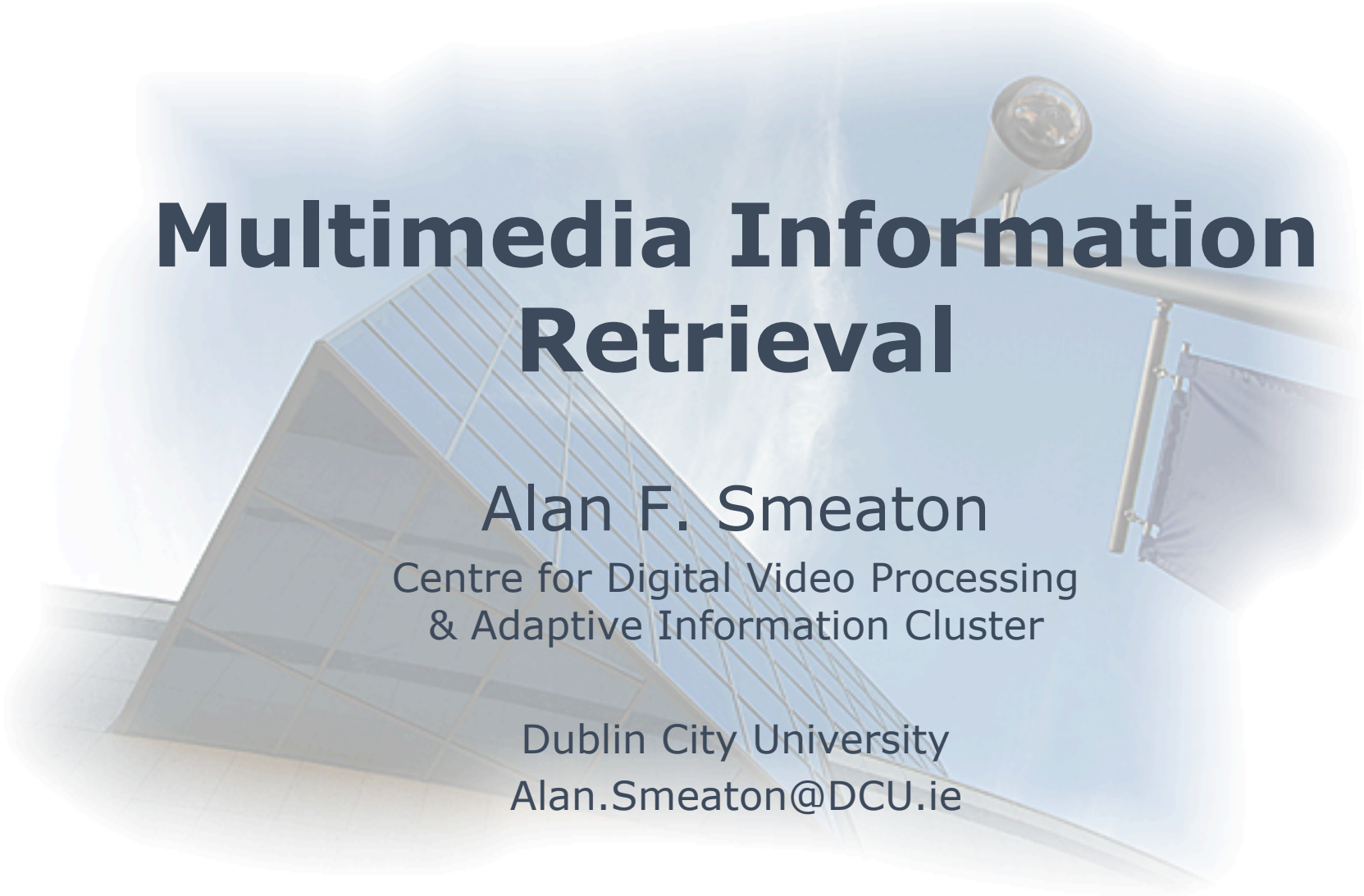


Multimedia Information Retrieval

Alan F. Smeaton
Centre for Digital Video Processing
& Adaptive Information Cluster

Dublin City University
Alan.Smeaton@DCU.ie

Contents

1. **IR vs. Information Seeking, and why it is hard**
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

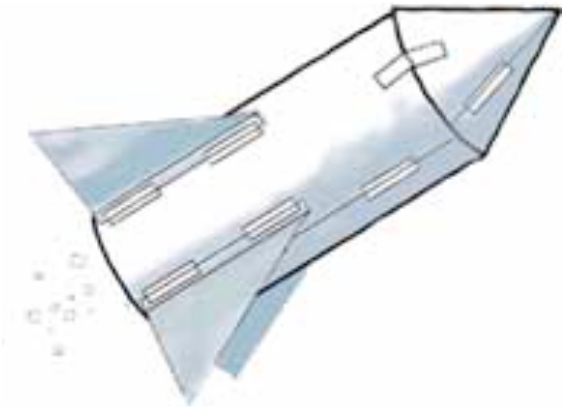
Information Retrieval

- Information retrieval looks easy
  - and everyone thinks they are good at it

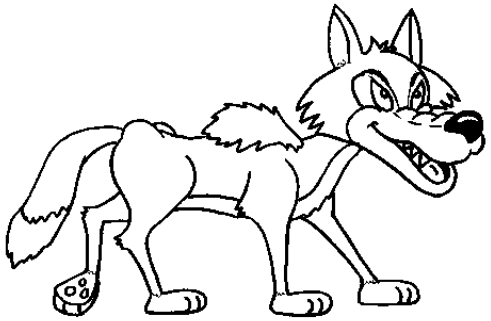
  - But accurate, high-quality IR is difficult
  - Yet we've grown tolerant of poor quality IR

Ian Ruthven

Sheep Attacks Rocket



What makes text IR hard is word polysemy



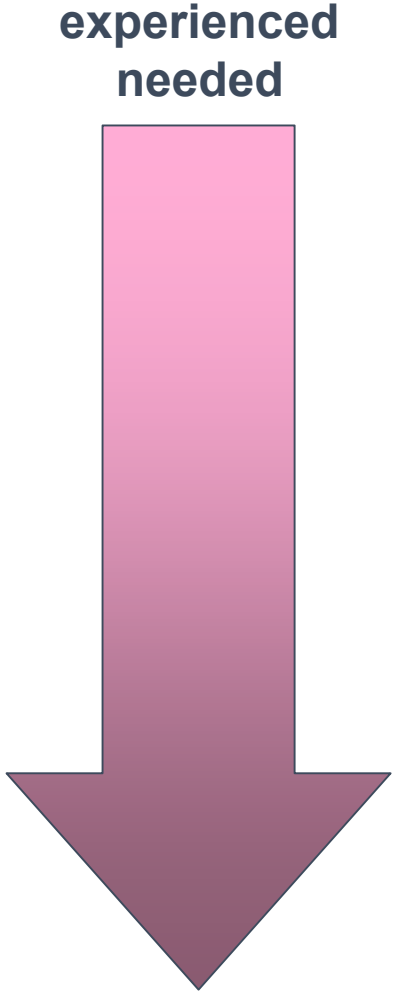
Nut screws washer and bolts

There's more to IS than IR

- Word polysemy is possibly **the** major issue in (text) IR
  - scale, interfaces, index freshness, etc. are all conquerable;
- Most (text) IR research, and IR evaluation, tries to find ways to work around this;
- Yet there is much more to evaluating IR …
  - what information do I want, what will I do with it, what information will be useful, how will I get it, what words should I use, what documents to save for later, how is this document different from the other 15 million that Google has retrieved, is there better information, do I trust this information, have I missed anything,…

ASPECTS OF Evaluation

- **System EVALUATION**
  - tests quality of IR system
  - high volume of queries
  - no user involvement
    - *Simulate* user
  - cheap and popular
  - highly controlled!
- **User EVALUATION**
  - tests quality of IR system + interface
  - (usually) low volume of queries
  - direct user involvement
  - artificial test
- **Operational EVALUATION**
  - similar to user but in real situations
  - expensive and difficult to run but very good test



IR and IS - The Emerging View

- Approaches to improving (text & MM) information access in general
  - computer science (system evaluation)
    - propose new (well-founded) solutions
    - evaluate them in evaluation campaigns to uncover
      - what benefits searchers and in what way
      - new questions for investigation information science
  - investigate searching behaviour from a human perspective (user evaluation)
    - identify generalities amongst searchers or search behaviour
    - identify meaningful differences between searchers or search behaviour
      - make recommendations to system designers

User-oriented evaluations

- IN PRACTICE, USER EVALUATIONS EVALUATE THE Whole system
  - algorithms plus interfaces
  - mostly comparative
    - e.g. two interfaces to same system
  - objective measures
    - e.g. number of relevant documents found, time to search
  - subjective measures
    - easy to search, easy to learn, popular
  - qualitative and quantitative analysis
    - also proposed are things like cost, quality of information, search satisfaction
  - less controlled than test collection evaluation
- .. but this is too expensive to do, so we do system evaluations instead.

Relevance in IR

- Relevance is important in the IS view, and is simplified in the CS view, of information retrieval;

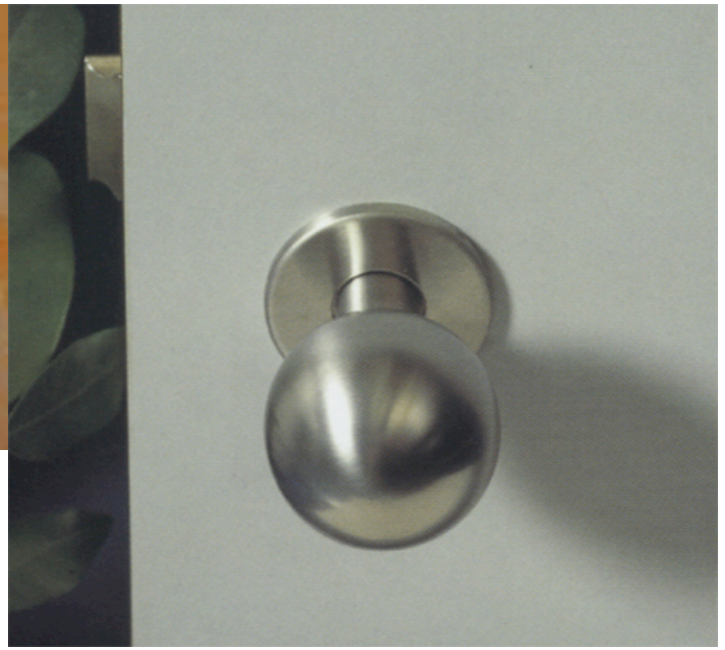
- It can mean many things (Saracevic)
  - ***system*** relevance
    - relation between query and document representation
  - ***topical*** relevance (aboutness)
    - relation between topic of query and topic of document
  - ***cognitive*** relevance (pertinence)
    - relation between user's state of knowledge and texts
  - ***situational*** relevance (utility)
    - relation between situation (problem) and texts
  - ***motivational*** relevance
    - relation between intention of user and texts

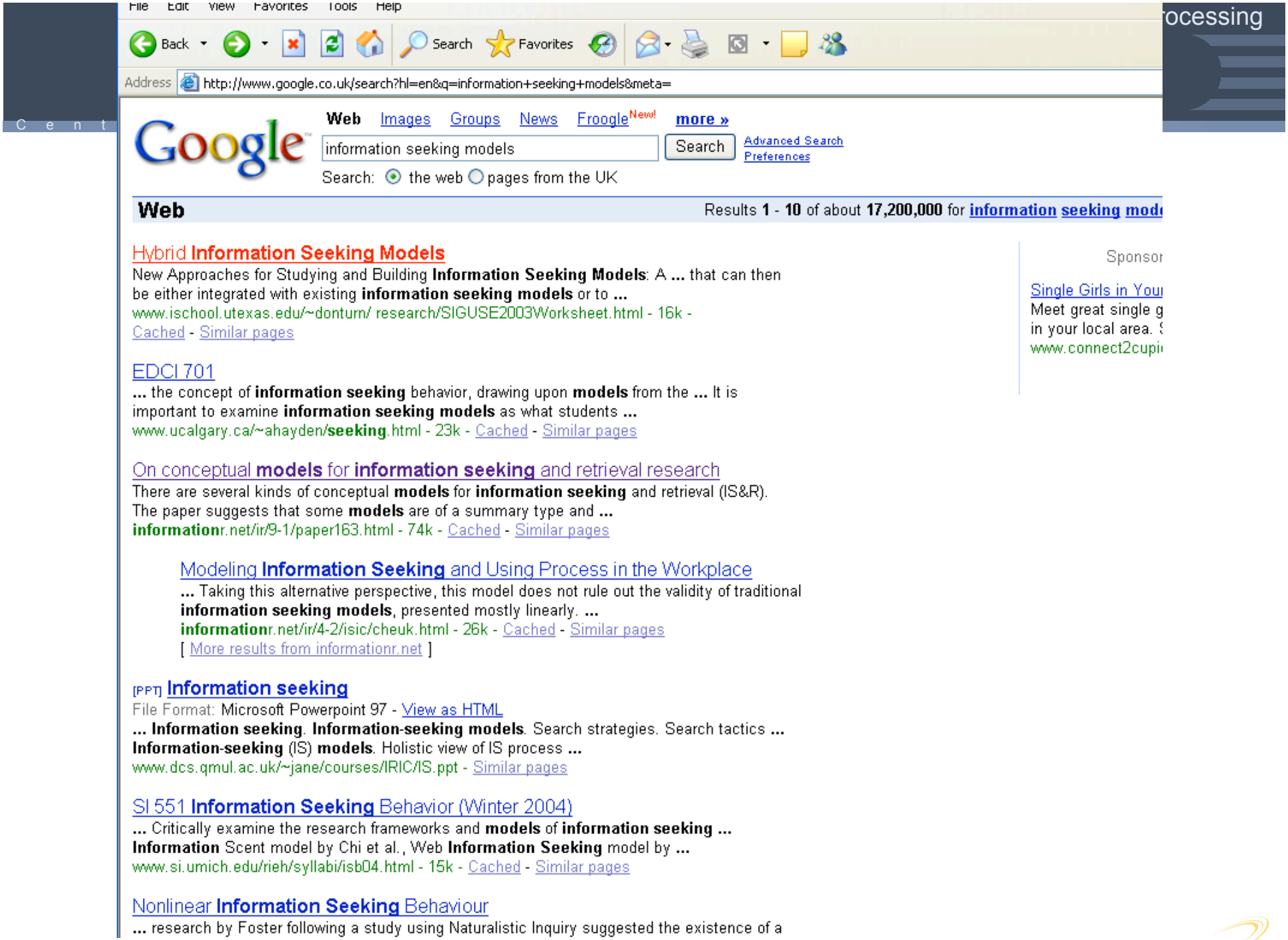
- In the CS view of IR, it is simplified, binary, non-complex whereas in the IS view of IR it is not;

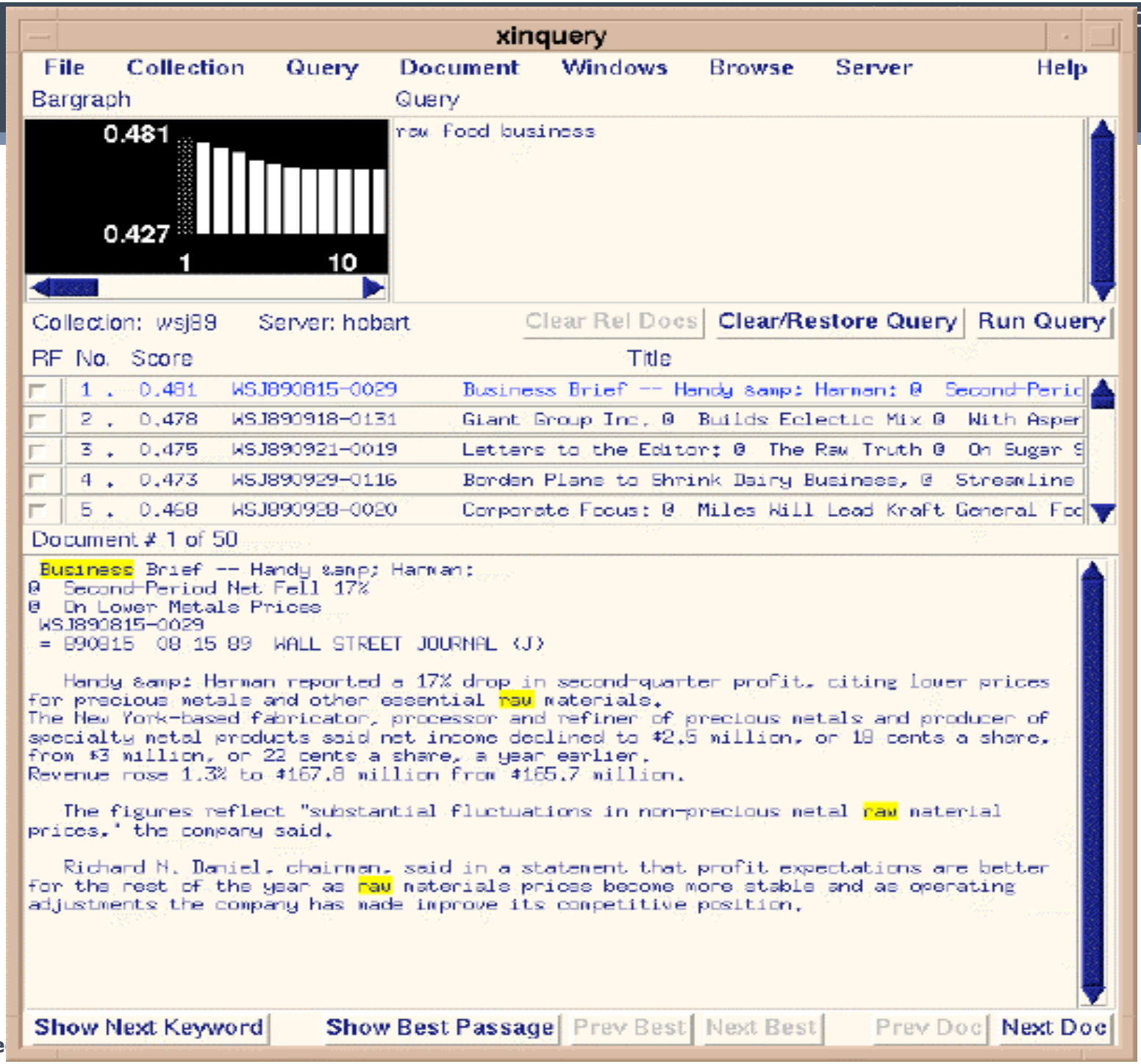
Don’t believe me ? IFs and IR/IS

- These things DO matter, but they are hard to address, so we don’t address them. Don’t believe it … look at SEARCH interfaces - features of common interfaces
  - what information seeking actions they support
  - what type of searches they support
  - what they do not support
- Interfaces are important, always are, and search interfaces determine actions the user can do, the system can support
- Interfaces make certain types of action more likely
  - generally known as the *affordance* of interfaces
- There is a difference between intended affordance and unintentional affordance
- For example, Google vs. Other system affordances?

Affordance of a door handle !




File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.google.co.uk/search?hl=en&q=information+seeking+models&meta=
Google
Web Images Groups News Froogle New! more »
information seeking models
Search
Advanced Search
Preferences
Search: the web pages from the UK
Web
Results 1 - 10 of about 17,200,000 for information seeking mod
Hybrid Information Seeking Models
New Approaches for Studying and Building Information Seeking Models: A ... that can then be either integrated with existing information seeking models or to ...
www.ischool.utexas.edu/~donturn/ research/SIGUSE2003Worksheet.html - 16k - Cached - Similar pages
EDCI 701
... the concept of information seeking behavior, drawing upon models from the ... It is important to examine information seeking models as what students ...
www.ucalgary.ca/~ahayden/seeking.html - 23k - Cached - Similar pages
On conceptual models for information seeking and retrieval research
There are several kinds of conceptual models for information seeking and retrieval (IS&R). The paper suggests that some models are of a summary type and ...
informationr.net/ir/9-1/paper163.html - 74k - Cached - Similar pages
Modeling Information Seeking and Using Process in the Workplace
... Taking this alternative perspective, this model does not rule out the validity of traditional information seeking models, presented mostly linearly. ...
informationr.net/ir/4-2/isic/cheuk.html - 26k - Cached - Similar pages
[More results from informationr.net]
[PPT] Information seeking
File Format: Microsoft Powerpoint 97 - View as HTML
... Information seeking. Information-seeking models. Search strategies. Search tactics ... Information-seeking (IS) models. Holistic view of IS process ...
www.dcs.qmul.ac.uk/~jane/courses/IRIC/IS.ppt - Similar pages
SI 551 Information Seeking Behavior (Winter 2004)
... Critically examine the research frameworks and models of information seeking ... Information Scent model by Chi et al., Web Information Seeking model by ...
www.si.umich.edu/rieh/syllabi/isb04.html - 15k - Cached - Similar pages
Nonlinear Information Seeking Behaviour
... research by Foster following a study using Naturalistic Inquiry suggested the existence of a
Sponsor
Single Girls in You
Meet great single g
in your local area.
www.connect2cupi


xinquery
File Collection Query Document Windows Browse Server Help
Bargraph
Query
0.481
0.427
1
10
raw food business
Collection: wsj89 Server: hobart
Clear Rel Docs
Clear/Restore Query
Run Query
RF No. Score Title
1 . 0.481 WSJ890815-0029 Business Brief -- Handy &amp: Harman: @ Second-Perio
2 . 0.478 WSJ890918-0131 Giant Group Inc. @ Builds Eclectic Mix @ With Asper
3 . 0.475 WSJ890921-0019 Letters to the Editor: @ The Raw Truth @ On Sugar S
4 . 0.473 WSJ890929-0116 Borden Plans to Shrink Dairy Business, @ Streamline
5 . 0.468 WSJ890928-0020 Corporate Focus: @ Miles Will Lead Kraft General Foo
Document # 1 of 50
Business Brief -- Handy & Harman:
@ Second-Period Net Fell 17%
@ On Lower Metals Prices
WSJ890815-0029
= 890815 08 15 89 WALL STREET JOURNAL (J)
Handy &amp: Harman reported a 17% drop in second-quarter profit, citing lower prices for precious metals and other essential raw materials.
The New York-based fabricator, processor and refiner of precious metals and producer of specialty metal products said net income declined to $2.5 million, or 18 cents a share, from $3 million, or 22 cents a share, a year earlier.
Revenue rose 1.3% to $167.8 million from $165.7 million.
The figures reflect "substantial fluctuations in non-precious metal raw material prices," the company said.
Richard N. Daniel, chairman, said in a statement that profit expectations are better for the rest of the year as raw materials prices become more stable and as operating adjustments the company has made improve its competitive position.
Show Next Keyword
Show Best Passage
Prev Best
Next Best
Prev Doc
Next Doc

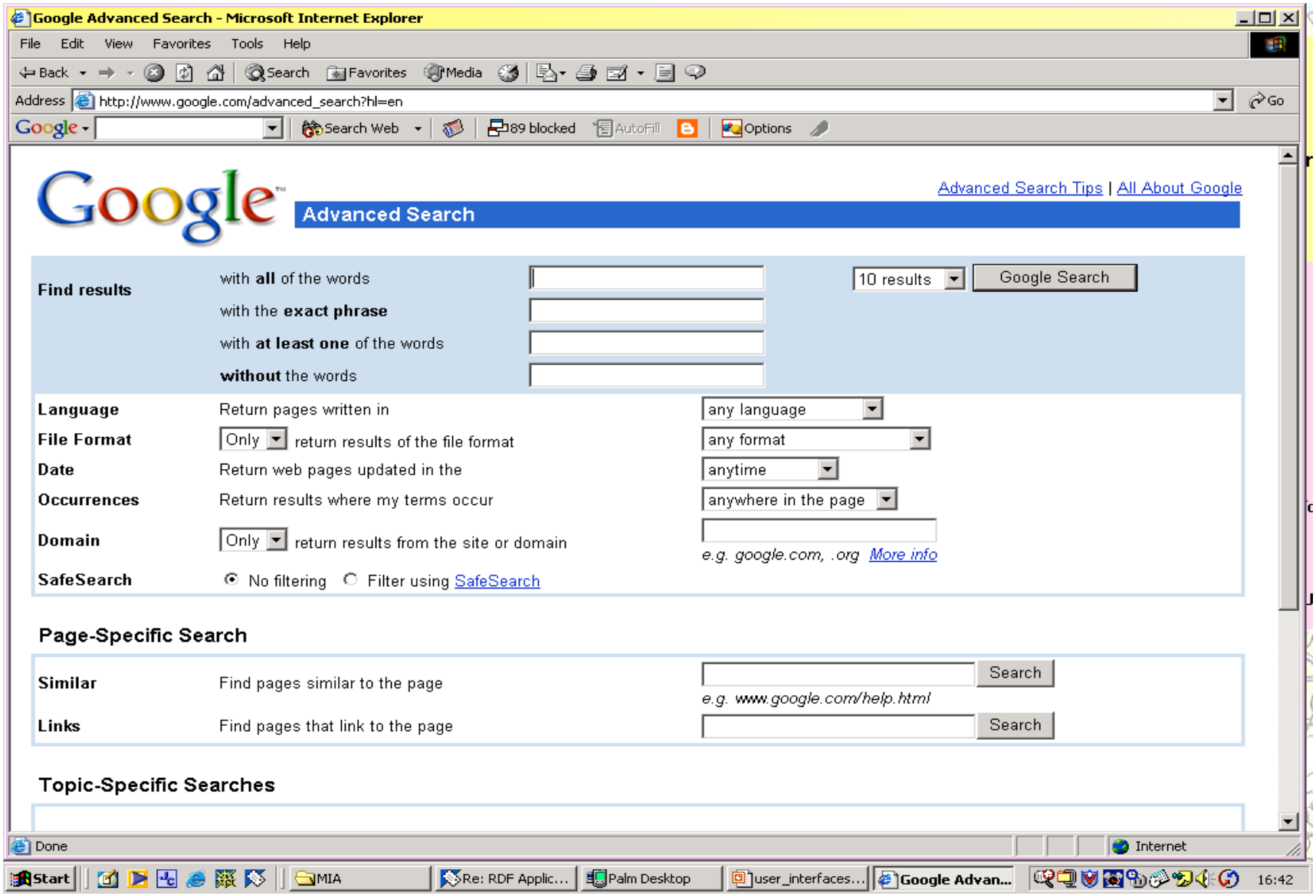
Key points

- standard interface allows for
  - query
  - examination of documents
    - all documents with at least one query term
    - top *n*
  - assessment of relevance
    - relevance feedback
      - binary – relevant/not relevant
      - partial – how relevant
      - range – relevant/partially relevant/partially non-relevant/not relevant

Key features ...

- supports cycle of activity
  - query -> document examination -> query/rf -> document examination
- documents and queries only entities
- good for rapid searching
  - quick analysis
- little support for
  - categorisation (types of documents, date of documents, etc.)
  - multiple querying in multitasking
  - storing search results
  - decision making
  - overview of collection

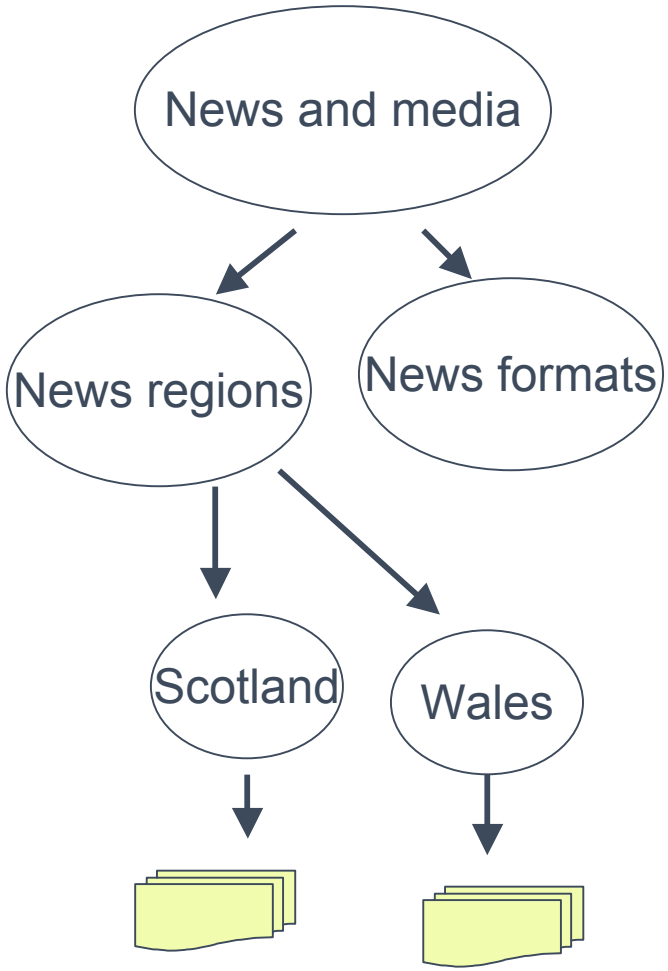
Form-based interface

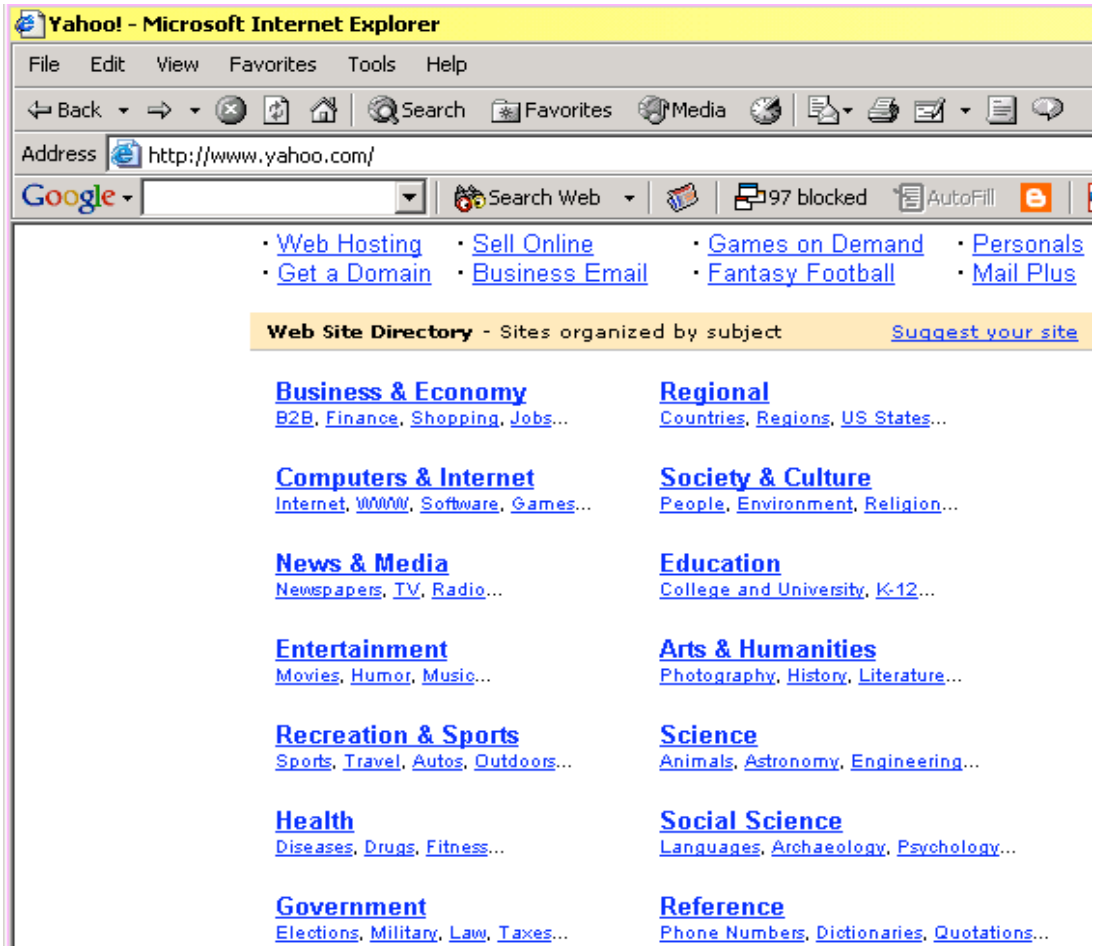


Main features ...

- emphasis on query
  - more support for complex queries
  - more support for precise queries
- support for meta-data
  - date, location, type
- but essentially narrowing options
- often cumbersome and tedious to use
  - especially for multimedia
- again little support for more complex actions
  - or for the difficult parts of searching

Category-based

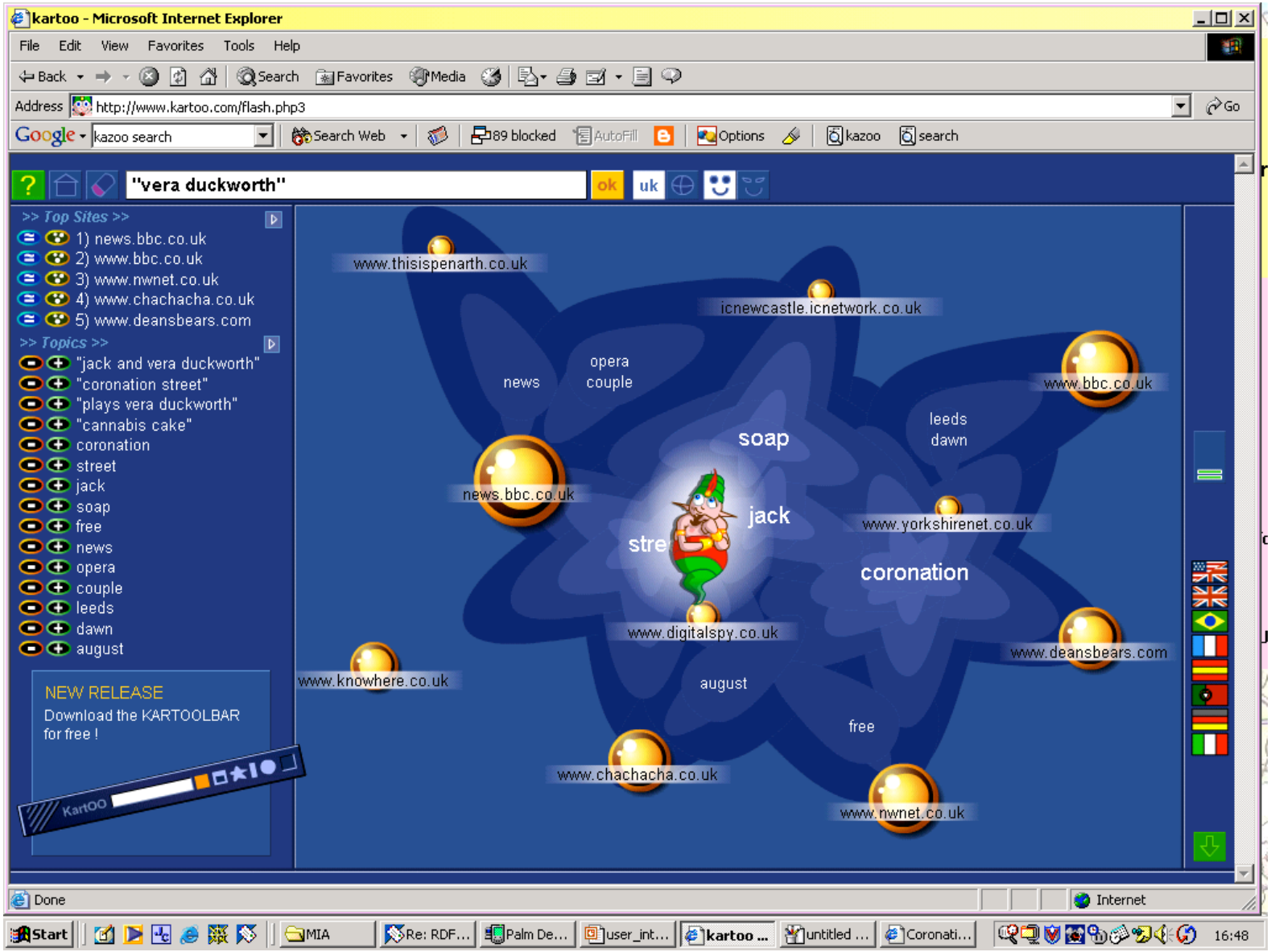




Main features ...

- possibly no search – browse only
- overview of collection
- but difficulties of categorisation
  - usually clustering
  - perhaps manual
- location issues
  - "where am I?"
  - "where is X?"
- some support for
  - completeness of search
  - organisation of analysis

Graphics-based – kartoo.com



Main features …

- specify relationships between concepts
  - terms, concepts, documents, websites, etc.
- overview of collection
  - main categories and concepts
- often, more direct access to information
  - less emphasis on documents
- still occasionally problem of defining queries
  - and little support….

Wrapping it up …

- IR, and now MMIR, is dominantly system-evaluation, ignoring context;
- Context is hard to capture, to use, and to evaluate;
- In this lecture we'll look at MMIR techniques and in tandem, how well/badly they work, using system evaluation (like everyone else);
- But really we should do more, if only we knew how.

Contents

1. IR vs. Information Seeking, and why it is hard
2. **MMIR**
   - **a) Key approaches and issues**
   - b) IR on Audio - spoken and musical
   - c) IR on Images
   - d) The nature of video
   - e) Commercial video IR systems
   - f) Experimental approaches (5)
     - a) Using video metadata
     - b) Using text
     - c) Keyframe image matching
     - d) Using semantic concepts and features
     - e) Using video objects
3. MMIR Evaluation
   - a) TREC
   - b) Copycat evaluations
   - c) TRECVid
4. Challenges and issues

Why MMIR?

- Search Engines allow (mostly) searching over textual data…
- Text IR is easy … match text query vs. text docs, but MMIR is harder 'cos we don't have queries !
- We could use text description of Q and D via user annotations but this has + and - (see later);

- What is MMIR?:
  - Find me the section of the movie where…
  - Forward to the goals in this match…
  - Just show me a summary of the F1 race, but I am only interested in Ferrari.
  - Find the person in this picture in this movie?

  - Much of it is visual, which is hard to express

Text vs. MM IR

→ ? → blah blah blah

→ Term term
term term
term term
term term

→ ? → ???????

→ ???? ????
???? ????
???? ????
???? ????

MMIR - key ideas

The trend in content-based retrieval systems for non-text is based on three key ideas:

1. Successful content-based retrieval systems (in image and video) are **domain-specific** and work best in those domains.

2. Automatic understanding tools are **difficult to develop**, so either:
   - Use simple techniques based on automatically extractable media primitives or;
   - Replace automatic techniques by interactive ones which involve users… i.e. manual mark-up.

3. Human involvement should be limited to **primitive tasks** that can be performed consistently.

MMIR – some issues

- Many underlying technologies are based on TEXT IR and exploit simple, proven methods.
  - Often text can be extracted from MM data
- Often a text IR based media retrieval system is augmented with media specific retrieval facilities operating over features of the underlying media.
- MM data has many possible interpretations, much more than text data.
- Hence we must develop a suite of retrieval techniques that can be used for these features.

Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

Retrieval of Spoken Documents

- From an IR perspective the retrieval of spoken documents is the most straightforward extension from text retrieval.
- Here we use text requests to search document content.
- However, the contents of spoken documents are not immediately available for indexing.
- An ideal solution would be to generate a perfect transcription of the spoken data then treat the document as text data and apply standard text information retrieval.
- But, perfect transcriptions are not available:
  - perfect automated transcription is not possible
  - full manual transcription is usually uneconomic.
- Speech recognition systems can be used to generate imperfect index information for the documents.
- Index errors arising from errors in speech recognition will reduce the performance, but overall retrieval can still be good enough to be useful.

Speech Recognition: An Introduction

- Human processing of spoken language is highly sophisticated, and unfortunately, currently poorly understood.
- Speech recognition can be thought of as a decoding task, specifically decoding air pressure wave signals into a written word signal.
- The speech signal is captured using a microphone and an audio amplifier. This signal is then sampled using an analogue-to-digital converter.
- A typical sample rate for telephone speech is 16,000 sample/second with 16 bits per sample, resulting in a data rate of 320,000 bits per second.
  - Data rates for hi-fi signals are much higher, e.g. CD uses 16-bits at 44.1kHz.
- A combination of efficient recognition algorithms and high-power computers are needed to achieve real-time speech processing.

Speaker Variability

- The speech patterns of different people show significant variations.
- Even for a given speaker every individual utterance of a word is unique.
- There may be changes in:
  - speed of delivery
  - the pitch of the voice
  - the pattern of stressing
  - the volume
  - background noise conditions
- Really it's remarkable that human listeners have the perception that the same word is spoken on each occasion!

Acoustic Ambiguity

- Some words are acoustically ambiguous. For example,
  - *to*, *too* and *two*

  are homophones which sound the same.

- Also, some words only have a very small acoustic distinction. For example,
  - *bee* and *pea*,

  while not true homophones, can be highly confusable, especially when the initial consonant is not well articulated.

- Early speech recognition was single, discrete words but current systems allow continuous speech recognition, i.e. recognition of flowing speech without silent gaps between words.

Context-Dependency

- All words can be broken down into a small set of constituent sounds, referred to as the phones of the language.
- English has about 45 distinct phones.
- For example, **speech** is composed of the phones **s p iy ch**
- Each phone can be produced a number of ways depending on its context, leading to slight variations in its sound.

Word vs Subword Recognition

- Current speech transcription systems typically have a vocabulary of around 80,000 words.
- Attempting to recognise 80,000 word models in parallel is not computationally possible.
- Also there are significant issues in availability of training data.
- The required recognition task can be dramatically reduced by performing recognition at the subword phone level. All words are recognised from their constituent phones.
  - This is not simply a phone level recogniser. A phonetic dictionary is used to constrain the allowed phone sequences to allow only those of real words.

Phonetic Modelling

- A basic approach to phone level recognition would be to simply have 45 HMM acoustic phone models and a phonetic dictionary showing the phonetic composition of each word.
  - e.g. the phrase ``Beat It!'' is represented by the phone sequence

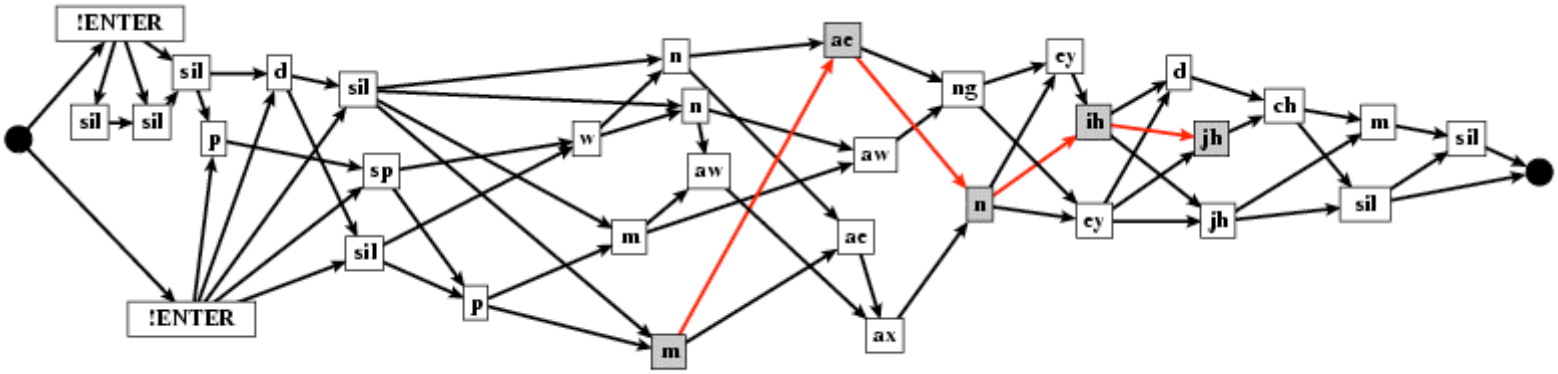
    sil b iy t ih t sil

- However, since phones are context-dependent a more sophisticated approach is required.

- The most common subword model in speech recognition is the triphone. In a triphone model a unique acoustic phone model is trained for each left and right phone combination.
- The previous example becomes

  sil sil-b+iy b-iy+t iy-t+ih t-ih+t ih-t+sil sil

Phone Lattice Spotting



Document Indexing - Phone Lattice Spotting

- The phone lattice is effectively a very constrained search-space of possible document contents.

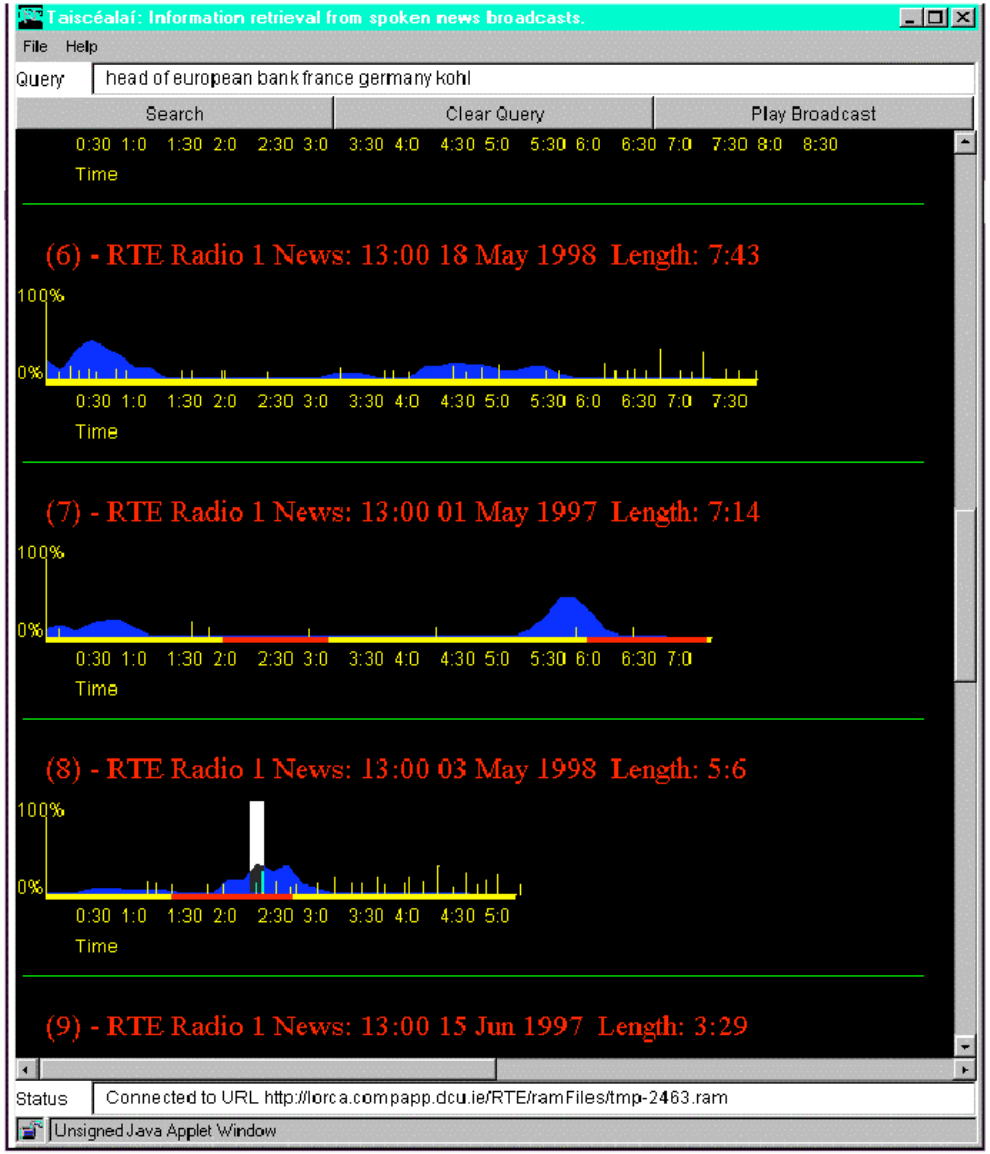
- Advantages:
  - open vocabulary - any word can be indexed at retrieval time.
  - very rapid indexing of any search word.

Example: Taiscealaí

- Our own Taiscealaí RTÉ news archive search did the same thing and was a real operational system which ran for 18 months.
- In the implementation, Taiscealaí, queries (text) were turned into phones via a dictionary lookup and the matching was a bag of tri-phones against bags of triphones.
- This project for RTÉ radio news broadcasts, recognised phones, chopped the stream of news into overlapping windows, and generated a document for each window as a retrievable unit of information (see online link to Taiscealaí paper).

Taiscealaí

C e n t e r f o r D i g i t a l V i d e o P r o c e s s i n g



Content based retrieval of music

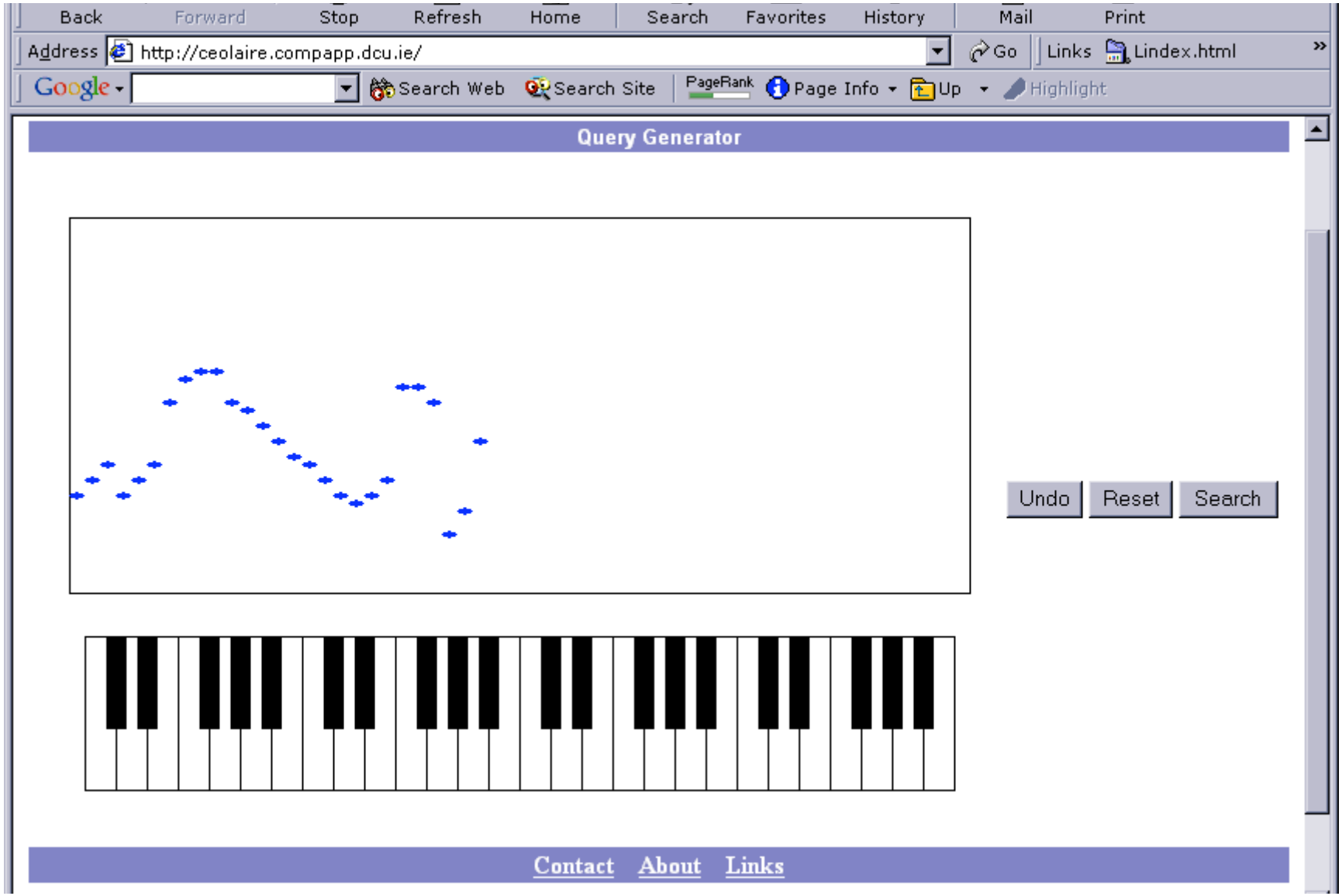
- Waveform matching?
  - It is difficult for a human to produce a perfect query which would support perfect note-note matching of audio documents, so some alternative is needed.
- MIDI is easy because…
  - Notes are available
  - Indexing terms are n-grams of notes…
- However if PCM or MP3 there are issues…
  - We must extract the notes…
    - Monophonic is not too difficult.
    - Polyphonic is difficult.
  - Once the notes are extracted we can then do conventional (text) retrieval.

Ceolaire

- CEOLAIRE supported IR on digital music by matching user's music queries directly against the melody content of recorded songs.
  - monophonic as opposed to polyphonic music.
  - CEOLAIRE processed music files stored in PCM or MP3.
- Songs were represented in CEOLAIRE in UDS format.
  - The sequence of note changes are used to represent both music and user's queries and matching occurs between these.

Ceolaire

C e n t e r f o r D i g i t a l V i d e o P r o c e s s i n g



UDS Notation

- UDS – Up, Down, Same
  - a perceived change in pitch is enough to match tunes from human memory.
  - Hence indexing music by UDS could work.
- The melody of a music file is defined by comparing the change from one note to the next.
  - irrespective of any timing information.
- UDS string n-grammed (size 6 worked best in Ceolaire)
  - Each n-gram is a term in the music document
  - Retrieval was using BM25… (text IR)

Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

IMAGE Retrieval

- Retrieval from image archives of:
  - Photos, technical drawings, legal documents, www images…
- Two types (for this lecture):
  - Text-based Image Search
    - Manual Annotation
    - Automatic Annotation, e.g. Google Image Search
  - Content-based Image Search
    - Low-level features
      - Colour, texture, etc…
    - Higher Level Attributes
      - Extracting semantic objects or meaning from images

Text Based Image Search

Google Images

Web **Images** Groups News **more »**

schloss Neuschwanstein Search Advanced Image Search Preferences

SafeSearch is off

Images Results **1** - **20** of about **2,270** for **schloss Neuschwanstein** . (**0.40** seconds)

Show: **All sizes** - Large - Medium - Small

schloss-neuschwanstein.jpg
400 x 300 pixels - 16k
www.familienfreizeittipps.de/ schloss-neuschwa...

umgebg2min.jpg
194 x 206 pixels - 9k
www.posthotel-ettal.de/html/ettal___bavaria.html

schloss-neuschwanstein.jpg
296 x 522 pixels - 32k
guide.supereva.it/.../2001/03/36683.shtml

neuschwanstein.jpg
350 x 458 pixels - 39k
www.bayerntours.com/.../neuschwanstein.htm

schloss_neuschwanstein.jpg
600 x 423 pixels - 51k
www.tourist-world.net/xxlnews/schloss_neuschw

neuschwan.jpg
280 x 350 pixels - 16k
www.ferienlandhaus.de/ fern.htm

neuschwa.jpg
240 x 207 pixels - 13k
www.mydaily.com/articles/2002/11/munich01.htm

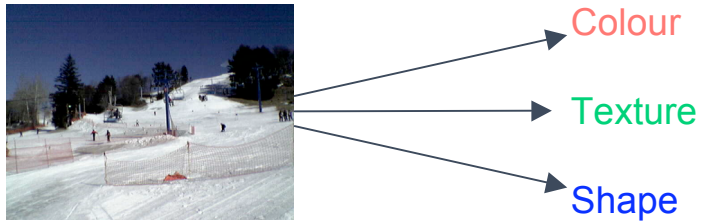
neuschwanstein.jpg
170 x 180 pixels - 14k
www.meteme.de/projekte/neuschwanstein.shtml

Automatic Text-based Indexing

- Using latent data from web page text to annotate images with textual description.
- Many possible sources:
  - Alternate text
  - Image URL
    - E.g. http://www.seriouswheels.com/ferrari/Ferrari-360-Modena-Coupe.jpg.
  - Neighbouring Text
  - Same Paragraph Text
  - Title
  - Heading
    - most recent heading prior to each image.
  - Anchor text
    - Windowing if necessary
  - Other terms
    - Any term on the web page

Content-based Image Search

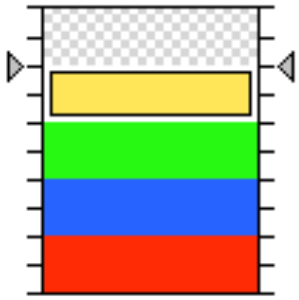
- Why?
  - Security, intellectual property, scientific imaging, large archives…
- How?
  - By examining the content of each image and extracting indexable data.



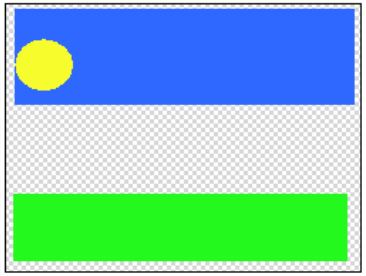
- Do this for each image in a collection.
- Do similar for any query image(s) and calculate distance between query and indexed images, thereby producing a ranked list.

How to query?

- How to query a content-based image retrieval system:
  - User query specification
    - required colours, textures, etc…
    - … or required features
    - … or keywords



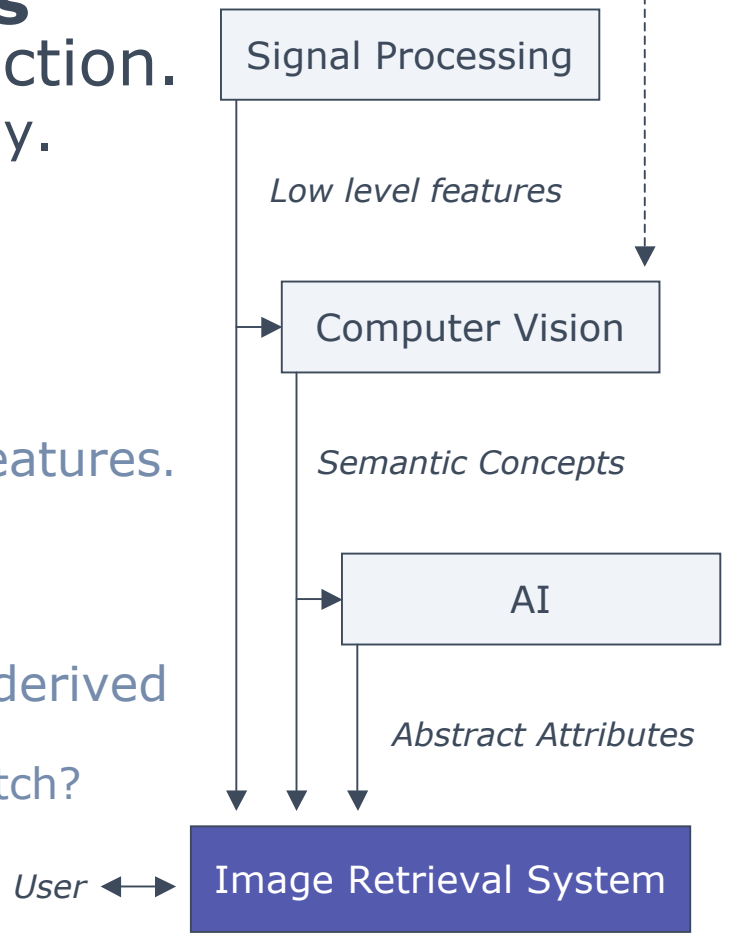
  - User drawing as input
    - Basic image requirements
    - Spatial arrangement



  - Query by example
    - Images with similar composition
    - **Relevance Feedback**

Two levels of Content-Based Retrieval

- Using **low-level features** based on automatic extraction.
  - Colour and Texture primarily.
- Using **Higher Level Attributes**
  - Semantic Concepts
    - Often based on low-level features.
    - Requires feature detectors.
  - Inferred abstract concepts.
    - Requires abstracting from derived features
      - E.g. what is a football match?



Using Low-level Features

- Low-level features are basically:
  - Colour
  - Texture
  - Shape
  - Spatial arrangement of the above
- Combining multiple sources together produces better results than one alone.
  - Weighted Fusion .. Usually based on training examples or query image categorisation.
- Using low-level features allows us to:
  - rank images w.r.t. a query image(s) based on visual similarity.
  - Rank images based on relevance feedback.

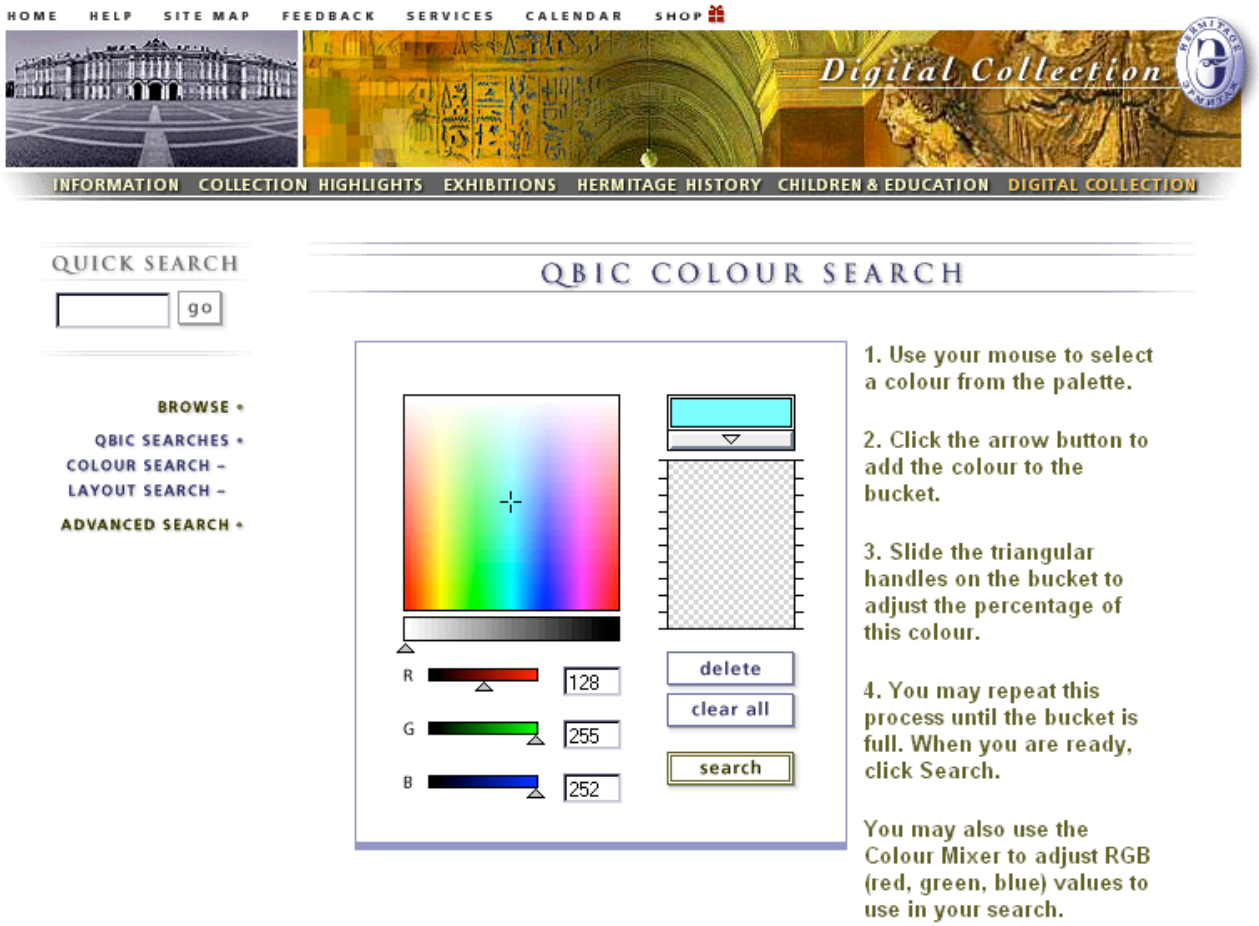
Example using Colour Histogram

- Take a sample image, convert to histogram and use to query the image index.

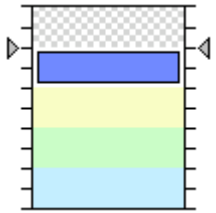
- Small variations in colour can lead to non-matches…
  - so we group perceptually similar colours together into colour bins.
  - A limited number of colour shades (10 to 15) are sufficient for visual discrimination.

Example Retrieval using Colour (QBIC)

http://www.hermitagemuseum.org/



Example Retrieval using Colour (QBIC)



1) View of the Valhalla near Regensburg

Klenze, Leo von
1836

2) Bouquet (Vase with Two Handles)

Matisse, Henri
1907

3) An Angler

Amaral, Tarsila do
Mid-1920s

4) Coast View with Apollo and the Cumaean Siby

Gellee, Claude (Le Lorrain) Between 1645 and 1649

5) Black Square

Malevich, Kazimir Severinovich
Circa 1930

6) Portrait of Suzanne Dufy, the Artist's Sister

Dufy, Raoul 1904

7) Pfaueninseln

Lutke, Peter Ludwig 1821

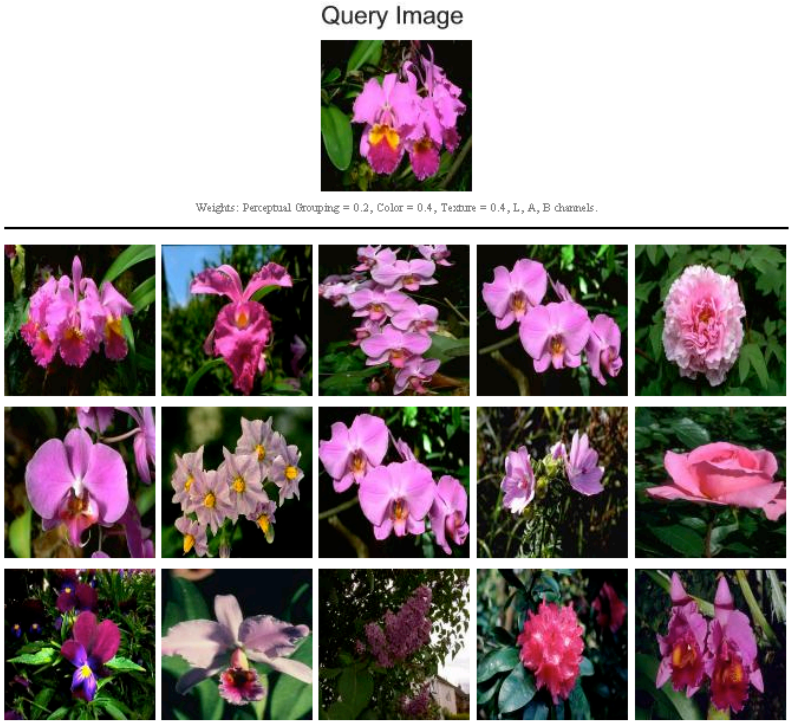
8) Piazza Santa Trinita, Firenze

Garibbo, Luigi (Louis) 1829

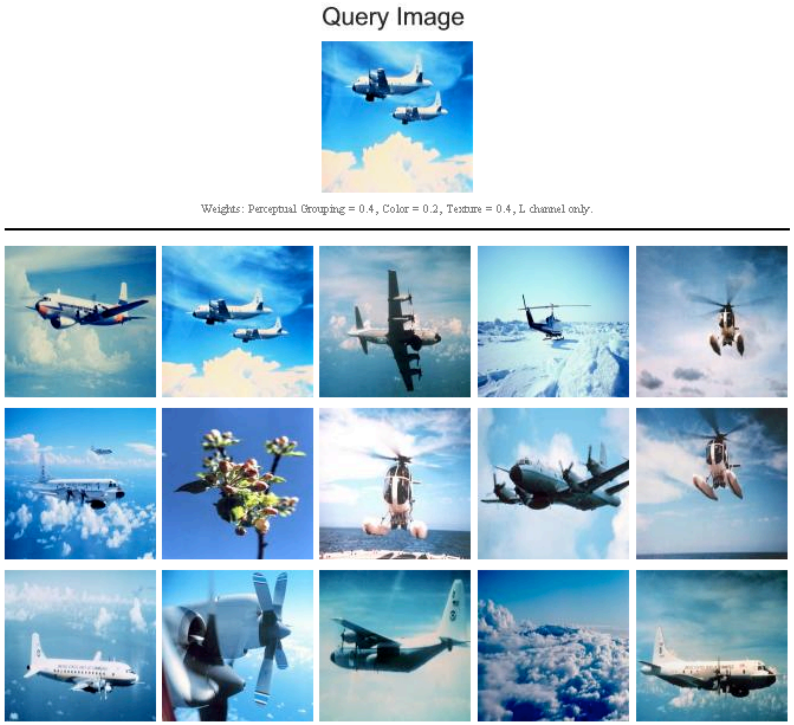
Example Feature Fusion

- Colour, texture and perceptual grouping (structure)

  http://amazon.ece.utexas.edu/~qasim/cires.htm



Perceptual Grouping=0.2, Colour=0.4, Texture=0.4



Perceptual Grouping=0.4, Colour=0.2, Texture=0.4

weighted fusion

Higher Level Attributes : Semantic Concepts

- Using low-level features is … UGH !
- Semantic Gap
  - A human interpretation of an image is not as colour, texture or any other low-level feature. Rather the human interpretation is naturally occurring semantic concepts.
- In images we have semantic concepts:
  - **objects** such as chairs, cars, crowds…
  - named specific individuals.
  - specific features of the image:
    - indoor/outdoor or landscape/cityscape.
- Retrieval is based on identifying and indexing images by concepts.

Try to find this using colour, texture & shape !

Semantic Concept Detectors

- Semantic concept detectors have been difficult to develop.
- Even in closed domains such as images from TV NEWS video, the detection rate for such features is not high.
- Annotation effort to help this process.

TV2004 Feature Extraction Results

	Boat	Albright	Clinton	Train	Beach	Basket	Airplane	Walking	Violence	Road
Avg	0.0506	0.0465	0.0921	0.0180	0.0174	0.2380	0.0426	0.0754	0.0264	0.0705
Mdn	0.0420	0.0495	0.0525	0.0020	0.0640	0.2630	0.0325	0.0610	0.0085	0.0680
High	0.1370	0.3040	0.2220	0.1090	0.0640	0.5610	0.1280	0.1700	0.0860	0.1470
Low	0.0010	0.0000	0.0010	0.0000	0.0010	0.0000	0.0020	0.0080	0.0000	0.0030

Higher Level Attributes : Inferred Concepts

- Detecting the meaning of an image…
  - A more natural representation of an image.
  - The semantic gap again!
- Inferred abstract attributes of an image:
  - do not correspond to any direct content in the image.
  - e.g. if we have football players and a goalpost and a football, then we have a football match.
- High-level to low-level image retrieval:
  - find images of a football match as opposed to;
  - find images of 2 footballers and a football as opposed to;
  - images of green with a grass texture, and two splashes of black&white in a striped arrangement with a black/white circular pattern in the middle of the two splashes

Overall Concept Detectors

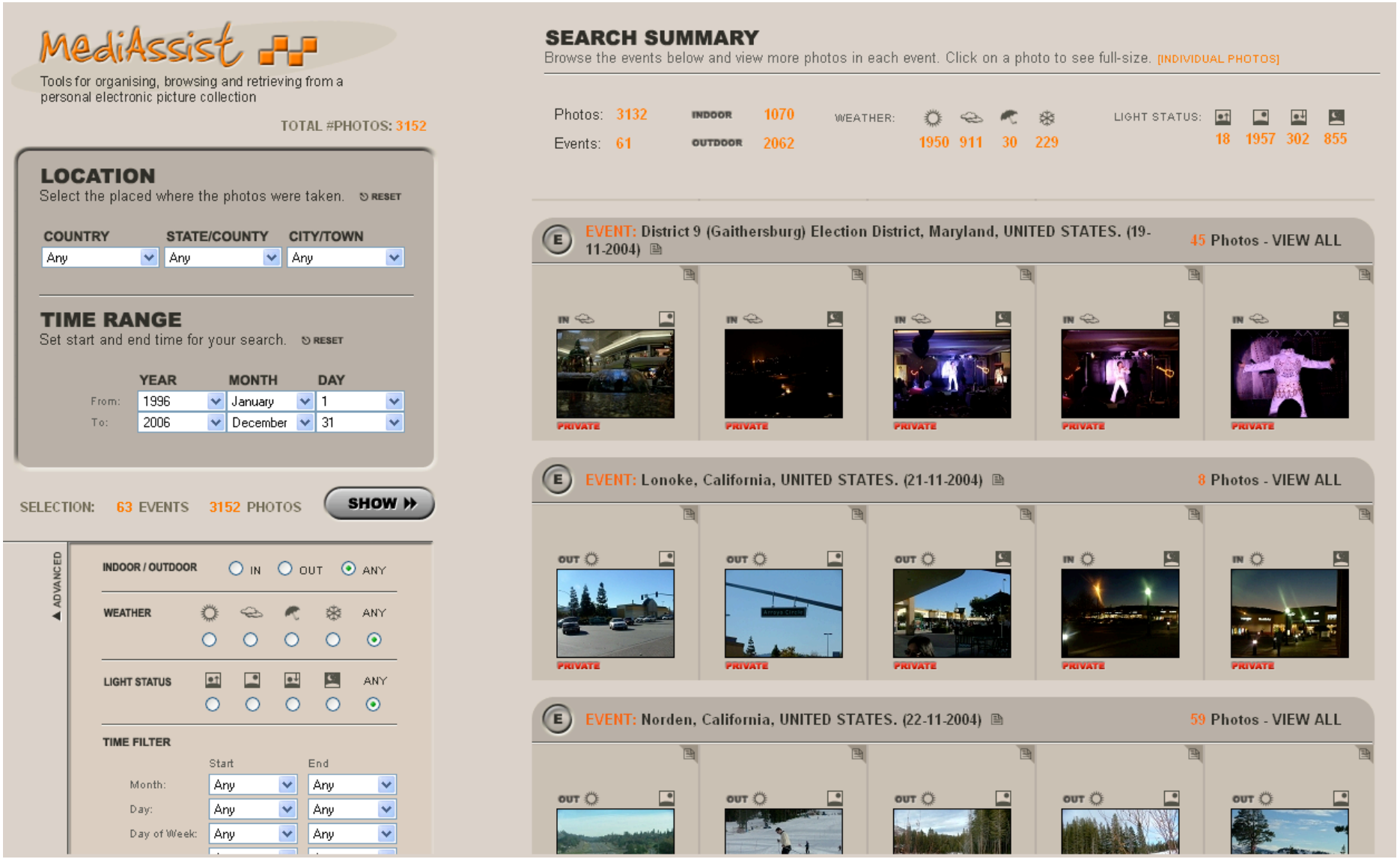
"Picture of an F1 motor race"

"Picture of a storm over a city"

Managing Digital Photos

- Of course complex content analysis is not always necessary.
  - E.g. digital photos.
    - Current software uses time/date and manual annotation.
  - BUT: Cameras capture context metadata
    - Time, flash, brightness, etc…
    - Adding location allows us to infer:
      - Daylight status and weather.
- Addition of feature detectors will help:
  - Feature detectors such as:
    - Known people, building and other object, scene, baby, and may more detectors.
- This gives us a very useful personal photo content management system

Example context-based photo tool



MediAssist

Tools for organising, browsing and retrieving from a personal electronic picture collection

TOTAL #PHOTOS: 1722

LOCATION

Select the placed where the photos were taken. RESET

COUNTRY	STATE/COUNTY	CITY/TOWN
UNITED STATE:	Maryland	Any

TIME RANGE

Set start and end time for your search. ▶ CALENDAR

2001 2002 2003 2004 2005

SELECTION: 5 EVENTS 26 PHOTOS SHOW

◀ ADVANCED

PHOTO DETAIL

This is a full-size photo. Amend annotation, or find similar photos. [BACK]

Date: 16-11-2004
Time: 11:14pm
Location: District 9 (Gaithersburg) Election District, Maryland, UNITED STATES

Faces Detected: 3

Hide Faces

Done

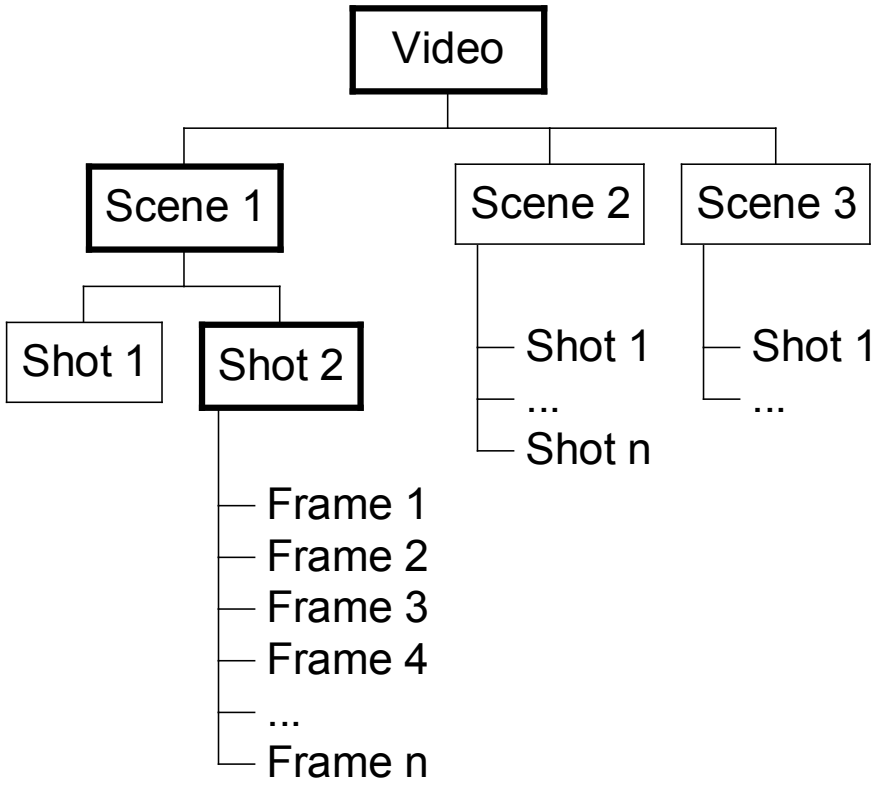
Local intranet

Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

What is Digital Video?

- Sequence of images (25-30 fps) often with audio.
- Mostly compressed:
  - Many encoding formats
- Video is structured:
  - So what is the unit of retrieval?
  - Dependent on task



The Nature of Video

- Video is a collation of images presented so fast they give the illusion of motion;
- A “programme” is divided into logical scenes, and a scene is composed of 1 or more camera shots;
- A shot is a single motion of a camera in time and can involve camera motion like panning, zooming, tracking or booming;
- Objects in a camera shot can also move;
- Shots can transition using fade in/out, fade to black, wipes, or the more usual hard cut;

- There’s an awful lot of video 10,000 TB = 1M hours of original TV and radio content in 2002;

Problems with Video ?

- Many of the issues associated with video in digital form are solved !
  - Capture, formatting, compression, storage, transmission, rendering on fixed and mobile
- Digital video is widespread and outstanding challenge(s) are managing video content;
- Key catalyst in this has been compression format standards;
- There are several (!) AVI, QuickTime, WMV, etc. but the ones that matter are MPEG;

MPEG Family

- MPEG-1
  - VCD (352x288 @ 25fps), for CD-ROMs and higher
- MPEG-2
  - DVD (720x576 @ 30fps), and higher
  - Better compression
- MPEG-4
  - Very low bit rates, better compression
  - Object based encoding
- MPEG-7
  - XML based video, audio and content description language… for describing content.

Managing Video

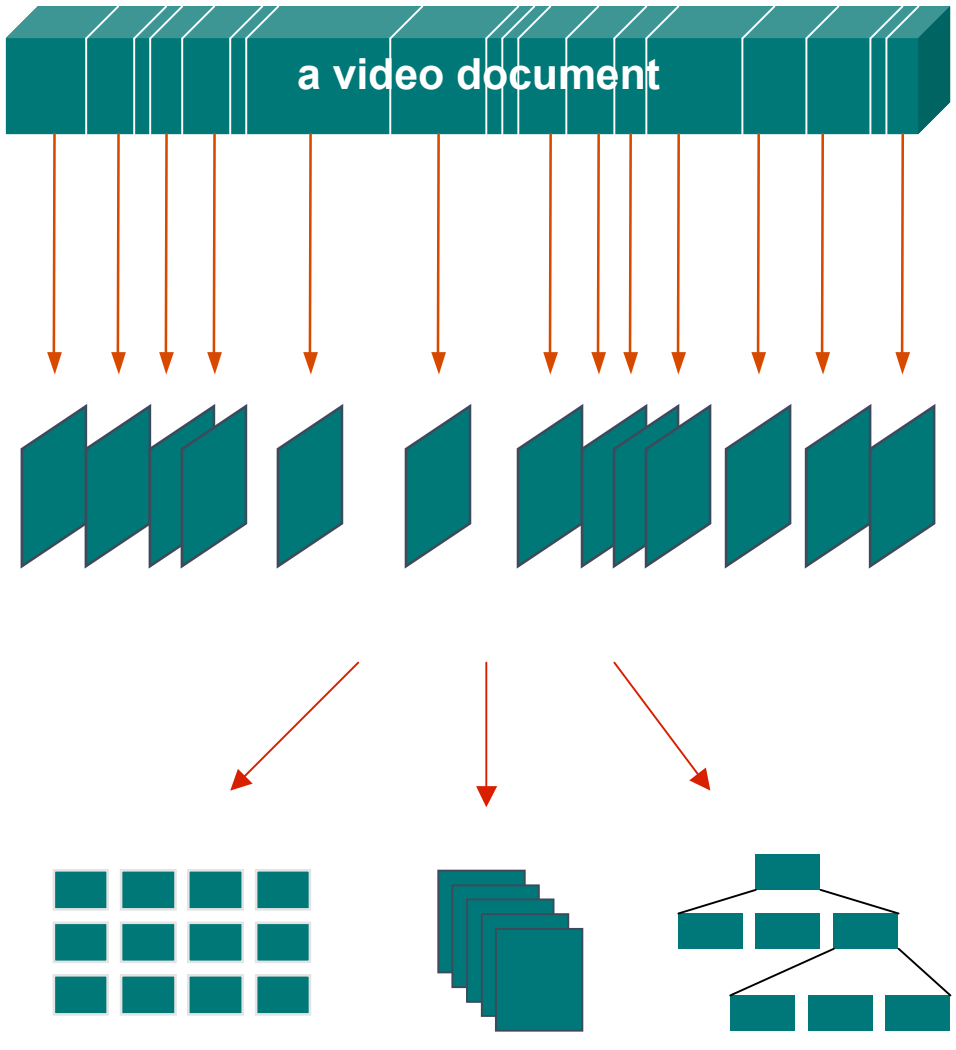
- Managing means many things
  - Analysis, indexing, summarising, aggregating, browsing and searching;
  - Some of these are very domain dependent eg sports vs. TV sitcom summarisation;
- Video analysis & search has much overlap with image
  - + easily scalable to larger volumes;
  - - doesn't take advantage of object/camera motion
- Good example of this is automatic structuring of video, necessary in order to make any progress in managing it - a.k.a. shot bound detection;

Shot Boundary Detection

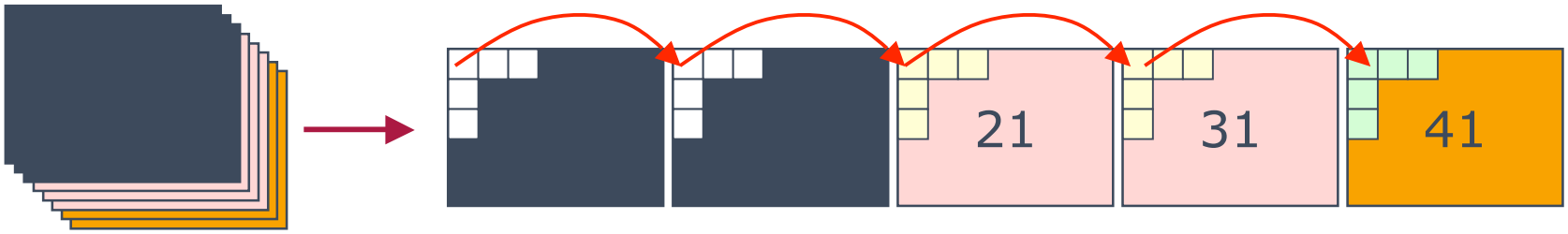
- A shot in video information is a sequence of continuous images (frames) from a single camera.
  - Shots are the retrieval units used by TREC for video IR
  - A shot boundary is crossed when a new camera is used, or a recording instance ends and a new one begins.

- The task of SBD or automatic video segmentation is to segment video into its constituent shots…
  - Why? To allow content-based operations over video at granularity of shot units.
    - Browsing
    - Searching
  - How? By examining exery X frames / adjacent frames to look for shot cuts.
    - Simple shot cuts are easy to process

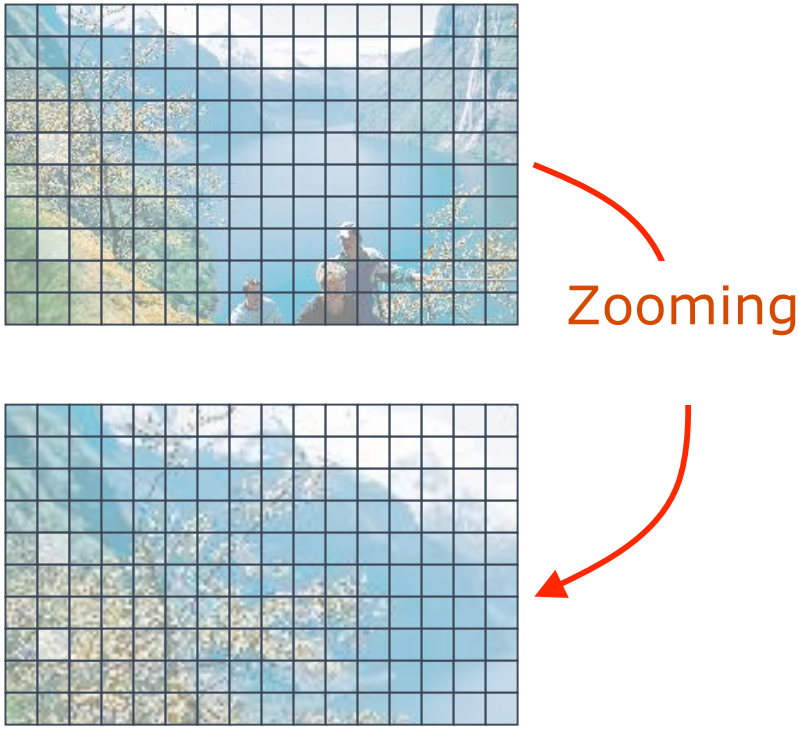
Automatic structuring of video



Shot Boundary Detection



- But it can be quite difficult because of camera tricks:
  - Dropped Shot Boundaries
    - fade-in and fade-out
    - dissolving
    - morphing
    - Wipes
    - Many other chromatic effects
  - False Shot Boundaries
    - zooming and panning
    - tilting,
    - booming and tracking
    - events in the content itself

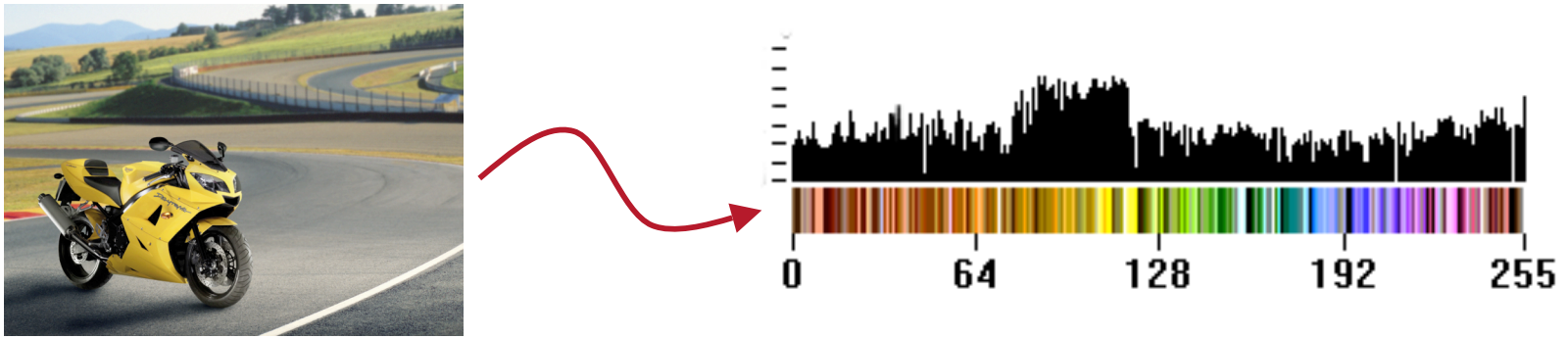


MPEG-7 Features

- Just as MPEG-1, -2 and -4 have catalysed video technology, MPEG-7 content description is enabling content applications;
- MPEG-7 describes basic characteristics of AV content e.g. shape, colour, texture, and user-defined;

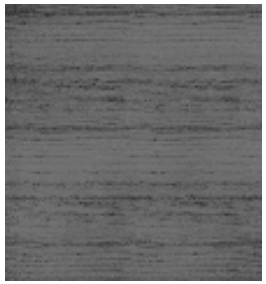
MPEG-7 Colour

- Dominant colour(s) for object, region or image/frame;
- Colour layout;
- Colour structure;
- Scalable colour (colour histogram)



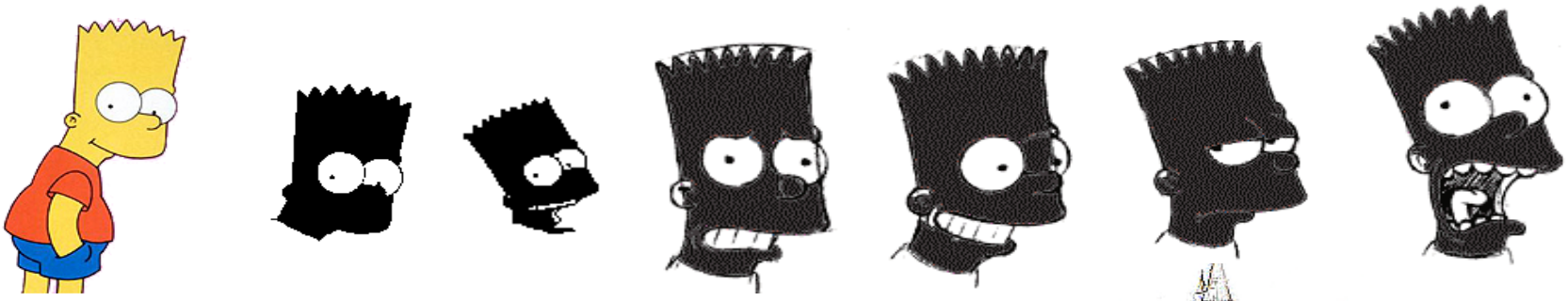
MPEG-7 Texture

- Texture is a measure of property such as smoothness, coarseness and regularity;
- Homogeneous texture and non-homogeneous texture (edge histograms) for regions, objects and frames/images;



MPEG-7 Shape & Motion

- Contour-based shapes for image/frame or regions;
- Camera motion, motion intensity, motion trajectory;



Complex shapes are difficult to process…!

Contents

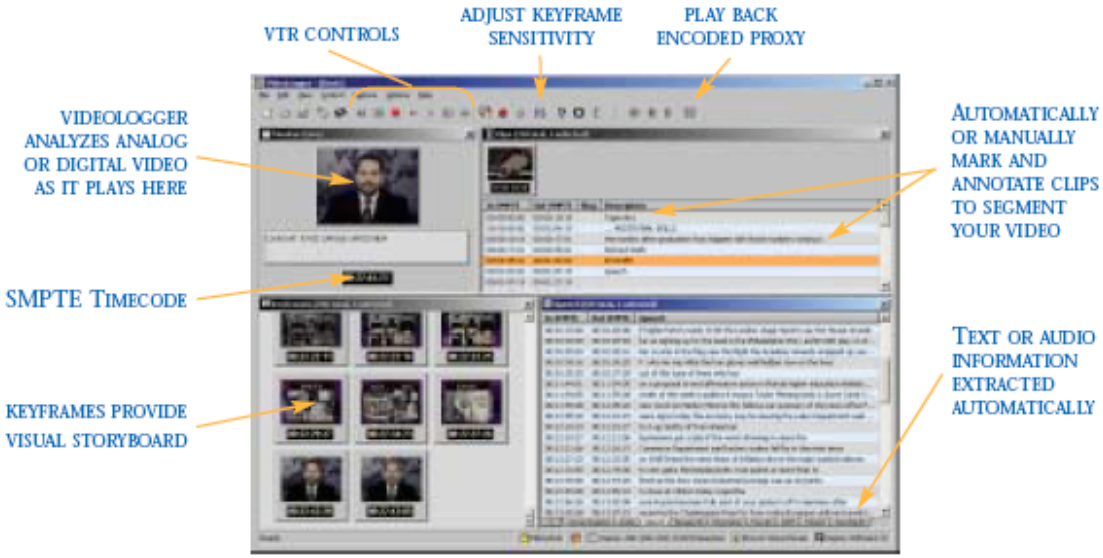
1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

VIDEO Retrieval

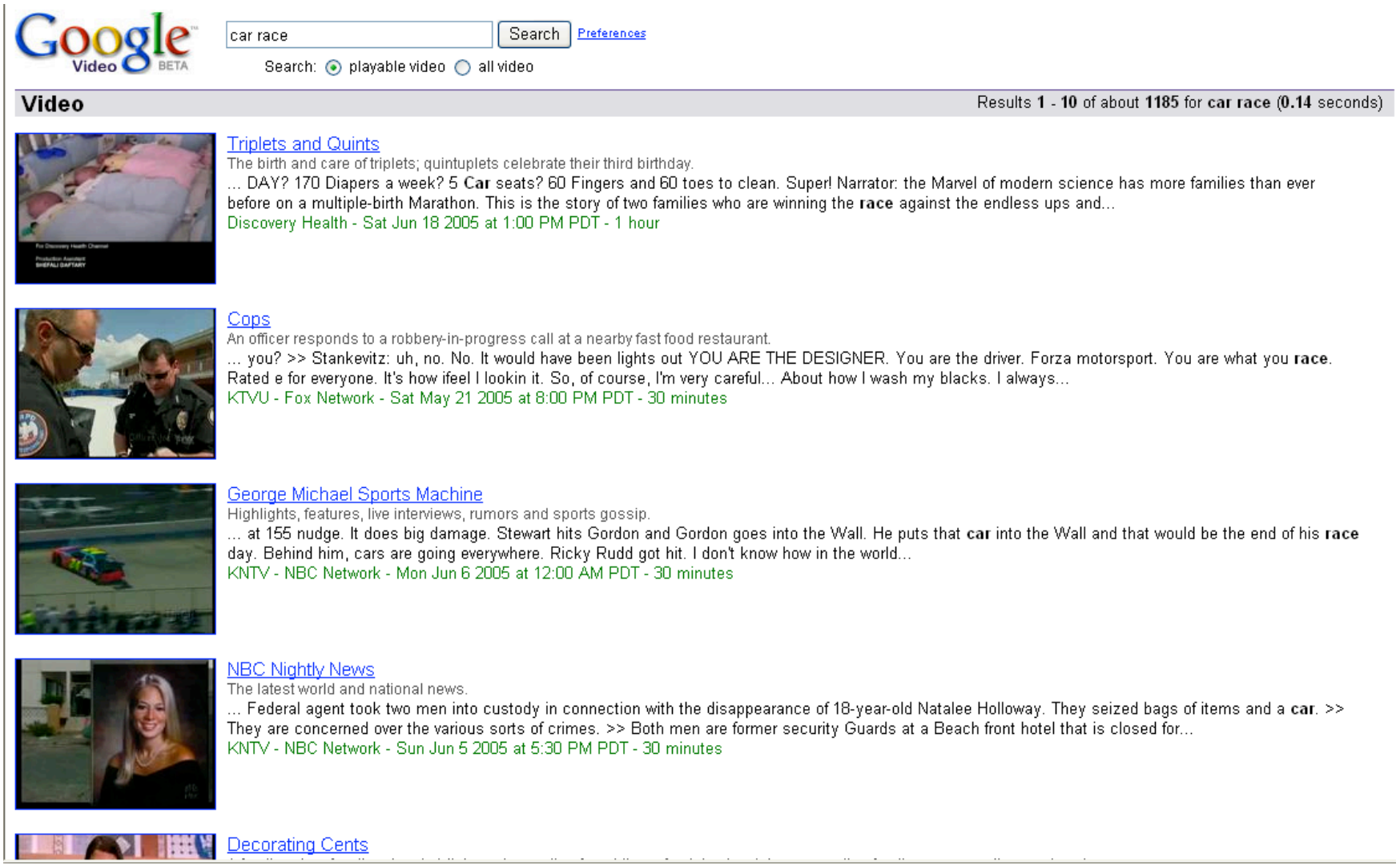
- A 'hot' topic for IR.
  - Annual TRECVid workshop driving research.
  - Vast quantities of digital video.
    - E.g. the BBC archive has +500k queries/year plus many millions of items.
- Why Video IR?
  - Large Archives
    - Broadcasters, Archivists, Personal Collections…
  - Manual Annotation is too tedious and time consuming.
    - Automatic Indexing is key.
- What video?
  - Television, NEWS, stock footage, movies, home videos, CCTV surveillance, music videos…
  - … and then there is recent public video like YouTube

Commercial Example : Virage

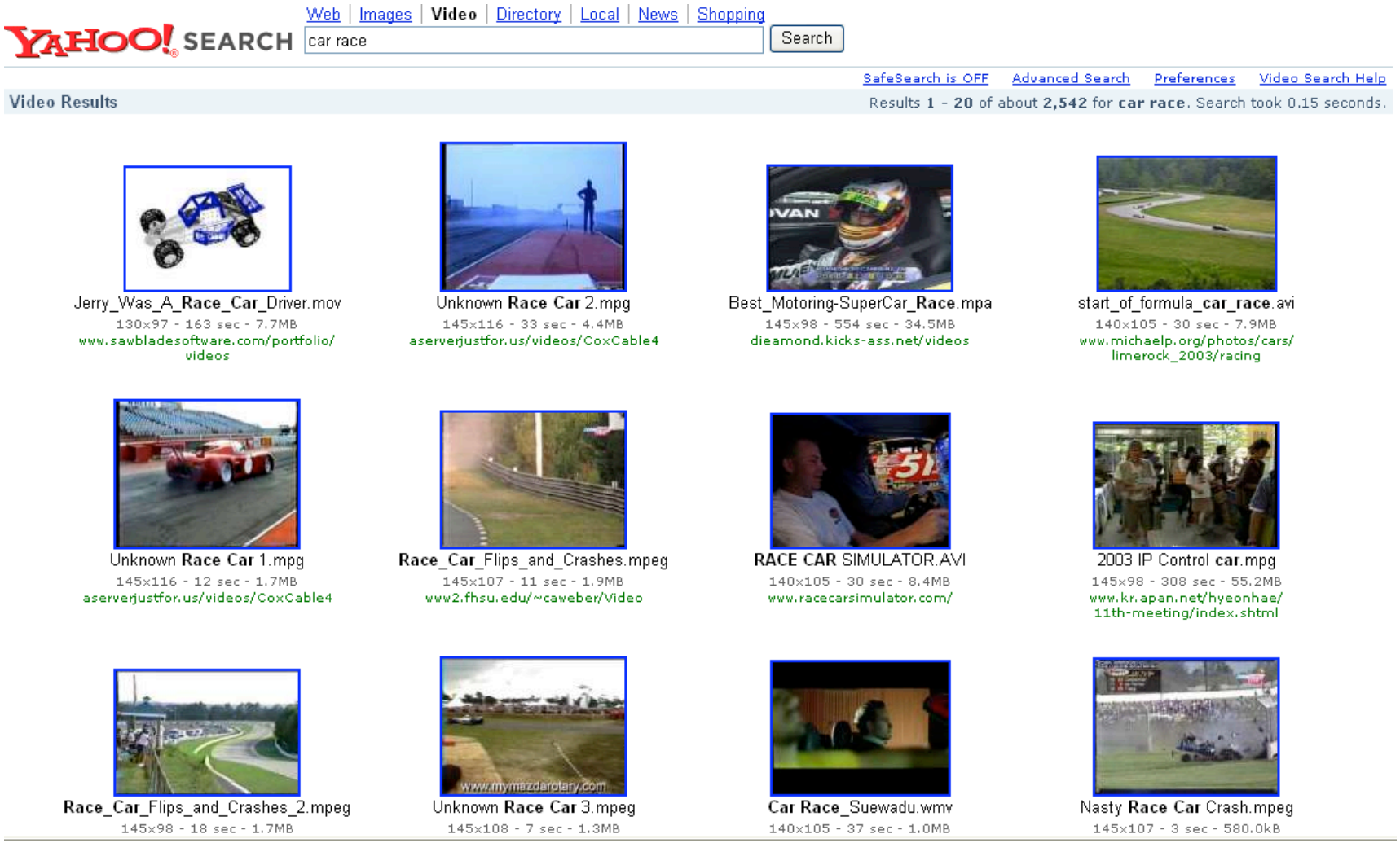
- Virage provide a suite of multimedia data management tools, e.g. VideoLogger:
  - Generates storyboards of video content
  - Extracts closed caption text
  - Can be expanded to:
    - Identify faces, voices, sounds.
    - Extract OCR (text on screen) and ASR (spoken) text.
  - Generates searchable archives of video content.



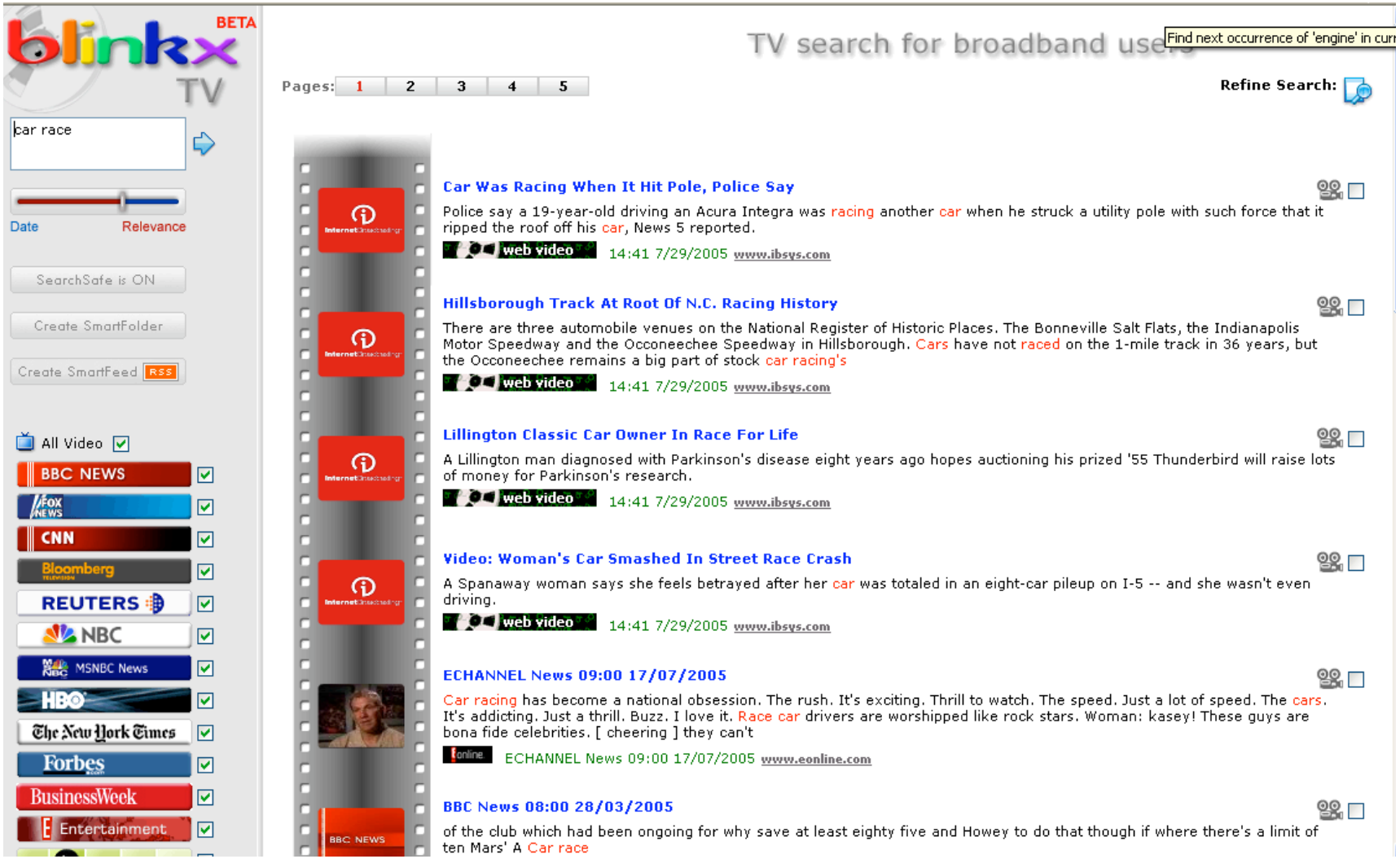
Example Web-based Systems



Example Web-based Systems



Example Web-based Systems



Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

Video Navigation Approaches

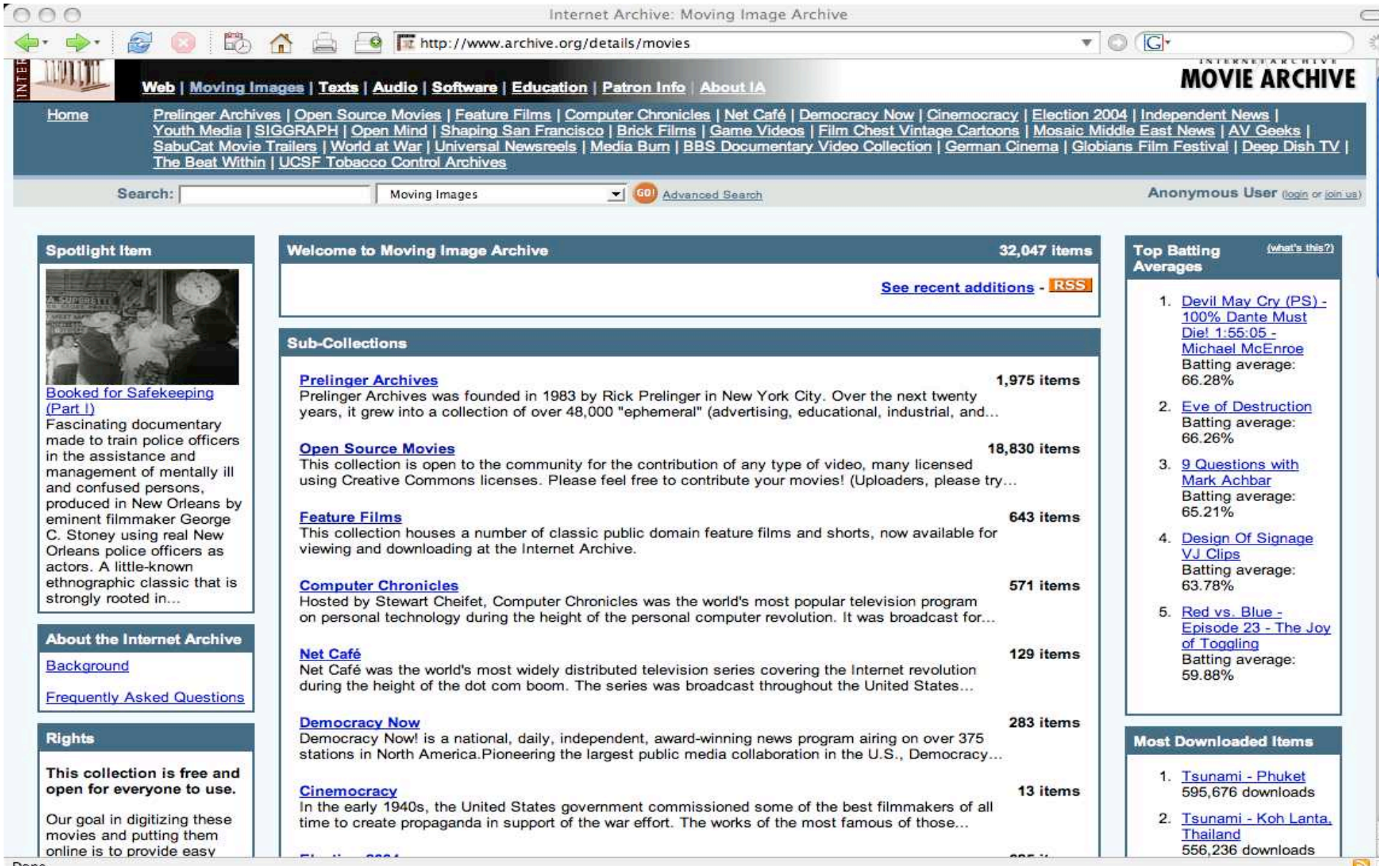
1. **Use metadata and browse keyframes**
2. Use text from speech - ASR/CC/in-video OCR
3. Match keyframes vs. query images
4. Use semantic video features
5. Use video/image objects as queries

1. Using Video Metadata

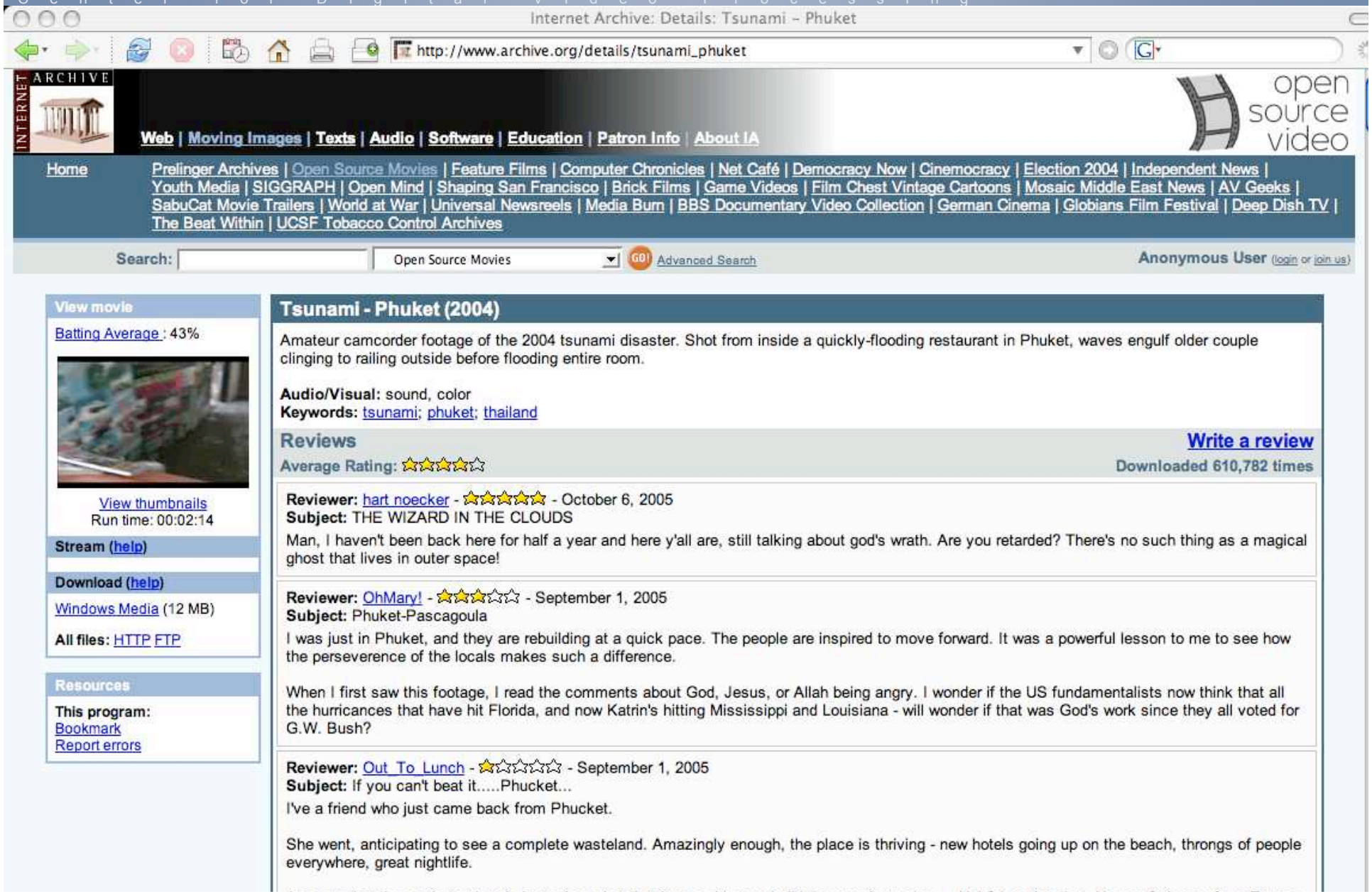
- Video metadata searching is a classic DBMS application;

- Metadata includes title, date, actor(s), producer(s), genre, running time, video format, reviews, ratings, ©, etc.
- Most coupled with keyframe / storyboard previews

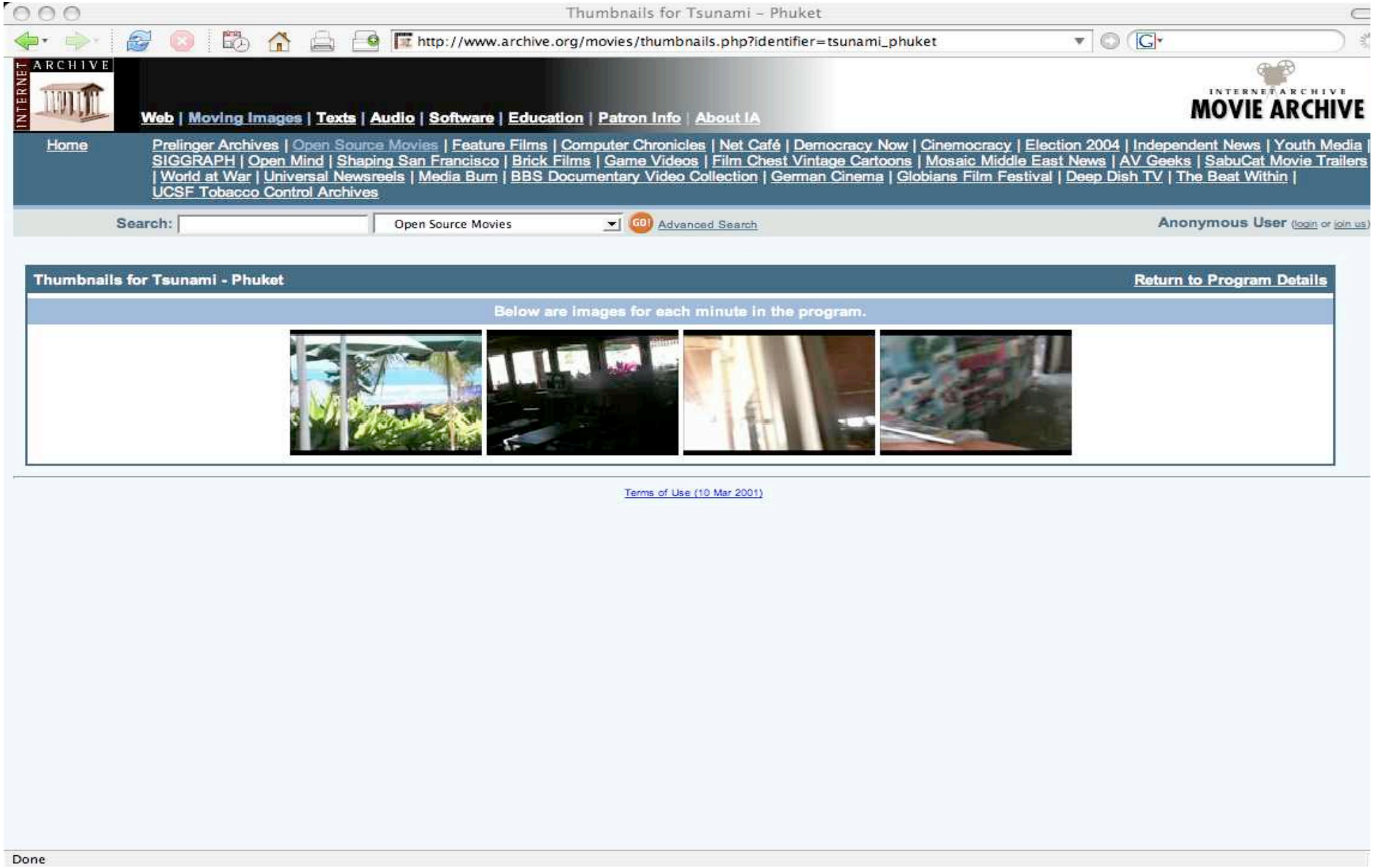
Internet Movie Archive



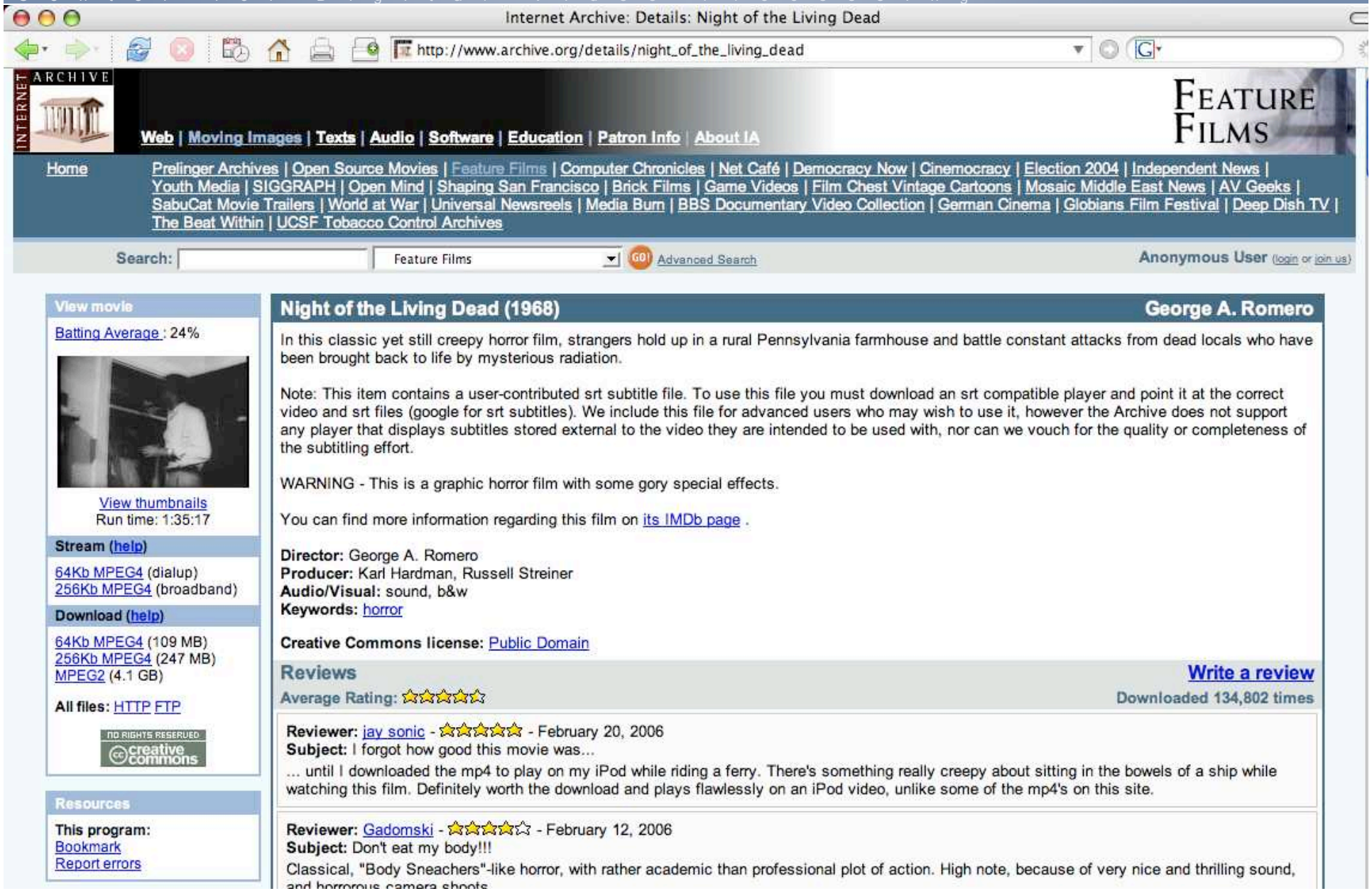
Internet Movie Archive



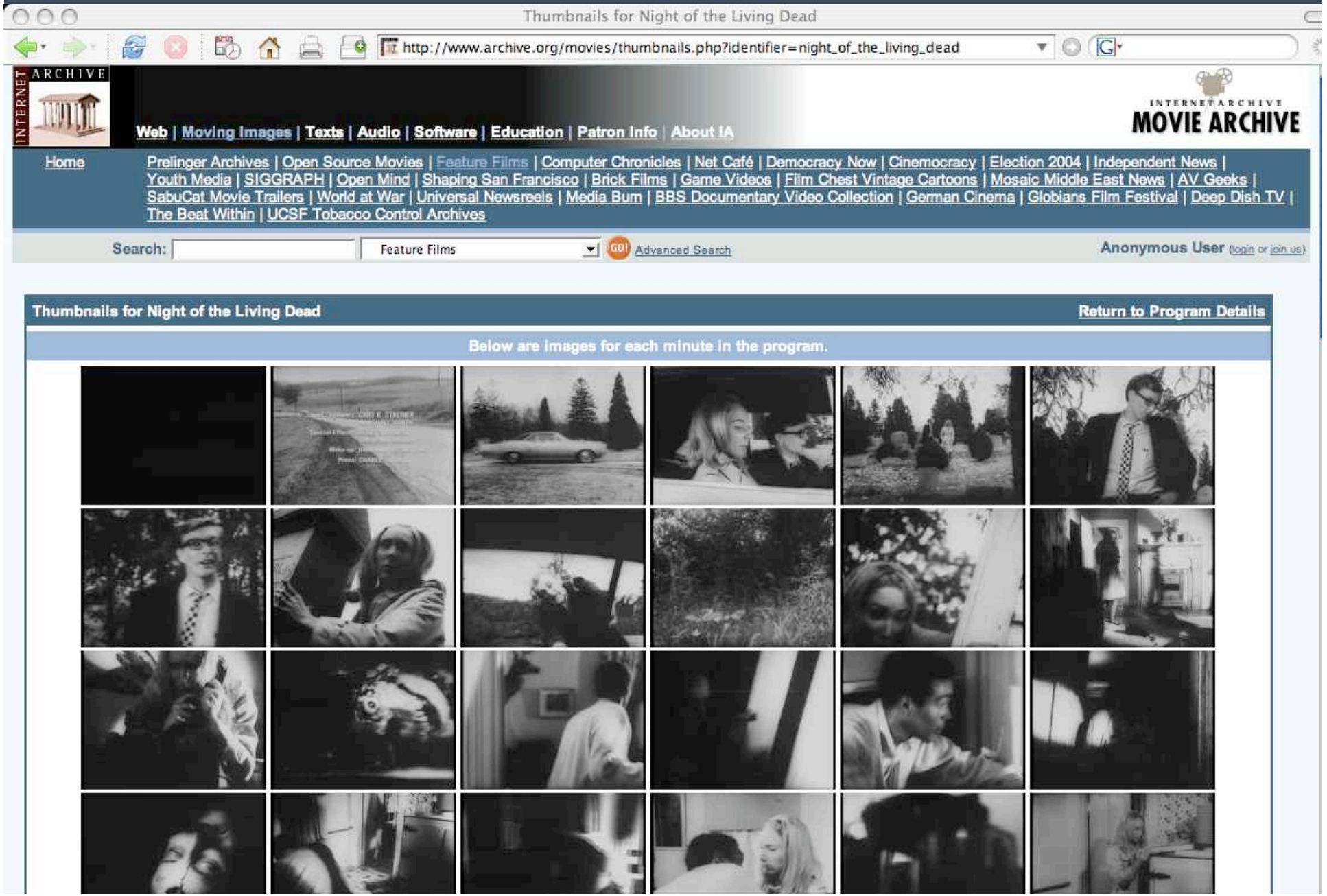
Internet Movie Archive



Internet Movie Archive



Internet Movie Archive



The Open Video Project

http://www.open-video.org/

THE OPEN VIDEO PROJECT
a shared digital video collection

- Home
- Contribute
- About

Search

Search

ex. "water" or "space shuttle"

Detailed Search

Browse

Genre

- Documentary [494]
- Educational [1072]
- Ephemeral [1139]
- Historical [187]
- Lecture [33]
- Other [5]
- Public Service [17]

Duration

- Less than 1 minute [259]
- 1 to 2 minutes [330]
- 2 to 5 minutes [669]
- 5 to 10 minutes [577]
- More than 10 minutes [1092]

Color

- In color [1908]
- In black & white [1039]

Sound

- With sound [2550]
- Silent [397]

Collections

- University of Maryland HCIL Open House Video Reports
- The Informedia Project at Carnegie Mellon University
- Internet Moving Images Archive
- 2001 TREC Video Retrieval Test Collection
- CHI Video Retrospective
- Digital Himalaya Project
- NASA K-16 Science Education Programs

Special Collection Spotlight

NASA K-16 Science Education Programs

NASA's Center for Distance Learning, at NASA Langley Research Center in Hampton, Virginia broadcasts five programs aimed at K-16 students, teachers, and parents.
This collection contains 580 videos.

other special collections...

Project News

- Teachers using video?
- CHI videos added
- Redesigned Project Web Site
- Understanding Video Symposium

more...

Featured Video

New

TIME: Three-dimensional Input, Modification and Evaluation

Other new videos

- ACM CHI 1984 Issue 12 - Selection-positioning Task Study
- ACM CHI 1985 Issue 18 - Designing a 'No Surprise Editor' prototype
- Tribute To World War II NISEI Veterans - Inouye

more...

Popular

Atom Bomb - Joe Bonica's Movie of the Month

Other popular videos

- Classic Television Commercials (Part II)
- NASA 25th Anniversary Show, Segment 01
- Bunjee

more...

Home • Search • Collections • Contribute • About • Project News

Done

The Open Video Project :: Video Results

http://www.open-video.org/results.php

THE OPEN VIDEO PROJECT
a shared digital video collection

- Home
- Contribute
- About

Modify Search

Search: All fields
for database
Genre: -- Any Genre --
Duration:
5 to 10 minutes
Format: MPEG-1
Color: Color B&W Either
Sound: Sound Silent Either
Search
Detailed Search

Page 1 **Search Results (5 videos found)**

Layout: Sort by: Relevance Results per page: 10

Visual information seeking using the FilmFinder (2000)
FilmFinder allows users to explore a large film database. By applying the dynamic queries approach to filtering information, a continuous starfield display of the films, and tight coupling among the...

Genre: Educational
Keywords: HCIL;
Duration: 00:06:12
Popularity (downloads): 697

Browsing Anatomical Image Databases: A Case Study of the Visible Human (1996)
Video demonstration from the 1996 CHI conference.

Genre: Educational
Keywords: Anatomy; Image; Database; Case; Visible; Human; CHI;
Duration: 00:06:10
Popularity (downloads): 899

Preview not available

Filter/Flow metaphors for Boolean queries (2000)
Evidence shows that users of database or information systems have difficulties specifying complex boolean queries. We present a novel visual presentation based on water filter-flow metaphors that reveals the effect...

Genre: Educational
Keywords: HCIL;
Duration: 00:06:36
Popularity (downloads): 305

ACM CHI 1994 Issue 97 - Visual Information seeking using the FilmFinder (1993)
The Filmfinder allow users to explore a large flm database. By applying the dynamic queries approach to filter information, a continuous starfield display of the films, and tight coupling...

Genre: Educational
Keywords: CHI; information retrieval; dynamic queries; video-on-demand;
Duration: 00:06:22
Popularity (downloads): 67

Snap together visualization (2000)
Information visualizations with multiple coordinated views enable users to rapidly explore complex data and discover

Done

The Open Video Project :: Video Details

http://www.open-video.org/details.php?videoid=708&surrogate=storyboard

THE OPEN VIDEO PROJECT
a shared digital video collection

- Home
- Contribute
- About

Search

ex. "water" or "space shuttle" Search

Detailed Search

Related Video

- **Video Grab Bag**

ACM CHI 1989 Issue 46 - Sun 386i on line help

Other random videos

- Easier Way, The
- Fluid Links for Informed and Incremental Hypertext Browsing
- New Indians, Segment 101

more...

- **Related keyword searches**
  - HCIL

Video Details

Visual information seeking using the FilmFinder

7-sec excerpt
Storyboard
FastForward

FilmFinder allows users to explore a large film database. By applying the dynamic queries approach to filtering information, a continuous starfield display of the films, and tight coupling among the components of the display, the FilmFinder environment encourages incremental and exploratory search.

Download: MPEG-1 • 61.00 MB

Video Information

Year:	2000
Genre:	Educational
Keywords:	HCIL;
Duration:	00:06:12
Color:	Yes
Sound:	Yes

Done

Metadata based video navigation is limited, but it is easy to do !

2. Using Text Dialogue

- Spoken dialogue **can** be of assistance in video navigation, and it may not be;
- Spoken dialogue can come from automatic speech recognition (ASR), or closed captions, or video OCR
- For certain genres of video, this can be sufficient

- Físchlar-News is a good example

Broadcast TV news

MPEG-1 encoding

An MPEG-1 encoded daily 9 o'clock news program (30 min)

Shot Boundary Detection

Shot segmented program

Advertisement Detection

Shot segmented, advert detected program

SVM (Support Vector Machine) with:

- Spoken diaogue indexing
- Speech vs. music discrimination
- Anchorperson detection
- Speaker segmentation & matching

Story segmented program

News story database

News story linkage analysis

Web application

User profile

Fischlár News: Browse - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Refresh Home Stop Address http://www.fischlar.dcu.ie/news/browse Links CDVP Google

Físchlár-News

Online Video Archive of Daily RTE1 9 o'clock News

Check the recommended news for you RECOMMENDED

SEARCH: GO

RTE News -- 16 SEPT 2003 20:56 (30min)

Following is the list of news stories in the selected news programme. You can browse each story in more detail or play the story. To help us recommend you the news stories you'd like, please indicate whether you like the news stories or not.

September 2003

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

August 2003

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

July 2003

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12

News Summary: **11** *stories*

News Story 1 (duration: 00:01:53)
800 new jobs have been promised by the online auctions website e-Bay which today confirmed it is to open a new facility at Blanchardstown in west Dublin... (PLAY THIS STORY)

News Story 2 (duration: 00:03:51)
Potential coalition partners in Government have disagreed tonight over the payment of benchmarking, under the current agreement, Sustaining Progress... (PLAY THIS STORY)

News Story 3 (duration: 00:01:44)
Health Min[illegible]r Micheal Martin has rejected criticisms that the Government is delaying its health refor[illegible] failing to appoint key independent personnel... (PLAY THIS STORY)

Browse this story

News Story 4 (duration: 00:01:51)
New figures show that patients in the the South East are waiting up to 7 years to see an Ear, Nose and Throat consultant and up to 5 years to see an Orthopaedic Surgeon... (PLAY THIS STORY)

News Story 5 (duration: 00:01:14)
Police in Sweden investigating the murder of the country's Foreign Minister, Anna Lindh, say they've identified and arrested her killer on Stockholm... (PLAY THIS STORY)

DIGITAL VIDEO Físchlár

EMAIL US WITH COMMENTS

MY ACCOUNT

http://www.fischlar.dcu.ie/news/browse/videoStory?content=17306&storyID=story17306_3

Trusted sites

Fischlár News: Browse - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Refresh Home Stop Address http://www.fischlar.dcu.ie/news/browse Links CDVP Google

Físchlár-News

Online Video Archive of Daily RTE1 9 o'clock News

Check the recommended news for you RECOMMENDED

RTE News -- 16 SEPT 2003 20:56 (30min)

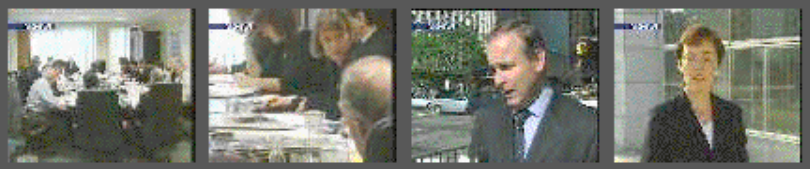
SEARCH: GO

News Story 3 (duration: 00:01:44)

STORY LIST

Health Minister Micheal Martin has rejected criticisms that the Government is delaying its health reform programme by failing to appoint key independent personnel...

But the next general election could be three years away, - Benchmarking could by then be a forgotten political issue. - Health Minister Micheal Martin has rejected criticisms - that the Government is delaying its health reform programme - by failing to appoint key independent personnel. - Earlier Professor Niamh Brennan, - the author of one of three reports shaping the changes, said she was alarmed that the staff members aren't yet in place. - Three months ago the Government launched its health reform programme It promised to abolish health boards and end fragmentation in decision-making. And greater financial controls. - Micheal Martin said the reforms would begin as quickly as possible. - The Minister for Finance said hospital reform - Chapter 11 of the report said - reforms should begin as soon as possible, - and progress reports should be every three months, - and an independent person employed. - Niamh Brennan said she was concerned about the lack of progress.



And concerned about momentum being lost. The Minister has rejected criticism of the delays. - I was surprised when I heard her comments - as it was always understood that a three-month consultation phase - would be implemented to help to clear staff. - Then we'd move on to the interim HSE board. The Brennan Commission believe that health reform can take two years. - The Department of Health believes it will take 10 years. - There's little indication of urgency. - But it's not just Ms Brennan who - believes that until an independent commission is involved, - there will be little progress.

September 2003

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

August 2003

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

July 2003

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12

DIGITAL VIDEO Físchlár

EMAIL US WITH COMMENTS

MY ACCOUNT

Trusted sites

Fischlár News: Browse - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Refresh Home Stop Address http://www.fischlar.dcu.ie/news/browse Links CDVP Google

Fischlár-News

Online Video Archive of Daily RTE1 9 o'clock News

Check the recommended news for you RECOMMENDED

SEARCH: Bremer GO

Query term(s): Bremer
Search found: **5 news stories**

Search found following stories which contain your query term(s). Browse and play the stories or look at other stories on the day following stories belong. To receive recommended stories, please indicate whether you like the news stories or not, using thumbs icons.

September 2003	S	M	T	W	T	F	S
		1	2	3	4	5	6
	7	8	9	10	11	12	13
	14	15	16	17	18	19	20
	21	22	23	24	25	26	27
	28	29	30				

August 2003	S	M	T	W	T	F	S
						1	2
	3	4	5	6	7	8	9
	10	11	12	13	14	15	16
	17	18	19	20	21	22	23
	24	25	26	27	28	29	30
	31						

July 2003	S	M	T	W	T	F	S
			1	2	3	4	5
	6	7	8	9	10	11	12

Search Result: 5 stories

News Story (duration: 00:02:24) (13 Jul 2003)
The first Iraqi council meets since Saddam's overthrow ... (PLAY THIS STORY)

News Story (duration: 00:02:01) (22 May 2003)
UN Security Council has voted by 14:0 to lift sanctions on Iraq and to recognise United States and Britain as... (PLAY THIS STORY)

News Story (duration: 00:01:53) (5 Jul 2003)
The chief US administrator in Iraq, Paul Bremer, has condemned an attack in the west of the country which killed seven recruits to a US-backed police force... (PLAY THIS STORY)

News Story (duration: 00:02:22) (23 Aug 2003)
At least nine people have been killed and more than 10 wounded in clashes between Kurds and Turkmen in Northern Iraq... (PLAY THIS STORY)

News Story (duration: 00:02:10) (17 Oct 2003)
The US House of Representatives has approved President Bush's request for $87bn in new spending for military operations and reconstruction efforts... (PLAY THIS STORY)

DIGITAL VIDEO Fischlár EMAIL US WITH COMMENTS

MY ACCOUNT

Done Trusted sites

Fischlár News: Browse - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Refresh Home Stop Address http://www.fischlar.dcu.ie/news/browse Links CDVP Google

Físchlár-News

Online Video Archive of Daily RTE1 9 o'clock News

Check the recommended news for you RECOMMENDED

SEARCH: Bremer GO

RTE News -- 23 AUG 2003 20:56 (30min)

I think those yearning for the return of Ba'athism will be disappointed. But Lebanese television showed a tape of a new militant group,

September 2003	S	M	T	W	T	F	S
		1	2	3	4	5	6
	7	8	9	10	11	12	13
	14	15	16	17	18	19	20
	21	22	23	24	25	26	27
	28	29	30				

August 2003	S	M	T	W	T	F	S
						1	2
	3	4	5	6	7	8	9
	10	11	12	13	14	15	16
	17	18	19	20	21	22	23
	24	25	26	27	28	29	30
	31						

July 2003	S	M	T	W	T	F	S
			1	2	3	4	5
	6	7	8	9	10	11	12
	13	14	15	16	17	18	19

Related Stories

News Story (duration: 00:04:06) (19 Aug 2003)
A top United Nations envoy to Iraq, Sergio Vieira de Mello, has been killed in a powerful truck bomb explosion at the UN headquarters in Baghdad. At least 14 others died in the blast and scores more have been injured... (PLAY THIS STORY)

News Story (duration: 00:01:55) (28 Aug 2003)
The United States has indicated that it may accept a UN-mandated international peacekeeping force, provided it is led by an American... (PLAY THIS STORY)

News Story (duration: 00:01:53) (23 May 2003)
UN secretary general Kofi Annan has named the human rights commissioner Sergio Vieira de Mello as his special representative for Iraq.... (PLAY THIS STORY)

News Story (duration: 00:01:53) (5 Jul 2003)
The chief US administrator in Iraq, Paul Bremer, has condemned an attack in the west of the country which killed seven recruits to a US-backed police force... (PLAY THIS STORY)

DIGITAL VIDEO Físchlár

EMAIL US WITH COMMENTS

MY ACCOUNT

Trusted sites

Fischlár News: Browse - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Refresh Home Stop Address http://www.fischlar.dcu.ie/news/browse Links CDVP Google

Físchlár-News

Online Video Archive of Daily RTE1 9 o'clock News

Check the recommended news for you RECOMMENDED

SEARCH: Bremer GO

Today : **5 November 2003**

You last viewed : **4 November 2003**

Following is the list of recommendable stories for you. **NEW** recommendations are stories you'll be interested in among newly appeared news since your last visit. **MISSED** recommendations are news stories you should have seen last time. You can change the thumbs icons below, which will further change the recommendations for you.

September 2003

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

August 2003

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

July 2003

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19

Recommended Stories

News Story (duration: 00:02:07) RECOMMENDED NEW (5 Nov 2003)
Gardai say the partially clothed body of a man found dead in a disused garage in Dublin yesterday afternoon was there for up to 18 hours... (PLAY THIS STORY)

News Story (duration: 00:01:57) RECOMMENDED MISSED (16 Jul 2003)
This afternoon the Israelis released Sean O'Muiireagain in a case of mistaken identity...
(PLAY THIS STORY)

News Story (duration: 00:01:33) RECOMMENDED MISSED (18 Jul 2003)
The period of detention of a man arrested yesterday as part of a cocaine seizure in Kilkenny city has been extended by 72 hours... (PLAY THIS STORY)

News Story (duration: 00:01:59) RECOMMENDED MISSED (18 Jul 2003)
The Planning Tribunal has heard that a former business associate of Liam Lawlor was blackmailed into paying hundreds of thousands of pounds by Liam Lawlor... (PLAY THIS STORY)

News Story (duration: 00:03:25) RECOMMENDED MISSED (22 Jul 2003)
The Planning Tribunal has heard that former Fianna Fail TD Liam Lawlor stands to make 2.5m euro if a development of 10,000 houses goes ahead at Adamstown, West Dublin ...

DIGITAL VIDEO Físchlár

EMAIL US WITH COMMENTS

MY ACCOUNT

Trusted sites

3. Keyframe Matching

- ASR based video retrieval is not visual;
- Keyframes extracted as shot representatives can be used for retrieval;
- Requires user to locate images/other keyframes, from browsing, or other search;

- Físchlár-TRECVid2004 is a good example
- The unit of retrieval is a shot

QUERY PANEL

Enter term(s) and click SEARCH button. Added shots will be used together for searching.
[**clear text**] [**clear all**]

rocket launch | SEARCH

GLOBAL COLOUR | LOCAL COLOUR
EDGE | MOTION
TEXTURE | FACE FILTER

had failed repeatedly and publicly and there're tests...

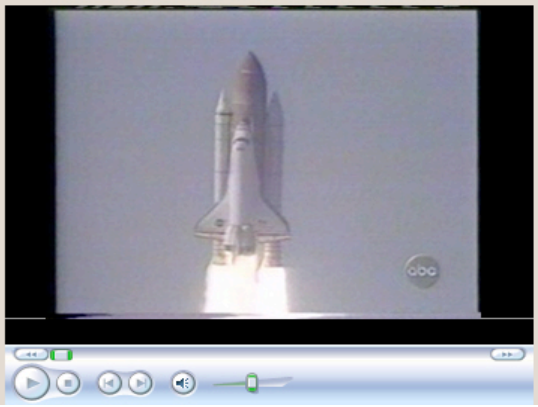
REMOVE

powerful rockets that in nineteen sixty one launched during...

REMOVE

are shuttle launch can be built in wrestling but...

REMOVE



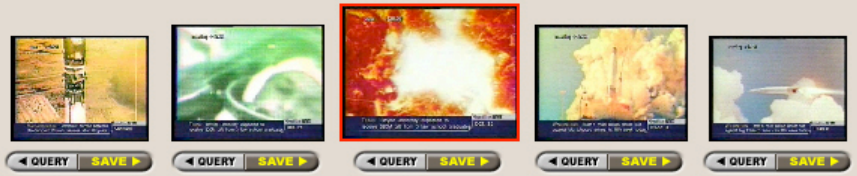
SEARCH RESULT

Search found **169** matching shots. Following is the **ranked list** of the search result. Click an image to play the segment, click SAVE button to add to your answers, or click QUERY button to improve subsequent query.

RESULT PAGE: 1 2 3 4 5

1: CNN News (26 Oct 1998)

> MORE MATCHES IN THIS BROADCAST


QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE

powerful rockets that in nineteen sixty one launched during the government the first person in space this all u. s. **launch** rockets had failed repeatedly and publicly and there're tests pressure was building devoted american investments

2: ABC News (30 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE

that manages about we've got big he'd be in on board discovery yesterday to **launch** would have looked like this when is in the middle of this picture are shuttle **launch** can be built in wrestling but it was far gentler than lives at least one we've that the astronauts back in space again that was quite different than a ride i got before of course back a long time ago

3: ABC News (26 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE

on wasn't whether one shot mission will been a permanent presence on mars because a mars vehicle would be too heavy to **launch** from earth it'd be assembled piece by piece in earth orbit half a dozen astronauts would then take six months to **rocket** the thirty four million miles to mars they'd spend the next year and a half exploring and conducting experiments that had all mission time two and a half years mas is really going to be the place where humans will

TASK 1

SEE DESCRIPTION

FINISH THIS TASK

SAVED SHOTS

Your answer to the task will be listed below. Add more, delete, or use them to help your subsequent query.

QUERY PANEL

Enter term(s) and click SEARCH button. Added shots will be used together for searching.
[**clear text**] [**clear all**]

rocket launch | SEARCH

GLOBAL COLOUR | LOCAL COLOUR
EDGE | MOTION
TEXTURE | FACE FILTER

had failed repeatedly and publicly and there're tests...

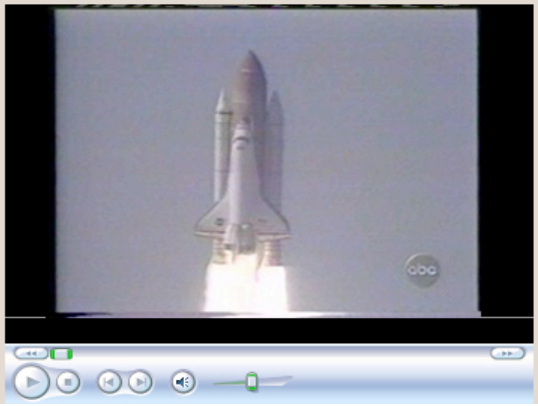
REMOVE

powerful rockets that in nineteen sixty one launched during...

REMOVE

are shuttle launch can be built in wrestling but...

REMOVE



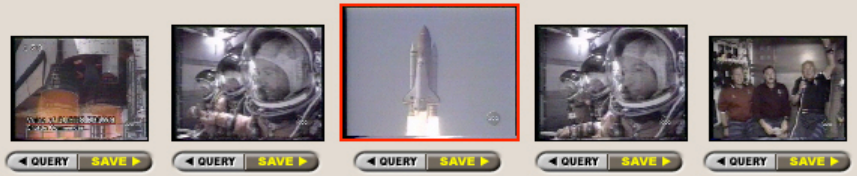
SEARCH RESULT

Search found **250** matching shots. Following is the **ranked list** of the search result. Click an image to play the segment, click SAVE button to add to your answers, or click QUERY button to improve subsequent query.

RESULT PAGE: 1 2 3 4 5

1: ABC News (30 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

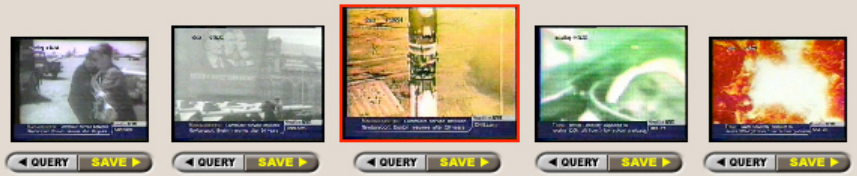


QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE

that manages about we've got big he'd be in on board discovery yesterday to **launch** would have looked like this when is in the middle of this picture are shuttle **launch** can be built in wrestling but it was far gentler than lives at least one we've that the astronauts back in space again that was quite different than a ride i got before of course back a long time ago

2: CNN News (26 Oct 1998)

> MORE MATCHES IN THIS BROADCAST



QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE

those were the coldest years of the cold war the soviet military arsenal included powerful rockets that in nineteen sixty one launched during the government the first person in space this all u. s. **launch** rockets

3: CNN News (26 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE | QUERY SAVE

the government the first person in space this all u. s. **launch** rockets had failed repeatedly and publicly and there're tests pressure was building devoted american investments these are the past there have been

4: CNN News (26 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

TASK 1

SEE DESCRIPTION

FINISH THIS TASK

SAVED SHOTS

Your answer to the task will be listed below. Add more, delete, or use them to help your subsequent query.

QUERY PANEL

Enter term(s) and click SEARCH button. Added shots will be used together for searching.
[**clear text**] [**clear all**]

rocket launch | SEARCH

- [x] GLOBAL COLOUR
- [ ] LOCAL COLOUR
- [x] EDGE
- [x] MOTION
- [ ] TEXTURE
- [ ] FACE FILTER

are shuttle launch can be built in wrestling but...

REMOVE ✕

private companies are racing to design of reusable rocket book...

REMOVE ✕

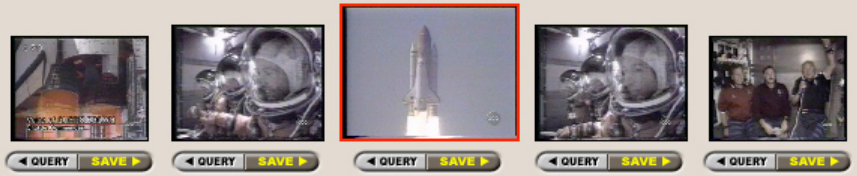
SEARCH RESULT

Search found **250** matching shots. Following is the **ranked list** of the search result. Click an image to play the segment, click SAVE button to add to your answers, or click QUERY button to improve subsequent query.

RESULT PAGE: 1 2 3 4 5

1: ABC News (30 Oct 1998)

> MORE MATCHES IN THIS BROADCAST



that manages about we've got big he'd be in on board discovery yesterday to **launch** would have looked like this when is in the middle of this picture are shuttle **launch** can be built in wrestling but it was far gentler than lives at least one we've that the astronauts back in space again that was quite different than a ride i got before of course back a long time ago

2: ABC News (27 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

everyone is a one shot one each problem bruno with its than once was before with the equivalent of a jetliner four so private companies are racing to design of reusable **rocket** book a very deep conrad commanded the apollo twelve mission to the moon he says the private sector can do it much cheaper

3: ABC News (27 Oct 1998)

> MORE MATCHES IN THIS BROADCAST

o'neal nypd blue a. b. c. tonight your discretion advised finally from us this evening the price tag one of the reasons nasa is so focus on public relations this week is plain and simple money there just isn't the nuf money government money to do everything the space agency would like solicit possible that the answer may be less government and more private enterprise is a. b. c.'s jane clayson maybe it can start space travel is expensive that the except for the space shuttle everyone is a one shot one each problem bruno with its than once was before with the equivalent

TASK 1

SEE DESCRIPTION

FINISH THIS TASK

SAVED SHOTS

You saved **5** shots, shown below - these are your **answers** to the current task. Add more, delete, or use them for subsequent query.

QUERY DELETE ✕

QUERY DELETE ✕

QUERY DELETE ✕

QUERY

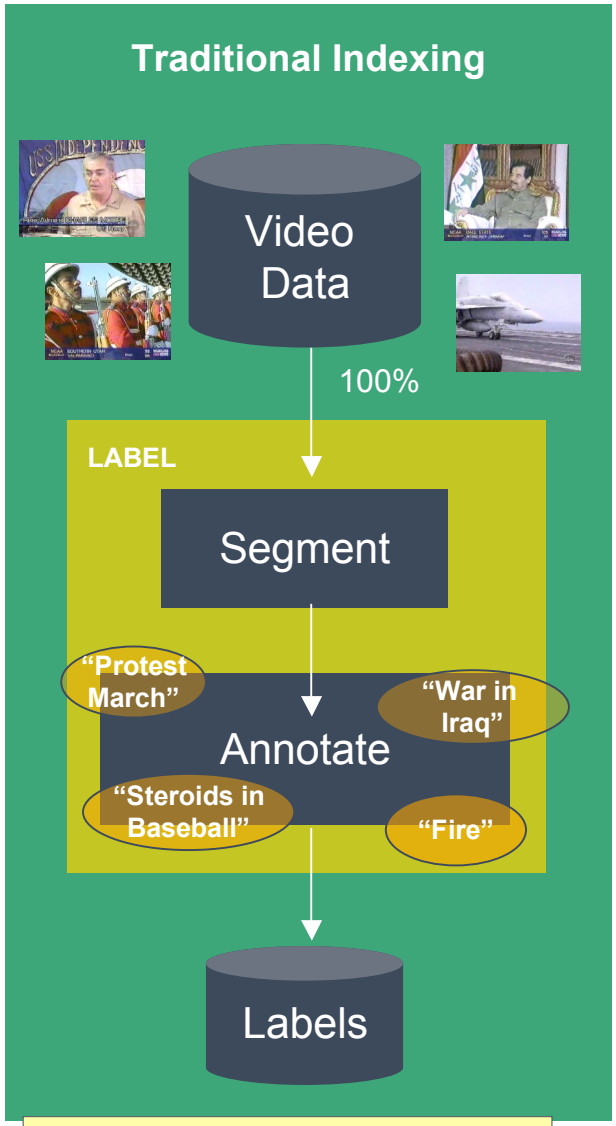
DELETE ✕

QUERY DELETE ✕

4. Semantic features for retrieval

- Based upon pre-processing video or keyframes to detect features;
- Features normally from an ontology and thus related;

Indexing by Semantic Features



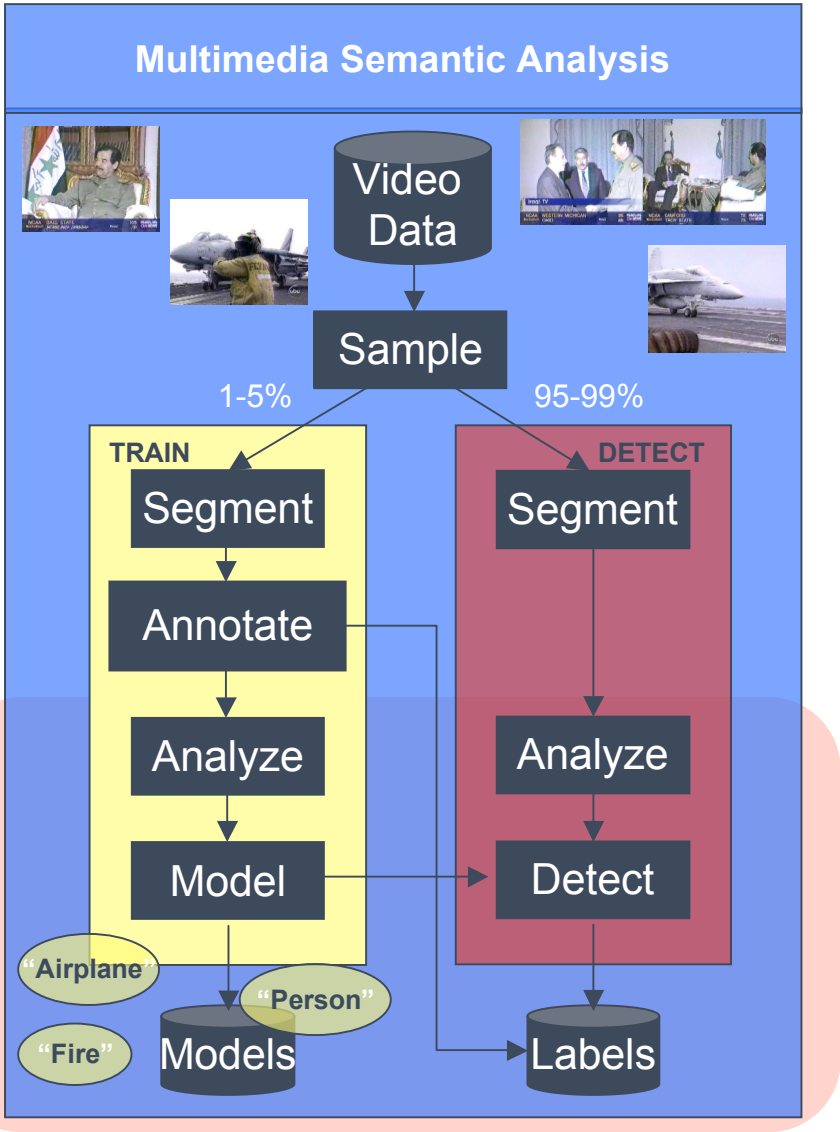
Thanks to John Smith, IBM, for slide

PROBLEM:

- Manual labeling is human resource intensive (10x)
- Results in incomplete & inconsistent annotations

SOLUTION:

- Replace 100% manual process with learning approach
- Annotate 1-5% of training data
- Build models from training data
- Adapt to annotation errors
- Propagate labels to video repository
- Assign confidence scores



Feature-based retrieval

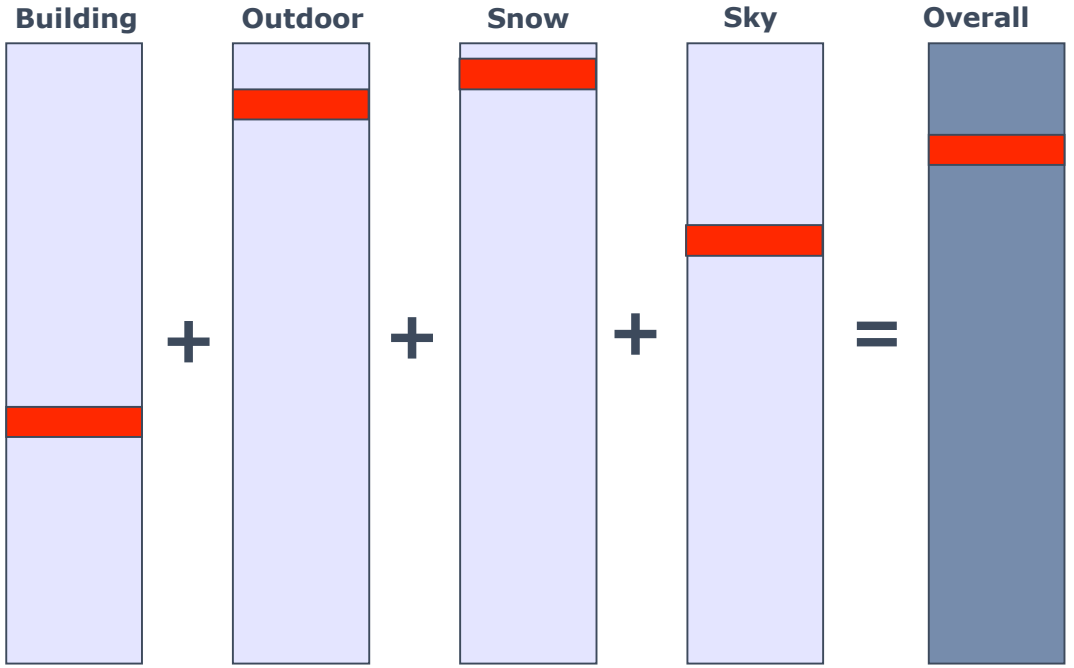
- To date, features in retrieval are independent;
- We must develop feature-based retrieval that uses features in combination;

0.2	Indoor
0.8	Outdoor
0.7	CityScape
0.3	Landscape
0.1	People
0.0	Face
0.8	Sky
0.2	Vegetation
0.7	Building

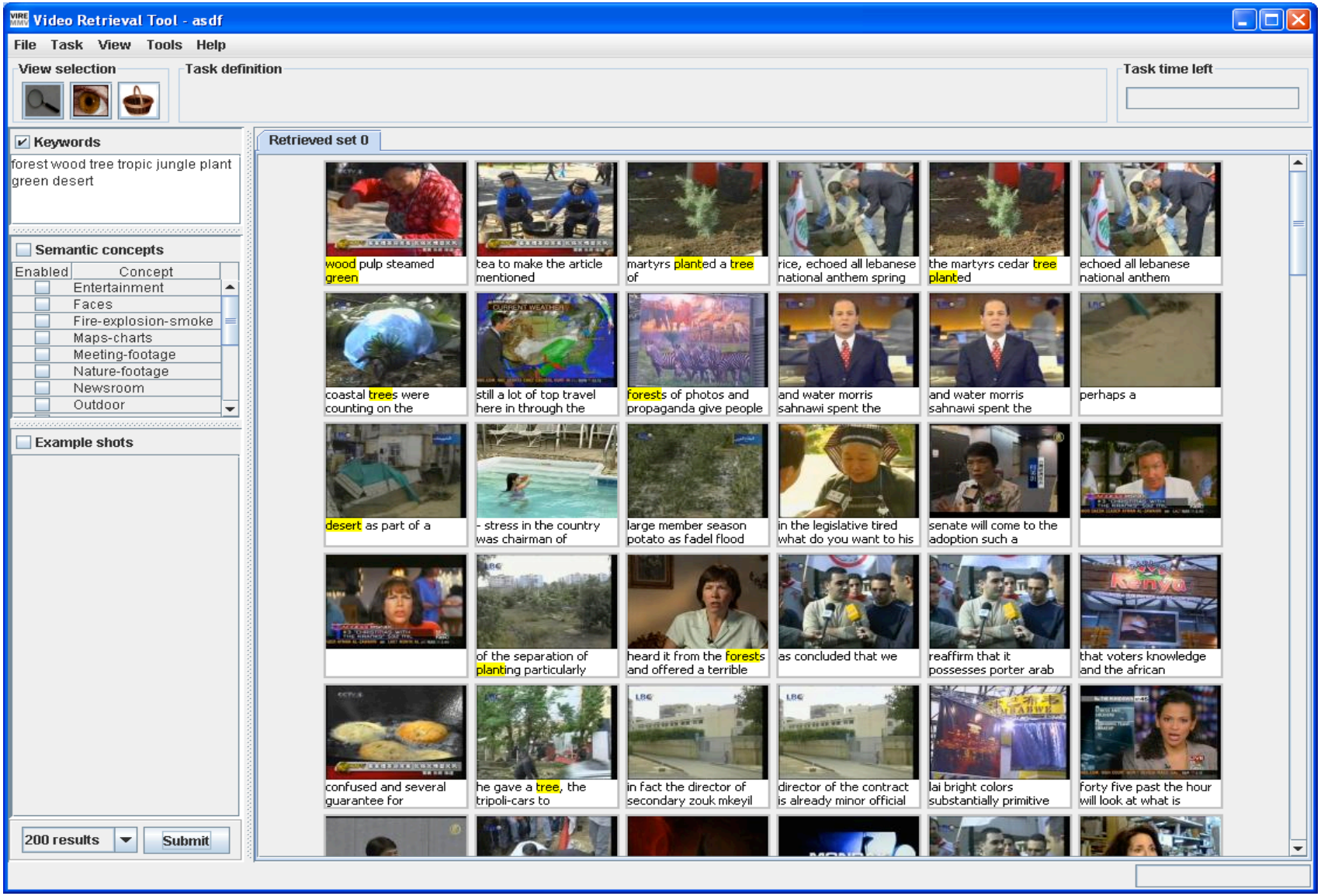
Feature-based retrieval

- Utilise the dependencies among and between features, from the ontology, into an **overall retrieval** combining individual evidences for each feature;

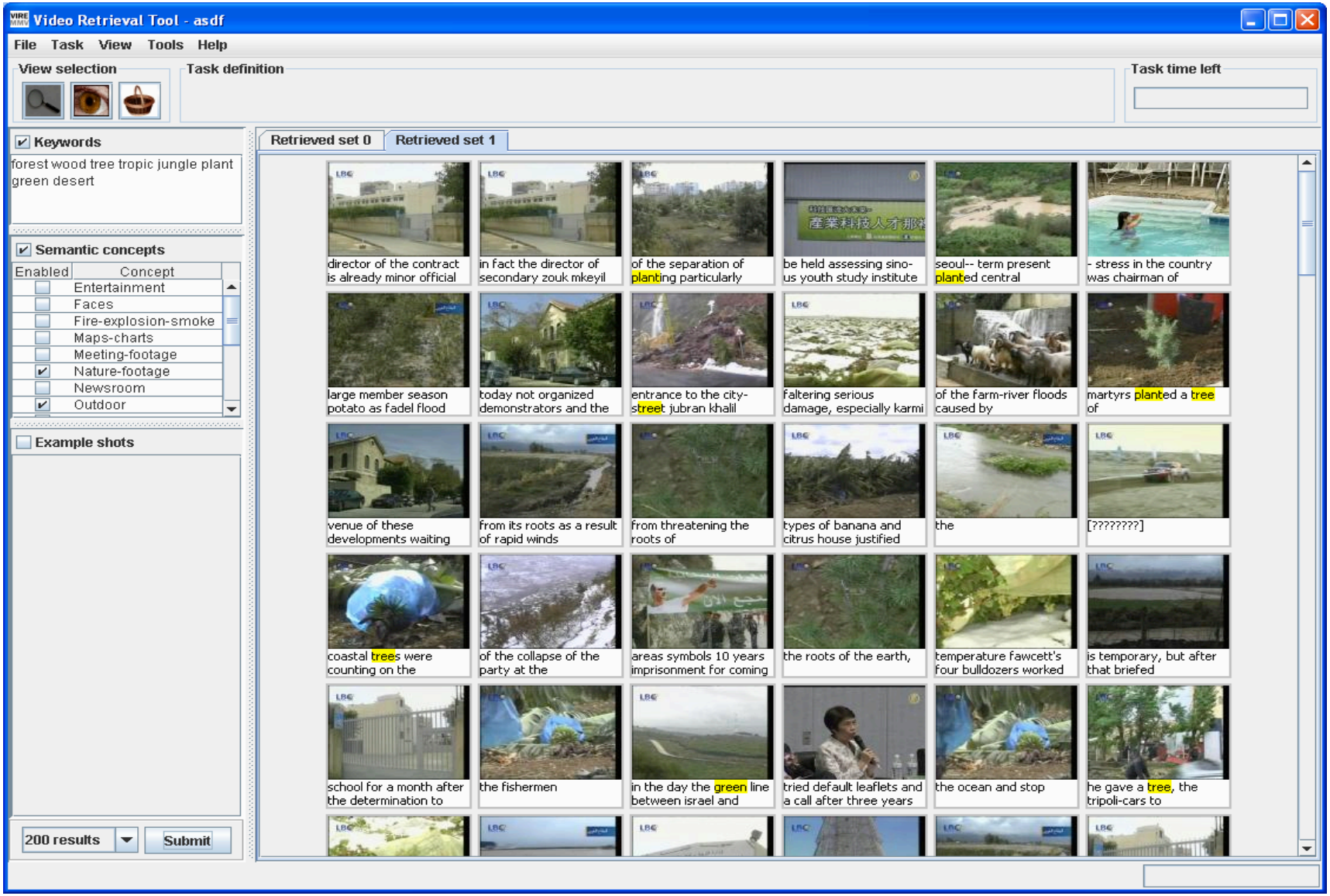
```
Building :      0.4
Outdoor :       0.8
Snow :          0.9
Sky :           0.6
```



TRECVid: Oulu



Query tool : text + concept search



Centre for Digital Video Processing

TRECVid Columbia U

C e n t e r f o r D i g i t a l V i d e o P r o c e s s i n g

CMU: Interface to Support Rich Visual Sets.

5. Object-based video retrieval

- For certain types of user query, with patience, and time, object-based retrieval can help;
- Another Físchlár system shows how it can be done

Automatic Understanding of Objects

- Take a simple object, a car
  - 4 wheels, fairly recognisable shape
  - Should not pose problems?

  - Well, it is not quite that easy…

Automatic Understanding of Objects

Automatic Understanding of Objects

QUERY PANEL

The images and objects below will be used for searching. Refresh the query panel or add/remove items below and click SEARCH button. **[ADD EXTERNAL IMG/OBJ]**

You can search objects when there is icon beside an image.

SEARCH

REMOVE

- [x] COLOUR
- [x] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

SEARCH RESULT

Following is the ranked list of the search result. An image with icon indicates it contains an object within it. Click on **QUERY** button to use an image/object as query, click on **SAVE** button to save this point in bookmark.

RESULT PAGE 1 2 3 4 5

QUERY SAVE

QUERY SAVE

QUERY SAVE

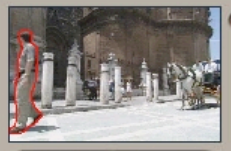
QUERY SAVE

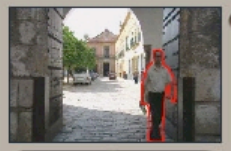
QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE


QUERY SAVE


QUERY SAVE

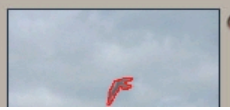
QUERY SAVE

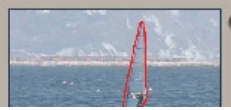
QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE





SAVED POINT

The saved points in the rush footage are displayed below.

QUERY PANEL

The images and objects below will be used for searching. Refresh the query panel or add/remove items below and click SEARCH button. **[ADD EXTERNAL IMG/OBJ]**

You can search objects when there is icon beside an image.

SEARCH

REMOVE

- [x] COLOUR
- [x] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

SEARCH RESULT

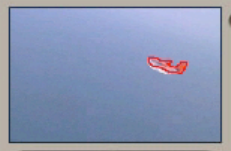
Following is the ranked list of the search result. An image with icon indicates it contains an object within it. Click on **QUERY** button to use an image/object as query, click on **SAVE** button to save this point in bookmark.

RESULT PAGE 1 2 3 4 5

QUERY SAVE

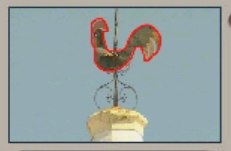
QUERY SAVE

QUERY SAVE



QUERY SAVE

QUERY SAVE



QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

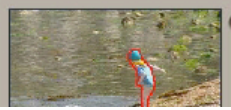
QUERY SAVE

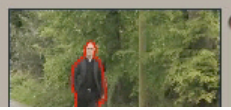
QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE





SAVED POINT

The saved points in the rush footage are displayed below.

REMOVE

REMOVE

REMOVE

QUERY PANEL

The images and objects below will be used for searching. Refresh the query panel or add/remove items below and click SEARCH button. **[ADD EXTERNAL IMG/OBJ]**

You can search objects when there is icon beside an image.

SEARCH

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

REMOVE

- [x] COLOUR
- [ ] TEXTURE
- [x] SHAPE
- [ ] COLOUR
- [ ] TEXTURE

SEARCH RESULT

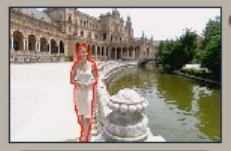
Following is the ranked list of the search result. An image with icon indicates it contains an object within it. Click on **QUERY** button to use an image/object as query, click on **SAVE** button to save this point in bookmark.

RESULT PAGE 1 2 3 4 5

QUERY SAVE

QUERY SAVE

QUERY SAVE



QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

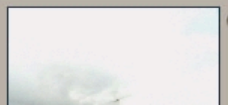
QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE

QUERY SAVE



SAVED POINT

The saved points in the rush footage are displayed below.

REMOVE

REMOVE

REMOVE

REMOVE

REMOVE

Video IR ?

- A variety of very different approaches … 5 of them in my categorisation;
- Each works well for certain **type** of query;
- Best search tool is a combination of them all
  - Let user choose the relative weighting;
  - Learn relative weightings based on query *types*;

Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. **MMIR Evaluation**
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

TREC - the original of the species !

- IR has always been empirically-based in terms of self-evaluation;
- TREC began in 1991 as a reaction to small collection sizes & need for coordinated evaluation;
- Run by NIST and funded by DoD/ARDA;
- Set out initially to benchmark the search & retrieval operation on text;
- Approach is always metrics-based - focus on evaluation of search performance
- Measurement is Precision & Recall;

TREC Operation

- Operation of TREC was established from the start:
  - Acquire data & distribute to participants;
  - Formulate search topics and release to participants en bloc;
  - After about 4 weeks accept submissions of top-1000 ranked documents per search topic;
  - Pool submissions - top100 - and use intelligence analysts to make binary relevance judgments;
  - Calculate P&R and other derived measures and distribute results;
  - Host workshop in NIST in November;
  - Make plans, and repeat annually … for the next 16 years !

TREC Tracks

- After initial search task there was strong interest in diversification;
- This led to the development of “tracks”;
  - See next slides

- Bonus is that TREC test collections are large enough so that they realistically model operational settings.
- Most of today's commercial search engines include technology first developed in TREC.

2006 TREC Tracks

- **Blog Track**

  A new track in 2006 to explore information seeking behavior in the blogosphere.
- **Enterprise Track**

  A new track in 2005 to study enterprise search: satisfying a user who is searching the data of an organization to complete some task.
- **Genomics Track**

  To study retrieval tasks in the domain of genomics data (broadly construed to include not just gene sequences but also supporting documentation such as research papers, lab reports, etc.) The genomics track first ran in 2003
- **Legal Track**

  A new track in 2006 to develop search technology that meets the needs of lawyers to engage in effective discovery in digital document collections
- **Question Answering Track**

  To take a step closer to information retrieval rather than document retrieval
- **SPAM Track**

  New in 2005 to provide a standard evaluation of current and proposed spam filtering approaches, thereby laying the foundation for the evaluation of more general email filtering and retrieval tasks.
- **Terabyte Track**

  First run in 2004 to investigate whether/how the IR community can scale traditional IR test-collection-based evaluation to significantly larger document collections

Past TREC Tracks

- **Cross-Language Track**
  Investigates the ability of retrieval systems to find documents that pertain to a topic regardless of the language in which the document is written.
- **Cross-Language Track**
  CL was run in 2002, now studied in both CLEF and the NTCIR workshops.
- **Filtering Track**
  User's information need is stable (and some relevant documents are known) but there is a stream of new documents and the system must make a binary decision as to whether the document should be retrieved
- **HARD**
  To achieve High Accuracy Retrieval from Documents by leveraging additional information about the searcher and/or the search context, through techniques such as passage retrieval and using very targeted interaction with the searcher.
- **Interactive Track**
  Studies user interaction with text retrieval systems, carry out studies with real users using a common collection and set of user queries.
- **Novelty Track**
  Investigate systems' abilities to locate new (i.e., non-redundant) information.
- **Robust Retrieval Track**
  Traditional ad hoc retrieval but with the focus on individual topic effectiveness rather than average effectiveness
- **Video Track**
  Video track devoted to research in retrieval of digital video.
- **Web Track**
  Featuring search on a document set that is a snapshot of the World Wide Web.

Number of participants per track and total number of distinct participants in each TREC

	TREC											
Track	92	93	94	95	96	97	98	99	00	01	02	03
Ad hoc	18	24	26	23	28	31	42	41	—	—	—	—
Routing	16	25	25	15	16	21	—	—	—	—	—	—
Interactive	—	—	3	11	2	9	8	7	6	6	6	—
Spanish	—	—	4	10	7	—	—	—	—	—	—	—
Confusion	—	—	—	4	5	—	—	—	—	—	—	—
Database merging	—	—	—	3	3	—	—	—	—	—	—	—
filtering	—	—	—	4	7	10	12	14	15	19	21	—
Chinese	—	—	—	—	9	12	—	—	—	—	—	—
NLP	—	—	—	—	4	2	—	—	—	—	—	—
Speech	—	—	—	—	—	13	10	10	3	—	—	—
Cross-language	—	—	—	—	—	13	9	13	16	10	9	—
High precision	—	—	—	—	—	5	4	—	—	—	—	—
Very large corpus	—	—	—	—	—	—	7	6	—	—	—	—
Query	—	—	—	—	—	—	2	5	6	—	—	—
Question answering	—	—	—	—	—	—	—	20	28	36	34	33
Web	—	—	—	—	—	—	—	17	23	30	23	27
Video	—	—	—	—	—	—	—	—	—	12	19	—
Novelty	—	—	—	—	—	—	—	—	—	—	13	14
Genome	—	—	—	—	—	—	—	—	—	—	—	29
HARD	—	—	—	—	—	—	—	—	—	—	—	14
Robust	—	—	—	—	—	—	—	—	—	—	—	16
Total participants	25	31	33	36	38	51	56	66	69	87	93	93

(Taken from URL on next slide)

TREC Contribution

- TREC was the original, it defined the modality, focused attention on evaluation campaigns and their usefulness, had real impact on the quality of (text) IR, reared a generation of IR researchers and spawned nearly 000's of papers;
- TREC also spun off a large number of copycat evaluation campaigns;
- TREC continues now, 15 years later, as strong as ever (c.100 participants)
- TREC Overview at http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=10667&mode=toc

Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

"Copycat" Evaluation Campaigns

- Following the Success of TREC, many evaluation campaigns have been launched;
- Not competitions, not just workshops, really "benchmarking evaluation campaigns";
- Most follow the TREC model, which in turn follows the Cranfield model;

- Related to video IR, in no particular order …

"Copycat" Evaluation Campaigns

- CLEF
- NTCIR
- INEX
- VACE
- ETISEO, PETS
- AMI
- ImagEval, Benchathlon
- CLEAR = VACE ∩ CHIL

- All are metrics-based, all use XML for data submission and exchange;

Cross-Language Evaluation Forum

- Tests aspects of mono- and cross-lingual IR;
- 2006 is the 7th in the annual series, DELOS funded, 74 groups;
- Proceedings online; 8 tracks in 2005
  - Mono-, bi- and ML doc retrieval on news;
  - Mono- and cross-lingual retrieval on structured scientific data;
  - Interactive CL information retrieval;
  - ML question-answering;
  - CL retrieval in image collections;
  - CL spoken document retrieval;
  - CL geographic retrieval;

Cross-Language Evaluation Forum

- 7 different document collections including 2M news documents, in 12 languages;
- Bulgarian, Dutch, English, Finnish, French, German, Hungarian, Italian, Portuguese, Russian, Spanish & Swedish;
- Herculean effort in securing permission for data provision;
- Task-specific details very much driven by the participants with loose control from the funders;
- Big team effort;

NTCIR

- Like CLEF, except Asian, and not as big !
- 2005 was 6th running;
- Like CLEF, NTCIR follows TREC quite faithfully;
- 4 tasks
  - ML, Bilingual and single language IR on Chinese, Korean and Japanese
  - Cross-lingual question-answering seeking named entities in Chinese, English, Japanese pairs
  - Patent retrieval and classification using Japanese and US patents
  - Question-answering beyond factoids and asking "why" - on Japanese

Initiative for the Evaluation of XML Retrieval (INEX)

- Started in 2002, 2006 is the 5th running, 80 participants;
- Aim is to establish infrastructure, collection & scoring for IR which exploits available structural information (XML elements) to yield more focused retrieval;
- XML IR may retrieve mixture of paragraphs, sections, etc.
- Collection is 659,300 Wikipedia articles from 113,483 categories with average 161 XML nodes from 5000 tagset - previously it was IEEE articles;
- Participants create candidate topics as content only or content-and-structure, gain access to document collection only then;
- Main task is ad hoc retrieval plus tasks in NL query, heterogeneous documents, interactive, document mining and Multimedia

Video Analysis and Content Extraction (VACE)

- VACE is a US DTO funding program for US groups, 14 funded participants;
- Address lack of tools to assist human analysts monitor and annotate video for indexing;
- Data is broadcast TV news, surveillance, UAV, meetings, ground reconnaissance;
- Tasks are detection a/o tracking of people, faces, vehicles and text in that data;
- VACE includes open evaluations with international participation
  - Increases competition, thus increases progress
  - Encourages peer review and information exchange, minimises wheel reinvention, focuses on common problems
  - Open evaluations include VACE, CLEAR, and TRECVid;

ETISEO

- Started Jan 2005, funded by French government, 23 participants;
- Aim to evaluate vision techniques for video surveillance applications;
- Video data used is single and multi-view surveillance (airport, car park, corridor, subway);
- Ground truth is annotations and classifications of persons, vehicles, groups;
- Tasks are detection, localisation, classification and tracking of physical objects, and event recognition

Performance Evaluation of Tracking & Surveillance

- 2006 is the 7th PETS cycle, funded by an EU FP6 project ISCAPS;
- Evaluates object detection and tracking for video surveillance, metrics based;
- Data is multi-view surveillance video (4 cameras);
- Task is event detection - left luggage in public place

Augmented Multi-Party Interaction (AMI)

- EU FP6 project covering meeting analysis from instrumented meeting rooms;
- Run by IDIAP, CH;
  - 2D multi-person tracking, head bounding box for each frame;
  - Head pose estimation;
  - Estimation of focus-of-attention (FoA) in meetings as table, documents, screen, or people using gaze

ImagEval

- French national evaluation campaign now open to other Europeans, 14? participants;
- 5 tasks related to content based image retrieval:
  - Recognition of (geometric) image transformations like rotation, projection, etc.;
  - Retrieval based on combining text and image;
  - Detect and extract text regions from images;
  - Detect objects in images - cars, planes, flowers, cats, churches, Eiffel tower, table, PC or TV, US flag;
  - (Semantic) feature detection - indoor, outdoor, people, night, day, etc.;
- Various data (image) sources, O(000's);

Benchathlon and CLEAR

- Benchathlon is a clearinghouse for data, annotations, evaluation measures, tools and architecture for CBIR;
- CLEAR is a cross-campaign collaboration concerned with getting consensus and crossover on evaluation of event classification evaluation;
- CLEAR = VACE ∩ CHIL

Consensus ?

- Multiple domains, not all IR;
- All very metrics-based with agreed evaluation procedures and data formats;
- All have manual self-annotation of ground truth vs. assessment of pooled results;
- All coordinate large volunteer efforts with little sponsorship funding;
- All have growing participation;
- All make their results public and data available to participants, for free;
- All have contributed to raising the profile of evaluation campaigns;
- These evaluation campaigns now exist in many other domains;

Contents

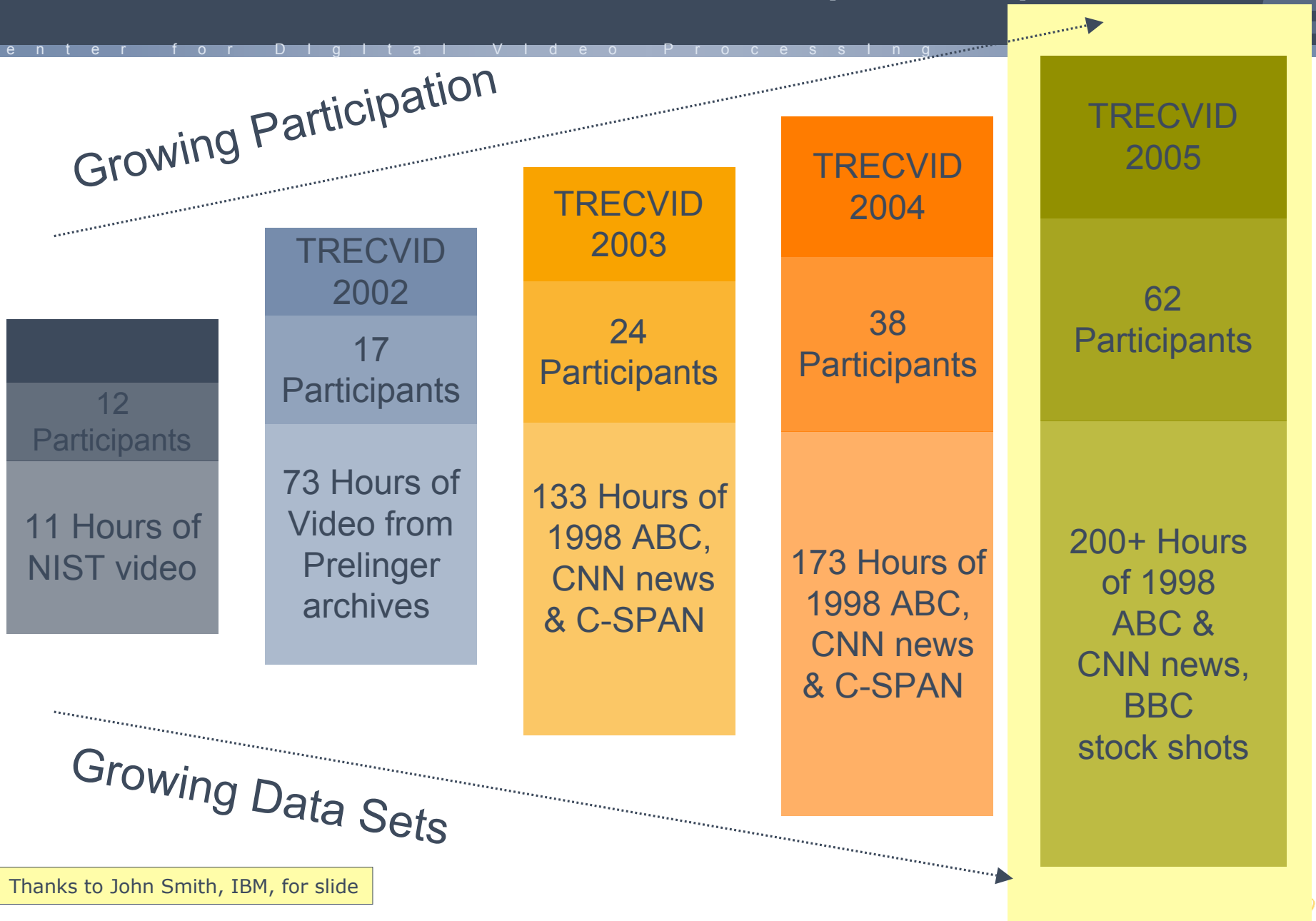
1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. Challenges and issues

TRECVid: Video IR Evaluation

- In 2001, “video retrieval” started as a TREC track;
- Usual TREC mode of operation (data-topics-search submissions-pooling-evaluation by metrics-workshop) but additional, related tasks besides search;

- In 2003 TRECVid separated from TREC because if was sufficiently different, and had enough participation, though TREC and TRECVid workshops are co-located;
- Started small, grew rapidly;

Evolution: data, tasks, participants



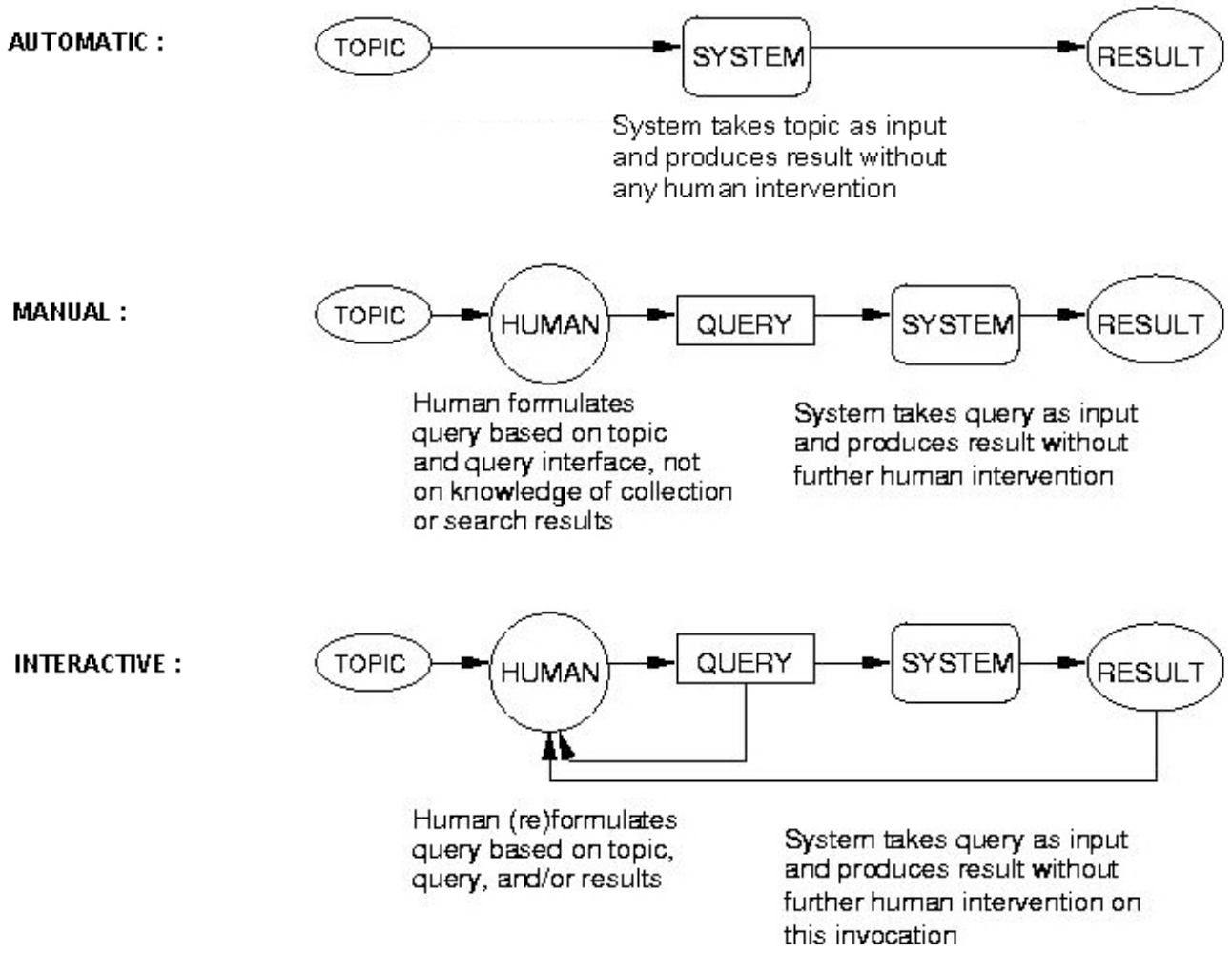
Thanks to John Smith, IBM, for slide

TRECVid Search

- 2006 will have 70+ participants, 250 hours of video (TV news and rushes);
- 2007 will have almost 1,000 hours of video;

- TRECVid tasks feature shot bound detection, concept detection and TRECVid and 3 kinds of search - automatic, manual and interactive;
- Task is to find shots (common bounds) which satisfy a user's information need;
- Evaluation is by Precision and Recall with manual judging of pooled submissions;
- Spoken dialogue is provided (from ASR/MT and CC), earlier we saw this as a basis for an approach to video IR;

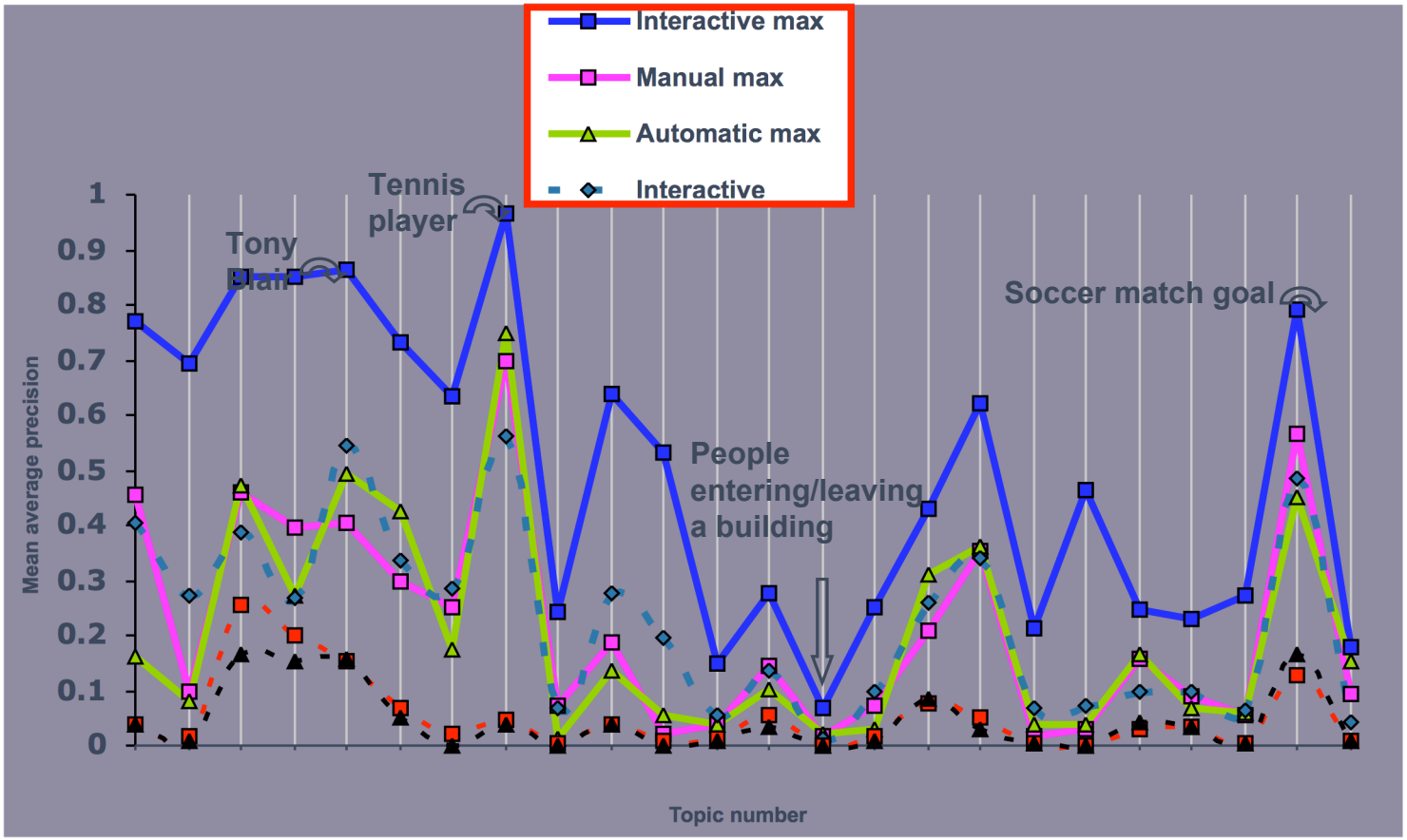
Interactive, Manual, Automatic



TRECVid Topics

149. Find shots of Condoleeza Rice [3, 6, 116]
150. Find shots of Iyad Allawi, the former prime minister of Iraq [3, 6, 13]
151. Find Find shots of Omar Karami, the former prime minister of Lebannon [2, 5, 301]
152. Find shots of Hu Jintao, president of the People's Republic of China [2, 9, 498]
153. Find shots of Tony Blair. [2, 4, 42]
154. Find shots of Mahmoud Abbas, also known as Abu Mazen, prime minister of the Palestinian Authority. [2, 9, 93]
155. Find shots of a graphic map of Iraq, location of Bagdhad marked – not a weather map [4, 10, 54]
156. Find shots of tennis players on the court – both players visible at the same time [2, 4, 55]
157. Find shots of people shaking hands [4, 10, 470]
158. Find shots of a helicopter in flight [2, 8, 63]
159. Find shots of George Bush entering or leaving a vehicle (e.g., car, van, airplane, helicopter, etc), he and vehicle both visible at the same time [2, 7, 29]
160. Find shots of something (e.g., vehicle, aircraft, building, etc.) on fire with flames and smoke visible [2, 9, 169]

2005: Mean avg. precision



TRECVid Features

- For (semantic) feature/concept identification - this is useful in retrieval and used within TRECVid for search task, and as a challenge itself;
- Initially this was done solo by groups
  - Annotate a small corpus;
  - Train a SVM as a feature detector;
- Problem was that this was not scalable to 000's of features;

TRECVid Features

- People walking/running: segment contains video of more than one person walking or running
- Explosion or fire: segment contains video of an explosion or fire
- Map: segment contains video of a map
- US flag: segment contains video of a US flag
- Building exterior: segment contains video of the exterior of a building
- Waterscape/waterfront: segment contains video of a waterscape or waterfront
- Mountain: segment contains video of a mountain or mountain range with slope(s) visible
- Prisoner: segment contains video of a captive person, e.g., imprisoned, behind bars, in jail, in handcuffs, etc.
- Sports: segment contains video of any sport in action

TRECVID-2005 Annotation

- 80 Hours Video (from development data set)
- 40 Features (from LSCOM-lite)

TRECVID Data

Up to 100%

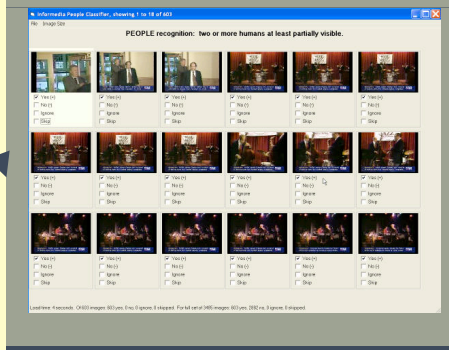
100%

Up to 100%

TRECVID Annotation Task

IBM Annotation Tool

35 concepts



CMU Annotation Tool

18 concepts

X%

100%

100-X%

TRECVID 2005 Common Training Data Set

Annotators

"Volunteers" from TRECVID participating groups

Thanks to John Smith, IBM, for slide

DCU

LSCOM-2006 Annotation

- 80 Hours Video
- 442 Features (from LSCOM 850)

TRECVID Data

LSCOM Annotation

IBM Annotation Tool

442 concepts

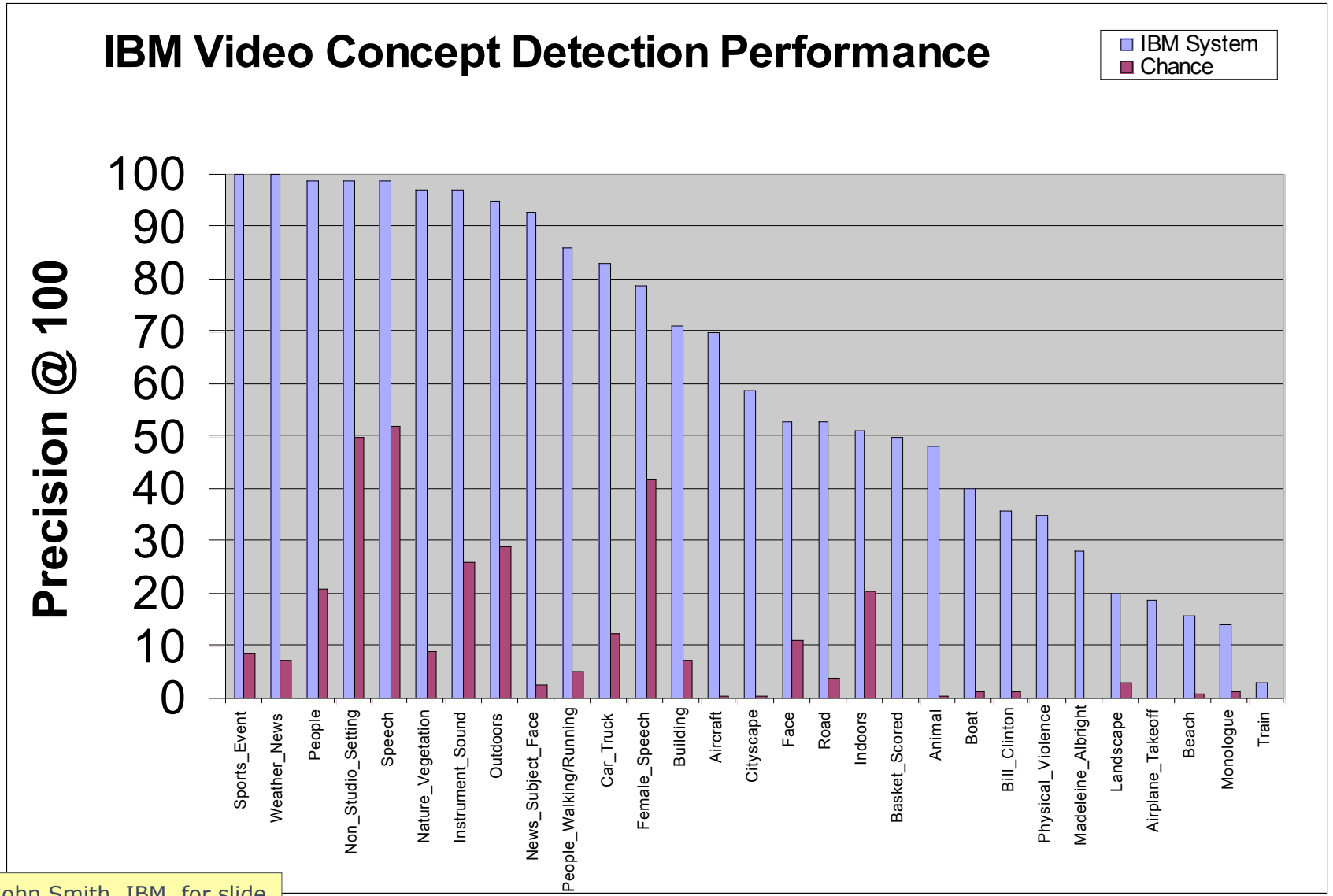
CMU Annotation Tool

442 concepts

28M shot annotations

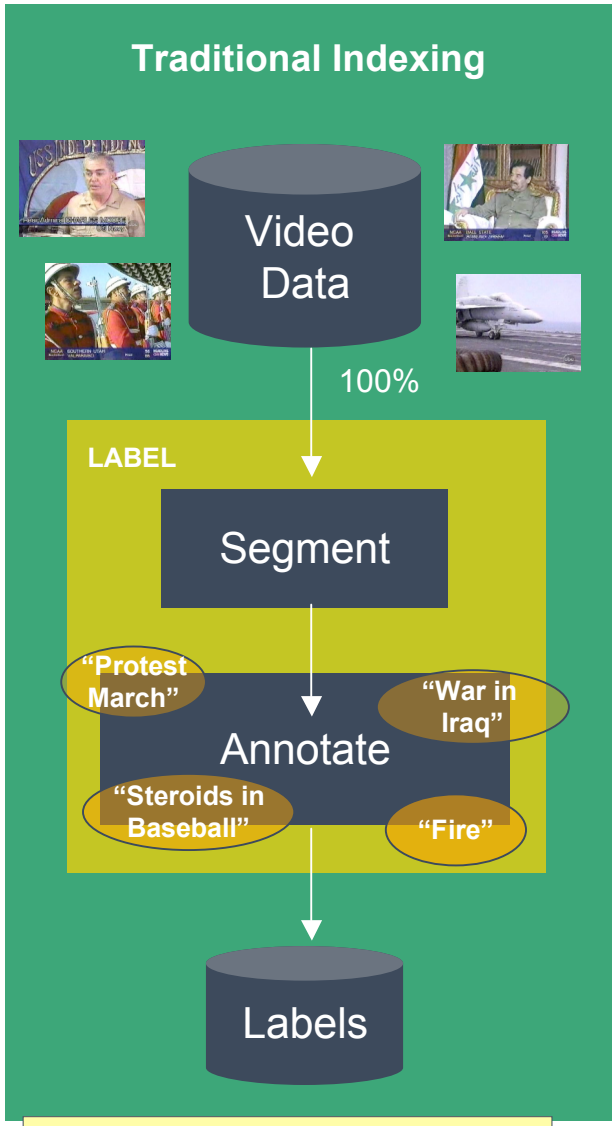
LSCOM
Funded activity

IBM Feature Detection Performance



Thanks to John Smith, IBM, for slide

Indexing by Semantic Features



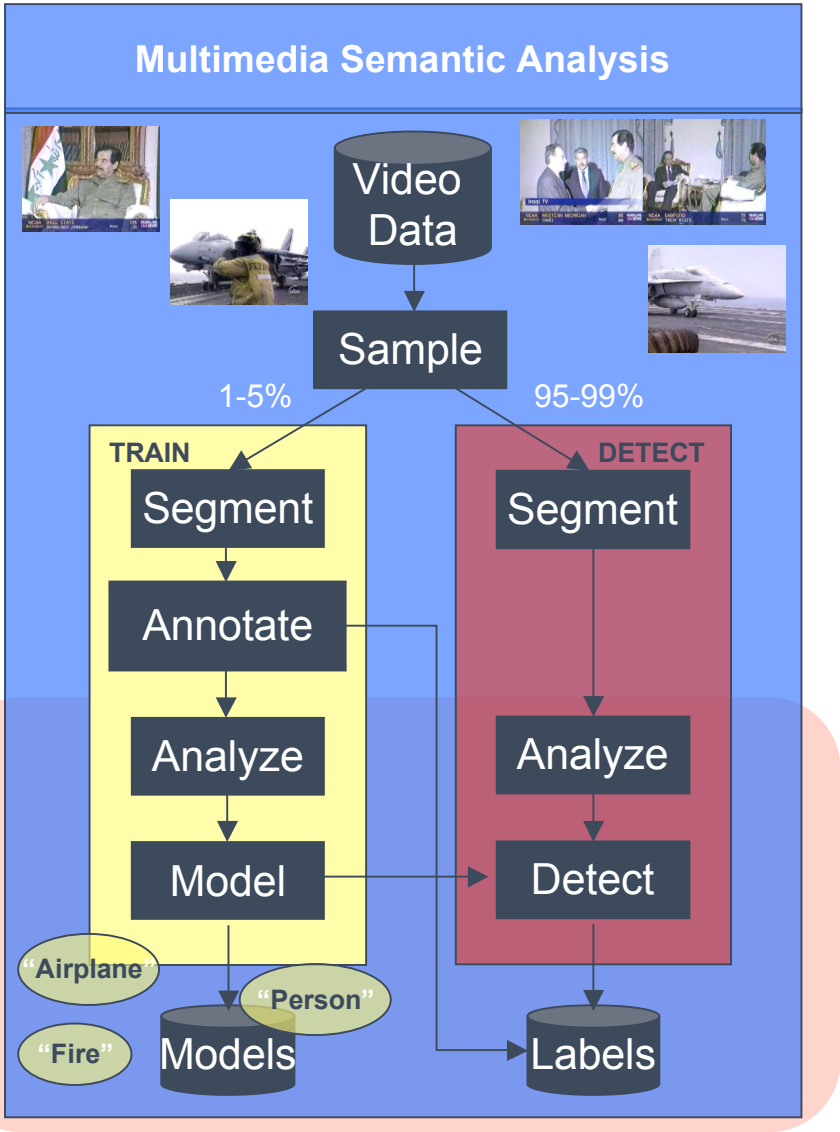
Thanks to John Smith, IBM, for slide

PROBLEM:

- Manual labeling is human resource intensive (10x)
- Results in incomplete & inconsistent annotations

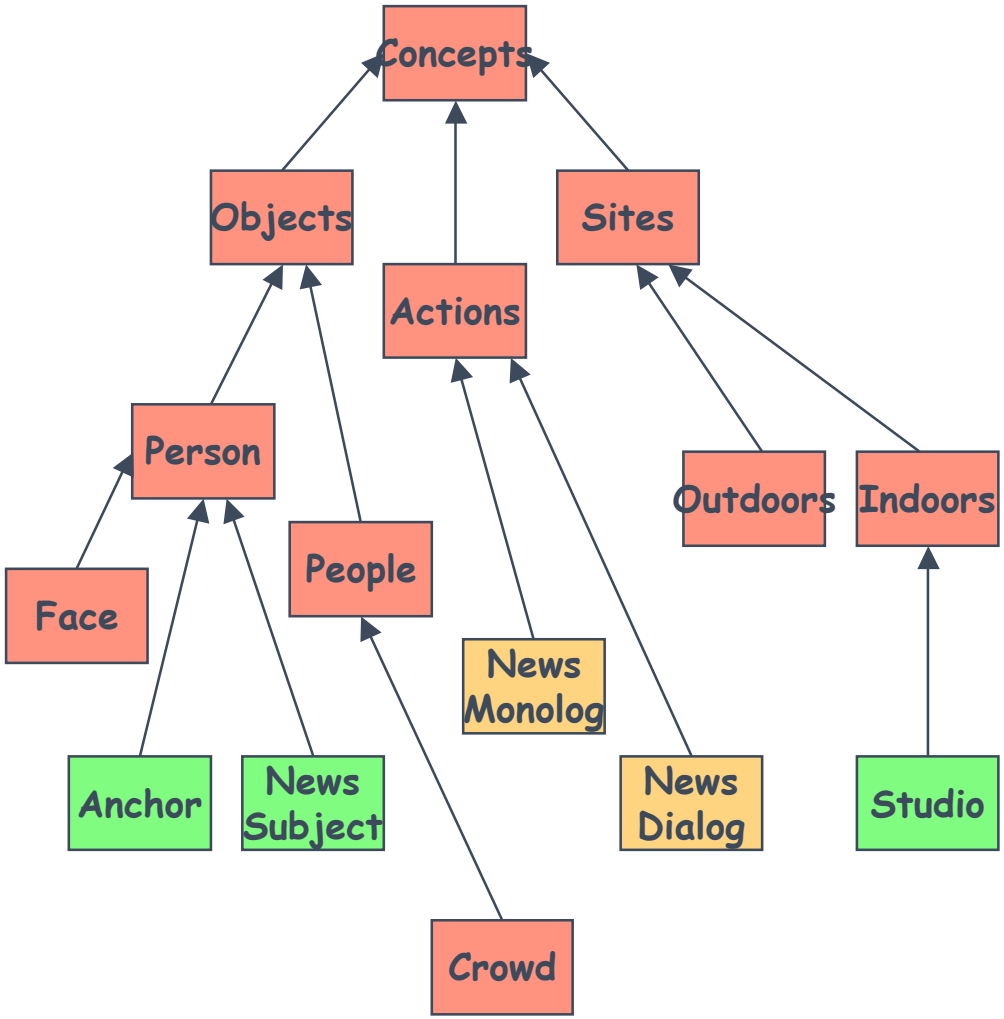
SOLUTION:

- Replace 100% manual process with learning approach
- Annotate 1-5% of training data
- Build models from training data
- Adapt to annotation errors
- Propagate labels to video repository
- Assign confidence scores



Feature Ontologies

- Previous work on feature detection treated features as independent;
- Now emphasis is on collections of features;
- Effort underway to define a 1000-concept ontology where concepts are related, dependent, and automatically computable;



Contents

1. IR vs. Information Seeking, and why it is hard
2. MMIR
   a) Key approaches and issues
   b) IR on Audio - spoken and musical
   c) IR on Images
   d) The nature of video
   e) Commercial video IR systems
   f) Experimental approaches (5)
      a) Using video metadata
      b) Using text
      c) Keyframe image matching
      d) Using semantic concepts and features
      e) Using video objects
3. MMIR Evaluation
   a) TREC
   b) Copycat evaluations
   c) TRECVid
4. **Challenges and issues**

Challenges

- Creation, capture, storage, transmission, rendering, are all solved problems in fixed and wireless domains. Challenge with MM is with access, including analysis for that purpose;
- Spoken documents & music have their obvious directions;
- New forms of MM data - combinations of sensed, biometric information - still wrestling with set-up and capture and then seeking user-pull in how it is to be accessed, what is it used for;
- Image and video is most problematic / challenging;
- We're still only doing keyframe/image and not video with motion of objects and cameras - need to break free of this convenience;
- We're still doing shot retrieval, not scene (+), or clip (-);
- Feature detection accuracy, scale-up to more features, relationships between features, move away from independent solo to ontology-based … all need to be pursued;
- Combining features, keyframe match, text and objects in a natural and usable way … those video IR systems sure aren't Google;
- Dynamically adjusting retrieval to the query/video type, but need to know more about the user's context;
- … then we can consider scale-up in volume;
- Aggregating results - synthesising "trailers" as answers;
- Summarising domain-specific video;