Collaborative Multi-Paradigm Exploratory Search

Michal Tvarožek
Institute of Informatics and Software Engineering
Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
tvarozek@fiit.stuba.sk

Mária Bieliková
Institute of Informatics and Software Engineering
Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
bielik@fiit.stuba.sk

ABSTRACT

New challenges for advanced web search interfaces and visualization tools arise as user needs shift from traditional lookup tasks towards more open ended search activities collectively described as exploratory search. Exploratory search opens new possibilities for employing social aspects for effective information retrieval. We facilitate exploratory search by providing users with an integrated search and navigation interface combining three search paradigms – full text search, view-based (faceted) search and content-based (query-by-example) search. Full text search is used for both domain data and metadata lookup, view-based search allows users to further refine/filter the respective result set, while content-based search orders or biases the results based on their similarity to a given set of sample results.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.5.2 [Information interfaces and presentation (e.g., HCI]: User Interfaces—*Graphical user interfaces (GUI)*; H.5.4 [Information interfaces and presentation (e.g., HCI]: Hypertext/ Hypermedia—*Navigation*

General Terms: Algorithms, Design, Experimentation

Keywords: exploratory search, personalized faceted browsing, collaborative search, navigation, visualization

1. INTRODUCTION

The Web along with many web-based resources have become a ubiquitous source of information and an environment for collaboration and interaction of many users and businesses. While the amount of available information and the quality and capabilities of information search and processing tools are growing at an incredible rate, so do the size and diversity of the Web's user base and the expectations and requirements of individual users.

Furthermore, people who grew up with the Web and the Internet, i.e. the "Net Generation", have a natural understanding of this new ubiquitous environment quite unlike their predecessors [8]. Consequently, they have (radically) new requirements, expectations and modes of operation compared to the previous generation of web users.

Presently, a prominent change is the shift from traditional lookup tasks (e.g., fact retrieval) towards more advanced and open ended learning and investigation tasks (e.g., knowledge acquisition, comparison, aggregation, analysis or planning) collectively described as *exploratory search* [6]. Furthermore, the trend towards more interaction and active (social) participation encourages the combination of approaches from human-computer interaction (HCI), information retrieval (IR) and adaptive web-based systems.

The challenges of managing and searching for information in the growing Web space were also identified within the Web Science initiative. A promising direction is the development of interactive search approaches integrating multiple search paradigms, such as keyword-based search, view-based search and content-based search, with successive focus on personalization, social interaction and overall usability and user experience. Moreover, considering the growing impact of the Semantic Web, search approaches should consider not only search in the actual web content but also metadata lookup.

In this paper, we present our ongoing work on the novel concept of exploratory search that empowers users with an integrated search and navigation interface combining three search paradigms – keyword-based, view-based and content-based search, amended by adaptation based on search behavior of individuals and communities. Full text search is used for both domain data and metadata lookup, faceted search allows users to further refine/filter the respective result set, while querying by example orders or biases the results based on their similarity to a given set of sample results. Our personalized faceted browser constitutes an exploratory search platform focused on search in large and complex metadata spaces represented by ontologies, in line with the view of Nigel Shadbolt, Wendy Hall and Tim Berners-Lee who identified the increasing need for shared semantics and a means for effective visualization and navigation in metadata, i.e. the huge connected RDF graph [11].

2. RELATED WORK

Much work has already been done in each of the respective subfields of IR. Keyword-based search is currently successfully used, e.g., in all major web search engines (e.g., Google, Live Search, Yahoo) thanks to its simplicity and ease of use, while its disadvantages include ambiguity, low expressiveness and the lack of guidance and interaction. Typical search queries are short (one to three words) though their length

WebScience'08, June XXX, 2008, Pittsburgh, Pennsylvania, USA.

varies depending on the domain [4], while advanced search forms are too complex to be practical [12]. Moreover, "guessing" the right keywords is difficult for many users though improvement can be achieved via context-based search where keywords comprise the search context (and the query) [7].

On the other hand, view-based search interactively guides users by presenting them with successive views of the respective information space and showing them the available options for further query refinement. In practice, view-based search is most commonly realized in faceted browsers often used, for example in online shops for product selection. Faceted browsers allow users to formulate queries via navigation by successively selecting metadata terms in a set of available facets, and to interactively browse the corresponding search results. Authors in [16] compare three major faceted browsers developed to explore new possibilities of view-based search – Flamenco, mSpace and RelationBrowser.

mSpace is a domain specific browser of RDF data, which provides users with a projection of high dimensional information spaces into a set of columns (filters) shown in the GUI [15]. These can be manually added, rearranged or removed by users. Column ordering in the GUI affects column contents – the contents of the next column are dynamically determined based on the selection in the previous columns. Flamenco stresses interface design and guides the user through the information seeking process from a high level overview through query refinement and results preview to the exploration of individual results [17]. The BrowseRDF faceted browser supports facet generation from arbitrary RDF data and extends the expressiveness of faceted browsing via additional operators (e.g., existential) [9].

While in Flamenco the facets are static and predefined, users can manually adapt columns in mSpace to match their needs. BrowseRDF automatically identifies facets in source data based on statistical measures, yet does not directly address issues of information overload or interface usability. Both Flamenco and mSpace support keyword-based search over the entire information space, however only mSpace supports keyword-based filtering in facets. Neither approach provides any kind of personalization or user adaptation.

Content-based (multimedia) IR is a interdisciplinary field which encompasses e.g. HCI, browsing and searching paradigms, such as query-by-example (QBE). The current state of the art in content based IR, its research directions and broader implications, are surveyed in [5]. Content-based approaches, such as QBE have been used in multimedia (e.g., image) domains where textual descriptions of instances are sparse, unavailable or inconsistent with user expectations.

While we stress the combination of multiple search paradigms in order to facilitate interactive exploratory search, for practical viability usability, visualization and personalization are also important. For example, graphical visualization of information can significantly improve user understanding and orientation within an information space compared to "classical" text based visualization.

TagSphere is an approach to visual presentation of search results obtained by QBE using collaborative tagging, originally developed for the digital image domain [1]. It divides the user interface into four parts – the tag visualization, the tag menu, the collection area and the suggestion display.

Users construct the search query interactively via QBE by selecting one or more sample images from the suggestion area and moving them to the collection area. Search results are presented visually in sets (denoted by circles) based on their associated tags in the tag visualization area, while the tag menu contains the list of all associated tags. For each set, its similarity to the query (i.e., the set of sample images) is shown via distance and the overlap between a tag search and a classifier search based on low-level image properties.

The annotated visual presentation in TagSphere provides users with a comprehensive overview of the results, as opposed to the plain attribute table most common in faceted browsers. Effectively, TagSphere provides a form of view-based search, though its expressive power is limited since only a list of tags can be used for querying in addition to QBE (i.e., only one flat facet).

Even though the described approaches present improvement in search mechanisms, there is still much space left for combining different approaches together and involving individual users and communities. Changing the way we search for information can change our lives since we would be able to use the new [*social adaptive semantic*] Web as a more effective means for information access and communication.

3. MULTI-PARADIGM EXPLORATORY SEARCH AND VISUALIZATION

We propose an extension of our original personalized faceted browser called Factic [13] using a combination of search paradigms and their extension with advanced personalization and visualization support. Factic provides personalized view-based navigation with support for dynamic facet generation based on data stored in the domain and user ontologies. Our design goal was to further improve search and navigation efficiency and overall user experience by providing greater expressiveness and flexibility in query construction. We build upon existing successful faceted browser interface concepts such as the overall interface layout (see Fig. 1).

Our approach gives users the following capabilities:

- *Keyword-based search* tools embedded into facets provide users with quick access to individual items (facets or restrictions) while also providing traditional full text search capabilities.
- *View-based search* for further query refinement. This includes both the use of facets and restrictions in our faceted browser and the selection of clusters/items in visual search results views (i.e., graphical visualization of search results based on e.g. hierarchical clustering).
- *Query-by-example* works by selecting positive and negative samples from the presented search results. These are then used to filter the result set and provide an ordering of individual information artefacts based on their (dis)similarity with positive/negative examples.
- *Collaborative search* includes personalization, recommendation and annotation of navigation and search results based on an individual's user model with respect to communities of practice or specific user groups.

3.1 Keyword-based search in facets

We extend the facets of "classical" faceted browsers with keyword-based search capabilities in three ways:

- *Full text search* over the entire information space selects the most relevant information artifacts as seen

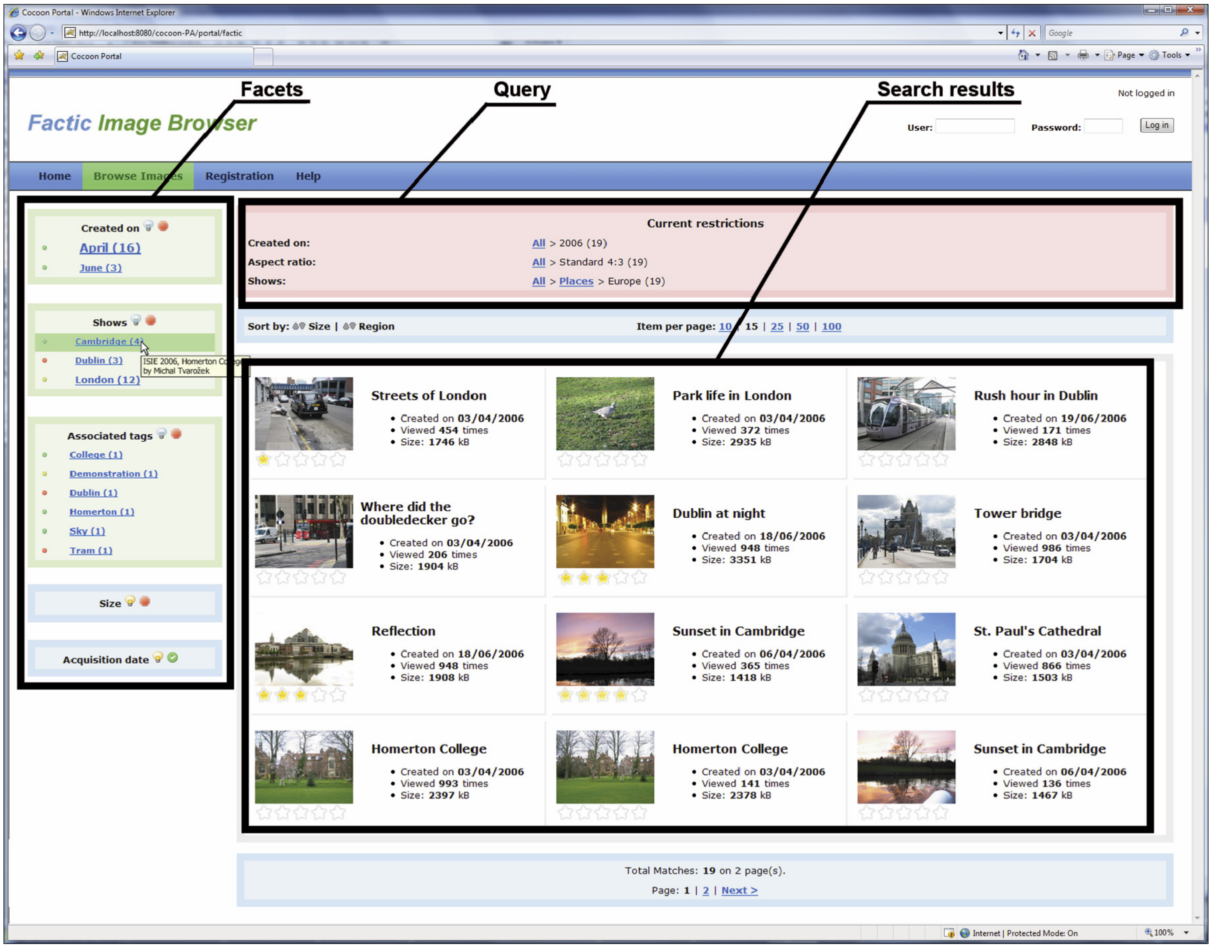

Figure 1: Example of our personalized faceted browser Factic in the digital image domain, employing the general faceted browser layout (facets on the left, query at the top, search results in the centre, optional manual search result customization, e.g. sorting, above search results).

in general web search engines. Though the search can be performed directly on ontological data (e.g., OWL instances), these contain only little textual information. Therefore we apply full text search primarily on documents associated with instances (e.g., full text of papers or job offers instead of ontological metadata).

- *Facet search* locates and makes available facets, which somehow correspond to the entered keywords. For each facet, its label (and potential synonyms), comment and annotations are examined. Optionally, the contents of facets (i.e., restrictions and their labels and comments) can be examined as well.

- *Restriction search* filters restrictions in facets based on a substring search using the supplied keywords. Furthermore, all levels of restriction hierarchy are examined offering quick and effective access also to restrictions normally hidden deep in the restriction hierarchy.

3.2 View-based search result visualization

We provide different personalized interface views for different types of data. The matrix view provides compact information about individual results and can be used to view many results at once (see Fig. 1). However, it is less suited to tasks, which require direct comparisons between attributes of individual results. The table view shows results in an attribute table where columns represent attributes and rows correspond to individual results. In both cases the presented attributes are selected based on user preferences and their estimated discriminative power (i.e., how well they allow users to distinguish "positive" instances from "negative" ones).

Furthermore, we propose a graphical overview of the returned search results based on a hierarchical clustering visualization (see Fig. 2). One central root cluster contains search results satisfying the current faceted query. One or more nested levels of clusters are shown corresponding to search results based either on a hierarchical attribute from the classification or on a custom clustering function. Individual clusters are annotated with short labels summarizing their contents (e.g., most prominent attributes or a summarization). Ultimately, clusters and results correspond to annotated nodes resulting in a form of graph visualization [10]. Examples of instances are shown as tooltips, while the size, relative layout and colour provide further information about instance counts, similarity, relevance (e.g., via social networks) and overall suitability (e.g., via user characteristics).

Apart from offering a visual overview of the whole information space, the cluster view also allows users to further refine their query by selecting a specific information subspace. Thus it works as a "visual facet" which can be constructed based on a (custom/predefined) hierarchical classification.

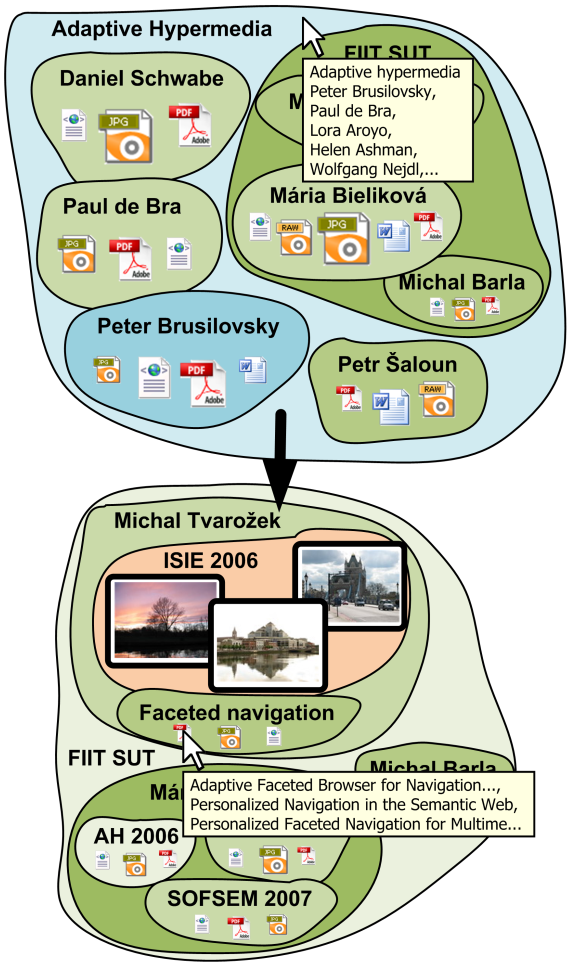

Figure 2: Graphical search result overview visualization via annotated hierarchical clusters on multi-typed data (papers, web documents, images). Background colour indicates relevance, icon sizes denote the result count, cluster distance denotes similarity.

3.3 Query refinement via positive and negative examples

Our original approach allowed users to provide feedback by rating individual search results. This feedback could then be used to personalize the ordering of items using external ordering tools [2]. In practice however, users might rate (positively or negatively) items, which are not related to their current search goal (e.g., they find a movie they like and rate it despite looking for something quite different).

Thus we additionally allow users to select positive and negative examples of instances for QBE search and present these along with the current faceted query at the top of the user interface (see Fig. 1). Subsequently, we filter/reorder the result set – instances that are similar to negative examples are filtered out, while those similar to positive examples are recommended. The overall search result ordering is computed based on an aggregate relevance value determined from QBE similarity (i.e., relevance to the search goals) and instance ratings (i.e., estimated overall user preferences).

3.4 Collaborative search

We facilitate collaborative search features in three ways:

- Implicitly by adding annotations to facets, restrictions and search results describing their use by related users. E.g., relevant community facets are ranked higher, restrictions are recommended showing their usage statistics, while search results include comments by others.
- Explicitly by allowing users to directly browse metadata about specific communities. E.g., only "good" resources related to the Adaptive Hypermedia and HCI communities can be selected, while "outsiders" might impersonate/follow (generic) community members.
- Actively by giving users the means to contribute to the community via comments, ratings and selective usage statistics/navigation tracking and sharing (e.g., recommending a paper to colleagues or writing a review).

3.5 Incremental visual exploration of search result properties

To improve user understanding of individual search results, we proposed three visualizations of instance details for further incremental exploration:

- *Textual attribute visualization* – attributes of instances are displayed in a nested table with a predefined, adaptive or expandable nesting depth used for complex data types, e.g. object properties in OWL ontologies.
- *Graphical neighbourhood visualization* – the context of a search result instance is displayed via a hierarchical cluster view of similar instances based on its properties (e.g., photos with the same author, job offers offered by the same company). Size, colours and relative positions of clusters are used for further annotation along with tooltips as with facets and restrictions.
- *Graphical attribute visualization* – the attributes of the search result instance are displayed as a graph, which can further be expanded to show attributes of associated instances (see Fig. 3). Simple data type attributes are shown directly in nodes, while object type attributes correspond to other nodes connected with each other via edges corresponding to properties (for OWL data). Furthermore, the centre can be moved from the original instance to another instance effectively resulting in a moving window showing the information space. The neighbourhood of individual instances can be shown, similarly to the previous visualization based on instance types (e.g., similar instances of the same class can be shown in hierarchical clusters).

4. CONCLUSIONS

Our main contribution is the enhancement of exploratory search via the combination of keyword-based, view-based and content-based search paradigms using social collaboration principles. We had developed a faceted browser Factic for search and navigation in Semantic Web repositories, and

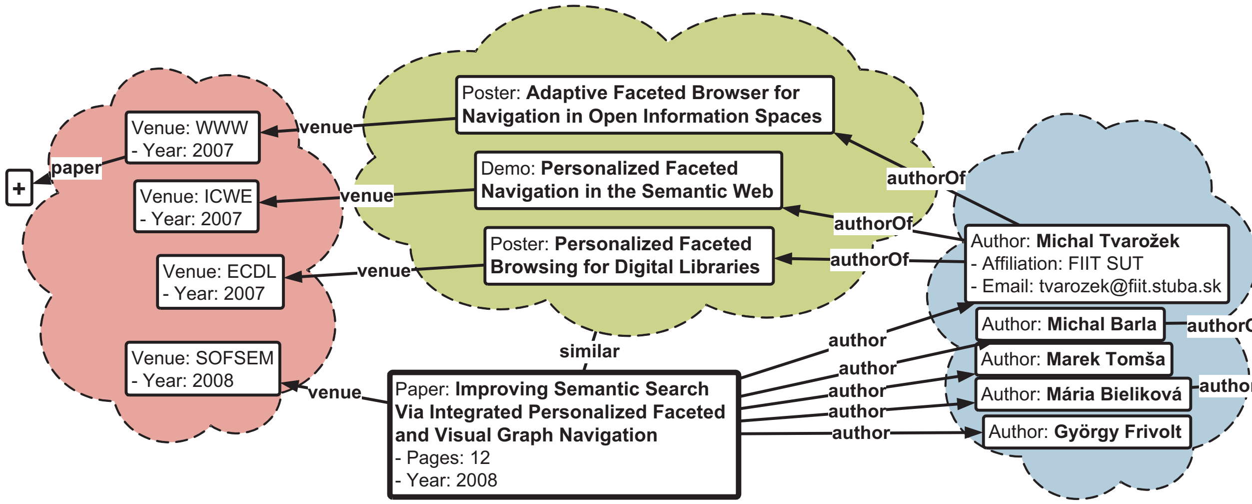

Figure 3: Graphical search result visualization showing domain ontology instances as nodes, object properties as edges between them and simple properties as node attributes. Clouds correspond to classes, the specific attributes and instances are adapted based on estimated user preferences and social networks. For example, the affiliation for authors is shown only for the first author while remaining hidden for the others.

enhanced it with selected features based on the proposed design. While our initial evaluation in three domains (job offers, `nazou.fiit.stuba.sk`; scientific publications and digital images, `mapekus.fiit.stuba.sk`) shows great potential for promising results, further work is needed to fully realize the proposed design and perform more user studies.

Thus, future work will include support for keyword-based and content-based search and improvement of collaborative features (e.g., sharing of data, enhanced user annotation and comments) and one or more comprehensive user studies. Moreover, the use of alternative visualizations for facet contents developed for OWL hierarchy visualization [14] might improve user understanding of deep facet hierarchies.

Acknowledgments

This work was partially supported by the Slovak Research and Development Agency under the contract No. APVT-20-00710 and the State programme of research and development “Establishing of Information Society” under No. 1025/04.

5. REFERENCES

[1] M. Aurnhammer, P. Hanappe, and L. Steels. Augmenting navigation for collaborative tagging with emergent semantics. In I. Cruz et al. [3], pages 58–71.

[2] P. Gurský, T. Horváth, R. Novotný, V. Vaneková, and P. Vojtáš. Upre: User preference based search system. In *Proc. of the Int. Conf. on Web Intelligence 2006*, pages 841–844, Washington, DC, USA, 2006. IEEE.

[3] I. Cruz et al., editor. *Proc. of the 5th Int. Semantic Web Conf.*, volume 4273 of *LNCS*. Springer, 2006.

[4] B. J. Jansen, A. Spink, and J. Pedersen. An analysis of multimedia searching on altavista. In *MIR ’03: Proc. of the 5th Int. Workshop on Multimedia Information Retrieval*, pages 186–192, NY, USA, 2003. ACM.

[5] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2(1):1–19, 2006.

[6] G. Marchionini. Exploratory search: from finding to understanding. *Comm. of the ACM*, 49(4):41–46, 2006.

[7] P. Návrat and T. Taraba. Context search. In *Web Intelligence Workshops*, pages 99–102. IEEE, 2007.

[8] D. Oblinger and J. Oblinger, editors. *Educating the net generation*. Educase, 2005.

[9] E. Oren, R. Delbru, and S. Decker. Extending faceted navigation for rdf data. In Cruz I. et al. [3], pages 559–572.

[10] H.-J. Schulz and H. Schumann. Visualizing Graphs - A Generalized View. *Tenth Int. Conf. on Information Visualisation, IV 2006*, 0:166–173, 2006.

[11] N. Shadbolt, W. Hall, and T. Berners-Lee. The semantic web revisited. *IEEE Intelligent Systems*, 49(May/June):96–101, 2006.

[12] Technical Advisory Service for Images: A Review of Image Search Engines. May 2006, http://www.tasi.ac.uk/resources/searchengines.html.

[13] M. Tvarožek and M. Bieliková. Adaptive faceted browser for navigation in open information spaces. In C. L. Williamson et al., editor, *WWW*, pages 1311–1312. ACM, 2007.

[14] T. D. Wang and B. Parsia. Cropcircles: Topology sensitive visualization of owl class hierarchies. In I. Cruz et al. [3], pages 695–708.

[15] M. L. Wilson and m.c. schraefel. mspace: What do numbers and totals mean in a flexible semantic browser. In *The 3rd Int. Semantic Web User Interaction Workshop at ISWC2006*, 2006.

[16] M. L. Wilson, m.c. schraefel, and R. W. White. Evaluating advanced search interfaces using established information-seeking models. *J. of the American Society for Inf. Sci. and Tech., to appear.*

[17] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proc. of the SIGCHI Conf. on Human factors in Computing Systems*, pages 401–408. ACM Press, 2003.