Towards a Flash Search Engine based on Expressive Semantics[ϒ]

Dawei Ding[1] Jun Yang[3] Qing Li[1] Liping Wang[1] Liu Wenyin[2]

{dwding, itqli}@cityu.edu.hk juny@cs.cmu.edu csliuwy@cityu.edu.hk

[1]Dept. of Computer Engineering and Information Technology, City University of Hong Kong, HKSAR, China

[2]Dept. of Computer Science, City University of Hong Kong, HKSAR, China

[3]Language Technology Institute, School of Computer Science, Carnegie Mellon University, USA

ABSTRACT

Flash, as a multimedia format, becomes more and more popular on the Web. However, previous works on Flash are totally based on low-level features, which make it unpractical to build a content-based Flash search engine. To address this problem, our paper proposes expressive semantics for bridging the gap between low-level features and user queries. To smoothly incorporate expressive semantics into a search engine, an Expressive Text Sensing (ETS) model is devised to map a user query to expressive semantics with the aid of link analysis method. Our experiment results confirm that expressive semantics is a promising approach to understanding and hence searching Flash movies more efficiently.

Keywords

Web application, Flash retrieval, search engine, expressive semantics, classification, ETS, Expressive Text Sensing

1. THE PROBLEM

Flash™ proposed by Macromedia Inc. is a new format of vector-based interactive movies, which can be embedded in web pages and delivered over the Web. Despite the fast growth of Flash media on the web, very limited research has been devoted to its retrieval. Systems like Google [1] use textual information embedded in and surrounding Flash movies to index and manage large movie collections. As the first endeavor in content-based Flash retrieval, Yang et al. [4] proposed a generic framework named FLAME, which embodies a 3-tier architecture for the retrieval of Flash movies on the Web. The semantic model [3] suggested by Ding et al. shows the potential of leveraging co-occurrence analysis for improving the performance of retrieval. However, these previous efforts consider only the low- or element-level features of Flash movies, leaving it still a challenging problem to fully meet the requirements of a Web-based Flash search engine.

2. EXPRESSING FLASH SEMANTICS

In the field of multimedia retrieval, the work of automatically analyzing high-level semantics of the media content appears to be a promising approach ([5]). Just like the film, Flash movies also have their own rules to express semantics by combining special effects. All the features mentioned below are computational in the sense that they can be automatically extracted from the raw data file of a Flash movie, or specifically from the features of the three basic movie elements (viz., *objects*, *actions*, and *interactions* [4]).

- Scene Complexity

We find that the scene complexity of a Flash movie generally indicates some of the producer's intention when making this movie. In many cases, whether a movie is composed of complex visual features may affect the expression of meaning. We define the scene complexity of a movie the average level of granularity and details available at the movie frames, such as *"video quality", "complex", "sketch", "rough"*.

- Interactivity

We define this feature to measure the interactivity of a Flash movie by calculating the amount and complexity of user interactions involved in a movie. Values include *"Interaction-driven", "Highly interactive", "Limited interaction",* and *"Passive"*. This feature describes the amount of actions involved in a movie within its length.

- Movie Mood

Scientists have found that certain colors were chosen to go with a mood-tone significantly and appear more frequently than remaining colors [8]. The psychological effect a Flash movie intends to give to the audience can be through its color. The values of this feature could include *"worm", "cold", "harmony", "unstable", "bright",* and *"vivid"*.

- Movie Rhythm

As in the film industry, Flash movie designers should attest to the existence of a rhythm of movie. Rhythm, also termed as pace, is the speed of progress of the story represented by a Flash movie. Three classes of rhythm are defined: *"Violent", "Fast", "Slow", "Static"*. The special pattern of timeline organization can, psychologically, give a unique impression to the viewers.

- Movie Category

Conceivably, category information is a very important type of semantic feature used to index Flash movies and to compose effective queries. Specifically, we define 7 categories of Flash movie primarily by their purposes and also by their appearance, namely *Game, Music TV (MTV), Cartoon, Interface, Banner & Logo, Intro,* and *Others*. The classification of Flash movies is conducted by applying a Bayesian classifier on a training set of manually classified movies to label unclassified movies. We select the following features of Flash movies to train the classifier: *Movie Info*, *Interactivity*, and *Scene Complexity*.

3. QUERY-BY-EXPRESSIVE SEMANTICS

Being more semantically-bound, expressive semantics can be employed to aid a textual query. For example, when a query "fight" is issued, a movie with strongly expressed pace may suggest it as a good result, even though there is no descriptive keyword of "fight" annotated/included in that movie. To leverage this, therefore, we present an **ETS** (Expressive Text Sensing) model to "sense" the underlying expressive semantics of textual information. Firstly, we define two concepts in our ETS model:

[ϒ] The work described in this paper was substantially supported by a grant from CityU (Project No. 7001457).

Common expression: When an expressive semantics occurs with a keyword in high frequency in a large collection of movies, we can regard the expressive semantics as a common expression of the textual keyword. The confidence of an expression is defined as $rank(e) = rank(F,t)$.

Well-designed Movie: When a movie is designed and created with the common sense of movie grammar [6], which means expressing the topic of the movie by properly making use of visual/aural elements, like color, motion, rhythm, we regard this movie as a well-designed movie. The confidence of a movie *m* to be well designed is defined as $rank(m)$.

Particularly, *common expression* and *well-designed movie* exhibit a mutually reinforcing relationship:

1. When an *expression* applies to more *well-designed movies*, it would gain a higher probability to be a *common expression*.
2. When a *movie* is composed of more *common expressions*, it would more likely be a *well-designed movie*.

From the above discussions, it is easy to see that this kind of textual/visual affinity can benefit us in twofold:

1. *Common expression* could help the search engine to expand the textual query with its common expressive semantics.
2. The knowledge about whether a movie is well-designed (or commonly-designed) could help the search engine to rank the movies on the web and return better results.

To exploit the aforementioned notions in our search model, an ***Expression Matrix*** *A* is constructed. In particular, each row corresponds to one of the possible expressions and each column is the whole expression vector of the corresponding movie. According to the relationship of expressions and movies, we suggest an extended iterative algorithm based on the HITS algorithm of link analysis [2] as follows:

$$rank^{(t+1)}(m_i) = \sum_{j:<F_j,t_j>\in m_i} rank^{(t)}(F_j,t_j)\times A_{i,j} \quad (1.)$$

$$rank^{(t+1)}(F_i,t_i) = \sum_{j:<F_i,t_i>\in m_j} rank^{(t)}(m_j)\times A_{j,i} \quad (2.)$$

4. SYSTEM EVALUATION

We have built an experimental search engine prototype supporting query-by-expressive semantics. To set up our evaluation, roughly 10 thousand Flash™ movies (1.98 GB) in total and in average 2 related Web pages of each movie have been crawled from the Web. To carry out experiment on the search performance, we make use of the Open Directory Project [7] web directory, where totally 20 first- or second-level category names are chosen as testing queries, such as *Art, Business, Computer, Game*... As a comparison, the results from Google [1] are selected as the baseline. Since Google has a far larger database than ours, we use the following procedure to do the result comparison.

1. Training the TES model using the initial Flash collection.
2. For a query *q*, we record the first 50 results *R* of Google's.
3. Add the result set *R* to our database, and then extract the expressive features of each movie in *R*.
4. Re-calculate the 50 movies' rank based on the TES model and the matching degree of descriptive content to the query.
5. After this step, the initial 50 movies in *R* are hence ordered in new rank. Compare our new rank of *R* with Google's.

After we have re-ranked the results, a user study is carried out to compare our rank with Google's. In each round, the users are given a query and the set of corresponding 50 Flash movies without any order. Then the user-chosen rank of the movies to the query is achieved and taken as the standard rank $< m_1, m_2, \ldots m_{50} >$, where m_i is the position of *i*th movie in the standard rank. If two results are ranked equally, they are given the equal rank in the list. The ranks of Google's and ours are compared by calculating the standard deviation: $SD(R) = \left(\sum_{1}^{50} (m_i^* - m_i)^2 \Big/ 50 \right)^{1/2}$, where m_i^* represents the new rank of *i*th movie. Obviously, the less *SD(R)* is, the better the rank *R* can be judged. Totally 20 queries are performed, and the comparison results of our result rank with the baseline are shown in Fig. 1. On average, a roughly 24% performance improvement on results ranking is achieved.

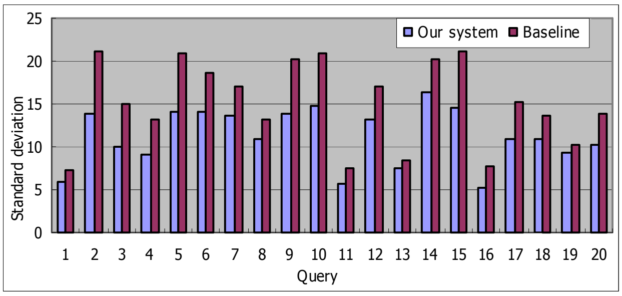

Figure 1: Comparison of Standard deviation

5. CONCLUSIONS

In this paper, we have proposed an expressive semantics-based search engine for Flash. As we have explored, the computational features derived from low-level features are more expressive than the latter and more reliable than high-level features. We claim that this kind of expressive semantics are crucial for bridging the gap between low and high level features. Targeting the most popular media on the web (viz., Flash), we have presented a search engine to demonstrate the effectiveness of expressive semantics, and proposed the method for utilizing and incorporating them into a practical search engine.

6. REFERENCES

[1] Google Inc. http://www.google.com

[2] Kleinberg, J, Authoritative Sources in a Hyperlinked Environment, in *Proc. 9th ACM SIAM Symposium on Discrete Algorithms*, 1998, pp.668-677.

[3] Dawei Ding, Qing Li, Bo Feng, and Liu Wenyin, A Semantic Model for Flash Retrieval Using Co-occurrence Analysis, in *Proc. ACM Multimedia 2003*, November 2003, Berkeley, CA.

[4] Jun Yang, Qing Li, Liu Wenyin, Yueting Zhuang, Search for Flash Movies on the Web, in *Proc. 3rd International Conference on Web Information Systems Engineering (workshop volume)*, 2002

[5] C. Dorai et al. Computational Media Aesthetics: Finding meaning beautiful. *IEEE Multimedia*, 8(4):10–12, 2001.

[6] D. Arijon, *Grammar of the film language*, Silman-James Press, 1976.

[7] Open Directory Project. http://www.dmoz.org/

[8] Itten J. "Art of Color (Kunst der Farbe)",Otto Maier Verlag, Ravensburg,Germany, 1961.