поисковыми системами либо разработчиками информационных ресурсов [1, 2].

Для построения структурно-лингвистического описания изображения, соответствующего концепции семантической обработки информации, которое может применяться в системах сбора информации, необходимо использовать методы структурно-лингвистической классификации и идентификации изображений.

Однако, известные методы [3, 4] не позволяют получить подобные описания ввиду того, что существует проблема создания универсальных (применяемых ко всем типам графических объектов распознавания) процедур, осуществляющих построение структур признаков классов распознавания, инвариантных аффинным преобразованиям и деформационным искажениям изображений.

Очевидно, что выделение подобных инвариантных признаков с использованием структурно-лингвистического метода должно сопровождаться преобразованием исходного изображения, представленного в одном из графических форматов, в семантическую структуру на основе применения формализма семантических сетей [5], что полностью соответствует концепции семантического Web-3.

Целью данной статьи является разработка способа сбора релевантной информации в компьютерных системах и сетях с использованием метода структурно-лингвистической классификации и идентификации изображений.

Результаты исследований

Рассмотрим общую схему поиска и сбора изображений в КСС с использованием метода структурно-лингвистической классификации и идентификации изображений (рис. 1).

Процесс поиска начинается с построения семантического запроса на основе анализа существующей глобальной проблемы получения необходимой информации (символ 1).

© В.В. Онищенко

Следующим шагом (символ 2) является формирование соответствующего семантического ключа (СК). Далее СК сопоставляется (символ 3) подструктуре семантической сети (СС), построенной заранее в процессе обучения системы распознавания [5].

На следующем шаге (символ 4) осуществляется выделение семантической сети распознавания (ССР), соответствующей сформированному СК. В случае если ССР не существует (символ 5), выполняется процедура обучения системы распознавания (символ 6). Процесс обучения осуществляется пользователем при помощи корректирующей обратной связи (КОС). Целью процесса обучения является коррекция СС путем синтеза конечных автоматов (КА) для распознавания новых объектов или же уточнение существующих КА (для известных объектов) [6].

Если ССР найдена, выполняется выбор (символ 7) изображений объектов по СК из слабоструктурированной базы данных (ССБД) или же простой выбор в случае неструктурированной базы данных (НСБД). Далее осуществляется непосредственно процесс классификации объектов изображения с использованием структурных и семантических потоков (символ 8). В случае положительного исхода (при достижении КА заключительных состояний) формируется список классифицированных изображений (СКИ). Если не вся БД проанализирована (символ 9), возвращаемся к символу 7, в противном случае анализируем СКИ (символ 10). Если СКИ не сформирован, то выдается сообщение (символ 11) об отсутствии на изображениях объектов, удовлетворяющих семантическому запросу пользователя, формируется пустая выходная страница браузера (символ 24) и алгоритм завершает свою работу.

На следующем шаге (символ 12) выясняется необходимость идентификации классифицированных объектов.

Если идентификация не требуется – выдаются найденные изображения (символ 13), содержащие интересующие объекты, осуществляется процедура сбора релевантных изображений (символ 22), формируются выходные страницы изображений браузера (символ 24), алгоритм завершается.

В противном случае, выбираются изображения и формируются для каждого объекта (символ 14) признаки идентификации (ПИ).

Если признаки идентификации отсутствуют (символ 15), выполняется процедура формирования ПИ (символ 16), осуществляемая пользователем при помощи КОС.

В противном случае, из СКИ выбираются изображения объектов для последующей идентификации (символ 17).

Результатом выполнения данной процедуры будет список идентифицированных изображений (СИИ).

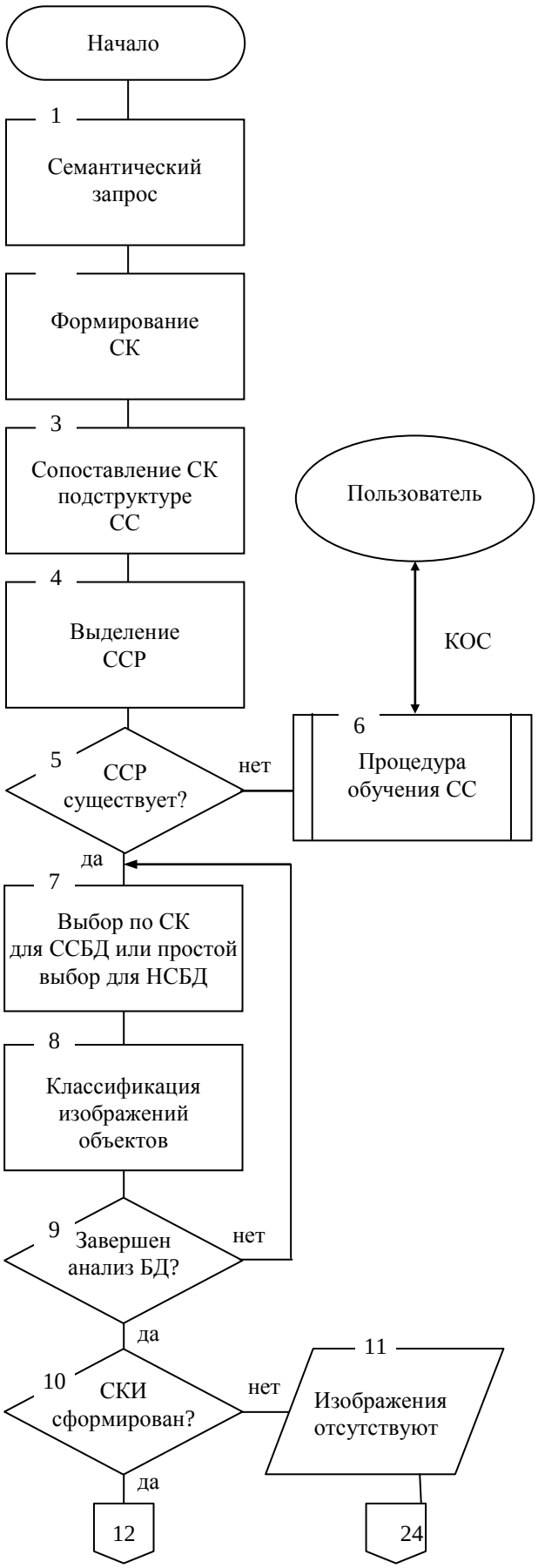

Рис. 1. Схема алгоритма поиска и сбора изображений в КСС

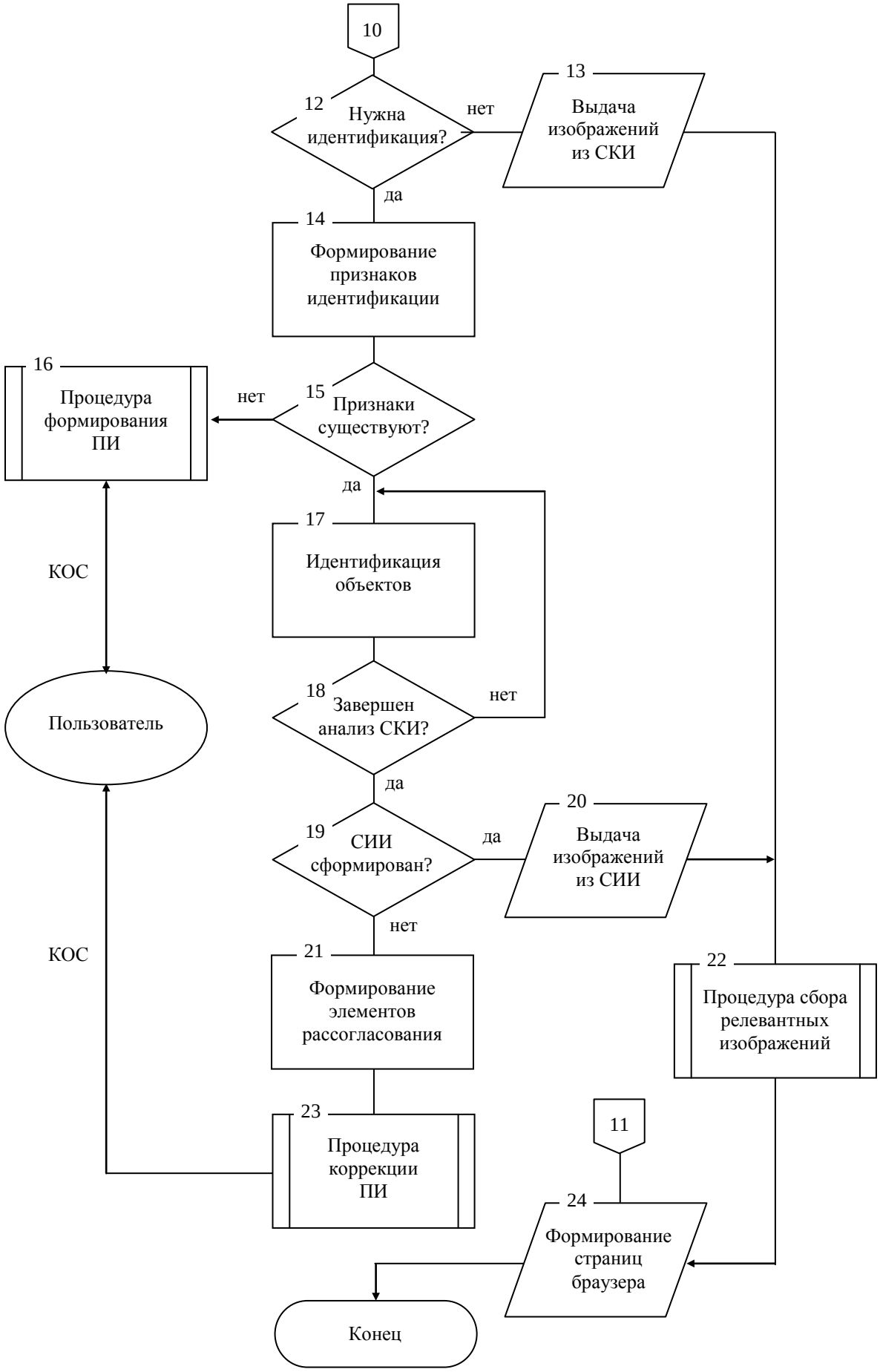

Рис. 1. Схема алгоритма поиска и сбора изображений в КСС (окончание)

Если не все изображения из СКИ (символ 18) проанализированы, управление передается на символ 17, в противном случае проверяется результат формирования СИИ (символ 19).

При положительном исходе – выдаются соответствующие изображения из СИИ (символ 20), осуществляется процедура сбора релевантных изображений (символ 22), формируются выходные страницы изображений браузера (символ 24) и алгоритм завершает свою работу.

В случае отрицательного результата формируются элементы рассогласования (символ 21) и осуществляется коррекция пользователем ПИ с помощью КОС (символ 23).

В общем случае при семантическом анализе поискового запроса должен формироваться состав этапов предварительной обработки изображений, а также осуществляться выбор подхода к распознаванию и используемых при этом методов. Спектр задач, которые приходится решать при распознавании изображений не может быть ограничен использованием какого-либо одного метода.

Выводы

Рассмотренный способ сбора релевантной поисковому запросу графической информации в компьютерных системах и сетях в результате применения метода структурно-лингвистической классификации и идентификации изображений, отличается от известных методов сбора тем, что, наряду с применением традиционных поисковых запросов по ключевым словам, предлагается использовать сортировку списка полученных ссылок на основе реализации процедур распознавания графических объектов, содержащихся в найденных документах.

Кроме того, предлагается в качестве признаков индексации источников применять не только статистические данные и совпадение по ключевым словам, но и структурно-лингвистические концепты классов распознавания, что ускорит и упростит процесс сбора релевантных графических изображений за счет использования распределенных поисковых систем с децентрализованной архитектурой.

Список литературы

1. *Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа: пер. с англ. / Д.В. Ландэ. – М.: Издательский дом "Вильямс", 2005. – 272 с.*
2. *Ландэ Д.В. Поисковые системы: поле боя – семантика / Д.В. Ландэ // Телеком. – 2004. – № 4. – С. 44-50.*
3. *Форсайт Д. Компьютерное зрение. Современный подход: пер. с англ. / Д. Форсайт, Ж. Понс. – М.: Издательский дом "Вильямс", 2004. – 928 с.*
4. *Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2006. – 1072 с.*
5. *Паржин Ю.В. Побудова семантичної мережі структурно-лінгвістичних концептів об'єктів розпізнавання / Ю.В. Паржин, Д.В. Гриньов, В.В. Онищенко // Системи управління, навігації та зв'язку. – К.: ЦНДІ НіУ, 2007. – Вип. 2. – С. 28-30.*
6. *Паржин Ю.В. Построение порождающих грамматик и конечных автоматов для семантической сети структурно-лингвистических концептов в процессе обучения системы распознавания / Ю.В. Паржин, Д.В. Гриньов, В.В. Онищенко // Системи обробки інформації: зб. наук. пр. – Х.: ХУПС, 2007. – Вип. 8 (66). – С. 66-71.*

Поступила в редколлегию 15.09.2010

Рецензент: д-р техн. наук, проф. В.А. Краснобаев, Харьковский национальный технический университет сельского хозяйства им. П. Василенко, Харьков.

СПОСІБ ЗБОРУ РЕЛЕВАНТНИХ ЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ МЕТОДУ СТРУКТУРНО-ЛІНГВІСТИЧНОЇ КЛАСИФІКАЦІЇ І ІДЕНТИФІКАЦІЇ

В.В. Оніщенко

У статті розглядається спосіб збору релевантною пошуковому запиту графічної інформації в комп'ютерних системах і мережах з використанням методу структурно-лінгвістичної класифікації і ідентифікації зображень, що відрізняється від відомих способів збору тим, що, разом із застосуванням традиційних пошукових запитів за ключовими словами, пропонується використовувати сортування списку отриманих посилань на основі реалізації процедур розпізнавання графічних об'єктів, що містяться в знайдених документах.

Ключові слова: *семантична мережа розпізнавання, ознаки класифікації і ідентифікації, пошук і збір зображень, пошуковий запит, релевантність.*

METHOD OF COLLECTION OF RELEVANT IMAGES WITH THE USE OF METHOD OF STRUCTURALLY-LINGUISTIC CLASSIFICATION AND AUTHENTICATION

V.V. Onischenko

In the article the method of collection is examined relevant the searching query of graphic information in the computersystems and networks with the use of method of structurally-linguistic classification and authentication of images, different om the known methods of collection тем, that, along with application of traditional searching queries on keywords, it is suggested to use sorting of list of the got references on the basis of realization of procedures of recognition of graphic objects, contained in the found documents

Keywords: *semantic network of recognition, signs of classification and authentication, search and collection of images, searching query, relevance.*