Temporal Filtering System to Reduce the Risk of Spoiling a User's Enjoyment

Satoshi Nakamura
Kyoto University
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan
Tel: 81-75-753-5969
E-mail: nakamura@dl.kuis.kyoto-u.ac.jp

Katsumi Tanaka
Kyoto University
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan
Tel: 81-75-753-5969
E-mail: tanaka@dl.kuis.kyoto-u.ac.jp

ABSTRACT
This paper proposes a temporal filtering system called the *Anti-Spoiler* system. The system changes filters dynamically based on user-specified preferences and the user's timetable. The system then blocks contents that would spoil the user's enjoyment of a previously unwatched content. The system analyzes a user-requested Web content, and then uses filters to prevent portions of the content being displayed that might spoil user's enjoyment. For example, the system hides the final score of football from the Web content before watching it on TV.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors.

Keywords: Temporal content filtering, Web, Anti-Spoiler

INTRODUCTION
Recent and remarkable advances in technology have enabled people to obtain unlimited amounts of information anytime and from a variety of sources, such as books, magazines, newspapers, television, radio, computer, and so on. Such media sources provide a huge variety of high-quality content; however, they sometimes include content that users might want to know at a later date.

Many systems and software applications are available to filter contents harmful to users. For example, antivirus software [1,2] detects contents, including viruses, worms, and spyware, by using pattern matching and blocks users from opening such contents to protect the user's computer. Popup killers found in Web browsers also detect harmful popup windows and close them automatically. Some filtering systems for children detect inappropriate adult content by referencing a blacklist of URLs and denying access to sites included on the list [3]. Spam filters [4] discard spam e-mails by using Bayesian filters or other such tools. However, these filtering systems only focus on full-time filtering of the contents. We think that in addition to this, a temporal filtering system, which filters the content that users simply do not want to see right now, is also need.

IUI'07, January 28–31, 2007, Honolulu, Hawaii, USA.

Situations where this system might be useful include sporting results. Seeing the final score often spoils the pleasure of a user who is waiting to watch a recording of this game on TV. Content that includes movie reviews may reduce a user's motivation to see a particular movie. Content, such as book reviews, that reveal the name of the criminal in a mystery novel may discourage a user from buying the novel or reading it later. However, these users will most likely want to see such content after they have watched the game, seen the movie, or read the book.

Considerable effort has been spent developing services that can recommend services based on user-specified preferences. A personalized Web portal service is one such recommendation service. Personalized Web portal sites are very useful to users because they provide the various contents that are suit-able for users all together. For example, a portal site recommends news that the user will be interested in. However, the user is sometimes informed of contents that he/she does not want to see right now. For example, a portal site may recommend a final score of a football match as high priority for the user if the site is aware he/she likes football. If the user waits to watch it on TV, knowing the results of the match spoils his/her pleasure in watching the match.

People want to control the access to such content themselves so as to prevent themselves from inadvertently watching contents until they have finished enjoying the target content. Unfortunately, people sometimes see or hear this information by accident while browsing the Web to check daily news, weblogs, and other information providing services. We believe that people sometimes think, "*I want to see it later, but I don't want to see it now!*"

Therefore, filtering systems need to take into account casual arrangements and plans, such as watching a football match on TV and reading books. Conventional filtering systems are not suitable for these uses.

Our goal is to create a novel filtering system that takes into account a user's daily life and user's enjoyment. Our system changes filters dynamically based on user-specified preferences and the user's timetable. The system then blocks contents that would spoil the user's enjoyment of a previously unwatched content. The proposed system described in this paper analyzes a user-requested Web content, and then uses the filters to prevent portions of the content being displayed that might spoil users enjoyment. We call our system *Anti-Spoiler*.

Note that in this paper, we use "*target content*" to refer to the content that the user is waiting to enjoy. Examples of target content are a football match, a TV program, books, and movies. The "*start date*" refers to the time from which people can receive the target content. Examples of the start date are the start time of the football match, the publishing date of the book, and the release date of a movie. In addition, we use "*undesirable part*" to refer to part of the content which spoils the user's motivation to enjoy the target content.

ANTI-SPOILER: TEMPRAL FILTERING SYSTEM

Concept

The *Anti-Spoiler* system is a temporal content filtering system. The system determines a user's intended action, such as waiting to watch a recorded TV program, to read a book, or to see a movie, based on some previous action, such as making a reservation to record a TV program, buying a book, or reserving a movie ticket. Then, to reduce the risk of spoiling the user's enjoyment, the system prevents any un-desirable parts from being displayed.

If the content has been updated before the start date of the target content, the content does not include the results of the target content, such as the final score of the game, a review of the book or movie, or the name of a criminal in the mystery novel or movie. The system ignores such contents. In addition, it starts running the filter for the target content when people are able to access the target content.

Figure 1 and 2 show the relationships between the user's activities, the system processes, the target content, and the related contents. In Figure 1, the system starts running the filter for program-A after the TV broadcast company begins to stream program-A. The system finishes running the filter when the user has finished watching program-A. In addition, the system filters the contents published after program-A starts. In Figure 2, the system starts running the filter for book-A when the user buys book-A. In this case, the system filters the contents published after the publishing company releases the book-A because many contents related to book-A are available on the Web.

Requirements

Creating the Anti-Spoiler system required us to implement the following requirements:

- **User detection:** The system has to detect the user who is in the vicinity of the system.
- **Schedule detection:** The system has to detect the user's favorite choices, casual schedules, and plans, such as waiting to watch a recorded football game on TV, or waiting to read a recently purchased book.
- **Activity detection:** The system has to detect a user's current activity, such as watching a recorded program, having finished watching it, reading the book, and having read it.
- **Filter creation:** The system has to create a filter automatically and has to set/unset the filter with the content browser.
- **Content division:** The system has to divide text into several parts, such as topic name (which are itemized in a news site), paragraphs of the text, and weblog entries.
- **Filtering:** The system has to detect undesirable parts in the content by using the preset filter.
- **Visualization:** The system has to prevent undesirable parts from being displayed to users.

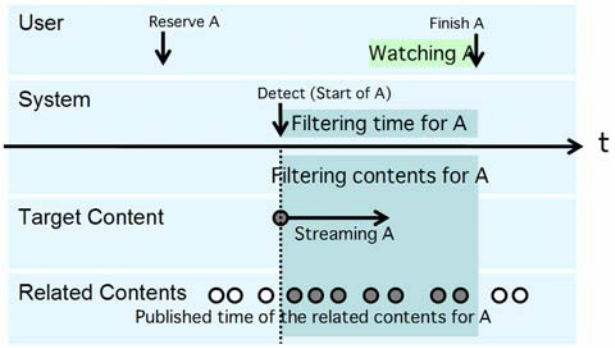

Figure 1: User reservation to record a TV program.

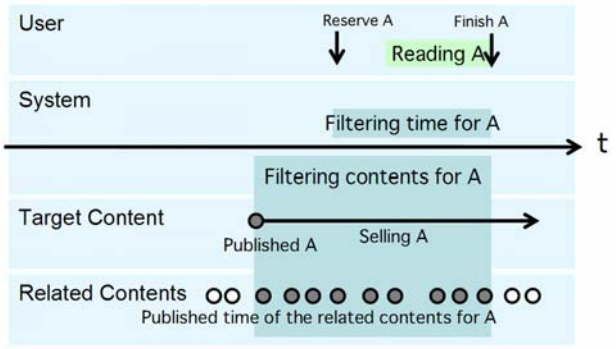

Figure 2: User buying a book.

DESIGN AND IMPLEMENTATION

User Detection

All users have their own ID card, each card has a unique identification number used by the system to determine which user is in the vicinity. Users initially register their ID in the user-ID database by using the user-ID registration system. We used RFID (radio-frequency identification) cards and an RFID reader in the current research. Once the user has entered their pre-registered user ID card in the ID reader, the system detects which user is present.

Detecting when the user leaves without actually observing the user's action is difficult. In this work, we required users to reenter their ID-card when leaving in the experiment so that system could determine who was present. Future work

will involve us implementing a mechanism to automatically detect the user's position without the user having to do anything.

Schedule Detection

We only focused on TV programs, books, and movies in the current research. Accordingly, the system detected the user's schedules, plans and favorite choices semi-automatically.

For example, if the user wants to watch a TV program, he/she selects a target program from a TV program guide (see Figure 3).

The system monitors the user's e-mail to see if any notifications from providers, such as Amazon.com, are received indicating recent purchases. The system then inserts the purchased book's title and publishing date into the user's schedule database. If the user purchases a book from another on-line site or at a store, the user has to manually input the title and publishing date into the schedule database by using an input form. If the user knows the books ISBN, he/she only needs to inputs it as the system automatically obtains all necessary information from the Web.

If the user wants to see a movie, they have to input its title and release date into the schedule database using an input form.

Activity Detection

Various devices, such as cameras and image processing, home appliances, and a range of sensors, can be used to detect the user's activity. Detecting a users activity discreetly/unobtrusively is important for the systems overall success. However, we did not consider automatic activity detection in the current research because this work is just the first step.

Users were required to indicate his/her activities by using an input activity form. The user can set the schedule by simply setting or unsetting an item (see Figure 4). If the user selects an item, the system updates the schedule database automatically.

Filter Creation

The system filters the Web content by using keywords and pattern matching. For example, if the target content is a football match, the system retrieves the team names and start dates from the TV program. A keyword list is created by the system to automatically filter information in the schedule database.

We prepared a sports team database, which included team name, owner's name, coach's name, players' names, and so on, using various football (soccer) leagues (*J-League*, *Serie-A*, *Premier League*, *Liga Espanola*) and national teams in the FIFA world cup 2006, American football (*NFL*), baseball (*NPB* and *MLB*), and basketball (*NBA*). The database also included the shortened name and the nickname of each team.

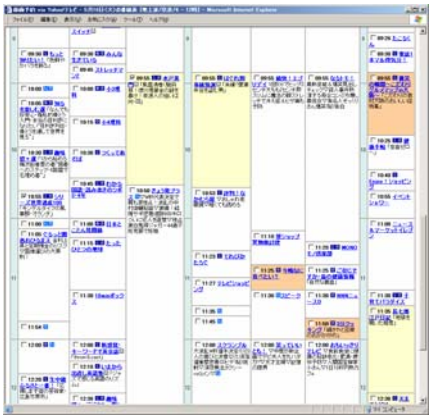

Figure 3: Interface system for choosing TV program.

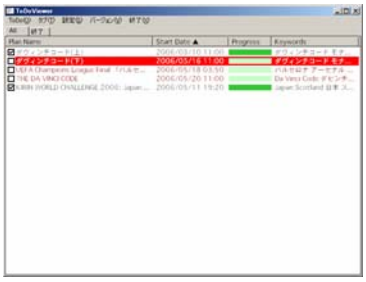

Figure 4: Schedule-management interface.

We also prepared a sports-result database, which included keywords related to the game's result, such as "win," "lose," "beat," and "upset". The result database included the pattern (regular expression) of the score (i.e., [\d]{1,3}-[\d]{1,3}).

Content Division

The system analyzes the received Web content and divides it into several parts, such as itemized news titles, paragraphs of long text, and entries of news and weblog sites. We used pattern matching of HTML tags in the current research.

There are many methods to segment Web content [5]. We plan to use such methods for content division in our future work.

Filtering

The system analyzes all Web contents by using a filter and sets a flag to indicate the degree of risk of spoiling the user's viewing pleasure by any undesirable parts.

The system checks the content on both the current Web page and the linked contents. For example, Web portal sites and Web news sites only show the title of the content to provide as much information as possible at the same time.

In addition, the system uses the last-updated date of the linked content and the start date of the target content to perform accurate filtering. If the last-updated date of the content is before the start date of the target content, the system ignores this content.

Visualization

The system prevents users from seeing the undesirable parts that they do not want to see now by setting the same color for the part of the text to the background (see Figure 5). The experimental system sets the color of undesirable parts based on the degree of the risk of spoiling the user's pleasure.

If the user wants to know the portions of content prevent from being displayed, he/she can check them by simply selecting them (see Figure 6).

DISCUSSION

We used our system during the 2006 FIFA world cup, and the system prevented all results from being displayed. However, the system sometimes prevented from being displayed the part of the Web content not related to the target contents because some Web contents had no updated information.

In addition, we found that some parts of the Web content unrelated to the target content were detected as undesirable if they were broadcast one or two weeks ago. For example, team A and B match results were contained in the target content (match). The user could not browse the information about team A and B until he/she watched the game. This was because our system always filtered the part of the content unrelated to the target content. We plan to introduce a mechanism to calculate a user's interest for target content.

Our system is very useful for browsing personalized Web portal sites and news sites because they often have a list of various contents as well as linked contents that include specific details.

The prototype system requires the user to indicate user's action such as wanting to watch some contents or finish to enjoy them. We plan to implement a system that can automatically detect some of the user's actions by using information supplied from appliances. For example, if a TV detects what the user is watching, when the user finishes watching the target program, the system will communicate with the server computer by using a network and will be able to automatically manage the user's filter.

The prototype system requires filtering keywords related to TV programs to be prepared. In the future, TV broadcasting companies will communicate these TV program keywords to users to encourage viewers to watch them. Especially users generally would not want to watch content after seeing any results of it.

As previously mentioned, a user has to manually register book information after they purchase a book. In the future, we believe that radio frequency identification (RFID) tags will be attached to books. This environment will enable people to easily register the book information and manage the status of their books.

Future work will involve us developing methods for the system to automatically discover related keywords, such as title, short title, and actors.

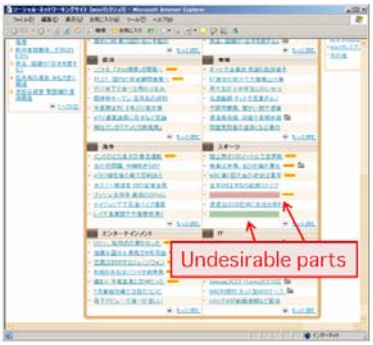

Figure 5: Example of display.

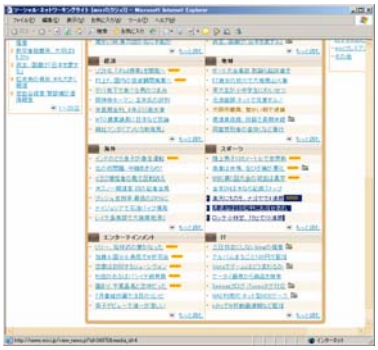

Figure 6: Reveal hidden parts by selecting them.

CONCLUSION

We proposed and implemented the Anti-Spoiler system to prevent portions of content from being displayed by using a user's casual schedules, plans, and activities. Our system blocks content thought to reduce users' enjoyment, such as game results and book and movie reviews. The daily usage of our system demonstrated its effectiveness.

Future work involves applying our current filtering system to other types of information sources, such as TV, radio, advertisement display, and word of mouth. For example, we can apply the filtering system to TV content by using the closed-caption function. In addition, we plan to improve the effectiveness of the filtering system.

REFERENCES

1. Symantec Corp., Norton Internet Security, http://www.symantec.com/.

2. McAfee Inc., Virus Scan, http://www.mcafee.com/.

3. DigitalArts Inc., i-FILTER, http://www.daj.co.jp/en/ir/.

4. Sahami, M., Dumais, S., Heckerman, D. and Horvitz EM., A Bayesian Approach to Filtering Junk Email. AAAI Workshop on Learning for Text Categorization, July 1998, AAAI Technical Report WS-98-05.

5. Deng, C., Shipeng, Y., Ji-Rong, and Wen. Wei-Ying, Ma., "VIPS: a Vision-based Page Segmentation Algorithm", Microsoft Technical Report (2003-79), 2003.