

УДК 004.42

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИСТЕМ ПОИСКА ГРАФИЧЕСКИХ ДАННЫХ

Шарапов Р.В., Варламов А.Д.

Муromский институт ФГБОУ ВПО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых», Муром, e-mail: mivlgu@mail.ru

В работе проводится сравнительный анализ существующих систем поиска графических данных. Система «Картинки Google» позволяет искать изображения по текстовым меткам и образцу. Поисковые системы «Поиск Yahoo! Картинки», «Bing Изображения» и «Яндекс картинки» позволяют осуществлять поиск по текстовым меткам. В системах «Поиск@Mail.ru картинки» и «Рамблер картинки» используются технологии Google и Яндекс соответственно. Поисковые системы в той или иной степени проводят анализ содержания найденных изображений. Подобный анализ используется для выбора изображений определенной цветовой гаммы, определенного содержания и т.д. Специализированная система «TinEye» осуществляет поиск изображений по заданному образцу. Задача создания поисковых систем, осуществляющих поиск графических данных по заданному образцу, остается открытой.

Ключевые слова: поисковая система, поиск изображений, поиск графических данных, поиск картинок

COMPARATIVE ANALYSIS OF GRAPHIC DATA SEARCH ENGINES

Sharapov R.V., Varlamov A.D.

Murom Institute of Vladimir State University, Murom, e-mail: mivlgu@mail.ru

In this paper we perform a comparative analysis of existing graphical data search engines. System «Google Images» allows you to search images by text labels and by pattern. Search engines «Search Yahoo! Images», «Bing Images» and «Yandex Images» allow you to search for text labels. In systems «Poisk@Mail.ru Images» and «Rambler Images» used technology of Google and Yandex. Search engines analyze the content of found images. A similar analysis is used to select a specific color images, specific content, etc. Specialized system «TinEye» can searches images according to pattern. Task of creating a graphical data search engine by pattern is actual.

Keywords: search engine, image retrieval, image search, graphical data retrieval

Поиск изображений последнее десятилетие стал достаточно популярен. Совершенствование цифровой техники, разрешающей способности мониторов и телевизоров, увеличение скорости передачи данных и ёмкости накопителей привело к неуклонному росту интереса цифровым фотографиям, доступным в сети Интернет. В связи с большим количеством доступных изображений все большую популярность стали приобретать различного вида системы поиска изображений.

Существует несколько направлений поиска по изображениям: поиск по содержанию (найти фотографию человека или изображение автомобиля), поиск по визуальному образцу (найти изображения, похожие на заданное), поиск по описаниям (найти изображение, помеченное заданными ключевыми словами) и т.д. Каждое из направлений поиска имеет свои особенности и сферы применения.

Цель работы – рассмотреть существующие системы поиска графических данных и провести их сравнительный анализ.

Картинки Google

Поиск по картинкам от корпорации Google (images.google.ru) можно считать самым крупным в мире. Поисковая система позволяет искать изображения по всем сайтам, занесенным в базу Google [5, 7].

Система предоставляет два варианта поиска:

- Поиск по текстовым меткам.
- Поиск по образцу.

При поиске по текстовым меткам система находит все изображения, так или иначе помеченные искомым словосочетанием. Для этого анализируются подписи к рисункам, замещающий текст (конструкция alt text), заголовок страницы (tag title), текст вблизи изображения, подписи ссылок на изображение и т.д.

Поиск по образцу можно произвести, указав путь к изображению (в сети Интернет) или выбрав изображение на локальном диске компьютера (или перетащив его в окно поиска). В этом случае Google позволяет найти не только похожие изображения, но и страницы сайтов, содержащие их. Кроме того, поисковая система пытается определить, какие текстовые метки относятся к изображению-образцу. Еще один вариант вызова поиска подобных изображений доступен при наведении указателя мыши на результаты поиска и выбора кнопки «Похожие».

Как дополнение к поиску по образцу, поисковая система поддерживает поиск идентичных изображений с другими размерами. При этом находятся изображения как с теми же самыми пропорциями, так и с обрезкой или расширением краёв.

Результаты поиска могут быть отфильтрованы по ряду параметров:

- Размер изображения (большие, средние, маленькие, больше чем, точный размер).
- Цвет (цветные, чёрно-белые, содержащий преобладающий цвет).
- Тип (лица, фотографии, клип-арт, ч/б рисунки).
- Тема.

- Время появления (за 24 часа, за неделю, за период).

Одной из особенностей поиска по картинкам Google является возможность фильтрации нежелательного контента в результатах поиска. Доступно три режима фильтрации: умеренная, строгая и без фильтрации. Кроме того, доступна опция «Пожаловаться на картинку».

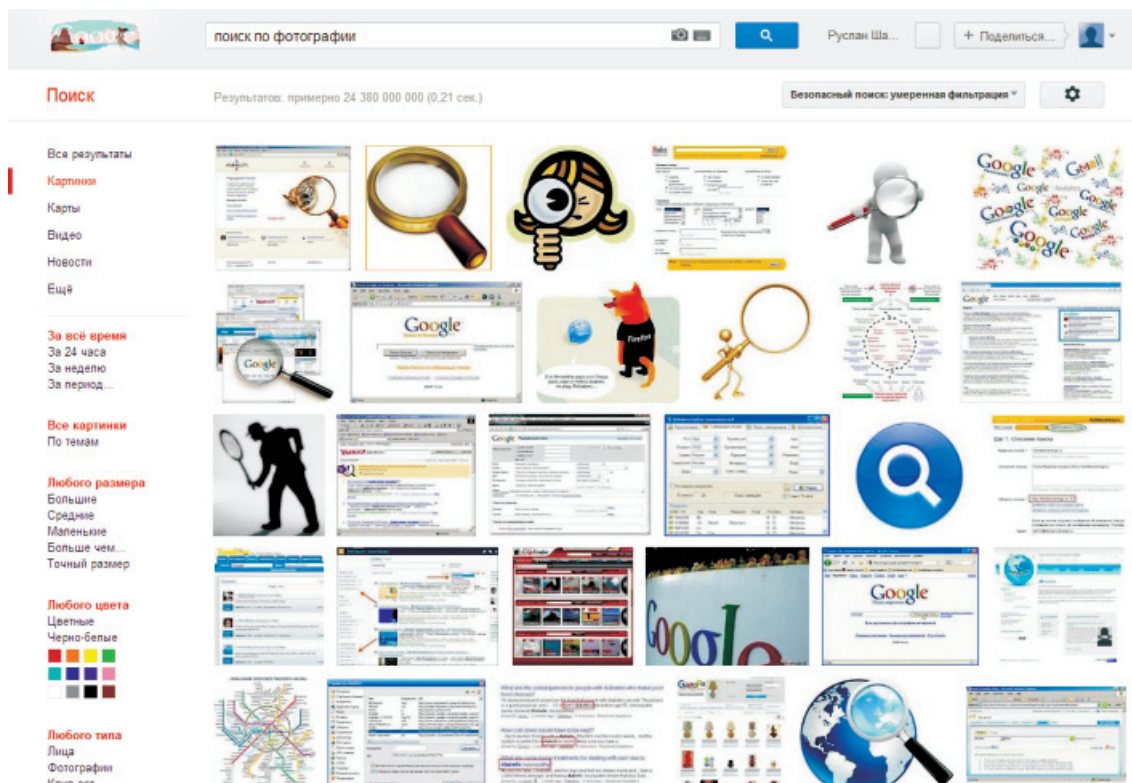


Рис. 1. Представление результатов поиска изображений в поисковой системе Google

Поиск Yahoo! Картинки

Компания Yahoo также предоставляет возможность осуществлять поиск по картинкам (images.search.yahoo.com). Поиск осуществляется только по текстовым меткам [3, 6].

Результаты поиска могут быть отфильтрованы по ряду параметров:

- Время появления (самые новые, все).
- Размер изображения (обои рабочего стола, большие, средние, значки).
- Цвет (цветные, чёрно-белые).

При просмотре найденных изображений сама картинка выдается в верхней части окна, а содержащая её страница – в нижней части.

Несмотря на достаточно скромный набор параметров поиска, сотрудниками и при участии Yahoo опубликовано большое количество работ, посвященных вопросам поиска изображений.

Bing Изображения

Компания Microsoft предоставляет поиск картинок по текстовым меткам в своей поисковой системе Bing (www.bing.com/images) [4, 8].

Результаты поиска могут быть отфильтрованы по ряду параметров:

- Размер изображения (маленькие, средние, большие, фоновые рисунки).
- Цвет (полноцветные, чёрно-белые, точный цвет).
- Стил (фотографии, картинки, рисунки линиями).
- Макет (квадратные, пейзажные, портретные).
- Люди (только лица, портрет по плечи).

Аналогично Google, поисковая система Bing позволяет фильтровать нежелательный контент в результатах поиска (так называемый, безопасный поиск).

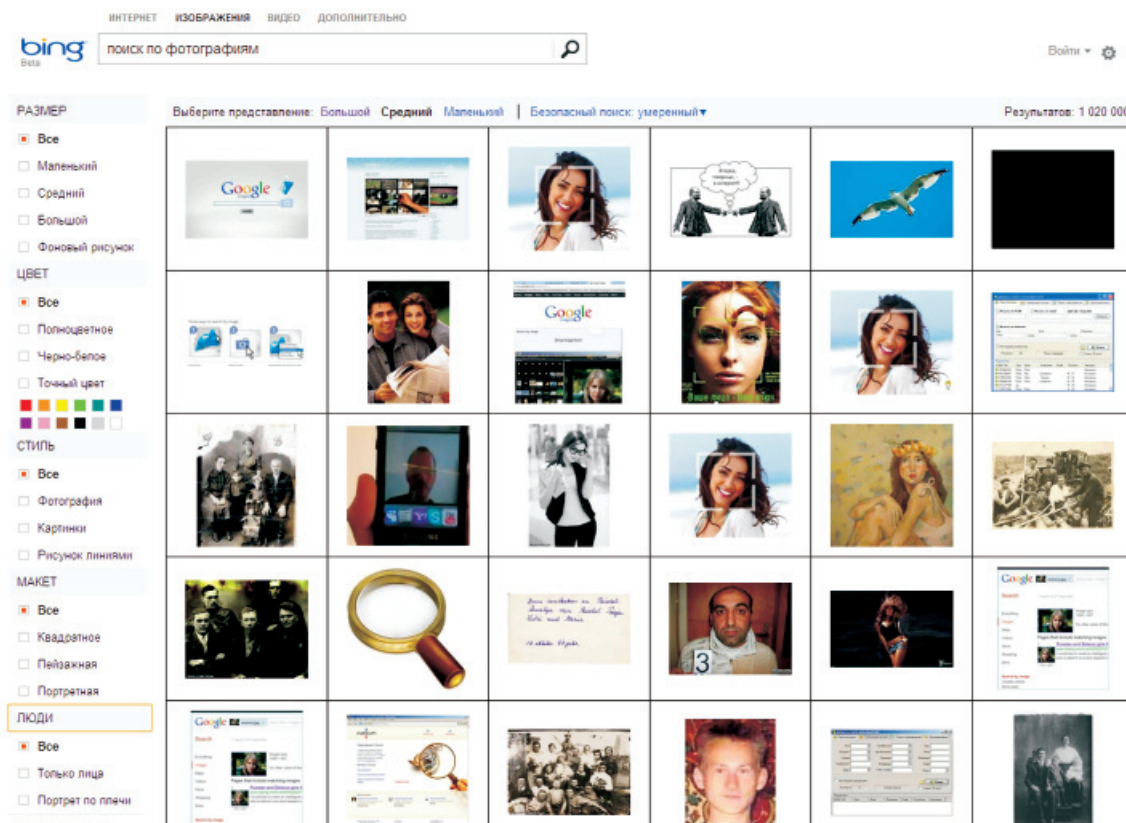


Рис. 2. Представление результатов поиска по изображениям в поисковой системе Bing

Для найденных изображений есть возможность поиска идентичных изображений с другими размерами.

Просмотр найденного изображения осуществляется в верхней части окна, а страницы, его содержащей – в нижней. В отличие от Yahoo для просмотра самой страницы отводится больше места. Сотрудниками компании Microsoft опубликовано большое количество работ, посвященных вопросам поиска и обработки изображений.

Яндекс картинки

Наиболее популярная в России поисковая система Яндекс предоставляет пользователям осуществлять поиск по текстовым меткам (images.yandex.ru).

Результаты поиска могут быть отфильтрованы по следующим параметрам:

- Размер изображения (большие, средние, маленькие, обои).
- Ориентация (горизонтальная, вертикальная, квадратная).
- Тип (фотографии, клипарт, рисунки и чертежи, лица, демотиваторы).
- Цвет (цветные, чёрно-белые, точный цвет).
- Формат файла (Jpeg, Png, Gif).
- Свежесть (за 3 дня, за неделю, за месяц).

- Поиск на определенном сайте.

На результаты поиска можно накладывать фильтры (семейный, умеренный, без ограничений) для ограничения показа нежелательного контента.

Просмотр изображений осуществляется непосредственно в результатах поиска. При этом пользователю выдается подпись изображения, адрес сайта, содержащего изображение, список похожих изображений с другими размерами. Отображение документа, содержащего найденное изображение, не осуществляется.

Поиск@Mail.ru картинки

Портал Mail.ru позволяет своим пользователям осуществлять поиск изображений по текстовым меткам (http://go.mail.ru/search_images).

Набор параметров фильтрации результатов поиска достаточно небольшой:

- Размер изображения (обои, большие, средние, маленькие).
- Цвет (точный цвет).

При просмотре результатов пользователю предоставляется само изображение и краткая информация о нем. Возможности фильтрации содержания от нежелательного контента не предусмотрено.

скими данными. Тем не менее, эти системы в той или иной степени проводят анализ содержания найденных изображений. Подоб-

ный анализ используется для выбора изображений определенной цветовой гаммы, определенного содержания и т.д.

Сравнение функций поисковых систем графических данных

Поисковая система	Поиск по меткам	Анализ содержания	Поиск по образцу	Поиск подобных	Просмотр оригинала	Поиск по дате	Фильтр
Google	+	+	+	+	+	+	+
Yahoo	+	+	—	—	+	+	+
Bing	+	+	—	+	+	—	+
Яндекс	+	+	—	+	—	+	+
Mail.ru *	+	+	—	—	—	—	—
Rambler *	+	+	—	+	—	+	+
TinEye	—	—	+	+	—	—	—

Примечание. * Используют сторонние технологии поиска.

Поиск изображений, похожих на заданный образец, реализован только в поисковой системе Google и специализированной системе TinEye. Тем не менее, в упрощенном виде такой поиск реализован в задаче определения подобных изображений (отличающихся размерами) в таких системах как Яндекс и Bing.

Таким образом, задача создания поисковых систем [1, 2], осуществляющих поиск графических данных по заданному образцу, остается открытой.

СПИСОК ЛИТЕРАТУРЫ

1. Варламов А.Д., Шарапов Р.В. Поиск визуально подобных изображений на основе машинного обучения // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: XIV Всероссийская научная конференция «RCDL/2012». Переславль-Залесский, 15-18 октября 2012 г.: труды конференции – Переславль-Залесский: Изд-во Университет города Переславля, 2012. – С. 152–159.

2. Сидоренко Е.В., Шарапов Р.В. Особенности построения поисковых систем по изображениям // Алгорит-

мы, методы и системы обработки данных. – 2002. – № 7. – С. 107–111.

3. Baeza-Yates R., Ruiz Verschaе R., Castillo C., Hurtado C. Content-Based Image Retrieval and Characterization on Specific Web Collections // Lecture Notes in Computer Science, Dublin, Ireland. – 2004. – Vol. 3115. – P. 189–198.

4. Cai D., He X., Li Z., Ma W., Wen J. Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information // MM'04, October 10–16. – New York, New York, USA, 2004.

5. Cui J., Wen F., Tang X. Real Time Google and Live Image Search Re-ranking // Proceedings of the 16th International Conference on Multimedia (MM'08), Vancouver, British Columbia. – Canada, October 26–31, 2008.

6. Hörster E., Lienhart R., Slaney M. Image retrieval on large-scale image databases // Proceedings of the 6th ACM International Conference on Image and Video Retrieval CIVR'07. – 2007.

7. Wan Y.C., Liu X.B., Bing J., Chen Y.P. Online Image Classifier Learning for Google Image Search Improvement // IEEE ICIA. – 2011.

8. Wu Z., Ke O., Isard M., Sun J. Bundling Features for Large Scale Partial-DuplicateWeb Image Search // CVPR. – 2009. – P. 25–32.