VITALAS at TRECVID 2009
Studying User Search Behavior with a Video Retrieval System

Henning Rode, Theodora Tsikrika, Arjen de Vries

CWI Amsterdam – The Netherlands – VITALAS Project

16th November 2009
Outline

1. **VITALAS System**
   - VITALAS GUI
   - Data Preparation

2. **Test Setup**
   - User Test Setup
   - Evaluation

3. **Results**
   - Questionnaire
   - Searchlogs

4. **Conclusions and Future Work**
The VITALAS System

The GUI offers different **search types**:
- text (ASR) search
- concept search
- similarity search
- fused search: AND/OR combination of the above search types

The GUI offers the following **actions** on the results:
- zoom keyframe or play video
- add item to result “lightbox”

The GUI offers **concept suggestions** related to the current query
Result View
Result View

The image shows the Result View interface of the VITALAS system. The query field is highlighted, indicating where users can input their search terms. Below the query field, there are keyword suggestions and concept suggestions, which can be used to refine the search. The results are displayed in a grid format, with images sorted by relevance. The interface also includes options for mosaic view, cluster view, and advanced visualization.
Result View

Concept Suggestions
Result View

Image Upload
Result View

[Image of the VITALAS GUI interface showing keyword suggestions, concept suggestions, and result keyframes]
Zoom View
Zoom View

Detailed View
Zoom View
Zoom View

Add to Results
Data Preparation

ASR
- machine translation of provided Limsi ASR to English
- shot-alignment of the English text

Concepts
- complement VITALAS HLFE concepts with provided MediaMill concepts
- index the top 5000 concept scores per concept

Visual Features
- extract and index SIFT features for similarity search
Test Users

- 8 + 2 test users, not involved in system development
- 4 professional *archivists* / 4 + 2 non-professional *novice* users
- professionals used our system for the first time
Training and Evaluation Methods

The users got

- tutorial and training topic
- **questionnaire** (3 parts: entry, search, exit)
- list of available concepts (only professional users)
- test instructions
  - *we did not ask to find as many results as possible*

User interaction **logging:**

- issued searches and results
- zoom and add-to-result actions
Topic Assignment

- each test user got 12 of TRECVID 2009 topics
- based on a latin squares arrangement
Research Objectives

- compare the use and effectiveness of the different search types: text (ASR) search, concept search, similarity search, fused search
- compare the behavior of the different user groups: archivists vs. novice users
General search experiences

- the users feel they had enough time for the search tasks
- novice users always had enough time, professionals sometimes
- novice users are slightly more confident about the quality of their search results than professionals
- professional users show a higher correlation between self-judged completeness of search and satisfaction with search time
## Satisfaction with search types

### Users judging the search types

<table>
<thead>
<tr>
<th>search type</th>
<th>archivists</th>
<th></th>
<th>novices</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ease</td>
<td>eff.</td>
<td>ease</td>
<td>eff.</td>
</tr>
<tr>
<td>text</td>
<td>4.25</td>
<td>3</td>
<td>4</td>
<td>3.16</td>
</tr>
<tr>
<td>concept</td>
<td>4.25</td>
<td>3.5</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>similarity</td>
<td>4.25</td>
<td>2</td>
<td>3.16</td>
<td>3</td>
</tr>
<tr>
<td>fused</td>
<td>4.25</td>
<td>2.75</td>
<td>3.66</td>
<td><strong>3.83</strong></td>
</tr>
</tbody>
</table>

questionnaire scale: 1 – 5

- All search types are easy to use for professionals; novice users have more difficulties with search types other than text search.
- Concept search and fused search are experienced to be slightly more effective than text search.
Querying behavior

- most queries are text queries, 20% contain concepts, 13% contain similarity searches, and 14% combine different search types
- entry searches are most often text searches
Actions on the search results

- professionals add on average 12.6 items per topic to the results, novices only add 6.6 items
- both user groups perform almost the same number of zoom actions
- professionals look deeper into the ranked results: median rank of zoomed/added items is double as high (34 for professionals vs. 18 for novices)
User Assessor Agreement

User vs. Assessor

- only 50% of the shots added by our users were judged by the assessors
- 50% of the judged added shots are marked by the assessors as irrelevant
- 40% of the judged zoomed-but-not-added shots are marked as relevant by the assessors
- 33% of the displayed-but-not-added shots are relevant

User vs. user

- proportion of common shots found by more than one user within all added shots for a topic is only 17%

Users missed many shots by looking only at the keyframe thumbnails
Search type effectiveness

System vs. user effectiveness

<table>
<thead>
<tr>
<th>search type</th>
<th>rel. displayed</th>
<th>added</th>
</tr>
</thead>
<tbody>
<tr>
<td>text only</td>
<td>2.76</td>
<td>0.98</td>
</tr>
<tr>
<td>concept only</td>
<td>13.47</td>
<td>2.69</td>
</tr>
<tr>
<td>similarity only</td>
<td>4.32</td>
<td>1.07</td>
</tr>
<tr>
<td>fused</td>
<td>7.88</td>
<td>1.79</td>
</tr>
</tbody>
</table>

- **system perspective**: similarity searches retrieve 2 times as many relevant results than text searches, fused searches 3 times, concept searches 5 times as many

- **user perspective**: concept search results in 2.5 as many add actions than a text search or similarity search
Conclusions and Future Work

Conclusions

• all search types are useful, concept search is most effective on TRECVID topics
• different search behavior:
  • professionals search longer and deeper in the retrieved list
  • professionals search more recall oriented
• GUI personalization is necessary to deal with different users
• keyframe thumbnails are not sufficient to judge a shot

Future work

• 2nd test phase with more test users is planned
• analyze search patterns, order of search actions
• analyze dependence between query type and search pattern