Advanced Management of Research Publications based on the Lightroom Paradigm

Matthias Geel, Michael Nebeling, Stefania Leone, and Moira C. Norrie

Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
{geel|nebeling|leone|norrie}@inf.ethz.ch

Abstract. Adobe Lightroom is an example of a domain-specific information management tool that has adopted many of the practices that have become well-established within the research community such as automatic metadata extraction, fast keyword-based retrieval and query-based collections for the management of images. It also supports the workflow of professional photographers. We will present a web-based system that transfers some of these advanced concepts to the realm of research publications. The new application introduces sophisticated facilities for classifying publications in different dimensions based on combined notions of flagging, tagging, ratings and colours in addition to manual organisation and sorting in collections. Retrieval and discovery in a potentially large publication corpus is achieved by providing efficient browsing techniques based on a faceted search interface and user-driven categorisation.
Keywords: domain-specific information management tools, faceted browsing, search interface

1 Introduction

Adobe Lightroom[1] is an example of a domain-specific information management tool for digital photographs. Many professional photographers use it to browse and manage large collections of digital photographs as well as performing post-production activities in an efficient way [9]. In order to support the management of thousands of pictures, it has adopted many of the practices that have become well-established within the research community such as automatic metadata extraction, fast keyword-based retrieval and query-based collections. Another popular example is Apple's iTunes[2] which provides similar features for managing music and video files. Both systems incorporate advanced organisation and retrieval techniques considered as emerging research trends in information management and human-computer interaction such as dynamic, faceted search [1, 2, 4, 14], sophisticated tagging systems [3, 11, 12] and the integration of web resources [7].

[1] http://www.adobe.com/products/photoshoplightroom
[2] http://www./apple.com/itunes

We aim to transfer the ideas and concepts introduced by state-of-the-art information management tools for digital media to other domains and different types of personal information. As a first step, we have developed a system for the management of research publications in the form of PDF documents and their associated metadata. Our system provides a faceted search interface similar to Adobe Lightroom, but adapted for effectively browsing research publications by taking into account both available metadata specific to publications as well as user-defined criteria. We will show how the system can support researchers in tasks and workflows related to scientific publishing, for example, when carrying out a literature review in order to collect and manage related work.

We will start with an overview of related work in Sect. 2, followed by a presentation of the use case that motivates the system and supported workflows in Sect. 3. In Sect. 4, we will outline the interaction with the system and describe the key elements of the user interface. Section 5 gives an overview of the implementation and our demonstration is summarised in Sect. 6.

2 Background

As well as supporting advanced means of organising and searching for data, Adobe Lightroom [9] differs from most information management tools in terms of its extensive support for the entire *information workflow* from the capture and storage of images, through the organisation and processing of them, to the publishing of images in a variety of formats such as slideshows and web pages. In addition, images can be published to Web 2.0 sites such as Facebook and Flickr. Existing publication management systems such as Mendeley Desktop or EndNote also provide search facilities and tools for managing publications using categories, labels and favourites. However, in constrast to Lightroom, they do not provide different views for different workflows and neither adapt the user interface to the current task the user is performing, nor do they provide helpful abstractions in different phases of a larger management and review process.

Other systems designed to improve information retrieval through the use of semantic data include Haystack [8] and iMecho [2]. Haystack allows users to store references to arbitrary objects of interest along with any other properties, i.e. attributes and relationships, that they consider to be important. These properties can then be used to support both faceted querying and associative browsing. iMecho takes these ideas further by building associative links between resources from implicit access patterns of user activity sequences. In addition to supporting various kinds of search services, it is important that personal information management systems can help users *organise* their information. The basic hierarchical model of the file system and the desktop metaphor continue to dominate even though many studies have shown that users often struggle to organise their resources in a way that suits their activities [15]. While many alternatives have been proposed over the years, including the document piles of Lifestreams [5] and the collection-based approach of MyLifeBits [6] that allows an information item to be categorised into arbitrary collections, these have had little influence

on desktop systems. However, alternatives to the hierarchical model of organisation can be seen, not only in various Web 2.0 applications such as Flickr, but also within desktop applications such as Adobe Lightroom.

While studies proclaim tagging as a viable substitute for file-based document management [12], so far no adequate user interface solution that supports a transition away from traditional document management has been proposed. One of the predominant problems with manual tagging is the increased effort and cognitive load for users. Facets on the other hand are usually easier to deal with than free-form tags, especially when they can be extracted semi-automatically. Note that facets do not have to be text-based, but can also incorporate colours, groups or numeric values such as ratings. Faceted search is still a very active area of research and several innovative visualisations have been proposed, e.g. [10, 13].

3 System and Use Case

We will illustrate the Lightroom information management paradigm at the core of our system by considering the example of managing scientific publications for a literature review. Researchers typically gather and compile their reading lists of relevant publications from various sources. They then read and classify them according to relevance and other attributes based on personal preferences in order to select the set of most relevant papers for a specific work. Given that a literature review may involve a huge number of publications, a publication management tool that provides a flexible and lightweight approach to efficient categorisation as well as fast searching and retrieval can help achieve that task more efficiently. By adopting the Lightroom model, we have designed the system illustrated in Fig. 1 that supports the management and classification of research publications along multiple degrees of freedom.

Figure 1 shows a screenshot of the application as we have used it to gather related work for this particular paper. The publications managed by the system have been imported from existing BibTeX files so that the system has access to the publication metadata such as *title*, *author*, *booktitle*, *year* and *location*. The tool offers a faceted search bar at the top where the bibliography can be filtered according to various attributes. For example, we can quickly filter and display only those publications that have been published at CAISE in previous years, and then continue to filter by authors as well as a specific year or location.

In the centre, all publications that meet the search criteria are displayed as boxes, with the title at the top followed by an excerpt of the abstract and additional metadata. In our example, we have organised the publications into three collections. The collection on the left-hand side contains all publications that are part of the reading list compiled for this paper. The collection in the middle is a hand-picked selection of publications that are related to the topic of faceted browsing and the collection on the right contains papers about existing personal information management tools. These collections were created directly in the browser simply by dragging and dropping publication entries from one

Fig. 1. Publication system populated with the related work from this paper

collection to another. Note that the faceted search bar on top filters the content of all collections, which is useful when searching across multiple publication libraries.

All publications currently displayed can also be directly manipulated using the in-place editing controls for colouring and rating at the bottom of each box. In our example, we have used these features to colour important publications in red and rate them according to relevance, as can be seen by the stars at the bottom of each entry. Making combined use of multiple and orthogonal classification dimensions allowed us to classify the relevant publications and quickly choose the ones to be referenced in this paper.

4 Lightroom Paradigm for Publication Management

The most important step in designing the proposed system was to decide which concepts from Lightroom are suitable for the management of research publications and how they would translate to the new domain and different metadata. Figure 2 highlights the key features of our search interface that align it with the use of metadata and faceted search in Lightroom, but it also illustrates the changes we had to make when moving from digital photographs to research publications.

Similar to Lightroom, we distinguish between metadata and user-driven attributes. Metadata can be extracted directly from the BibTeX source and define the publications in terms of various dimensions which can then be used as facets in faceted browsing. Suitable candidates for facets are attributes where the same values appear multiple times and ideally cluster the information space into sig-

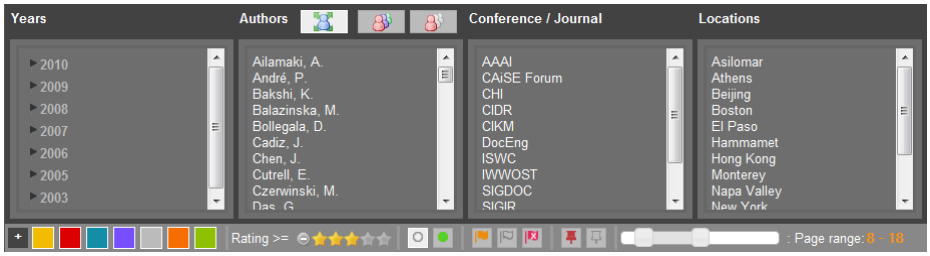

Fig. 2. Faceted search interface for research publications

nificantly large groups. For each of these facets, the user interface offers a list of attribute values inferred from the data that can be selected and combined to perform faceted browsing. In terms of research publications, the candidates we have chosen are *Authors*, *Conference/Journal*, *Locations* and *Year* (top of Fig. 2). Note that the publication title is not a suitable candidate as the title is usually unique and better used in combination with keyword-based retrieval.

On the other hand, user-driven attributes introduce a way for users to classify publications, which allows them to quickly organise, manage and browse their publication corpus. Examples of user-driven attributes in our system include colours, ratings and different kinds of flagging, e.g. to mark read/unread papers. These user-driven attributes can then also be queried with corresponding toggle buttons and sliders that hide or show matching publications (bottom of Fig. 2).

The main advantage of faceted browsing is that users can only perform selections that actually yield results. Additionally, facets can be re-calculated after every refinement to immediately give the user some feedback about the result size of further selections. All selections and restrictions are performed in real-time and the results are immediately shown to the user. When designing the faceted search interface for the publication domain, it was important to consider the types of queries a user may want to perform. Since publications are often the combined effort of more than one author, our system provides several search modes for the *Authors* facet. In the first mode, several authors can be selected simultaneously, selecting all the publications that have been written by any of the selected authors. The second mode is similar to the first one, but performs a conjunction of selected authors resulting in all publications that have been written by selected authors collaboratively. The last mode allows publications to be filtered by first author only.

Another important difference when moving to research publications relates to how Lightroom visualises the contents of photo collections where it makes extensive use of thumbnails that can be directly created from the pictures. For other types of information, it is typically required to dynamically look up or even create thumbnails derived from metadata, e.g. album art for MP3s or individual frames from movies. While, in our case, the generation of thumbnail images from the PDF is technically possible, it was considered impractical as the scaled down version of a text document might be barely readable and often provides

only a snapshot of the first page. In addition, the two-column layout used by many research publications provides very few visual cues that might help users remember a particular publication based on its thumbnail. That is why we opted for an approach where we create an appropriate representation based on the publication's metadata such as title, authors and abstract. In Fig. 3, an example of our visual representation of a single publication is given. These "thumbnails" are enhanced with small, unobtrusive user interface elements which enable users to directly manipulate some key attributes as mentioned before.

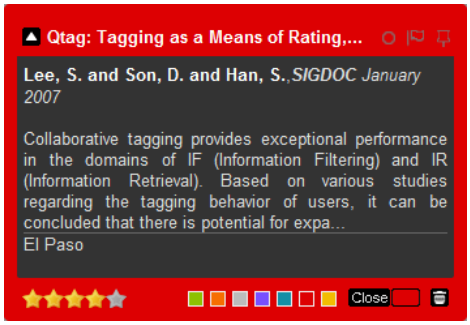

Fig. 3. Visual representation of a single publication with direct attribute manipulation

Finally, our system also supports the creation and maintenance of different collections of research publications. These can, for example, be used to maintain personal or shared reading lists, individual publication lists as well as all publications of a research group. Our solution allows ad-hoc lists to be quickly created that can be laid out spatially as illustrated in Fig. 4. Using this approach, users are free to arrange the publications according to their preferences since individual publications can be moved, not only within a collection, but also between collections using simple drag-and-drop interactions.

5 Implementation

The system is based on a rich client web architecture and was created based on the popular jQuery web framework[3] in order to support rich and responsive interactions as well as a look-and-feel similar to the original *Adobe Lightroom*. The server-side components were implemented based on CakePHP[4]—a PHP web development framework with well-suited abstractions along the MVC design pattern—and are responsible for user account registration and login as well as the storage and retrieval of publication collections. Facet calculation and filtering in the client is based on AJAX and HTML DOM manipulation techniques

[3] http://www.jquery.com
[4] http://www.cakephp.org

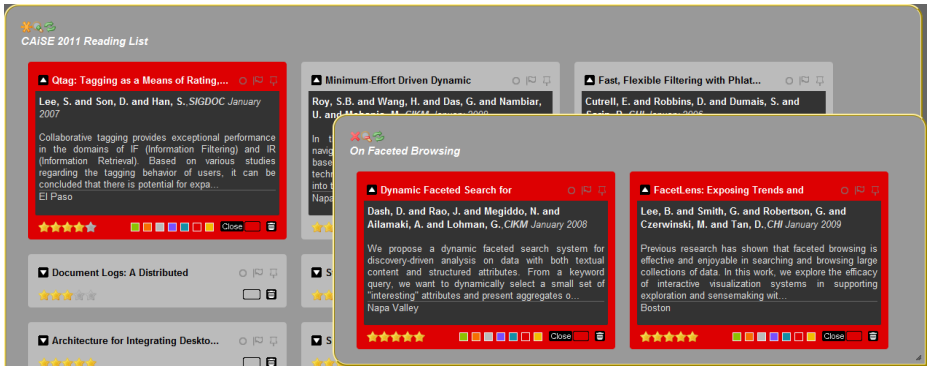

Fig. 4. Management and sorting of research publications in collections

that allow us to reduce loading time of publication entries and dynamically show and hide them according to whether they match the search criteria. We have also developed a number of tools for converting BiBTeX to XML and RSS formats, allowing for easy import of existing bibliographies as well as exporting publication lists managed with our system. Moreover, a lightweight version of the system, with only faceted search rather than publication management capabilities enabled, can be integrated with existing web sites, e.g. those of research groups, and allow visitors to quickly filter and browse publication databases.

6 Demonstration

In our demonstration, we will show the publication management system described in this paper. We will provide several example publication databases, including personal and shared reading lists for different topics, but we will also give the opportunity for interested parties to explore the novel management facilities using their own bibliographies imported from BibTeX. In this way, visitors will be able to experience a Lightroom-like publication management system for themselves and test whether it enables them to browse and search individual or groups of publications more efficiently.

Acknowledgements

This work was supported by the SNF under research grant 200021_121847.

References

1. Basu Roy, S., Wang, H., Das, G., Nambiar, U., Mohania, M.: Minimum-Effort Driven Dynamic Faceted Search in Structured Databases. In: Proc. of 17th ACM Conf. on Information and Knowledge Management (CIKM 2008) (2008)

2. Chen, J., Guo, H., Wu, W., Wang, W.: iMecho: An Associative Memory based Desktop Search System. In: Proc. 18th ACM Conf. on Information and Knowledge Management (CIKM'09) (2009)
3. Cutrell, E., Robbins, D., Dumais, S., Sarin, R.: Fast, Flexible Filtering with Phlat. In: Proc. of Conf. on Human Factors in Computing Systems (CHI 2006) (2006)
4. Dash, D., Rao, J., Megiddo, N., Ailamaki, A., Lohman, G.: Dynamic Faceted Search for Discovery-Driven Analysis. In: Proc. of 17th ACM Conf. on Information and Knowledge Management (CIKM 2008) (2008)
5. Freeman, E., Gelernter, D.: Lifestreams: a Storage Model for Personal Data. SIGMOD Record 25(1) (1996)
6. Gemmell, J., Bell, G., Lueder, R.: MyLifeBits: a Personal Database for Everything. Comm. ACM 49(1) (2006)
7. Groza, T., Handschuh, S., Moeller, K., Grimnes, G., Sauermann, L., Minack, E., Mesnage, C., Jazayeri, M., Reif, G., Gudjonsdottir, R.: The NEPOMUK Project - On the way to the Social Semantic Desktop. In: Proc. I-Semantics'07 (2007)
8. Karger, D., Bakshi, K., Huynh, D., Quan, D., Sinha, V.: Haystack: A General-Purpose Information Management Tool for End Users Based on Semistructured Data. In: Proc. Intl. Conf. on Innovative Data Systems Research (CIDR 2005) (2005)
9. Kim, G.K.: Early Research Strategies in Context: Adobe Photoshop Lightroom. In: Proc. ACM Intl. Conf. on Human-Computer Interaction (CHI Extended Abstracts 2007) (2007)
10. Lee, B., Smith, G., Robertson, G., Czerwinski, M., Tan, D.: FacetLens: Exposing Trends and Relationships to Support Sensemaking within Faceted Datasets. In: Proc. of Conf. on Human Factors in Computing Systems (CHI 2009) (2009)
11. Lee, S., Son, D., Han, S.: Qtag: Tagging as a Means of Rating, Opinion-Expressing, Sharing and Visualizing. In: Proc. of 25th Annual ACM International Conf. on Design of Communication (SIGDOC 2007) (2007)
12. Oleksik, G., Wilson, M., Tashman, C., Mendes Rodrigues, E., Kazai, G., Smyth, G., Milic-Frayling, N., Jones, R.: Lightweight Tagging Expands Information and Activity Management Practices. In: Proc. of Conf. on Human Factors in Computing Systems (CHI 2009) (2009)
13. Smith, G., Czerwinski, M., Meyers, B., Robbins, D., G. Robertson, G., Tan, D.: FacetMap: A Scalable Search and Browse Visualization. IEEE Trans. on Visualization and Computer Graphics 12(5) (2006)
14. Wilson, M., André, P.: Backward Highlighting: Enhancing Faceted Search. In: Proc. of 21st annual ACM Symposium on User Interface Software and Technology (UIST 2008) (2008)
15. Zacchi, A., Shipman, F.: Personal Environment Management. In: Proc. 11th European Conf. on Research and Advanced Technologies for Digital Libraries (ECDL 2007) (2007)