More than a thousand words...

Stefan Rüger
Multimedia and Information Systems
Knowledge Media Institute
The Open University
http://kmi.open.ac.uk/mmis
Projects

Projects | Hot
SPOTLIGHTED HOT PROJECTS

Service Web 3.0
Future Internet  Semantic Web and Knowledge Services
Coordinating the research, standardization, and dissemination activities creating the internet of billions of web services.

ECOSENSUS
Knowledge Management
Electronic/Ecological Collaborative Sensemaking Support System

The UK Multimedia Knowledge Management Network
Multimedia and Information Systems
Enhance communication between the experts in both academia and industry

KM Tools

KMi Tools

Future Internet

Knowledge Management

Multimedia & Information Systems

Narrative Hypermedia

New Media Systems

Semantic Web & Knowledge Services

Social Software

The Open University
Since 1995: 104 projects & 55 technologies

Current year

30 live projects (£2.3m ext, £1.2m internal)

- 17 EU
- 7 UK
- 1 US
- 5 internal (iTunes U, SocialLearn, ...
PHAROS Project

“Lighting the way for European audiovisual search”

Funded IP by EU under IST-FP6

Platform for search of Audiovisual Resources across Online Spaces
Built by the monks and nuns of the Nipponzan Myohoji, this was the first Peace Pagoda to be built in the western hemisphere and enshrines sacred relics of Lord Buddha. The Inauguration ceremony, on 21st September 1980, was presided over by the late most Venerable Nichidattsu Fujii, founder and ...
Snap.Send.Get™

Snap → Send → Get

[© 2007 SNAPTELL Inc, reproduced with permission]
### New search types

<table>
<thead>
<tr>
<th>text</th>
<th>video</th>
<th>Images</th>
<th>speech</th>
<th>sound</th>
<th>sketches</th>
<th>multimedia</th>
<th>query</th>
<th>doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>text</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>video</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>images</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sound</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sketches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>multimedia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example**

you roar and get a wildlife documentary
The semantic gap

1m pixels with a spatial colour distribution

faces & vase-like object

victory, triumph, ...
Piggy-back retrieval

<table>
<thead>
<tr>
<th></th>
<th>text</th>
<th>video</th>
<th>images</th>
<th>speech</th>
<th>music</th>
<th>sketches</th>
<th>multimedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>query</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>doc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>text</th>
<th>video</th>
<th>images</th>
<th>speech</th>
<th>music</th>
<th>sketches</th>
<th>multimedia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>text</td>
<td>video</td>
<td>images</td>
<td>speech</td>
<td>music</td>
<td>sketches</td>
<td>multimedia</td>
</tr>
</tbody>
</table>
Content-based retrieval

<table>
<thead>
<tr>
<th></th>
<th>text</th>
<th>video</th>
<th>images</th>
<th>speech</th>
<th>music</th>
<th>sketches</th>
<th>multimedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>query</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>text</td>
<td>video</td>
<td>images</td>
<td>speech</td>
<td>music</td>
<td>sketches</td>
<td>multimedia</td>
</tr>
</tbody>
</table>

The table illustrates the relationships between different types of content (text, video, images, speech, music, sketches) and how they can be used as queries or documents in content-based retrieval systems.
Features and distances

Feature space
Content of an image

145 173 201 253 245 245
153 151 213 251 247 247
181 159 225 255 255 255
165 149 173 141 93 97
167 185 157 79 109 97
121 187 161 97 117 115
Colour histogram
Texture

coarseness  contrast  directionality
Texture histograms

Gabor filter design

Query

[with Howarth, CLEF 2004]
Medical image collection, 8725 images, 25 topics
mean average precision 34.5%, 3rd within 1% of 1st
Tiled histograms
Points of interest

Many PoI, ie, many feature vectors
Quantised feature vectors \(\approx\) words
Bag of word model \(\approx\) text retrieval
[Vlad Tanasescu: Anticipation, SCiFi trailer, 2007]
Annotation

{Snow, ice, bear, grass, ...} ↔ {City, water, building, sky, ...}
Automated annotation as machine translation

water  grass  trees

the beautiful sun

le soleil beau
Automated annotation as machine learning

Probabilistic models:
maximum entropy models
models for joint and conditional probabilities
evidence combination with Support Vector Machines

[with Magalhães, SIGIR 2005]
[with Yavlinsky and Schofield, CIVR 2005]
[with Yavlinsky, Heesch and Pickering: ICASSP May 2004]
[with Yavlinsky et al CIVR 2005]
[with Yavlinsky SPIE 2007]
[with Magalhães CIVR 2007, best paper]
A simple Bayesian classifier

\[
P(w|I) = \frac{P(w, I)}{P(I)} = \frac{\sum_J P(w, I|J)P(J)}{\sum_J P(I|J)P(J)}
\]

\[
= \frac{\sum_J P(I|w, J)P(w|J)P(J)}{\sum_J \sum_w P(I|w, J)P(w|J)P(J)}
\]

Use training data \( J \) and annotations \( w \)

\( P(w|I) \) is probability of word \( w \) given unseen image \( I \)

The model is an empirical distribution \( (w,J) \)
Example: grass classifier

- very likely
- may be
- probably not
Modelling semantic concepts

Bayesian networks from training data

outdoor → town → crowd

sky → grass → tarmac → skin
Automated annotation

[with Yavlinsky et al CIVR 2005]
[with Yavlinsky SPIE 2007]
[with Magalhães CIVR 2007, best paper]

Automated: water buildings city sunset aerial

[Corel Gallery 380,000]
The good door

[beholdsearch.com, 19.07.2007, now behold.cc (Yavlinksy)]
[images: Flickr creative commons]
The ugly

[beholdsearch.com, 19.07.2007, now behold.cc (Yavlinksy)]
[images: Flickr creative commons]
State of the art

Effectiveness

- (Yavlinsky, Schofield, Rüger 2005)
- (Feng, Lavrenko, Manmatha 2004)
- (Barnard, Forsyth 2001)
- (Duygulu et al 2002)
- (Carneiro, Vasconcelos 2005)
- (Magalhães, Rüger 2007)
- (Amir, Argillander et al 2005)
- (Snoek, Worrning et al 2006)
- (Snoek, Gemert et al 2006)

Flexibility

- (Blei, Jordan 2003)
- (Lavrenko, Manmatha, Jeon 2003)
- (Jeon, Lavrenko, Manmatha 2003)

Efficiency
More than a thousand words...

Stefan Rüger
Multimedia and Information Systems
Knowledge Media Institute
The Open University
http://kmi.open.ac.uk/mmis