Social Information Access for the Rest of Us: An Exploration of Social YouTube

Maurice Coyle[1], Jill Freyne[1], Peter Brusilovsky[2], and Barry Smyth[1]

[1] School of Computer Science and Informatics,
University College Dublin, Belfield, Dublin 4, Ireland
{maurice.coyle,jill.freyne,barry.smyth}@ucd.ie
[2] University of Pittsburgh, Pittsburgh, USA
peterb@sis.pitt.edu

Abstract. The motivation behind many Information Retrieval systems is to identify and present relevant information to people given their current goals and needs. Learning about user preferences and access patterns recent technologies make it possible to model user information needs and adapt services to meet these needs. In previous work we have presented ASSIST, a general-purpose platform which incorporates various types of social support into existing information access systems and reported on our deployment experience in a highly goal driven environment (ACM Digital Library). In this work we present our experiences in applying ASSIST to a domain where goals are less focused and where casual exploration is more dominant; YouTube. We present a general study of YouTube access patterns and detail how the ASSIST architecture affected the access patterns of users in this domain.

1 Introduction

For many people, access to online information is a pervasive feature of everyday life. A recent report by ComScore [1] found that more than 61 billion searches were carried out in August 2007 (with Google search properties accounting for over 37 billion of these searches). At the same time, however, users are frequently finding it increasingly difficult to access the right information at the right time; for example, recent research points to search failure rates of 50% [10]. Personalization and recommendation techniques are often proposed as potential solutions to these information access difficulties. By learning about users' preferences and interests, these technologies make it possible to model their information needs with a view to adapting modern information services in response to these needs. To date there has been considerable research when it comes to the development of the "algorithmics of recommendation" (the development of the core algorithms that underpin recommendation engines) but relatively little attention has been paid to the interfaces that are needed to deliver recommendations to end-users.

Recently we have developed ASSIST, a general-purpose platform that can be used to incorporate various types of *social cues* into existing information access systems. ASSIST is a proxy-based architecture that facilitates the tracking of

user information access requests in order to build a repository of community preferences that can then be used to enhance the interactions of future users as they search and navigate for information. ASSIST capitalizes on two earlier streams of research in the field of social information access: social search and social navigation. Social search systems such as AntWorld [7], I-SPY [11], or SERF [6] archived past search successes of their users to recommend relevant resources to the future users, who are looking for similar information. Social navigation support systems such as Footprints [12], CoWeb [4], and Knowledge Sea II [2] archived past browsing traces of their users and visualized them to help new users to make navigation decisions. ASSIST is, however, unique among these systems in that it archives and integrates both navigation and search cues to provide users with information access *hints* that reflect the past experiences of other users. Combining the search and browsing experiences of past users allows ASSIST to generate more reliable social recommendations and provide coherent support in the context of integrated search and browsing [9].

To date a number of studies have been conducted to evaluate the benefits of ASSIST in the context of traditional information access services. For example, in [5] we focus on the application of ASSIST to the ACM Digital Library, where users search and navigate for particular items of information that fulfill specific information needs; in this context ASSIST is used in the support of *goal-driven* information access. This type of access is typical for information systems aimed at professional users; however, it is not a dominant form of information access for the average Web user. Nowadays, the majority of Web information systems such as Web stores, news agencies, and entertainment services, strive to support both goal-driven and *exploratory* information access. It reflects the fact that many of their users do not have a specific information goal in mind and thus are not in a position to express a detailed information need. To support both kinds of information access, modern systems pay attention to both search and browsing support, which, as we expected provide a good application context for our framework. YouTube serves as an excellent example of a modern Web information system. It supports search, but also provides many opportunities for exploration through several types of featured videos connected to the system home page and rich opportunities to navigate from one video to related videos.

In this paper, we describe our attempt to explore the applicability of the ASSIST framework in this context. We chose YouTube as our target and implemented the ASSIST-YouTube proxy, which provides social recommendation for YouTube users. In the following sections, we describe how the ASSIST platform can be used to capture user interactions and the ways in which this interaction data can be used to adapt the YouTube interface. We also present the results of a recent live-user trial of ASSIST-YouTube. In particular we will examine the influence that ASSIST's social recommendations have on the manner in which YouTube users search for information and how they explore the YouTube information space. We will also consider how these results differ from those found for the application of ASSIST to the ACM Digital Library, highlighting a number of

features that emphasise important differences between the interaction patterns that are commonplace in these different information access scenarios.

2 ASSIST Engine

ASSIST is a proxy-based architecture for social information access. It resides between a user and an information system (such as YouTube), intercepts user requests to an information repository, and enhances the source of the returned pages with social guidance features such as re-ranking lists of *related videos* or recommending a specific item to the user (Figure 2). A store of past user interactions with the system maintained by ASSIST constitutes a store of "community wisdom", which is used to bring forward content to users in order to allow more informed relevance judgements to be made and to recommend content from beyond the current page. For reasons of space we have omitted many of the technical details of the ASSIST architecture; for more details please refer to [3].

2.1 Monitoring User Interactions

ASSIST records three different types of click behaviour: search result selections, simple browsing clicks, and contextual browsing clicks by capturing implicit relevancy feedback through user click behaviour within the system. A search result selection occurs each time a user selects a result, I_T, from a result-list generated by ASSIST in response to a query. A record of this selection is noted in the search hit-matrix maintained by ASSIST (Figure 1). A simple browsing click occurs when a user views a video whose link is contained within a "Featured" or
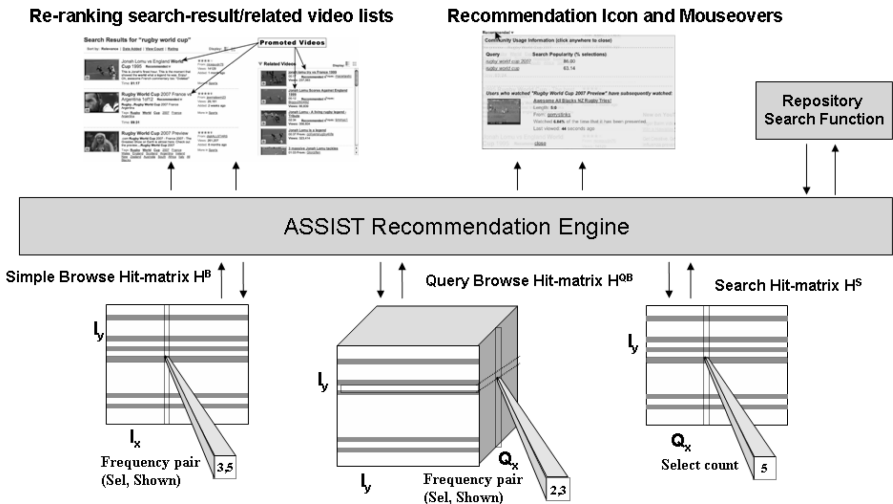

Fig. 1. ASSIST Architecture

"Most rated" list. Contextual browsing clicks are clicks which are made in the context of a previous selection or query submission. An item I_T is considered to have an associated item context I_C if I_C contains a hyperlink to I_T and a contextual click is recorded if a user follows the link. An item I_T is seen to have a *query browsing context*, Q_C and a *browsing item context* I_C if the path being navigated by the user started at the result-list for q_C and the user has navigated away from the result-list to an item I_C which contains a hyperlink to I_T.

2.2   Exploiting User Interactions for Recommendations

ASSIST helps the users of an information system by providing both active and passive social guidance (recommendations). Both types of guidance provided by the ASSIST engine are based on the past search and browsing interactions of community members. In the context of the YouTube system, social guidance provided by ASSIST-YouTube offers a number of enhancements to the standard YouTube interface, providing both improved search and browsing capabilities.

To provide active recommendations, ASSIST re-ranks lists of videos offered by YouTube to reflect accumulated community preferences. In the search context, ASSIST re-ranks the search results returned by YouTube in response to a user query Q according to their *relevance* to Q. ASSIST leverages the search hit-matrix data to assign relevance scores to videos based on their search interactions. The relevance of video item I to query Q can be calculated by calculating the number of times I has been selected in response to Q as a percentage of the total number of selections across all items for Q. ASSIST also identifies videos which have been selected for *similar* queries (using a simple term-overlap similarity metric) and weights their relevance to their associated query Q_i by the similarity of Q_i to Q. These *promotion candidates* are ranked according to their weighted relevance score and placed at the top of the result-list for the query Q.

In a browsing context, ASSIST re-ranks the list of YouTube generated related videos which are displayed alongside a video which is being watched. This list is a valuable source of complementary content for engaging in browsing activities and thus the position of videos within the list is important. ASSIST reranks the related video list according to the items *contextual browsing popularity*.

To provide passive recommendation, ASSIST augments hyperlinks to content with visual *social cues* throughout the interface, highlighting areas of interest and suggesting paths through the space. The presence of these cues indicates previous encounters by community members with the content behind the link. If the user mouses over the icon, they are presented with the item's search and browsing history with community members (Figure 2). The search history information in a mouseover aims to convey to users that the associated content has been chosen by a community member in relation to a query and also the strength of the item-query relationship (i.e. the relevance score). This mouseover includes a list of all queries which have led to the selection of the video (see Figure 2). Users may click on these queries to commence a new search, which essentially allows them to query YouTube for 'more videos like this' with very little effort. The query list is ordered by the strength of the item-query relationship.

Fig. 2. By mousing over the 'recommended' icon presented alongside recommended videos, users may access recommendations for queries or additional video content

The mouseovers are also used to provide the user with *contextual recommendations* to provide the user with an Amazon-style "*users who watched this video subsequently watched these*" feature. As mentioned in Section 2.1, if previous users engaged in browsing behaviour after viewing a particular video (i.e. they selected a related video), this fact is recorded in the browse hit matrices. By recommending videos that were subsequently watched in the mouseover provided alongside a hyperlink, the user may choose to skip watching the top-level video and go straight to one of the recommended videos.

3   ASSIST-YouTube User Studies

In the context of the ASSIST-YouTube project we ran two user studies. The first study was a short 7 week monitoring of students using the official YouTube video sharing site (http://www.youtube.com). This study was used for data collection and observation and was performed before ASSIST-YouTube was implemented, to analyse activity patterns in YouTube. The results of the study were reported in [3]. The data, amongst other findings, uncovered the presence of long navigational trails through the repository, motivating the need for social support in the domain. The second study, reported below, attempted to assess the effect of social guidance provided by ASSIST-YouTube. This study took place over 14 weeks in the winter semester of 2007. The trial monitored 21 participants from the School of Computer Science and Informatics in University College Dublin in their regular activities with YouTube. All participants communicated with

YouTube through the ASSIST-YouTube proxy server, which enhanced their interaction with active and passive social support as described above. This study pursued two goals. First, we were interested in analyzing patterns of user interaction in YouTube and investigating the need for social support beyond the results of our earlier smaller-scale study. Second, we wanted to investigate how the social support provided by ASSIST-YouTube influenced user interaction with YouTube. The following two sections address each of these issues.

4   YouTube Usage Analysis

The most important thing, which we discovered when analyzing YouTube user logs is the differing nature of YouTube usage in comparison with more traditional information systems, such as the ACM Digital Library, which we explored in the process of evaluation of ASSIST [5]. A typical ASSIST-ACM user came to the ACM Digital Library with a reasonably well-defined goal in mind: to find papers on a specific topic. The vast majority of user sessions started with search, while browsing from a paper to related papers was most popular as a search follow-up. This was the context for which the original ASSIST system was engineered.

As we discovered, a similar type of access (searching for a video on a specific topic or with specific features) happens in YouTube as well, but it accounts only for one (and by far not the dominant) type of YouTube usage. Out of 1230 sessions recorded in the ASSIST-YouTube logs, only 366 (i.e, less than 30%) started with search. These goal-directed sessions displayed similar characteristics to those observed in ASSIST-ACM: searching for videos, which matched their goal, the users were eager to examine related videos creating session trails. 47% of sessions initiated by search activity resulted in a trail and the average length of the trail was quite considerable (3.07 clicks). To clarify, the submission of a search query was not counted as part of the trail.

The majority of sessions (864) started directly with video browsing. Most of these sessions are unlikely to have been driven by a specific goal (i.e. an attempt to find a video on a specific topic). Only 125 of these browsing sessions resulted in a trail and the average trail length was shorter than for search-initiated sessions (2.69 clicks). The remaining 739 browsing sessions (> half of all sessions) produced no trail. To shed light on the nature of user browsing-initiated activity in YouTube, we classified all sessions which started with browsing by the origin of the session. The data uncovered another eye-opening fact: 47% of browsing-initiated sessions (or about 30% of all sessions) were external accesses to YouTube through a specific video URL. These URLs can be considered as direct social recommendations, which the user most likely received in an email from another user (44%) or found on such social sites such as FaceBook, MySpace and Bebo, inside blogs and other sites (3%), which allow users to embed videos on their pages. This shows the collaborative nature of YouTube and highlights the social recommendation potential in this context.

The remaining browsing-initiated activity could be classified as casual browsing. Here the users were not trying to find something specific, but were simply trying to watch interesting videos with no apparent goal in mind. Surprisingly,

user casual browsing was not dominated by exploring links, which were specifically engineered by YouTube to support this kind of browsing (what's being watched right now, featured videos and the menus on the videos page). These links only contribute to 11% of clicks showing their relatively low value to our users. These menu lists are generated by YouTube as recommendations to all of their users (in the case of the featured and directors videos) and as a response to general popularity figures (in the case of "what's being watched right now" and the popular links on the videos page). The content of these lists inspired our users less frequently than might be expected, which is another fact motivating the need for support at the level of communities and groups in YouTube. At the same time, the users were quite eager to follow various kinds of related links from the videos they liked. For example, 7% of all browsing sessions were started by follow-up links shown by YouTube at the end of watching a movie. It hints that navigating through related videos is a valuable approach not only for goal-directed, but also for casual browsing.

Overall, our analysis uncovered three major types of user behavior in YouTube: traditional goal-directed search, direct browsing (following an externally recommended link) and casual browsing (watching interesting, but not specific videos). While the ASSIST-YouTube social recommendation engine was designed to assist only the first type of activity, the nature of its browsing support component makes it also quite useful for social support of casual browsing. However, social support of casual browsing may be more challenging than social support of goal-directed browsing. While the search goals of the users of a specific community have some reasonable overlap [10], their casual browsing is driven by their general interests, not goals. Since these general interests could be much more diverse even in a small community, it may be hard to expect that users in a small community will see movies recommended by other community members during their search and browsing (the context supported by ASSIST-YouTube). Indeed, the users in our trial watched 1257 unique videos 2027 times in total. It gives a relatively low watch repetition rate of 1.6. To leverage social support in this context, alternative social recommendation tools for casual browsing should be considered, such as a list of recent popular movies in the community. The next section will analyse to what extent we can demonstrate the success of social navigation support in a relatively small group in this new context.

5   Socially Supported Exploration in YouTube

In previous trials using ASSIST [5] we examined the speed, accuracy and effort exerted of the users using the traditional versus the socially enhanced versions of the repository in question. We also set the users specific tasks which were representative of how the systems are generally used and monitored their performance in terms of the task at hand. The emphasis of this trial differed in two ways. Firstly we moved from a focus driven environment into a more leisure-oriented environment which was undoubtedly going to produce differing results and secondly we allowed users to use the system without setting an agenda or task to be

completed. In order to observe the user in their natural interaction mode with the YouTube system we opted for non disruptive feedback methodologies and opted to consider two implicit indicators *success rates* and *view percentages* when comparing the performance of YouTube and ASSIST-YouTube. When measuring the quality of ranked lists of videos, we calculate the *success rate* of a set of lists to be the percentage of lists that had at least one item selected (that is a list is *successful* if the user finds at least one apparently interesting item). The second metric of performance is the *view percentage* of a video, since it seems likely that the proportion of a video that is watched by a user could be used as a proxy for the user's opinion of the video content. We will also examine the effort exerted by users as they navigate, with comment on how this compares to ASSIST deployments in a more goal-focussed domain.

Effects of Social Recommendation on Search. Figure 3(a) graphs the success rates for different types of search sessions; we can see that search sessions that had some results with social explanations attached (*Expln*) had at least 1 result selected 11% more often than sessions without explanations (!*Expln*). When we examine sessions in which the searcher accessed more detailed explanatory information by mousing over an icon (*MO*), we see that the difference is more pronounced still, with a 31% increase observed. These findings suggest that
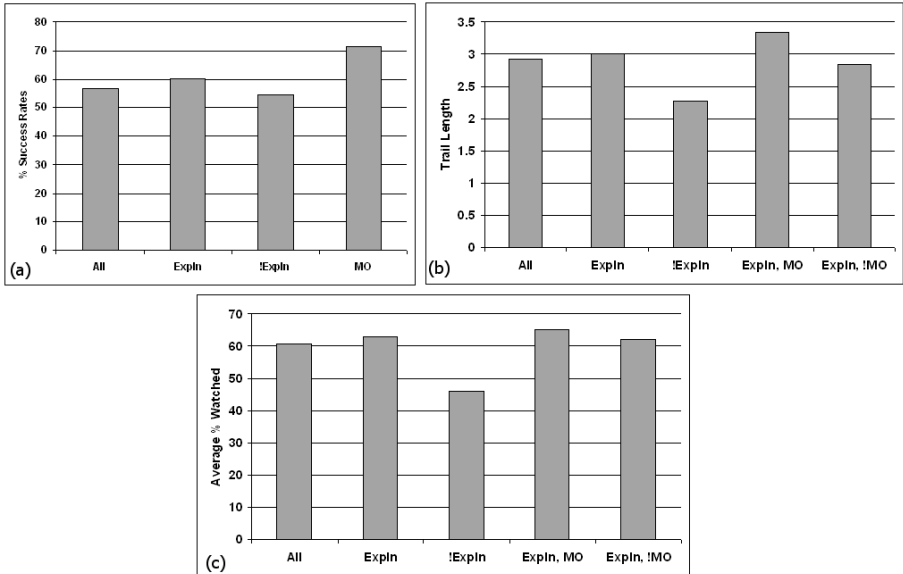

Fig. 3. (a)Success rates of search sessions with explanations (*Expln*), without explanations (!*Expln*) and with explanation icon mouseovers (*MO*). (b) Average trail length when explanations are encountered (c) Average percentage of videos along a trail that were watched when explanations were encountered.

augmenting search result lists in the YouTube repository with social recommendations results in a greater likelihood of the user finding a video of interest and speaks to the utility of reusing such community preference information.

Effects of Social Recommendation on Exploratory Behaviour. In an environment such as the ACM DL where users are likely to be searching for a specific paper or a number of papers on a given topic, we can assume that users want to get to these articles as quickly possible. Indeed, the ASSIST-ACM system received positive qualitative feedback regarding the ease and speed of finding information while empirical data showed an overall reduction in user effort both in search and browsing contexts (see Farzan et al. [5]). However, as we can see from Figure 3(b), ASSIST-YouTube caused users to exert *more* effort when navigating through the site. For navigation trails which had explanations presented along the way, users travelled 33% further on average than they did in trails where no explanations were present. Indeed, when the user moused over an explanation icon at some point in a trail, they browsed $\sim 47\%$ further on average than on trails with no explanations. To investigate the reason for this increase in effort exerted, we turn our attention to the average view percentages for different trail types, as graphed in Figure 3(c). As we can see, the presence of explanations along a trail corresponds to an increase in view percentages over trails with no explanations of $\sim 37\%$, on average, with a slightly higher increase ($\sim 42\%$) observed when a mouseover occurs along the trail. Assuming that users watch more of videos that they are interested in, we can claim that the presence of ASSIST's social enhancements improves the user experience by aiding them in selecting videos which they are likely to watch more of.

6   Conclusions

Evaluating ASSIST within a multimedia site with a casual and leisure-oriented focus has enriched our insight into the value of social support in information spaces. The results of the current study suggest that the effects and type of social enhancements have to be engineered to match the user task and target repository. The original social support in ASSIST was engineered for a goal-driven search task which needs to be done with minimal effort versus an entertaining exploratory task with fewer time constraints. Social support of casual browsing may require some modifications to the ASSIST approach. For example, a high number of items examined in a goal-driven search task could reflect user dissatisfaction. In the context of casual browsing, a high number of examined items can reflect continued user interest in the retrieved items. In addition, the need to support casual browsing calls for additional social recommendation tools such as the most popular or currently watched videos within a community.

Our work also demonstrated that the users are eager to share links to interesting videos with others as well as to follow such direct social recommendations. While the majority of modern research focuses on indirect recommendations, our data shows that in a multimedia domain it could be wise to return to the roots of collaborative recommender system research [8] and to embed a direct

recommendation mechanism into the system. In addition to the convenience provided to the recommending users, it allows the exploring user to take advantage of the recommendation in the context of using the system and while spending time with the system. As our result suggests, receiving a video through an email or external site can interfere with the user's current task and will result in less exploration of the system while social recommendation provided in the system encourages even higher exploration. We hope to address some of these ideas in our future research with ASSIST-YouTube.

References

1. 61 Billion Searches Conducted Worldwide in August (2008),
   `http://www.comscore.com/press/release.asp?press=1802`
2. Brusilovsky, P., Chavan, G., Farzan, R.: Social Adaptive Navigation Support for Open Corpus Electronic Textbooks. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137. Springer, Heidelberg (2004)
3. Coyle, M., Freyne, J., Farzan, R., Smyth, B., Brusilovsky, P.: Reducing Click Distance through Social Adaptive Interfacing. In: Proceedings ReColl 2008 - International Workshop on Recommendation and Collaboration. In conjunction with International Conference on Intelligent User Interfaces (2008)
4. Dieberger, A., Guzdial, M.: CoWeb–Experiences with Collaborative Web Spaces. Computer Supported Cooperative Work. In: From Usenet to CoWebs: Interacting with Social Information Spaces, pp. 155–166. Springer, Heidelberg (2003)
5. Farzan, R., Coyle, M., Freyne, J., Brusilovsky, P., Smyth, B.: Adaptive Social Support for Information Space Traversal. In: Proceedings of the 18th ACM Conference on Hypertext and Hypermedia, Manchester, UK, pp. 199–208 (2007)
6. Jung, S., Harris, K., Webster, J., Herlocker, J.L.: SERF: integrating human recommendations with search. In: Proceedings of the Thirteenth ACM conference on Information and knowledge management, pp. 571–580 (2004)
7. Kantor, P.B., Boros, E., Melamed, B., Meñkov, V., Shapira, B., Neu, D.J.: Capturing human intelligence in the net. Communications of the ACM 43(8), 112–115 (2000)
8. Maltz, D., Ehrlich, K.: Pointing the Way: Active Collaborative Filtering. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 202–209 (1995)
9. Olston, C., Chi, E.D.H.: ScentTrails: Integrating Browsing and Searching on the Web. ACM Transactions on Computer-Human Interaction 10(3), 177–197 (2003)
10. Smyth, B., Balfe, E., Boydell, O., Bradley, K., Briggs, P., Coyle, M., Freyne, J.: A Live-User Evaluation of Collaborative Web Search. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005), Edinburgh, Scotland, pp. 1419–1424 (2005)
11. Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., Boydell, O.: Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine. User Modeling and User-Adapted Interaction: The Journal of Personalization Research 14(5), 383–423 (2004)
12. Wexelblat, A., Maes, P.: Footprints: history-rich tools for information foraging. In: Proceedings of the SIGCHI conference on Human factors in computing systems: The CHI is the limit, pp. 270–277 (1999)