Conceptions of Features and Semantic Clusters as Search Mechanisms: A Pilot Study[1]

Barbara M. Wildemuth[*], Meng Yang[*], Gary Geisler[†],
Tom Tolleson[*], Jon Elsas[*], Jei Luo[*], and Gary Marchionini[*]
[*]Open Video Project, School of Information and Library Science
University of North Carolina at Chapel Hill
[†]Graduate School of Library and Information Science, Simmons College

TREC VID 2004 Notebook Paper
October 2004

Abstract

Search mechanisms are the interface tools that people can use to search the items in a collection; a common example is a text box for entering search terms. In this pilot study, the Open Video team investigated two mechanisms for providing access to a collection of news videos. The first provided direct access to the subsets of the collection containing particular video features, as identified by other TREC VID participants; the second provided direct access to the subsets of the collection that were identified as semantic clusters, using latent semantic indexing approaches to analyzing the video transcripts. Each access mechanism was depicted in the interface as a series of labeled checkboxes. In the pilot study, four team members completed the TREC VID topic searches and responded to measures of their perceptions of the experience of using each search mechanism. Reasonably high precision was achieved on the user searches across all three systems (0.67-0.74), but none of the systems achieved high recall (0.10-0.11). Mean average precision across three runs (as calculated by NIST) ranged from 0.06-0.09. Searches were completed in approximately 8 minutes across all three systems. User satisfaction with the two experimental systems was mixed. Lessons learned from conducting this pilot study will contribute to the design of a follow-up study investigating the ways in which users of digital video retrieval systems conceptualize search mechanisms that incorporate access to subsets of the collection based on video features or semantic clusters of transcript content.

1 Introduction

For any information retrieval system, including a digital video retrieval system, a mechanism must be developed by which the user can communicate some type of specification of a query to the system. Historically, retrieval systems have asked that users specify their queries as search terms – specific words that will be processed by the system in relation to the words included in the items in the collection. For example, in a digital video retrieval system, a search term might be matched against the words present in a transcript of the words spoken on the video. It is expected that users will be able to easily transfer the skills they have developed in searching document collections to the task of searching digital video retrieval systems. This premise provides a foundation for concept-based image retrieval generally [6], including digital video retrieval.

[1] This work is supported by National Science Foundation (NSF) Grant IIS 0099638. In addition, we would like to acknowledge the contributions of David Eichmann, University of Iowa, who shared his story segmentation work with us.

Digital video as the content of a retrieval system, however, suggests additional possibilities for the specification of queries. The features of the images within a video can be analyzed and made available for searching, as demonstrated by the feature extraction task of TREC VID. Zooming, panning, and other aspects of video can be leveraged as part of a query specification. Detection of shot boundaries or story boundaries can be used in the processes supporting digital video retrieval. (For a review of a variety of these approaches, see Brunelli, Mich, and Modena [3].) This line of research is referred to as content-based image retrieval [6].

In spite of the potential of these video indexing and representation methods, it is not clear that users of digital video retrieval systems will understand them well enough to employ them in developing an effective search strategy [9]. For example, if users want to retrieve videos that include a lot of action, how can they specify such a query? If they want to retrieve videos of Washington, DC, with a lot of traffic heading toward the Capitol, how can they specify such a query?

The study described here was a pilot test of methods for examining users' conceptions of novel search mechanisms that could be developed for a digital video retrieval system. The methods developed during the pilot study will enable a later, more thorough investigation of the ways in which users conceptualize these search mechanisms and incorporate them into their search strategies.

2 Background

Two types of evidence supported the design of the search mechanisms that are the focus of this study: the theoretical literature related to people's understanding of video and results from empirical studies of video searching systems. Each of these two types of evidence was briefly reviewed in Wildemuth et al. [16]. Our experience in TREC VID 2003 also informed the design of this study. In 2003, study participants interacted with a search mechanism incorporating selection of particular video features, as well as more traditional text searching of words from the transcript. While this system did not outperform the alternatives to which it was compared, it did perform as well as the more familiar text-only system. These results encouraged us to further explore the incorporation of features data into a digital video search mechanism. Such a search mechanism would represent subsets of the video collection, based on the presence of a particular feature in that subset, and allow users to select particular subsets for inclusion in their search results.

Clustering the database into subsets based on the text associated with particular items is a parallel endeavor, from the perspective of the search mechanisms it affords. Such clustering was suggested by Salton as early as 1968 [13], and has been explored in interactive retrieval systems more recently, beginning with work on the Scatter/Gather method by Hearst and Pedersen [10] and continued up to the present [e.g., 11]. While Hearst and Pedersen found people to be successful in using a retrieval system based on the Scatter/Gather method, it is not clear how people understand how clusters are formed or the potential impact of clusters on their search strategies. For the purposes of the current study, clusters of news stories (i.e., subsets of the database) were represented in a way that is similar to the representation used in the Scatter/Gather method.

The current study was conducted as a pilot study, to test the methods to be used in a full-scale study of people's conceptions of two experimental search mechanisms. The first mechanism was based on the features extracted as part of the TREC VID 2004 activities. The user's search could be limited to particular subsets of the collection that were identified as containing particular features. The second search mechanism was based on the text of video (news story) transcripts. Using a latent semantic analysis of the terms in the transcripts, stories were clustered and users could select particular clusters to include in their searches. These two experimental search mechanisms were compared with a simple text-only search mechanism, in terms of user performance and satisfaction.

3 The Search Systems

Three search systems were evaluated, each with a different search mechanism. The baseline system allowed only text searching of the transcript. A second system augmented text searching with access to the subsets of the collection that included each of the TRECVID 2004 features. A third system augmented text searching with access to subsets of the collection that represented semantic clusters of stories. Each of these three search mechanisms will be described below.

3.1 Organizing shots into stories

All three search systems represented the items retrieved as stories (rather than shots). Because UNC's Open Video project is not a participant in the TREC VID story segmentation task, we enlisted the assistance of David Eichmann (University of Iowa). In preparation for TREC VID 2004, Dr. Eichmann provided us with the 2003 story segmentation data he had submitted [5] and, in 2004, provided us with his current story segmentation as it was submitted to NIST. These story boundaries support our work in developing the cluster-based search mechanism and the display of results for all three systems.

3.2 The search mechanisms

The text-only system allowed users to search the ASR transcripts of the stories in the video collection via a text box for search term entry. The MySQL full text search engine was used for this study; their default list of stopwords was accepted and the research team set the minimum word length at three characters. In computing a relevance score, MySQL takes into account the number of words in a record, the number of unique words in that record, the total number of words in the collection, and the number of records that contain a particular word. The search results were ranked based on the relevance score computed by MySQL.

The features-based system allowed users to select, from a list of features, those that might be relevant for inclusion in the search. The ten features (road, train, beach, boat or ship, airplane taking off, physical violence, people walking or running, basketball basket scored, Madeleine Albright, and Bill Clinton) were represented to users in a list with checkboxes (see Figure 1). The study participants were allowed to check as many features as they wished, or they could use just the text box at the upper right corner of the screen.

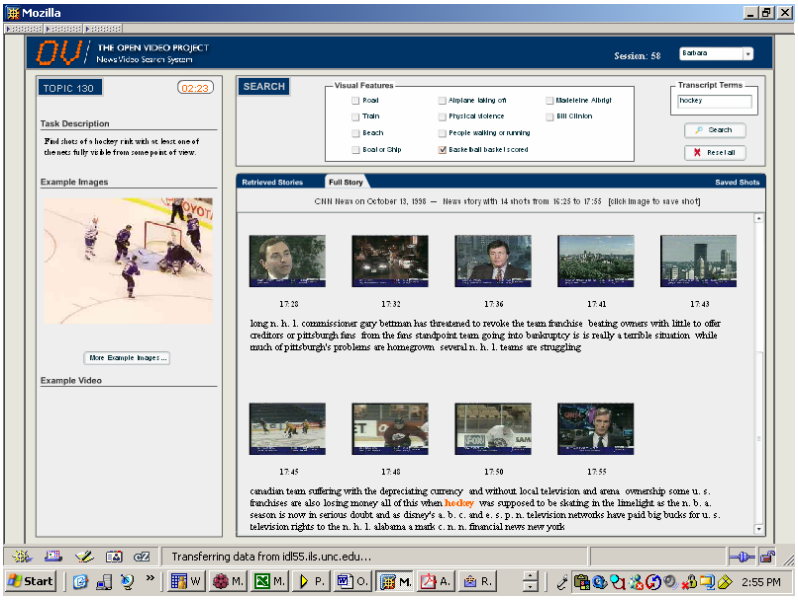

Figure 1. Features-based system, displaying full stories

The features data was based on IBM's results for the TREC VID 2004 features extraction task, selected because of their consistently-high performance in TREC VID 2003 [1]. If at least half of the ten runs submitted to NIST contained a feature, then the shot was considered to contain that feature in our database (and given a feature score of 1). The feature scores for shots were averaged over the shots within a story to obtain a feature score for each story. The average feature score for each story, where the average was based on the story's scores for just those features selected by the user when submitting the query, was then added to the MySQL relevance score for the story (doubled, to add weight to the text search) in order to rank the retrieved stories.

The third system was based on semantic clusters of stories, using the terms in the story transcripts to form clusters. Like the features-based system, the semantic clusters were represented to users in a list with checkboxes (see Figure 2). Each cluster was described with a few terms selected from those terms with high logs-odds-ratios or high frequencies within the story. In addition, the number of stories in each cluster was included in the cluster description.

To develop this system, the transcripts provided were used to create a transcript-by-term matrix, where each row represents a transcript, each column a term, and the matrix entries term counts. The DTIC/Verity stopword list [4], augmented with words from the corpus deemed by the researchers to be not topically informative, was applied to the transcript representations. These additions to the stop list included frequently-occurring proper names ("calloway", "gibson") and other common terms specific to the corpus ("hello", "nightline", "headline"). Additionally, terms that occurred in less than 22 documents were removed in order to retain only the most general topical terms in the collection.

Latent Semantic Analysis [8] was applied to the matrix, and the first 50 principal components were retained. The EM algorithm was then applied to this modified matrix [12], with the number of clusters (k) ranging from 12 to 19, a range determined by the researchers to be useful for the targeted interface. To evaluate the clusters, the transcripts within each cluster were inspected, as well as the ranking of terms on each cluster, by term frequency within clusters and by log-odds-ratio.

The average cluster score for each story, where the average was based on the story's cluster scores for just those clusters selected by the user when submitting the query, was then added to the MySQL relevance score in order to rank the retrieved shots.

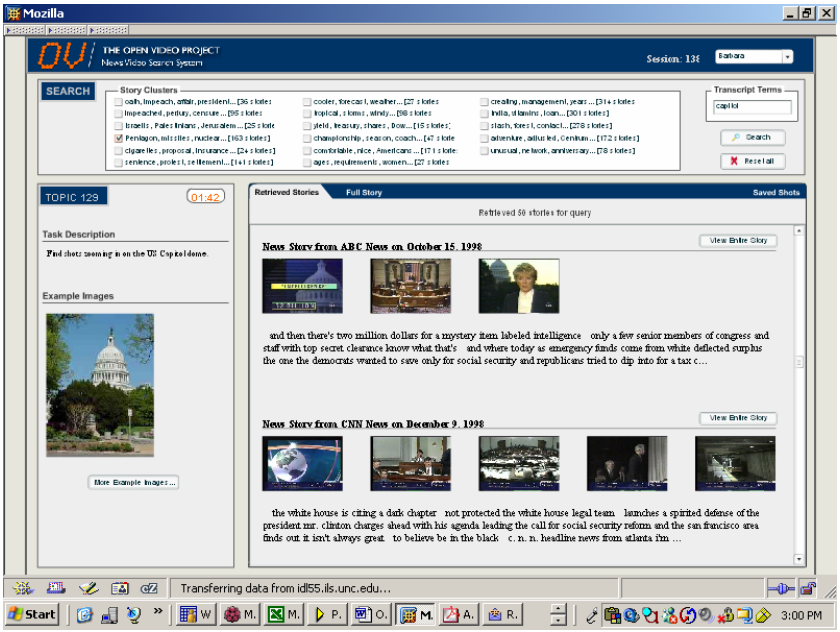

Figure 2. Cluster-based system, with display of initial results list

3.3 Display of the results

In all three systems, the results were displayed by story (not by shot). Poster frames were displayed for each of the first five shots of each story retrieved (see Figure 2). In addition, the text for the beginning of the story was displayed. The story title indicated the source of the story (CNN or ABC) and the date on which it was originally broadcast.

If a user believed that a story might include relevant shots, the full story could be viewed by clicking on a button labeled, "View Entire Story", placed just to the right of the story title. By clicking on this button, the user was moved to the "Full Story" tab of the interface (see Figure 1). Only by viewing the entire story could particular shots be selected for submission as relevant for the purposes of the TREC VID evaluation.

4 Study Methods

As noted in the introduction, the current study was a pilot study, conducted for the purpose of developing methods that could be used in a fuller study of people's conceptions of features and semantic clusters as potential search mechanisms. For this reason, only members of the Open Video team participated in this study. (It is expected that a later study will include a small number of people – not on the research team – interested in using digital video collections.)

Four members of the Open Video team completed the study protocol. Two participants were faculty researchers and two were doctoral students. While three participants participated in development of the system to be studied, one participant had joined the team in August and had not participated in the system development or study design activities.

A within-subjects research design was used, so that each of the four participants was exposed to all three search systems. Each person completed the following activities:

- a pre-session demographic questionnaire, including both questions suggested for TREC studies and questions used in previous Open Video studies,
- 8 search topics on the text-only system, each followed by a post-search questionnaire using questions suggested by TREC VID,
- a questionnaire about the text-only system, using questions suggested by TRECVID and measures of perceived usefulness, perceived ease of use, and flow,
- 8 search topics on one of the experimental systems, each followed by the post-search questionnaire,
- a post-system questionnaire about the experimental system,
- 8 search tasks on the other experimental system, each followed by the post-search questionnaire,
- a post-system questionnaire about the second experimental system, and
- a brief post-session questionnaire, using questions suggested for TREC VID.

The order of the two experimental systems (text plus features and text plus semantic clusters) and the three sets of search topics were counter-balanced among the four participants. Because all of the participants were familiar with the system, no training topics were used.

User performance on the three systems was compared by calculating the average precision and the average recall achieved with each system, across all topics and all users. These calculations were based on the relevance assessments provided by NIST. Calculation of recall [14] was based on the assumption that the full set of relevant items in the collection is represented by the set of all relevant items identified by NIST assessors. While this approach to performance measurement is somewhat unusual within the context of TREC and other traditional information retrieval experiments, we believe that it is more able to take into account the variability in searcher performance, which can be

as great as an order of magnitude [2]. In addition to these performance comparisons, the mean average precision and precision achieved at several levels of retrieval were compared across systems, based on the data provided by NIST.

The user satisfaction measures recommended by NIST were augmented with measurements of participants' perceived usefulness, perceived ease of use, and flow (enjoyment and concentration), measured after use of each of the three search systems. Each of these measures was used in TREC VID 2003 and is described in Wildemuth et al. [16].

Each of the three systems was evaluated, using both performance and satisfaction data. Because this is only a pilot study, these data are of little long-term interest. Of more importance is what they can reveal about the feasibility of conducting a follow-up study with a broader audience; such methodological findings will be discussed.

5 Results

5.1 Characteristics of the participants

Three women and one man participated in the study. The age distribution among the participants was bimodal; two of the participants were in their mid-20's and two were in their mid-50's. As noted above, two were faculty researchers and two were doctoral students. All were members of the Open Video project team.

The two graduate students had an average of 8 years of online searching experience, and the two faculty members had an average of 26.5 years of online searching experience. All four participants use computers daily and conduct online searches daily. They considered themselves very experienced with point-and-click interfaces, online library catalogs, and Web search, but less experienced with searching CD ROMs, searching commercial online systems, and searching the Open Video collection (see Table 1).

Table 1. Participant characteristics

	Mean	*s.d.*
Experience with searching/tools (1, no experience, to 5, a great deal of experience)		
Point-and-click interface	5.0	*0.00*
Searching an online catalog	5.0	*0.00*
Searching on CD ROM	3.8	*0.96*
Searching on commercial online systems	3.5	*0.58*
Searching on the web	5.0	*0.00*
Searching on the Open Video database	4.3	*0.96*
Experience with news and current affairs (1, not at all, to 5, more than once a day)		
Frequency of watching TV news	3.5	*0.58*
Knowledge of current affairs	3.5	*0.58*
Experience with digital video (1, never, to 5, daily)		
Frequency of watching videos/films	3.8	*0.50*
Frequency of using a digital video retrieval system	2.8	*0.96*

The participants watch TV news fairly frequently and have a moderate knowledge of current affairs in general. They watch videos or films fairly frequently, but search using a digital video retrieval system less often. When they do search for videos/films, three indicated they go online, two indicated they search film archives, one indicated they search the video rental store, and one indicated they browse channels. Three participants search for videos by title, three by topic, and one by author or actor. The purposes for conducting searches included entertainment, academic review purposes, and research.

5.2 Performance results

Performance on the three systems was evaluated in terms of the average precision the study participants achieved with each system, the average recall the study participants achieved with each system, and the average amount of time used per search on each system. These results are summarized in Table 2. Since the sample size was so small, no tests of statistical significance were undertaken. However, these preliminary results suggest that the system augmented with features searching may improve precision, but require more time per search. In addition, it should be noted that precision and recall were moderately positively correlated ($r = 0.33$, $p = 0.0065$).

Table 2. Summary of performance, by system

	Precision		Recall		Time per search	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
Text-only	0.67	*0.350*	0.11	*0.140*	7.7	*2.73*
Text + features	0.74	*0.329*	0.10	*0.119*	8.2	*3.02*
Text + semantic clusters	0.69	*0.343*	0.11	*0.107*	7.8	*2.76*

Note: 21-23 searches contributed to each of the means. These data include all the search runs submitted to NIST, including the supplemental run.

A subset of search results (three runs) was submitted to NIST to be evaluated in terms of mean average precision. Each of the three runs represented one of the three systems. The results of those analyses, as reported by NIST, were then averaged by system, and are shown in Table 3. The general trend in these results is that the two augmented systems (incorporating features or semantic clusters in addition to text searching) outperformed the text-only system.

Table 3. Results reported by NIST, aggregated by system

	Average precision	Precision at 10 shots	Precision at 30 shots	Precision at 100 shots	Precision at 1000 shots
Text-only	0.055	0.4400	0.1756	0.0527	0.0053
Text+features	0.092	0.4818	0.1939	0.0582	0.0058
Text+semantic clusters	0.070	0.4882	0.2118	0.0635	0.0064

5.3 Satisfaction results

In addition to the measures suggested by NIST, measures of user perceptions of usefulness, ease of use, and flow (enjoyment and concentration) were taken in relation to each of the three search systems. Each set of user perceptions will be described, in turn.

The questionnaire recommended by NIST was used to collect data on user perceptions immediately after each search. In addition, a question was added to address users' perceptions of the two experimental systems (administered only when those systems were being used for the search). The results from this questionnaire, aggregated by system, are shown in Table 4.

Table 4. User perceptions, based on post-search questionnaire (1, not at all, to 5, very much)

	Text-only		Text+features		Text+clusters	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
I was familiar with this topic before I did the search.	3.8	*1.14*	3.8	*0.97*	3.7	*0.74*
The example images/videos given with the topic description were useful for searching.	2.8	*1.59*	3.4	*1.11*	3.0	*1.35*
I found that it was easy to find shots that are relevant for this topic.	2.7	*1.35*	2.7	*1.06*	2.9	*1.28*
The ability to search by particular features of the videos was useful.			2.4	*1.24*		
The ability to search by clusters of videos was useful.					2.4	*1.20*
For this particular topic I was satisfied with the results of my search.	2.6	*1.34*	2.9	*1.21*	2.6	*1.48*
For this topic, I had enough time to find enough answer shots.	3.5	*0.97*	2.6	*1.09*	3.2	*1.24*

As with the previous analyses, the statistical significance of any differences between systems was not tested. However, from these data it appears that the users may have perceived the system augmented with features to be less efficient than the other systems (see the last row of Table 4).

After completing the eight assigned searches for each system, each participant completed measures of usefulness (6 items), ease of use (6 items), and two dimensions of flow (4 items each). The results from these measures are shown in Table 5 (note that lower scores indicate more positive attitudes). While the users did not rate the systems very differently on ease of use, it appears that the familiar text-only system was perceived as less useful and did not result in as strong an experience of flow (enjoyment and concentration) as the two experimental systems.

Table 5. User perceptions, based on post-system measures

	Text-only		Text+features		Text+clusters	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
Perceived ease of use	3.8	*0.65*	3.3	*0.79*	3.6	*0.38*
Perceived usefulness	3.6	*0.55*	3.1	*0.63*	3.2	*0.84*
Flow (enjoyment)	4.7	*1.98*	4.0	*0.90*	3.9	*1.13*
Flow (concentration)	4.4	*1.97*	3.0	*0.75*	3.3	*1.18*

Note: Lower scores indicate more positive attitudes.

After working with all three systems, three of the four participants completed one additional questionnaire (a series of five-point scales). The users' responses indicated that the system's response times were too slow (mean = 2.7, s.d. = 1.15), and that they did not support efficient searching (mean = 2.7, s.d. = 0.58). However, participants did find it relatively easy to learn to use the systems (all rating it at 4). The participants had a good understanding of the searching task (mean = 4.3, s.d. = 0.58), even though they found it only moderately similar to other searching tasks (mean = 3.0, s.d. = 1.00). They found the three systems to be only moderately different from each other (mean = 3.0, s.d. = 1.00).

The participants were then asked for direct comparisons of the three systems; their responses are shown in Table 6. While the text-only system was easier to learn to use, it did not maintain its advantage on other criteria.

Table 6. Comparison of systems in post-session questionnaire (number of participants giving each response)

	Text-only	Text+features	Text+clusters
Easier to learn to use	2	1	
Easier to use		2	1
Liked the best overall		2	1

6 Discussion

Because the current study was based on a very small number of study participants, all of whom were members of the design team, no conclusions can be drawn about the relative value of the two experimental search mechanisms. However, lessons were learned through the pilot study that can be used to develop a strong study design to investigate "real" users' conceptions of feature-based or text-based clusters as they are made available through search mechanisms.

The first lesson learned was concerned with coverage of the complete database of videos with the features extracted for TREC VID 2004 and the clusters created. This year, a smaller set of features was extracted than last year. Of the 2548 stories that formed the full database (based on the story segmentation provided by Eichmann, as noted earlier), only 1610 (63%) contained ANY features and only 2012 (79%) were included in ANY semantic cluster. Thus, if a user selects any features or clusters as part of their search strategy, they have truncated the size of the database and, thus, potentially lowered recall. Further investigation is needed to see whether there are differences in the effects on potential recall across the TREC VID topics, due to exclusion of some stories from the coverage provided by the features search mechanism or the semantic clusters search mechanism. Those TREC VID topics suffering the least effects from this problem will be selected for use in the planned full-scale study.

In order to make a fair comparison of the effectiveness of the two experimental search mechanisms and their comparison with the baseline text-only search mechanism, some control may need to be exerted over user behavior when interacting with these search mechanisms. In the pilot study, participants were allowed to incorporate features or semantic clusters in their search strategies, but could also conduct text-only searches with those systems. In the planned user study, we will require that users incorporate at least one feature in each search on the features-based system and at least one cluster on the cluster-based system. The implications of this requirement for system processing speed will need to be estimated, to ensure that response time is reasonable (given that it was already perceived as somewhat slow in the pilot study).

In the systems used in the pilot study, the story was used as the unit of retrieval. While the search algorithms for features acted on individual shots (in the features-based system), the system retrieved and displayed the story of which each relevant shot was a part. In the cluster-based system, the story was the unit analyzed to form the clusters; thus, the search system retrieved and displayed stories. The selection of the story as the unit of retrieval leads to a dependence on the quality of the story segmentation process. It affects the text that is processed for clustering, and it affects the representation of data in the search results. From the pilot study data, no conclusions can be drawn about the effectiveness of using the story as the unit of retrieval, in terms of users' understanding of the search results. This question will be more thoroughly explored in the planned user study.

7 Conclusion

We hope to conduct a full user study of these two experimental search mechanisms in spring 2005. While data will be gathered concerning their effectiveness (as signified by user performance in conducting searches) and user satisfaction with the interaction, the planned study will focus on people's understanding of the search mechanisms themselves, incorporating think-aloud protocols [7,

15] to elicit these understandings. When users are confronted with a listing of video features from which they can select, how do they conceptualize those features and how do they understand how selection of particular features will affect their search outcomes? When users are confronted with a listing of semantic clusters, represented by a string of words and the size of the set, how do they conceptualize those clusters and how do they understand how selection of particular clusters will affect their search outcomes? The context of TREC VID 2004 has helped us to develop plans for this future user study.

8 References

[1] Amir, A., Iyengar, G., Lon, C.-Y., Dorai, C., Naphade. M., Natsev, A., Neti, C., Nock, H., Sachdev, I., Smith, J., Wu, Y., Tseng, B., & Zhang, D. (2003). The IBM semantic concept detection framework. TREC VID 2003 slides. IBM Corporation.

[2] Borgman, C. L. (1989). All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management, 25*(3), 237-251.

[3] Brunelli, R., Mich, O., & Modena, C. M. (1999). A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation, 10,* 78-112.

[4] Defense Virtual Information Architecture. (2000, October). DVL/Verity stop word list. Ft. Belvoir, VA: Defense Technical Information Center. http://dvl.dtic.mil/stop_list.html.

[5] Eichmann, D., & Park, D.-J. (2003). Experiments in boundary recognition at the University of Iowa. TREC VID 2003 notebook paper. http://www-nlpir.nist.gov/projects/tvpubs/papers/uiowa.paper.pdf.

[6] Enser, P. (2000). Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science, 26*(4), 199-210.

[7] Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data.* Cambridge, MA: MIT Press.

[8] Furnas, G. G., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., & Lochbaum, K. E. (1988, May). Information retrieval using a singular value decomposition model of latent semantic structure. *Proceedings of SIGIR '88,* 465-480.

[9] Greisdorf, H., & O'Connor, B. (2002). Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation, 58*(1), 6-29.

[10] Hearst, M. A., & Pedersen, J. O. (1996). Reexamining the cluster hypothesis: Scatter/Gather on retrieval results. *Proceedings of SIGIR '96,* 76-84.

[11] Muresan, G., & Harper, D. J. (2004). Topic modeling for mediated access to very large document collections. *Journal of the American Society for Information Science & Technology, 55*(10), 892-910.

[12] Nigam, K., McCallum, A. K., Thrun, S., Mitchell, T. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning, 39*(2-3), 103-134.

[13] Salton, G. (1968). *Automatic Information Organization and Retrieval.* New York: McGraw-Hill.

[14] Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science, 39*(3): 161-176.

[15] van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes.* London: Academic Press.

[16] Wildemuth, B. M., Yang, M., Hughes, A., Gruss, R., Geisler, G., & Marchionini, G. (2003). Access via features versus access via transcripts: user performance and satisfaction. TREC VID 2003 Notebook Paper. SILS Technical Report 2003-05. http://ils.unc.edu/ils/research/reports/TR-2003-05.pdf or http://www-nlpir.nist.gov/projects/tvpubs/papers/unc.paper.pdf.