ABSTRACT
Item-oriented Web sites maintain repositories of information about things such as books, games, or products. Many of these Web sites offer discussion forums. However, these forums are often disconnected from the rich data available in the item repositories. We describe a system, movie linking, that bridges a movie recommendation Web site and a movie-oriented discussion forum. Through automatic detection and an interactive component, the system recognizes references to movies in the forum and adds recommendation data to the forums and conversation threads to movie pages. An eight week observational study shows that the system was able to identify movie references with precision of .93 and recall of .78. Though users reported that the feature was useful, their behavior indicates that the feature was more successful at enriching the interface than at integrating the system.

Author Keywords
online asynchronous discussion, user interface, recommender system

ACM Classification Keywords
H.5.2 User Interfaces

INTRODUCTION
Many Web sites are repositories of information about collections of items: Project Gutenberg offers free ebooks, GameSpot reviews video games, IMDb contains information on thousands of movies, eBay allows auctions for all sorts of things. Furthermore, each of these sites has rich data that gives users deeper insights into the content of each book, GameSpot has user ratings and wish-lists as well as features (e.g., product name, author, current lowest price), eBay has purchase data as well as features (e.g., a detailed category, a photo). Each of these sites also has one or more forums for users to discuss relevant topics. This is a common practice, and research shows that adding forums to an existing Web site has been shown to be correlated with higher user activity and improved user experience [5].

However, in each instance, item-land and forum-land are two separate worlds. When looking at items, typically there is no evidence of forums, or at most a generic link such as “Discuss this item.” Similarly, when viewing a forum discussion, there is little or no evidence of the rich data available for items being discussed. Often forums have a different look and feel and are implemented with different software.

Bridging these two different worlds could bring rich item-specific data to users in forum-land, and provide additional information-gathering opportunities for users in item-land. For example, users in the GameSpot forums might be shown review scores next to games being discussed, and users reading GameSpot product reviews of a game might be shown parts of conversations about that game. In each case, the user is given links to transition from one world to the other. Two keys to bridging these worlds are unambiguously identifying items in discussions and designing interfaces that effectively present this information to the user.

Our research group maintains an item-oriented site, MovieLens, where users can rate movies and get recommendations. We added forums and a novel bridging technique called movie linking. Movie linking is a way to identify when a movie is mentioned in forum posts. Our system automatically detects these references, and allows users to identify them interactively. We also investigated several interface designs that use this bridging information: linking appropriate discussion threads on movie detail pages, and showing appropriate movie information next to discussion posts.

Our research questions are:

RQ1: Were users able to use movie linking to create bridges between item-land and forum-land?

RQ2: Were users drawn across the bridges between item-land and forum-land?

RQ3: Did users find it useful to have information from one world shown in the other?
In the rest of this paper, we discuss the design of the movie linking system, including two tools that help users explicitly identify movie references in their posts, and several interfaces that leverage movie references. We then discuss results of an observational study of user behavior using MovieLens forums with movie linking.

**SYSTEM OVERVIEW**

MovieLens uses an automated collaborative filtering algorithm [4] to generate movie recommendations. Users rate movies and view movie recommendations through a search interface, which also displays a limited amount of information about each movie (e.g. actors, director, release date, genre). We will call the version of MovieLens that preceded our modifications “MovieLens Classic.”

We integrated forum software [3] into MovieLens. We were motivated to add forums to MovieLens because results from a previous study [2] and user feedback indicated that MovieLens users were interested in using forums. MovieLens forums provide a basic, non-threaded forum interface.

Without further integration, we have nothing more than two complementary systems that share a look and feel. This work develops movie linking, a new approach to integrating an information-rich Web site with discussion forums. Movie linking extracts references to movies in users’ posts to build bridges that connect MovieLens Classic and the forums in a way that enriches the user experience.

**Creating New Links**

The first challenge in bridging forum-land with item-land is identifying references to items in conversation. To address this challenge, we modified the MovieLens forum software, which includes a markup syntax for formatting posts (e.g. bold text is surrounded by [b] and [/b] tags). We extended this markup to identify a recognized movie reference by [movie id = N] and [/movie] tags, where the id attribute identifies the movie’s unique id number. We provide two interfaces for tagging movies.

First, we developed **movie chooser**, an AJAX-enabled text entry field that supports the manual addition of movie links. As a user types in the field, a background process gathers movie titles that match the input. A short list of matching movie titles is displayed. If the user selects a movie from the list, markup is inserted in the message composition area. A similar interface is used by Google Suggest [1].

Second, we developed **movie finder** to automatically detect and tag movie references in posts. If the user enables this option, clicking the **Submit** button in the posting interface will trigger an algorithm to search the post for movie titles. If movie titles are found, users are taken to a preview screen, where they are given the chance to manually add movie references that the algorithm missed, or delete movie references that were incorrectly added.

The movie finder algorithm uses regular expression matching. At system startup or when the movie database changes, a large OR expression is generated from thousands of movie titles in the MovieLens database.

We built a number of domain-specific optimizations into our algorithm. To improve the positive matching rate, we built the regular expression to account for different ways a user might refer to a movie. For example, a user might type either “A Clockwork Orange” or “Clockwork Orange” to refer to the movie “Clockwork Orange, A (1971).” We include multiple string matches for each title. The elements of the regular expression are sorted by the length of the string to ensure that longer titles matched before shorter titles. This ensures that the text “Piano Teacher” matches to “Piano Teacher, The (2001)” rather than “Piano, The (1993).” To limit the number of false positive matches, we employed a stop list to omit some titles from the regular expression. The stop list contains hundreds of strings such as “It” and “Big.” Finally, we dealt with the issue of ambiguous movie references. Movie titles such as “The Professional” have multiple releases, and strings such as “Star Trek” correspond to many movies in a series. When the matcher discovers an ambiguity, it inserts a link for each movie that matched, with instructions to help the user to remove the unwanted references.

**Using Links Across MovieLens**

Several new MovieLens features take advantage of tagged movie references. These features are designed to provide opportunities for users to easily transition between forum-land and item-land and to inject content from each of the two worlds into the other.

First, movie references in posts are hyperlinked to MovieLens Classic. This allows users to quickly find out about any movie mentioned in the forums. To add recommendations directly into the forums, each movie reference adds a ratings box to the side of the post. The ratings box shows the reader his or her predicted rating for the referenced movie, and gives that user the opportunity to rate the movie and add the movie to a wish-list. When the user’s mouse hovers over the hyperlinked movie reference, the corresponding ratings box is highlighted. This helps the user locate the movie of interest in posts containing many references. Figure 1 shows a post with two movie references.

Second, we added movie details pages to MovieLens Classic. These pages serve as the destination for users who click on a linked movie reference in the forums. Each movie in the system has a movie details page. These pages contain information such as the movie’s actors and languages as well...
as descriptive statistics such as the movie’s average rating. Users are able to rate the movie and add the movie to their wish-list from the details page. Movie details pages also display a hyperlinked list of ten movies with similar ratings. This creates a new way to explore the collection of movies.

To integrate forums into the recommender system, movie details pages include links to forum messages that reference the movie. To create more browsing opportunities, and improve coverage across rarely watched movies, movie details pages also contain links to the most recent messages that mention one or more of the movies shown in the similar movies list. We also added links that allow users to start a new forum thread about the movie. Clicking this link brings the user to the interface for creating a new post, with a movie reference and some stub text included in the message body. Figure 2 shows the movie details page.

EVALUATION

To evaluate movie linking, we conducted an eight week observational study of user behavior in MovieLens. The study consisted of two distinct, consecutive phases of four weeks each. The first phase began with the introduction of movie details pages (without links to forum messages) and non-linked forums into MovieLens Classic. This interface contained a handful of static links connecting MovieLens to the forums, but did not include the movie linking features. We created five initial forum threads to stimulate discussion.

The second phase began with the introduction of the movie linking feature. An announcement on the forum publicized the availability of this feature. We also added movie links to the initial five forum threads to demonstrate the feature to our users.

Through both phases, we collected data on all users in the MovieLens system concerning their behavior in the system. Also, at the end of each phase, we surveyed users about their perceptions of MovieLens and the forums. At the end of the second phase, we asked users specifically about the movie linking feature. This data is summarized in the next section, with respect to our research questions.

DISCUSSION

RQ1: Were users able to use movie linking to create bridges between item-land and forum-land?

To answer this question, we examine how people used the movie chooser and movie finder tools. We are specifically concerned with the amount of use of each feature and the accuracy of the movie finder algorithm.

To analyze the performance of the movie linker, we created an “ideally-linked” version of the forum data for comparison. To generate the ideally-linked data set, we first ran the movie finder algorithm on all posts from the two observational phases. We then looked at each post for errors of omission and addition, and corrected these errors by hand. We compared this ideally-linked data set to the real data set to investigate the success of this feature.

There were 282 posts (about 10 per day) made during the second observational phase. 185 of these posts (66%) contain movie references, for a total of 455 references (about 2.4 per post). 403 different movies were mentioned.

The ideally-linked data set contained 527 movie references. Using this for comparison, we found that there were 117 missing movie references and 30 extra movie references in the real data set. Of the 117 missing references, 67 would have been found if the user had not turned the movie finder algorithm off. 17 were the result of slang or shorthand references to movies, such as using the text “Plan 9” to refer to the movie “Plan 9 from Outer Space (1959).” 25 were not linked because they were contained in the stop list. This points out a downside of the stop list method, which accurately prevents false-positives, but fails to catch actual references. The remainder of the missing references were attributable to users’ spelling errors or typos.

Of the 30 extra movie links, 16 were the result of words that should have been added to the stop list, such as the word “silence” matching to the movie “The Silence (Tystnaden)” (1963). Two were the result of the intended reference not existing in the MovieLens database, and the text matching to another movie in the system. The remainder of the extra movie references were due to multiple matches (e.g. “Star Wars”) that were not edited out of the post. Figure 3 summarizes these results.

Using the standard information retrieval metrics precision and recall, the movie linking algorithm had a precision of .93 and a recall of .78. A .93 precision means that movie linking had a low false-positive rate, which was one goal of our system design. Because some users chose to disable the movie finder, a .78 recall shows that users were able to successfully use the movie linker the majority of the time. If all users had enabled the movie finder, 477 references would have been found (out of a possible 527), for a recall of .90.
We expect that this technique could be successfully applied to other domains. References to product names, book titles, and video game titles, for example, could all be extracted using a method similar to the one described here. Of course, each of these applications would require it’s own domain-specific optimizations.

RQ2: Were users drawn across the bridges between item-land and forum-land?

We used session information to create a model of users’ transitions between MovieLens classic, movie details pages, and the forum. The addition of movie linking corresponded with increased traffic between forum pages and movie detail pages, though this traffic was only a small percentage of the overall number of transitions. Initially, without movie linking, there were only 39 transitions from forum pages to movie details pages, out of 10,773 total actions initiated from forum pages, roughly 0.36%. Adding movie linking grew this to 237 out of 8,494, or 2.79%. In the other direction, without movie linking there were 86 transitions from movie details pages to forum pages, out of 10,225 total, or 0.84%. Movie linking grew this to 342 out of 7,905, or 4.33%.

The relatively low integration may be attributable to our interface design, which allowed users to rate and view recommendations directly in the forum. Because ratings and recommendations are the most useful features of the recommender system, adding these to the forum interface reduced the need to visit movie details pages. This may indicate a trade-off between enriching forum-land with data from item-land and tightly integrating the two areas of the system.

RQ3: Did users find it useful to have information from one world shown in the other?

To address this question, we administered online surveys to forum users at the end of both observational phases. In order to identify the changes brought by movie linking, we analyzed Likert scale data from the 32 users who took both surveys using a paired Wilcoxon signed rank test. The addition of movie linking caused users to agree more strongly that the forums help them know if they’ll enjoy a movie ($v = 32, p < 0.01$). Finally, the second survey included a question asking users how strongly they agree that movie linking made the forums more enjoyable to use. The average response was 3.97 on a 5 point Likert scale (1 = strongly disagree, 5 = strongly agree). One user reported, when asked about movie linking:

_It is a very valuable tool, because it gives me a much faster access to films I might be interested in. If there was no linking I might not be as motivated to check out certain films._

CONCLUSION AND FUTURE WORK

In this paper, we report on the design of a movie linking system, which identifies references to movies in users’ forum posts and enhances the user interface with this information. We found that users were able to effectively use our system to link movies, and used it often. The level of integration between MovieLens Classic and the MovieLens forums was not as high as we had expected, though this may be attributed to including the most relevant recommendation data in the forum interface.

We are continuing to improve the movie linking interface to encourage greater traffic between the forums and MovieLens Classic. We hope this new interface will allow us to conduct more controlled studies of how movie linking affects new user behavior, levels of user activity, and the interconnectedness of the site. Movie linking also has provided us with technical infrastructure necessary to replace static, chronological views of messages with personalized views of conversation generated using data from the recommender system. We hope to address issues of information overload in discussion forums by recommending conversations to our users.

ACKNOWLEDGEMENTS

Special thanks to Dan Cosley for his contributions during the writing process. This work was funded by the National Science Foundation, grant IIS 03-24851.

REFERENCES


