A Web Media Agent

Zheng Chen, Liu Wenyin, Rui Yang, Mingjing Li, and Hongjiang Zhang
Microsoft Research China
49 Zhichun Road, Haidian District, Beijing 100080, P.R.China
(86-10)62617711~3190/5769
{zhengc, wyliu, mjli, hjzhang}@microsoft.com

ABSTRACT
A web media agent is presented, which can make a user's web surfing experience more productive. Once the user visits a web page, semantic descriptions of the media objects on the page are automatically collected and used as their semantic indexes, which can help the user quickly find relevant media objects later.

Keywords
Multimedia information retrieval, web information retrieval, web experience, media agent, WWW

1. INTRODUCTION
The popularity of multimedia data, including video/audio clips, images, and other documents, is rapidly increasing due to improving multimedia technologies and convenient availability facilitated by the Internet. However, how to manage/utilize these multimedia data is non-trivial. Among other issues, how to find multimedia data that are relevant to users' intents is one of the most difficult problems.

Although people have invented some content-based media search approaches [2][3], they are currently at very low performance level. The best way to search multimedia documents is still by their semantics, which is usually represented by some natural language text descriptions. Hence, semantic indexing of these multimedia data by text descriptions is critical for later search.

Fortunately, as the World Wide Web has now become a rich source of multimedia data, it is very common that, on many web pages, the text content (e.g., ALT text, anchor text, and surrounding text) is semantically related to the media objects on the same pages to some extent, and can therefore be used to describe the semantics of these media objects. There are already quite a few papers [1][4] that have taken advantages of the relevancy of surrounding text and embedded images to do semantics-based retrieval of web images.

Based on the above observation, we have invented an agent-like system, which can, on behalf of a specific user, automatically collect the semantics of web media objects when the user accesses them. The collected semantics are used to index media objects and facilitate later retrieval. Hence, the web media agent makes the user's web surfing experience more productive. Furthermore, the agent can also provides other kinds of assistances in using these media objects, as we present in the rest of this paper.

2. THE WEB MEDIA AGENT SYSTEM
The architecture of the web media agent is shown in **Figure 1**. There are six components in the web media agent: the agent trigger, the semantics collection module, the feature extraction/indexing module, the media search engine, the suggestion module (UI), and the learning module.

The trigger is used to monitor users' actions and activate the media agent when necessary. The actions that may trigger the media agent include visiting a URL or saving/downloading a media object file from the web.

The collection module is used to collect the media objects' URLs and related text information from the web. When the user accesses some media objects, the media agent collects their URLs and their semantic descriptions and stores them in the media database. The semantic descriptions are text information on the same web pages containing the media objects. Potential sources of these semantic descriptions include the alternative text (ALT tag), the anchor text, the page titles, URLs, filenames, and surrounding text, etc.

The feature extraction and indexing module then extracts suitable semantic features from the text information and index the media objects using these features. The text information obtained as potential semantics descriptions is usually in some coarse form containing much redundant information. Some semantic features, which are concise keywords representing the semantics of the original text information, have to be extracted. These semantic features are used to index these media objects for quick retrieval in later re-use.

The media search engine is used to find relevant media objects according to the query, which is either explicitly specified by the user or automatically guessed by the media agent. The search engine uses a matching algorithm to find out the most relevant media objects that match to the user's need represented by the query. In the matching algorithm of our experiment, we calculate the semantic similarity between the query and each media object in the media database. The semantic similarity is calculated using the dot product of the query's semantic feature vector and the media object's semantic feature vector.

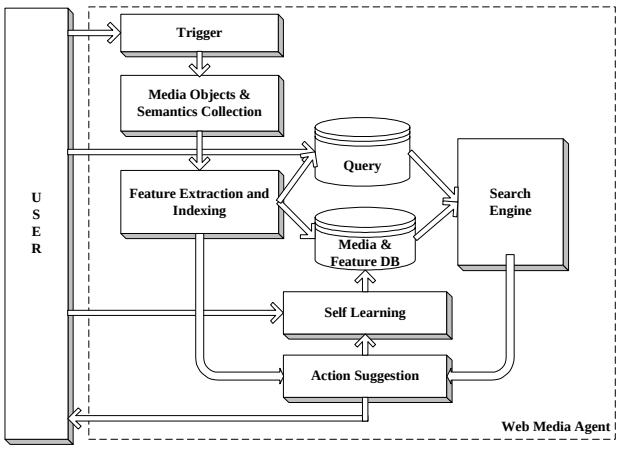

Figure 1. Architecture of the web media agent system.

The suggestion module is used to show the search result to the user in a ranked list given by the search engine for the user to make decisions. The retrieved media objects are sorted by their semantic similarity to the query. Each object is displayed with a short paragraph of text or a few keywords to describe its semantics. The user may choose some from the list. The user may also be able to modify the semantic descriptions of the media objects. There is another suggestion case in the web media agent. When the user wants to save or download a media object from the web, the web media agent is able to provide a set of suggested filenames according to the semantic feature of the media object. The user can choose one from them, and if necessary, modify it a little bit, and use it as the filename of the saved media object on the local machine. This scenario is exemplified by the "Save As" Advisor, as we show in the experiments in the next section.

When the user selects and accepts one or more of the suggestions, the learning module will automatically update the semantic indexing of the media objects. The media agent is also able to learn from the user's response to the suggestions and update the semantic features accordingly. The learning ability is a very important aspect of the intelligence of the media agent. For instance, the modified filename is probably the most relevant keyword(s) in the semantic feature and therefore should be assigned with a much higher weight. The chosen media object among the retrieval results is also considered very relevant to the query. Hence, the query should also take a higher weight in the semantic feature of the chosen media object. Since the web media agent has collected all media objects the user has accessed, they reveal to some extent the user's habitual actions and favorites. Hence, an offline crawler may be activated at the system idle time to search the web for those media objects relevant to the user's favorites and to collect their URLs and semantic features into the media database.

3. EXPERIMENTS AND EVALUATIONS

We have built a baseline configuration of the wed media agent system by implementing key ideas presented in this paper, including media object collection, retrieval, and suggestion. We only handle html, jpeg, gif, and mp3 files in the current system, though we can also extend the system to handle other forms of media files. In this section, presents experiments and evaluations of two main modules that can show the effectiveness of the web media agents. They are the "Save As" Advisor and the "Media Player" Reminder.

The "Save As" Advisor is activated when a user wants to save/download a media object from a web page. The "Save As" Advisor automatically collects and extracts some keywords from the web page and suggests a list of concise and precise filenames for the user to select. In the meantime, the semantic feature of the media object is also saved into the local database. The "Media Player" Reminder is used to help the user to locate the media objects that have been used before. The user only needs to submit a few related keywords to the "Media Player" Reminder, rather than to remember the exact location on the local disk or the URL on the web.

In order to test the efficacy and efficiency of the web media agent, we have used the "Save As" Advisor to download some media objects, including some mp3, html, wav, and jpeg files from the web. About 1000 media files are collected from 20 web sites. The media agent extracts all the related semantic features, including filenames, surrounding text, page titles, hyperlinks, etc., and then indexes these media objects with these features. We have calculated the statistics of the suggestion accuracy (based on the users' subjective judgment) of the "Save As" Advisor. The accuracy of the top 1 suggestion is nearly 40% and the accuracy of top 5 suggestions is as high as nearly 80%, as shown in **Figure 2**.

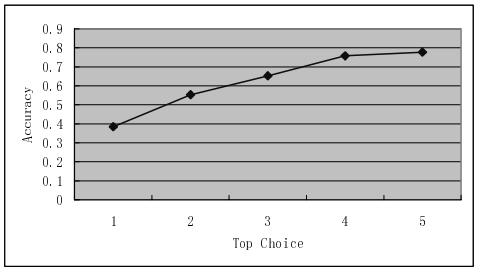

Figure 2. Accuracy of the "Save As" Advisor's Suggestions.

It is difficult for us to calculate the recall of the search engine of the media agent because it is a tedious job to browse the entire media database and specify the ground truth manually. Therefore, we only choose 11 queries to demonstrate the performance of the "Media Player" Reminder. The average precision-recall is fairly good with two representative pairs being (0.4, 0.7) and (0.6, 0.3), approximately.

4. CONCLUDING REMARKS

In this paper, we have presented the web media agent, which is an intelligent system to automatically collect semantic descriptions of multimedia data on behalf of users whenever and wherever they access these multimedia data from the web and provide necessary suggestions when users want to use these multimedia data again. As we show in the experiments, the web media agent is effective in gathering relevant semantics for media objects and is able to help users to quickly find relevant media objects.

In addition to making the user's web experience more productive, the web media agent is a very personalized administrator of multimedia data. The collection of media objects of each user is quite different from others since each user has different preferences in web surfing. Moreover, the web media agent can learn the user's preferences and profile more precisely after mining from action records of the user of a long period and is therefore able to take the initiative to provides the user with more satisfactory services.

5. ACKNOWLEDGMENTS

The authors would like to thank Z. Su, W. Zeng, and F. Zhou for their help as users of the web media agent in our experiments.

6. REFERENCES

[1] Z Chen et al. Web Mining for Web Image Retrieval. To appear in *Journal of the American Society for Information Science*, 2001.

[2] M. Flickner et al. Query by Image and Video Content. *IEEE Computer* 28(9): 23-32.

[3] N. Kosugi et al. A Practical Query-By-Humming System for a Large Music Database. *Multimedia'2000*, pp 333-342.

[4] H.T. Shen, B.C. Ooi, and K.L. Tan. Giving Meanings to WWW Images. *Multimedia'2000*, pp 39-48.