

Dublin City University (DCU-CLARITY-iAD)

TRECVID 2010 Interactive Search

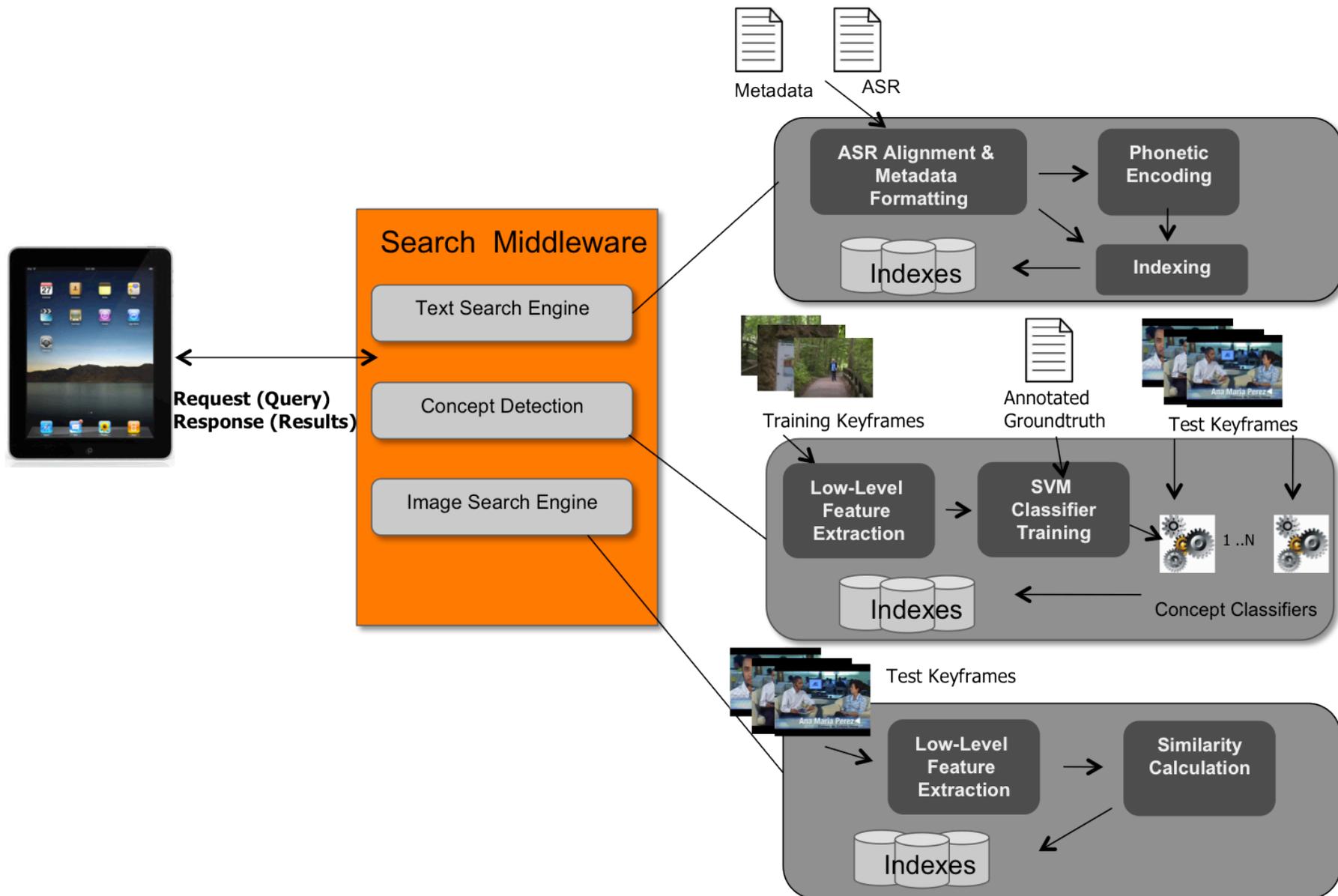
Colum Foley, Jinlin Guo, David Scott,
Paul Ferguson, Pete Wilkins, Kealan Mc
Cusker, Emma Sesmero Diaz, Cathal
Gurrin, Alan F. Smeaton

November 2010

DCU – CLARITY – iAD @ TV KIS 2010

- Our submission in summary
 - Simple, intuitive iPad interface
 - Supports the 3 common modes of interaction
 - Text search
 - Concept search
 - Image search
 - Evaluated the performance of novices versus experts
 - 18 users in total
 - Novice users were business management students from BI School of Management in Oslo
 - Official results show no difference in performance between user groups

DCU 2010 Interactive KIS System Overview



Webservice Overview

- Constructed using .NET web service as underlying technology
- Communication with iPad through HTTP POST
 - Text and Concept Search
 - Image Similarity Search
 - Shot timing request
 - Validity checking for known-item

Webservice Overview II

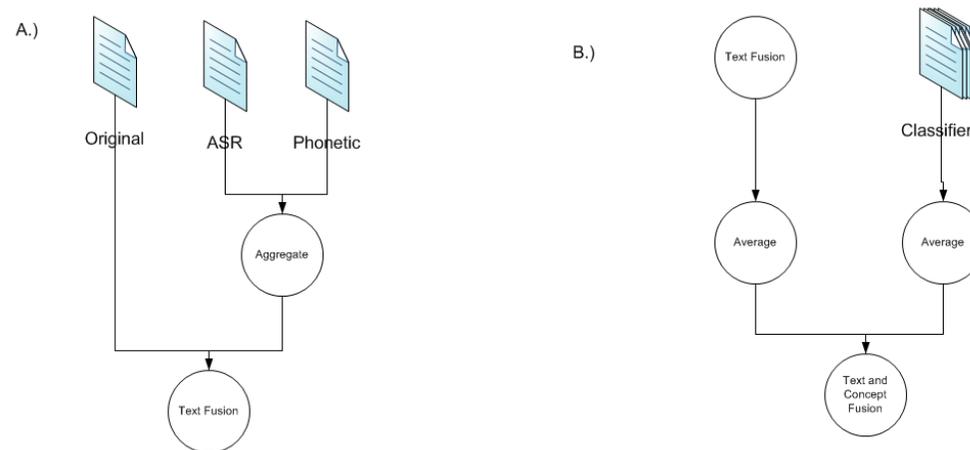
- Returned results in the form of XML Document:
 - Top 100 videos for text & concept search
 - Top 50 similar keyframes for similarity search
- Logging
 - Keeping a record of all interactions with the system
 - NIST record for each topic

Text Search

- Terrier search engine used as underlying text search engine (University of Glasgow)
- Three indexes created:
 - Original Metadata (title, description, keywords)
 - Automatic Speech Recognition
 - Phonetic Translation
- Both phonetic and ASR indexes over shot level, needed to be aggregated to video level

Fusion Overview

- Weighted CombSUM was used to fuse three text sources with weights set by experiments over training topics
 - Meta(6),ASR(2),Phonetic(1)
- Text and concepts are fused by using concepts in a boosting technique over an original text list



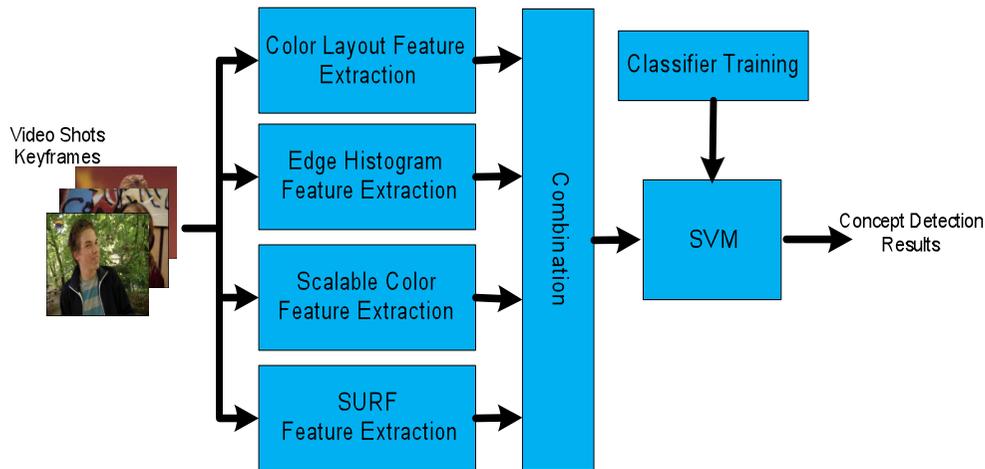


Semantic Concept Detection

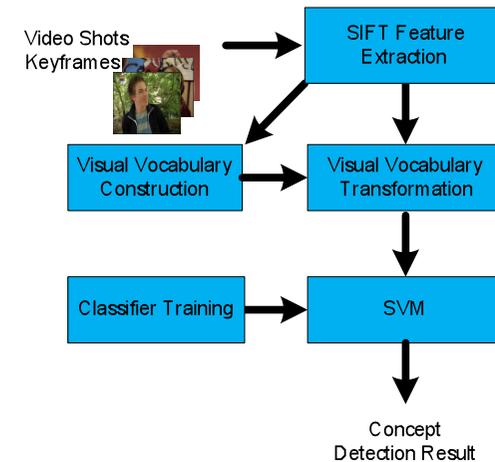
Concept Detection

- We evaluated two SVM classification frameworks for concept detection in our system

Concept detection based on MPEG-7 descriptors and SURF



Concept detection based on Bag of Words (BoW) model

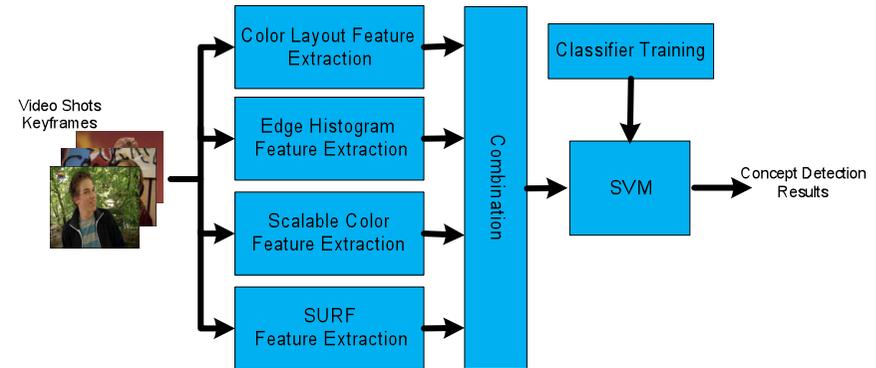


Concept Detection Based on MPEG-7 Descriptors and SURF

– Low-level Feature Extraction

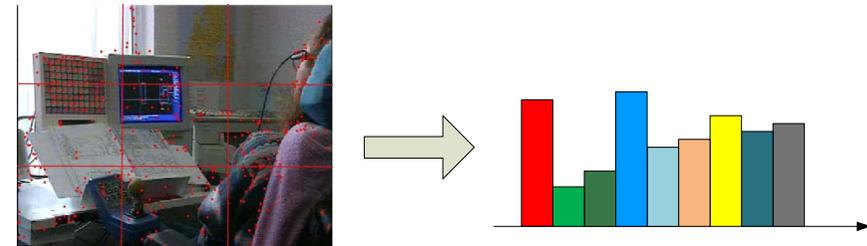
- MPEG-7 Colour and Texture Descriptors:

- ❖ Colour Layout
- ❖ Scalable Colour
- ❖ Edge Histogram



- Local Interest Point Descriptors:

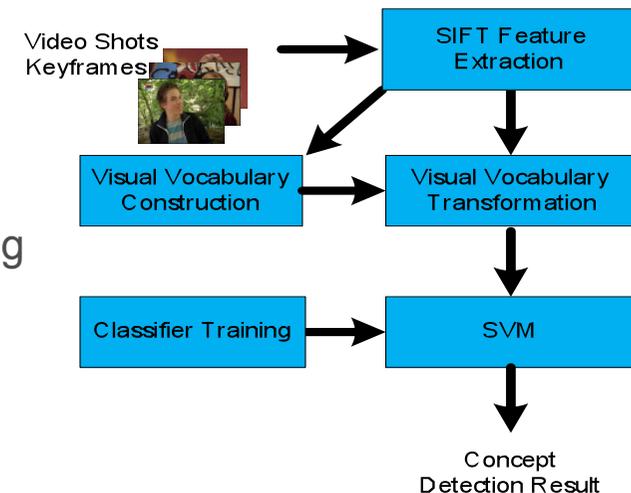
SURF is a scale- and rotation-invariant spatial descriptor based on Haar wavelet responses and has been proven effective in detecting objects.



SURF feature Extraction

Concept Detection Based on BoW Model

- In recent years the Bag-of-Visual-Word (BoW) model produced good results on several large-scale content based image and video retrieval benchmarks
- Low-Level Feature:
 - Scale-invariant feature transform (SIFT)
- Visual vocabulary constructed using K-means clustering with a vocabulary of 1296 visual words
- Visual vocabulary transformation technique:
 - Soft assignment is employed
 - For each SIFT point the top-100 nearest visual words selected and weights assigned based on distance



Implementation Details

- Radial Basis Function (RBF) kernel is adopted for both classification frameworks
 - Better classification results than polynomial or linear kernels
- In the final system we developed 33 concepts based on types of concepts used in the training topics

Animal, beach, beard, Black and White video, boat/ship, building, bus, car charts, cityscape, computers, computer screen, crowd, daytime outdoor, face, flower, ground vehicle, indoor, indoor sports, landscape, map, meeting, military, nighttime, office, outdoor, person, road, sky, snow, stadium, tree, vegetarian

Image Similarity Engine

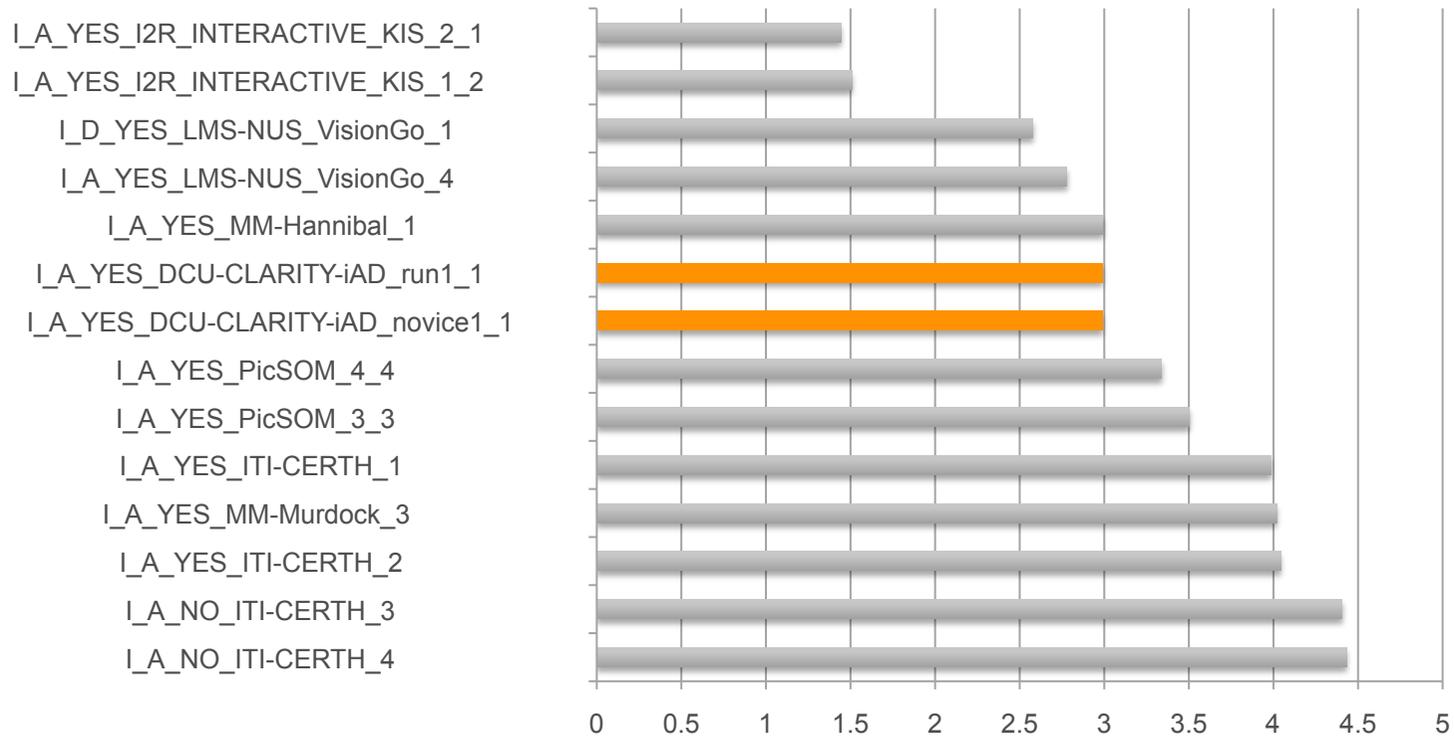
- Used by the user to find similar keyframes to a query image (from search results)
- Based on the MPEG-7 Colour and Texture Descriptors:
 - ❖ Colour Layout
 - ❖ Scalable Colour
 - ❖ Edge Histogram
- For each keyframe in the collection we calculated the similarity to each other keyframe for each feature
- Multiple features are combined using CombSUM



iPad User Interface ...Video

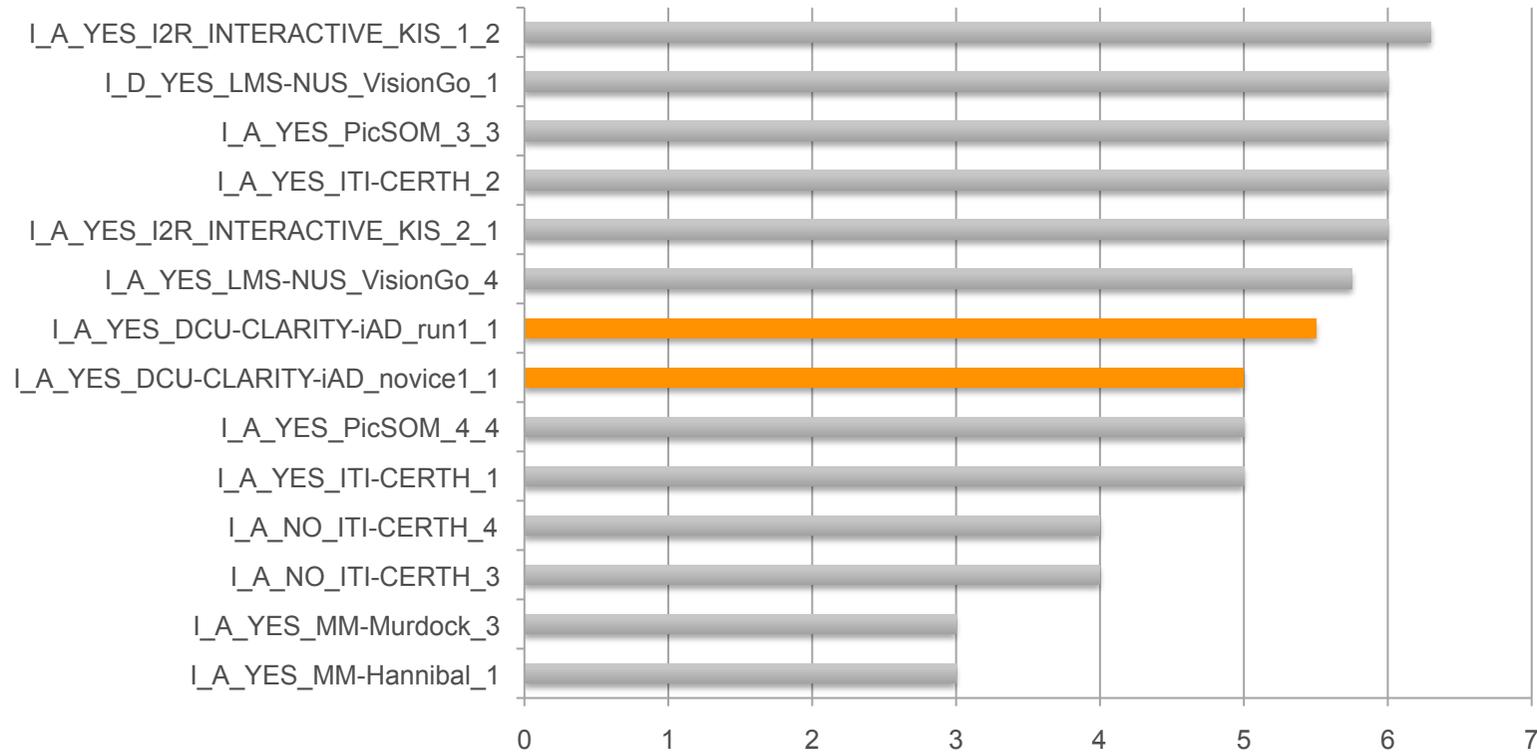
TV Results: Mean Elapsed Time

Mean Elapsed Time (mins)



TV Results: User Satisfaction

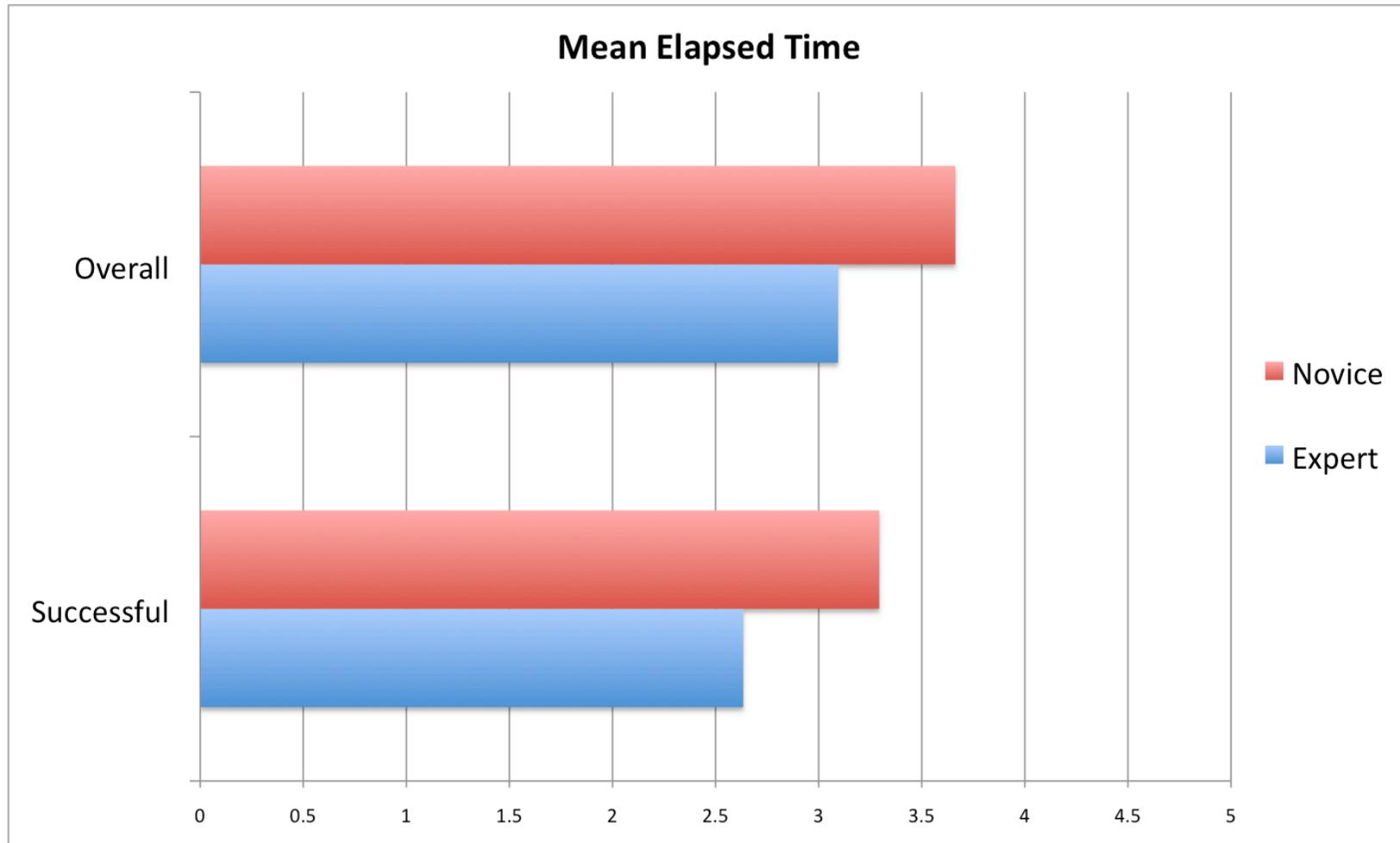
User Satisfaction



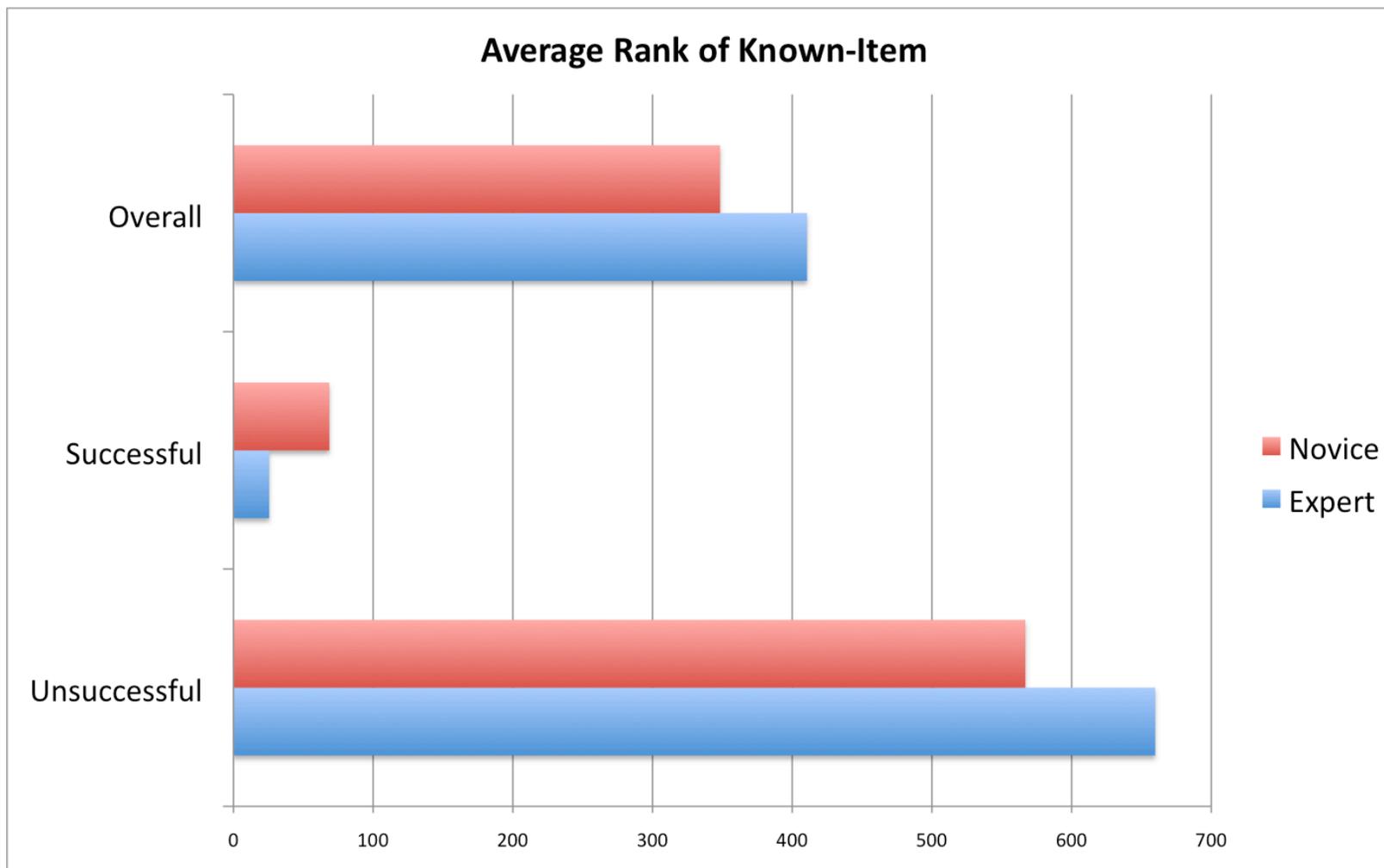
Post-Experiment Analysis

- Examined the interaction logs and questionnaires
- Wanted to examine both the performance of novice versus experts in more detail
 - Overall search performance
 - Different search strategies used
- Also wanted to compare performance of three text indexes

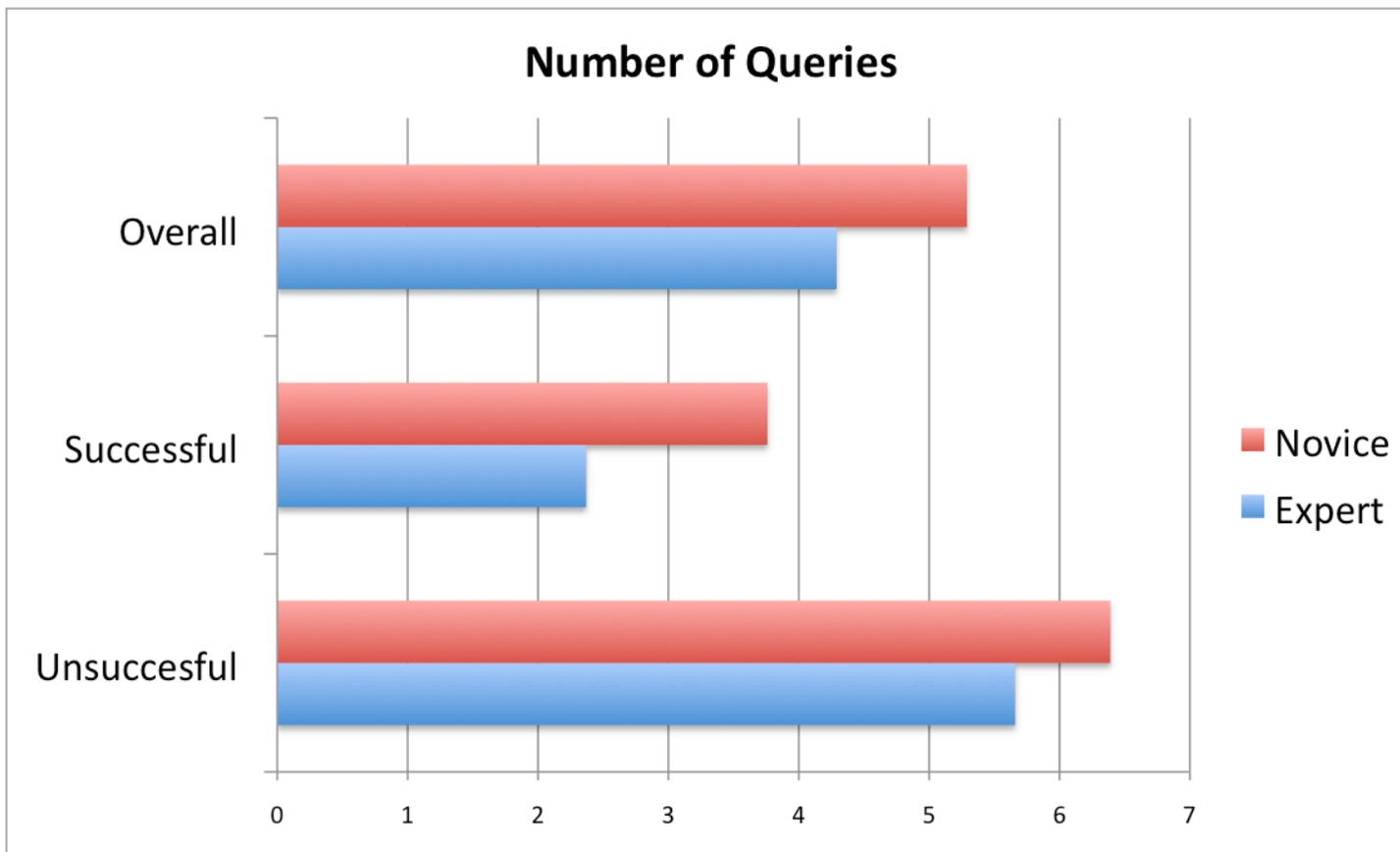
Overall Search Performance



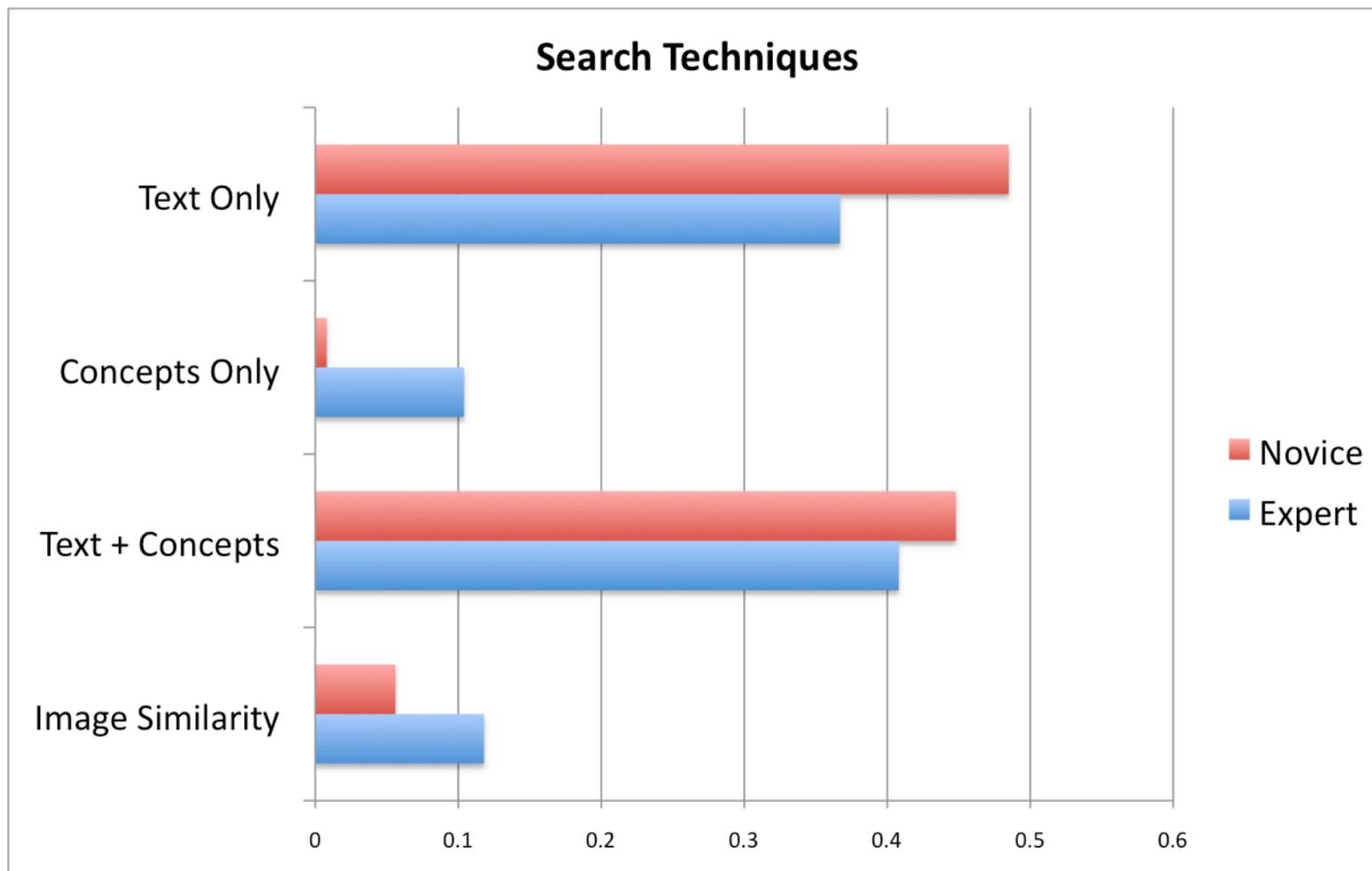
Rank of Known-item in ranked lists returned



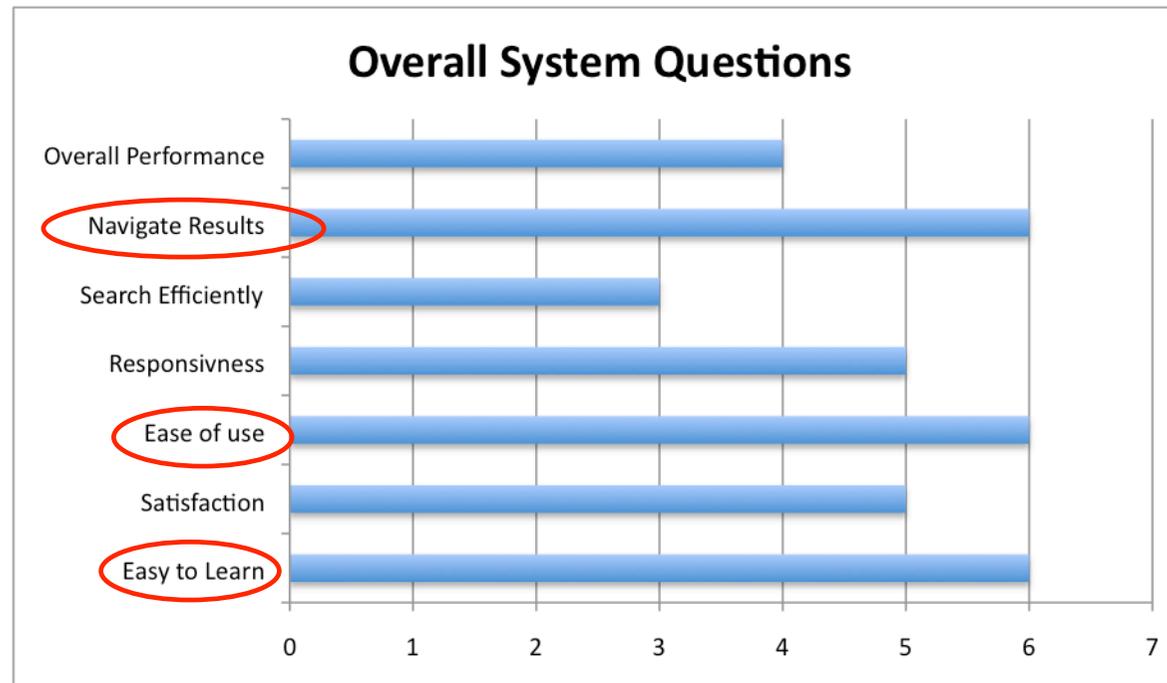
Number of Queries Per Search Task



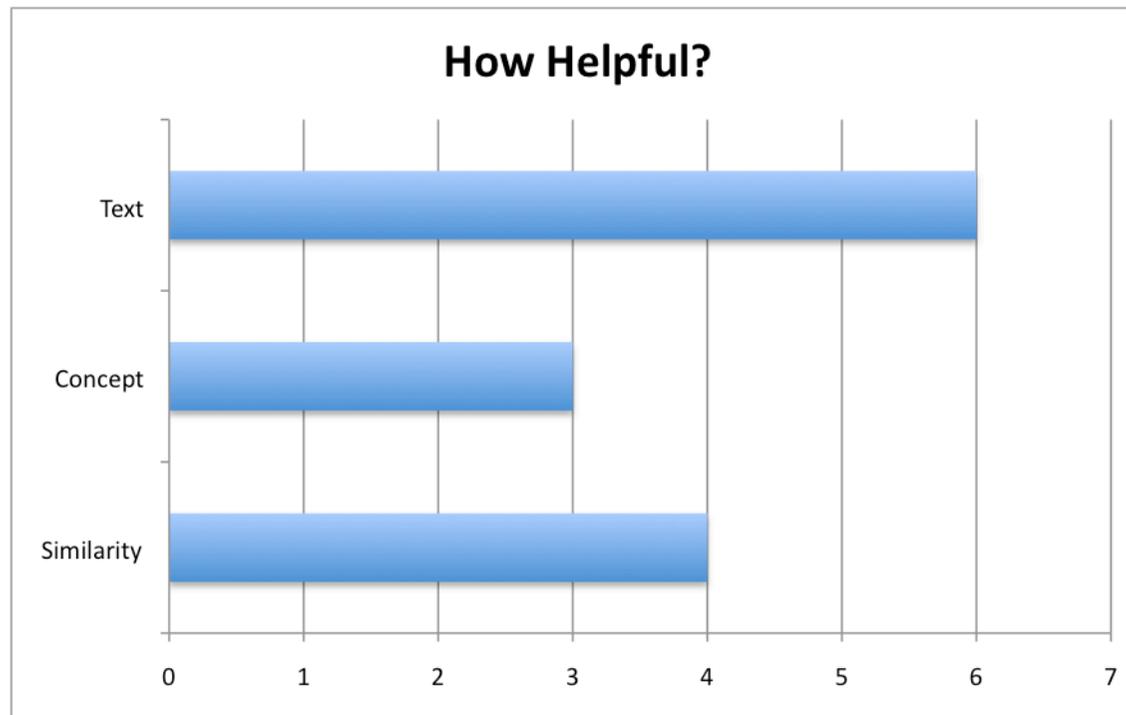
Search Techniques



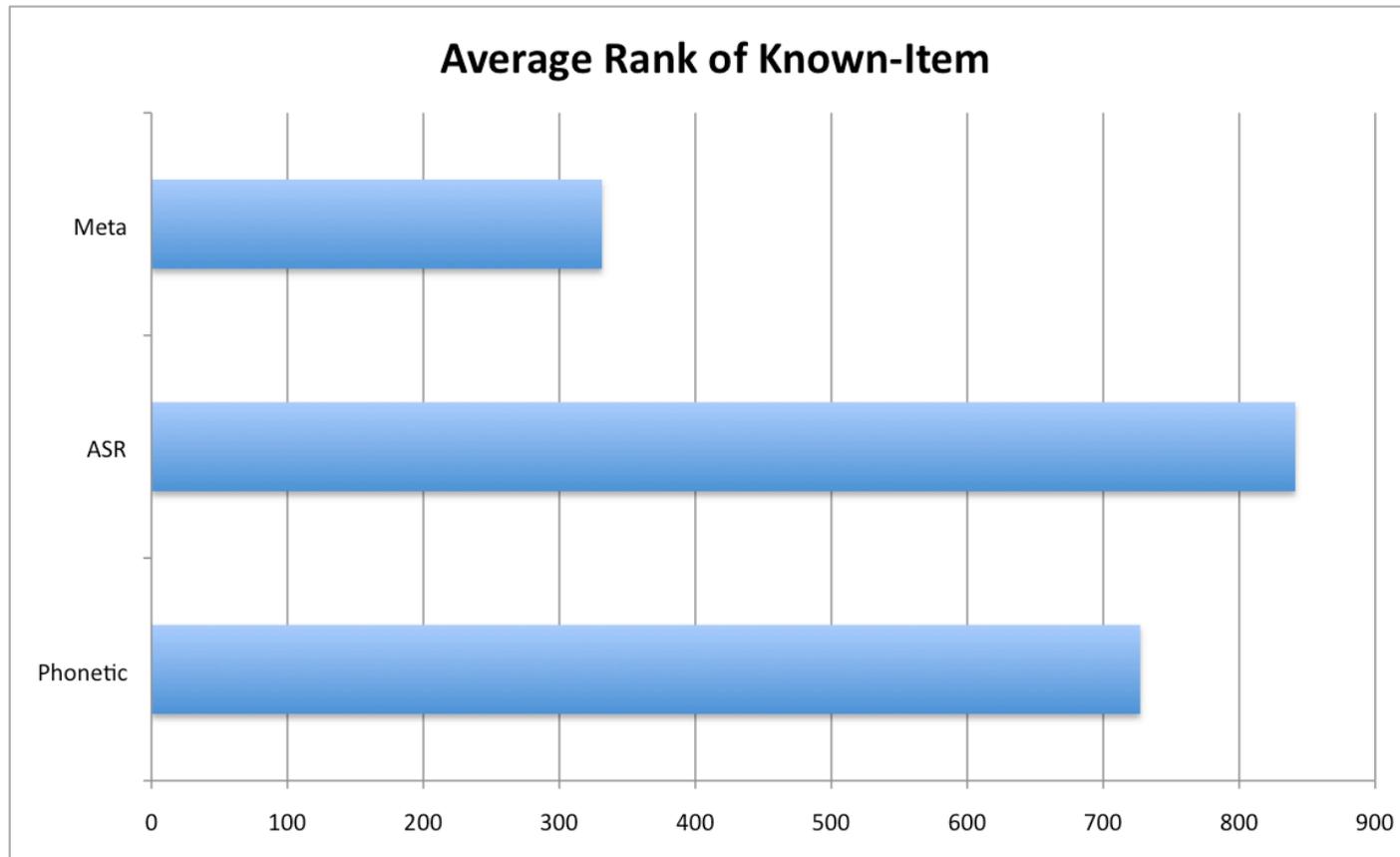
Questionnaires 1 – Overall System Questions



Questionnaire – Different Search Techniques



Text Search Performance



Conclusions

- Developed a simple, intuitive iPad app as a front-end to a video search engine
- Official runs show novices and experts performed the same
- Post-experiment analysis of our extended runs shows experts performed better overall in terms of mean elapsed time
 - However average rank of known-item was better for novices

Conclusions II

- Experts used fewer queries than novices
- Text search most popular querying technique
 - Text & Concepts most popular for experts
 - Text alone most popular for novices
 - Content based techniques used more by experts
- Novice users found our system easy to use and easy to learn how to use

Thank You!

– Questions?