universität koblenz
weiter:denken

Web Retrieval
Multimedia Search

Frank Hopfgartner
Institute for Web Science and Technologies

Intended Learning Outcomes

universität koblenz
weiter:denken

At the end of this lecture, you will be able to:

- Describe the differences between metadata based and content-based retrieval
- Describe different types of image, audio and video retrieval
- Describe, at a high level, how content-based search of image, audio and video works
- Discuss the advantages and disadvantages of content-based search, for images, audio and video content

Outline

universität koblenz
weiter:denken

- Motivation and Challenges of Multimedia Search
- Image Retrieval
- Audio Retrieval
- Video Retrieval
- Summary

universität koblenz
weiter:denken

1. Motivation and Challenges of Multimedia Search

The “Grand Challenge”

universität koblenz
weiter:denken

*“Given a query in **any medium** and any language, select relevant items from a multilingual **multimedia collection** which can be in **any medium** and any language, and present them in the style or order most likely to be useful to the querier, with identical or near identical objects in **different media** or languages appropriately identified.”*

Douglas W. Oard and David Hull, AAAI Symposium on Cross-Language IR, Spring 1997, Stanford, USA

What is multimedia?

universität koblenz
weiter:denken

- Often defined as a combination of two or more media
  - E.g. combination of audio and moving images into a video, many webpages (text + images) can also be considered as “multimedia”
  - See Jaimes et. al. (2005) for some definitions from a number of prominent researchers in the field
- In IR, “multimedia” is often used to refer to the retrieval of non-text items

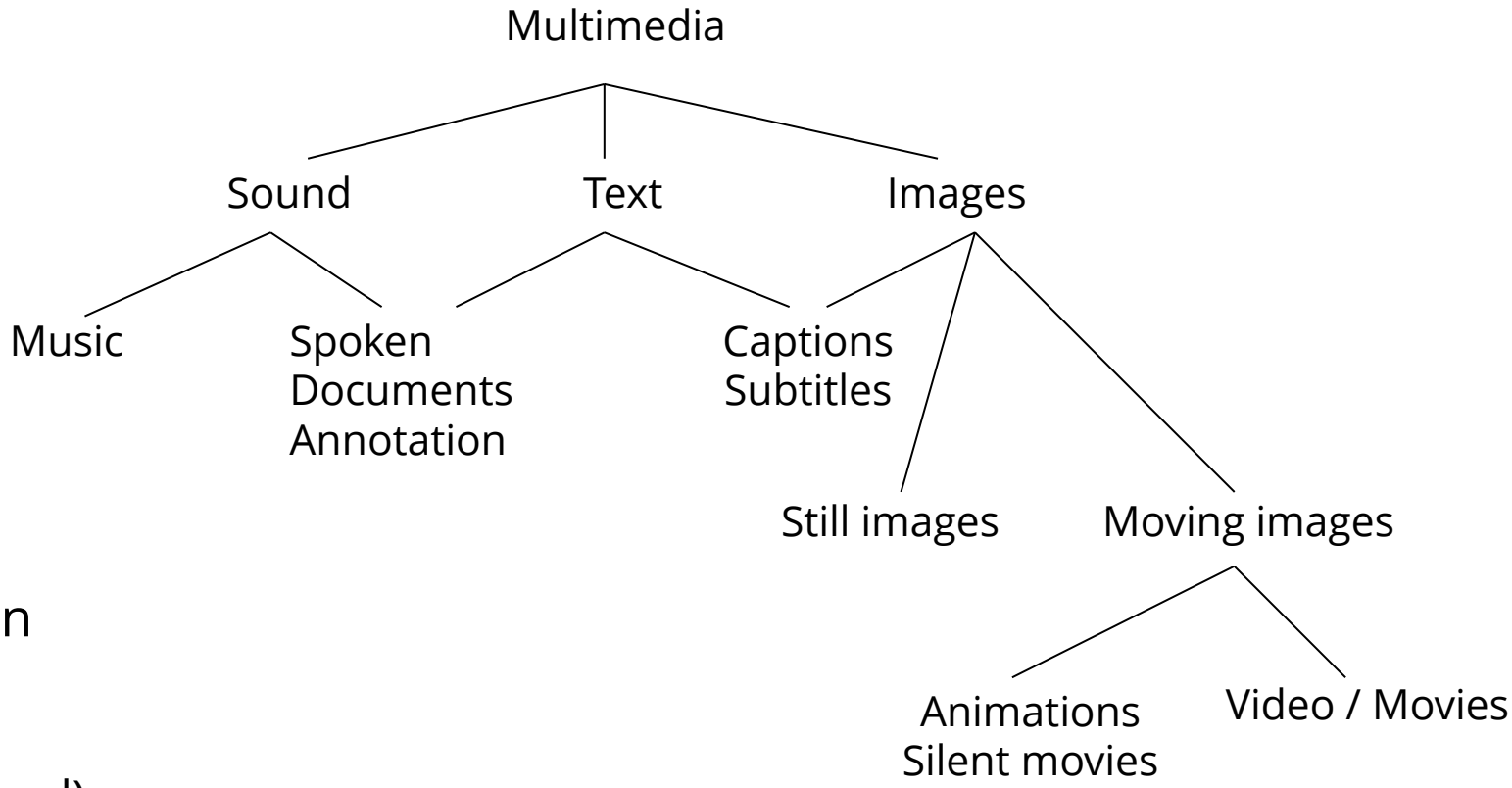
Why multimedia?

universität koblenz
weiter:denken

- Digital multimedia content is growing rapidly in both domestic and commercial sectors
  - driving new forms of interaction with images, speech, video, text and other forms of unstructured data
- Domestic consumer market fuelling the creation, use and sharing of digital multimedia
  - video (camcorders, digital cameras)
  - images (digital cameras, mobile phones)
  - audio (music and speech)
- Applications include
  - entertainment, e-commerce, digital photos, music downloads, e-learning, mobile media

How many multimedia devices do you own?

Multimedia information



Access methods depend on information available

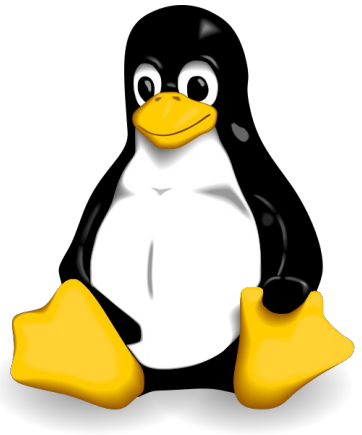
- about an object (metadata)
- within an object (feature-based)

Chu, H. (2006:43) Information Representation and Retrieval in the Digital Age, Information Today Inc (August 2003), ISBN: 1573871729.

2. Image Retrieval

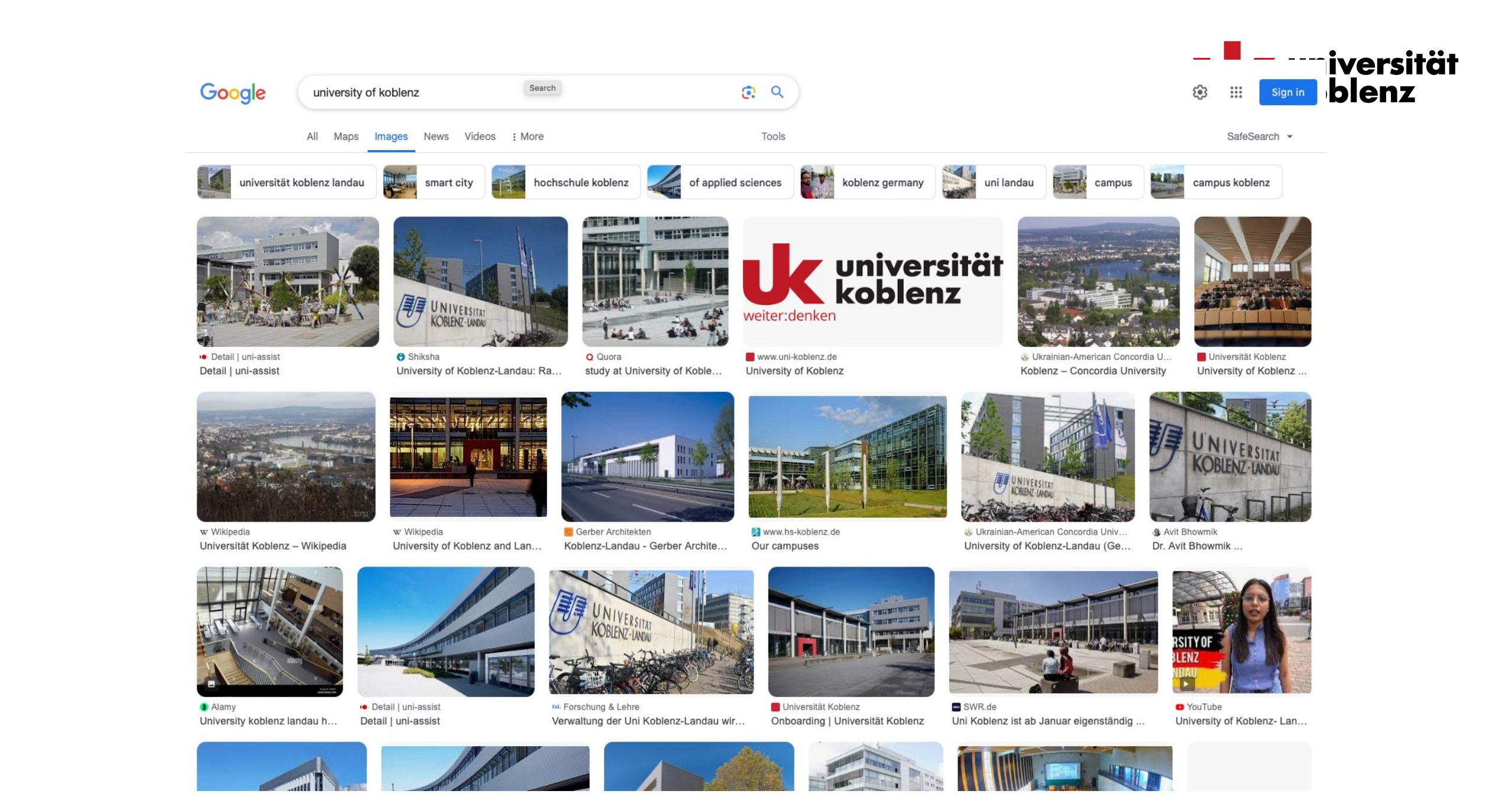
Images

- Bitmap or raster
  - Image is a grid of pixels (dots)
  - Produced by digital cameras and software such as Adobe Photoshop or Microsoft paint
- Vector
  - Image is a sequence of drawing instructions
  - Produced by drawing software such as Adobe Illustrator



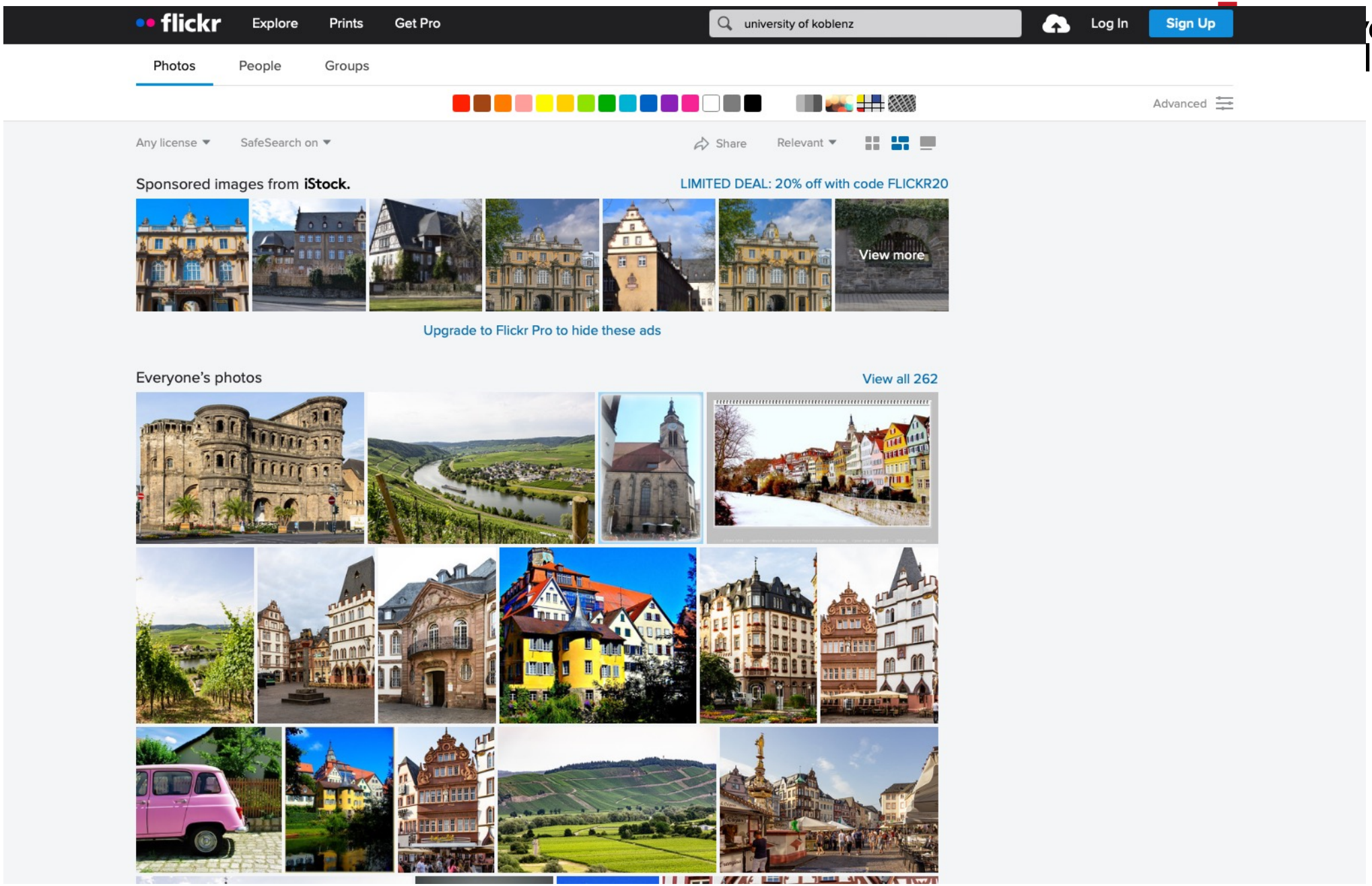
universität koblenz
weiter:denken

Have you ever carried out an image search using Google images, Bing images, or similar?


Google
university of koblenz
Search
Sign in
All
Maps
Images
News
Videos
More
Tools
SafeSearch
universität koblenz landau
smart city
hochschule koblenz
of applied sciences
koblenz germany
uni landau
campus
campus koblenz
Detail | uni-assist
Detail | uni-assist
Shiksha
University of Koblenz-Landau: Ra…
Quora
study at University of Koble…
www.uni-koblenz.de
University of Koblenz
Ukrainian-American Concordia U…
Koblenz – Concordia University
Universität Koblenz
University of Koblenz …
Wikipedia
Universität Koblenz – Wikipedia
Wikipedia
University of Koblenz and Lan…
Gerber Architekten
Koblenz-Landau - Gerber Archite…
www.hs-koblenz.de
Our campuses
Ukrainian-American Concordia Univ…
University of Koblenz-Landau (Ge…
Avit Bhowmik
Dr. Avit Bhowmik …
Alamy
University koblenz landau h…
Detail | uni-assist
Detail | uni-assist
Forschung & Lehre
Verwaltung der Uni Koblenz-Landau wir…
Universität Koblenz
Onboarding | Universität Koblenz
SWR.de
Uni Koblenz ist ab Januar eigenständig …
YouTube
University of Koblenz- Lan…

universität koblenz
weiter:denken

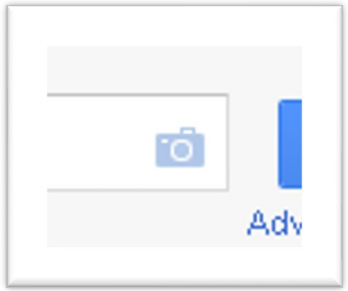
Have you ever carried out an image search using Flickr, Getty images, or similar?


flickr
Explore
Prints
Get Pro
university of koblenz
Log In
Sign Up
Photos
People
Groups
Advanced
Any license
SafeSearch on
Share
Relevant
Sponsored images from iStock.
LIMITED DEAL: 20% off with code FLICKR20
View more
Upgrade to Flickr Pro to hide these ads
Everyone's photos
View all 262

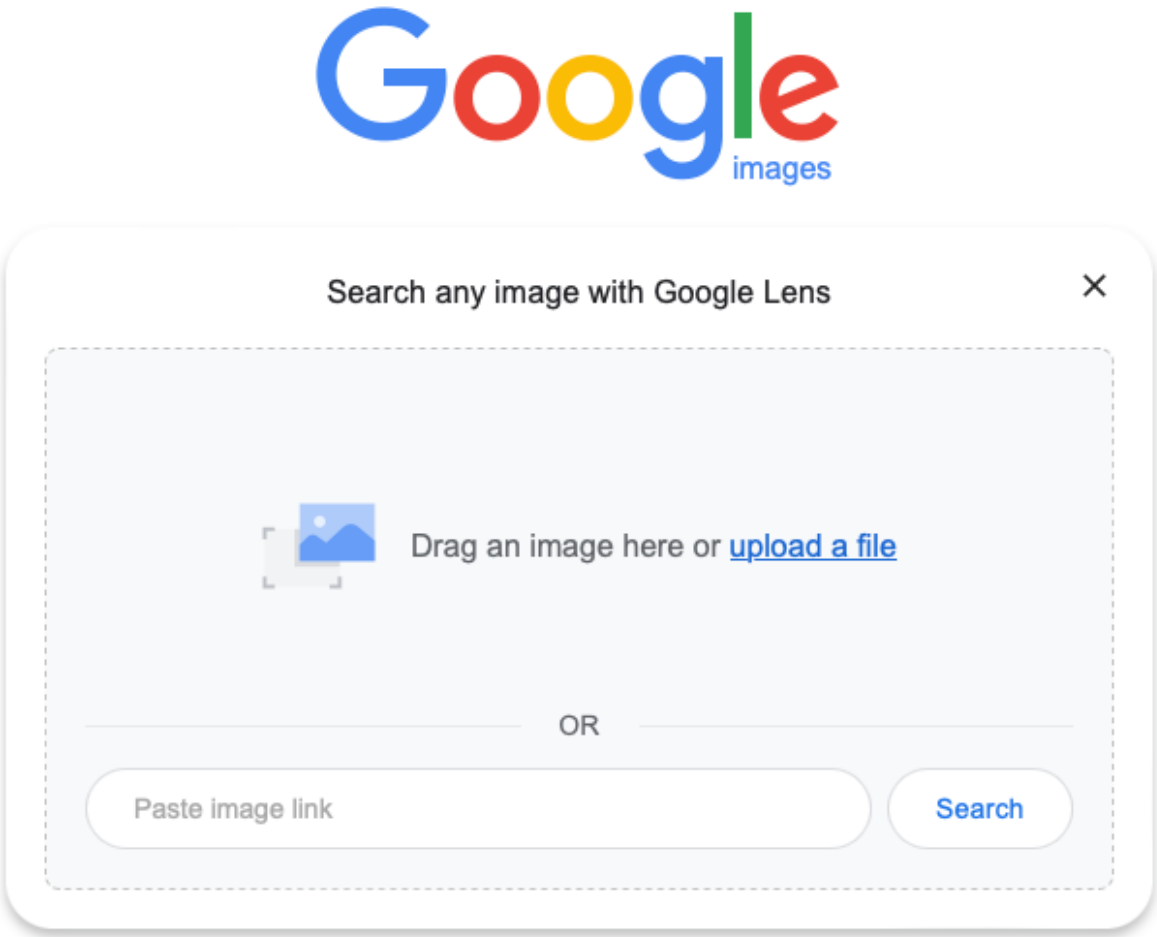
universität koblenz
weiter:denken

Have you ever carried out an image search using TinEye or Google Lens?

Aside: Google Lens: the little blue camera icon on the right of the search bar



universität koblenz
weiter:denken



Google

Upload

Sign in

Universität Koblenz

4.2

University

Search

Related searches

University of Koblenz and Landau

Koblenz

Wikipedia

File:UK Logo CMYK.png - Wikipedia

See exact matches

Universität Koblenz

Fachtag Demokratie | Universität Koblenz

Top Universities

Jan Kochanowski University in Kielce :...

LinkedIn

Aravindhan S P - Graduate Engineering...

Universität Koblenz

Kulturelles Erbe digital erfassen - Kultur ist...

X

Universität Koblenz on X: "Die neue...

LinkedIn

Sadaaf Chowdhury - Software Engineer - Ci...

Did you find these results useful?

Yes

No

Different types of image search

universität koblenz
weiter:denken

1. Search for images embedded in textual or hypertext documents
   E.g. Google/Bing image search

2. Search metadata associated with images
   E.g. Flickr/Getty images

3. Search using the image content
   E.g. TinEye, Google Lens

Representing an image by its content

universität koblenz
weiter:denken

- If an image is part of a larger document, such as a web page, we can describe it using features such as:
  - The link text which points to the image
  - “Alt” text which is provided in the HTML
  - The filename of the image
  - The text which surrounds the image
  - etc…

universität koblenz
weiter:denken

Menü Anmelden EN

universität koblenz

Research

WeST members perform research on a large number of topics that evolve around the World Wide Web and how people use it. The web is a virtual space for people to interact and share experiences. A core focus is on the analysis of data-driven applications, their application in the context of digital transformation, and their wide ranging and often disruptive impact on society.

The WeST Institute has a strong record of collaborations with national and international partners from the scientific community and industry. These fruitful collaborations allow us to bring our cutting-edge technologies into the world.

For our research, we can rely on an IaaS cloud infrastructure capable of handling big data. In addition, we are in progress of setting up a state-of-the-art usability lab to enable further research on data-driven applications.

Transparency, reproducibility, and open access are at the core of our work. Most of our research papers can be downloaded from open access repositories. In addition, we have made available various research prototypes and datasets under open licenses. Researchers all over the world use resources that got prepared and published by WeST.

Research Projects

https://www.uni-koblenz.de/de/informatik/west

universität koblenz
weiter:denken

Menü Anmelden EN

universität koblenz

Research

WeST members perform research on a large number of topics that evolve around the World Wide Web and how people use it. The web is a virtual space for people to interact and share experiences. A core focus is on the analysis of data-driven applications, their application in the context of digital transformation, and their wide ranging and often disruptive impact on society.

The WeST Institute has a strong record of collaborations with national and international partners from the scientific community and industry. These fruitful collaborations allow us to bring our cutting-edge technologies into the world.

For our research, we can rely on an IaaS cloud infrastructure capable of handling big data. In addition, we are in progress of setting up a state-of-the-art usability lab to enable further research on data-driven applications.

Transparency, reproducibility, and open access are at the core of our work. Most of our research papers can be downloaded from open access repositories. In addition, we have made available various research prototypes and datasets under open licenses. Researchers all over the world use resources that got prepared and published by WeST.

Research Projects

https://www.uni-koblenz.de/de/informatik/west

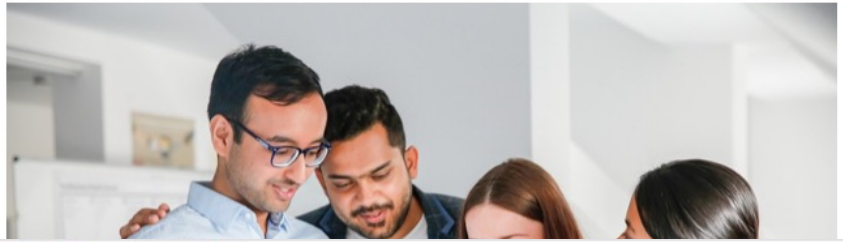
Menü Anmelden EN

universität koblenz

Note: the content of the image is not used…

Research

WeST members perform research on a large number of topics that evolve around the World Wide Web and how people use it. The web is a virtual space for people to interact and share experiences. A core focus is on the analysis of data-driven applications, their application in the context of digital transformation, and their wide ranging and often disruptive impact on society.



The WeST Institute has a strong record of collaborations with national and international partners from the scientific community and industry. These

```
<div class="six wide computer twelve wide mobile six wide tablet column">…</div>
<div class="six wide computer twelve wide mobile six wide tablet column">
  <div>
    <p class="block image align center">
      <img class="responsive" src="/de/informatik/west/uni-ko-6.jpg/@@images/image-1300-614310bd4fada5a2ea4b9a7a818847c2.jpeg" width="1300" height="867"
      srcset="/de/informatik/west/uni-ko-6.jpg/@@images/image-32-11162572e…mages/image-1000-7a30a7ec9ee116ec93cb0e5a6de90fff.jpeg 1000w" loading="lazy"
      decoding="async" alt=""> event
    </p>
    <br>
  </div>
</div>
```

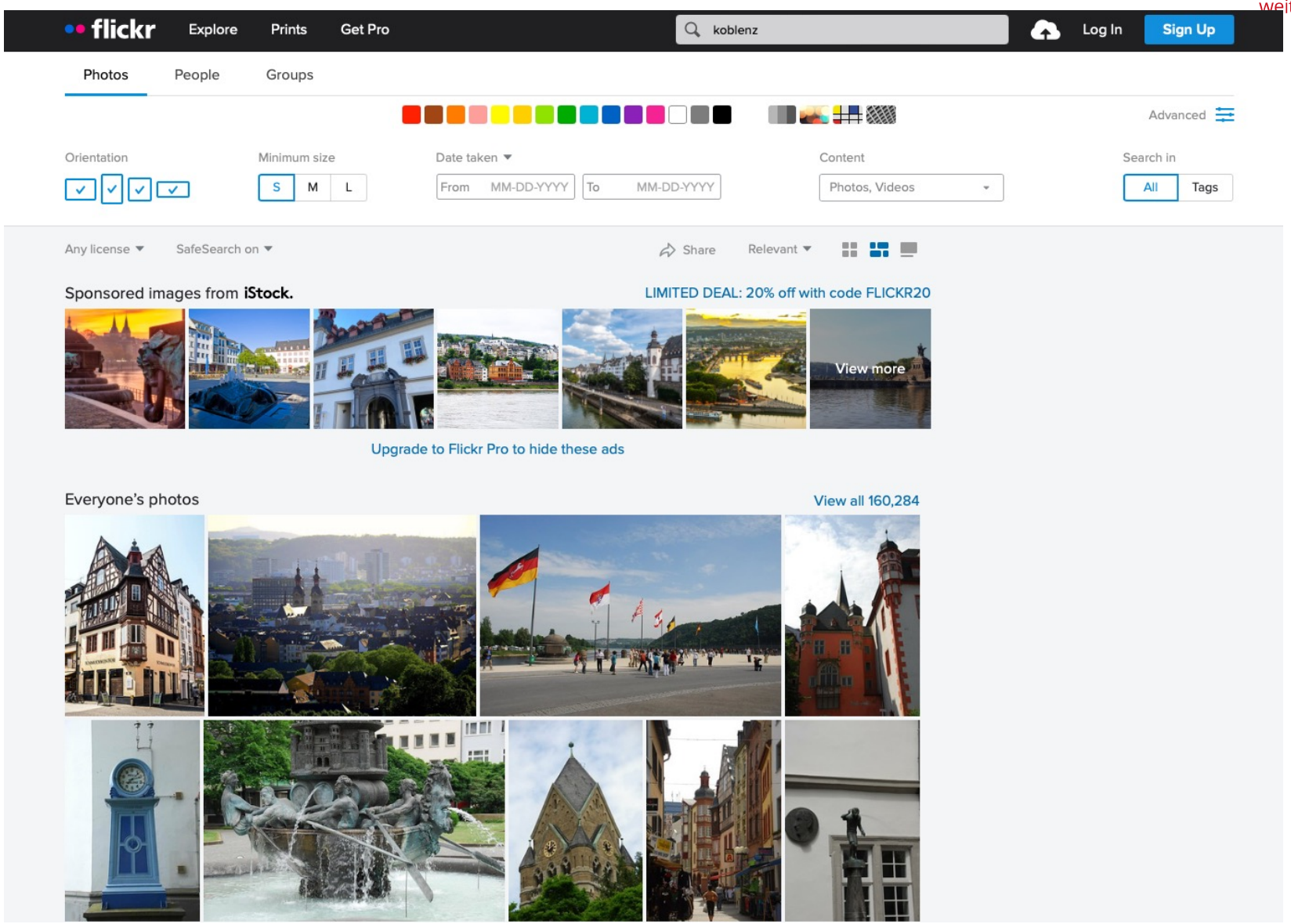
resources that got prepared and published by WeST.

Research Projects

https://www.uni-koblenz.de/de/informatik/west

Search metadata associated with images

universität koblenz
weiter:denken



Metadata-based image retrieval

universität koblenz
weiter:denken

- Images indexed using associated text (metadata)
- Metadata is “data about data”
  - Controlled vocabularies
  - Uncontrolled vocabularies (folksonomies)
  - Embedded data (e.g. Exif)
  - Textual descriptions
  - File (and folder) names
- Typically assigned manually
- Retrieval uses traditional text retrieval or database approaches
- This approach is popular online (e.g. Flickr) where there are many image uploaders who can add metadata to their own images

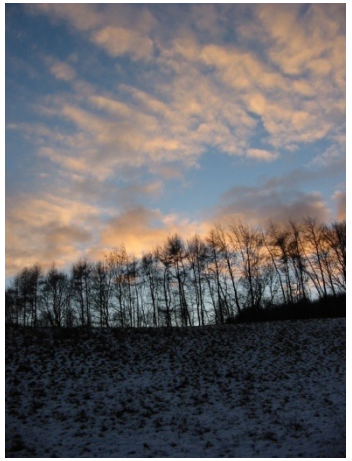


Photo
Trees
Sunset
Dusk

Example metadata

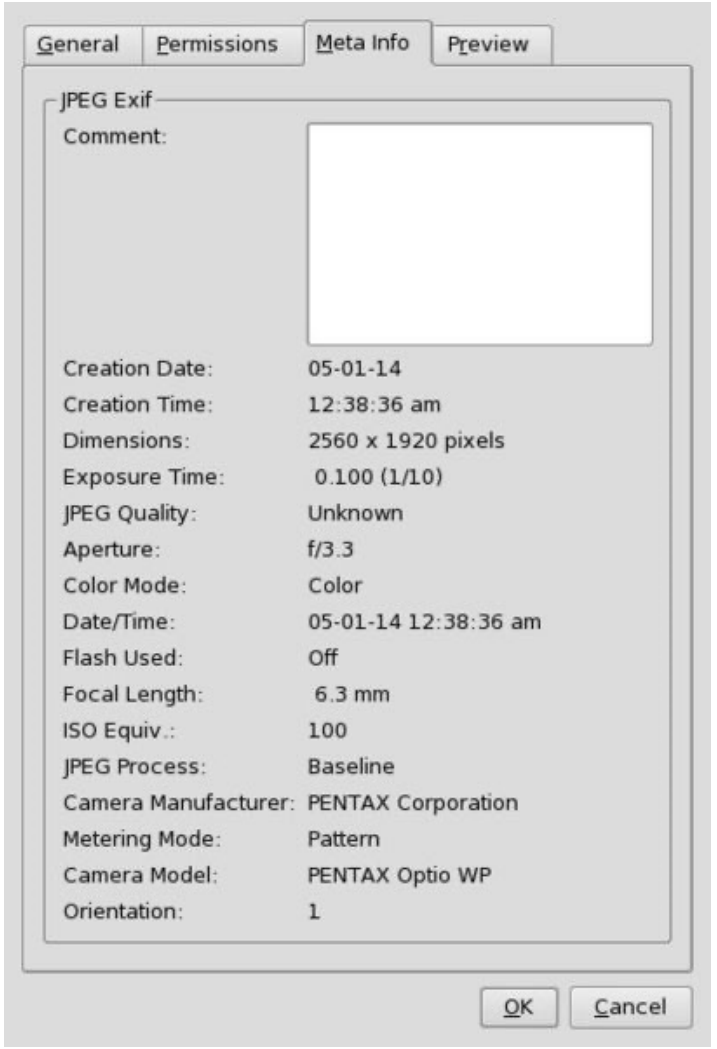
universität koblenz
weiter:denken

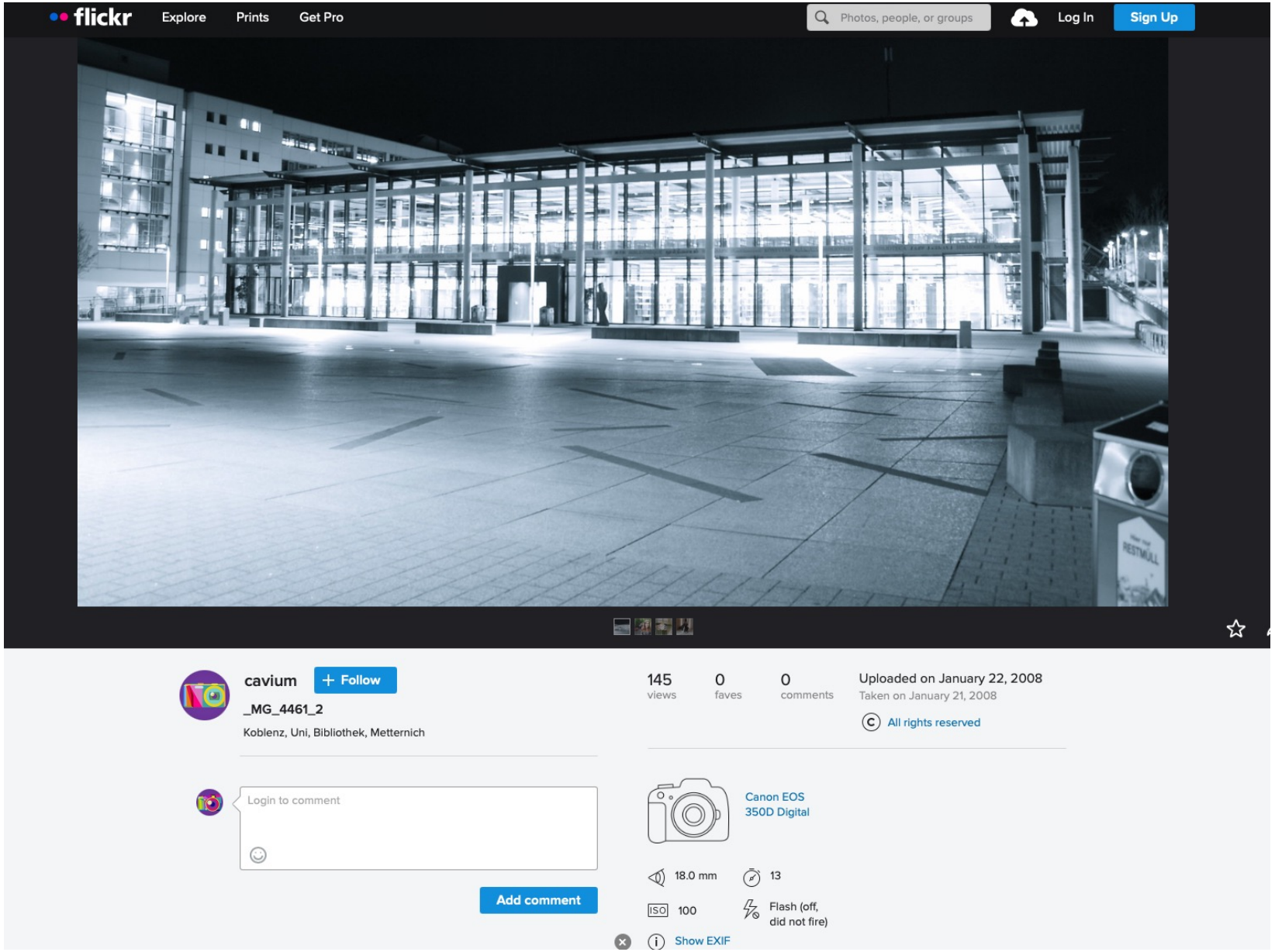
Record ID:	JV-.044809
Short title:	The Smeaton Tower, Plymouth.
Long title:	Plymouth Hoe. The Smeaton [Lighthouse] Tower.
Location:	Devonshire, England
Description:	Red and white striped lighthouse on coastal cliff with harbour and town beyond, and substantial building on cliff terrace below.
Date:	Registered 1904
Photographer:	J Valentine & Co
Categories:	[lighthouses][beacons & lighthouses][Devon all views][Collection - J Valentine & Co]
Notes:	JV-44809 pc/mb(or possibly 44810)TECH: Coloured.

Automatically assigned metadata: EXIF

universität koblenz
weiter:denken

- Exchangeable Image File Format (EXIF)
  - A *standard* for storing interchange information *in* image files, especially those using JPEG compression
- Most digital cameras now use EXIF format to store
  - Date and time information
  - Camera settings (e.g. orientation, aperture, shutter speed, focal length)
  - Geolocation (based on GPS)





Canon EOS
350D Digital

18.0 mm

13

ISO 100

Flash (off,
did not fire)

Hide EXIF

Compression - JPEG (old-style)

Make - Canon

Orientation - Horizontal (normal)

X-Resolution - 72 dpi

Y-Resolution - 72 dpi

Software - QuickTime 7.4

Date and Time (Modified) - 2008:01:22 00:00:41

Host Computer - Mac OS X 10.5.1

YCbCr Positioning - Centered

ISO Speed - 100

Exif Version - 0220

Date and Time (Original) - 2008:01:21 21:03:13

Date and Time (Digitized) - 2008:01:21 21:03:13

Components Configuration - Y, Cb, Cr, -

Exposure Bias - +2/3 EV

Metering Mode - Multi-segment

Flashpix Version - 0100

Color Space - Uncalibrated

Camera ID - 31

Camera Type - Digital SLR

https://www.flickr.com/photos/23001067@N05/2210075187/

Problems with metadata-based image retrieval

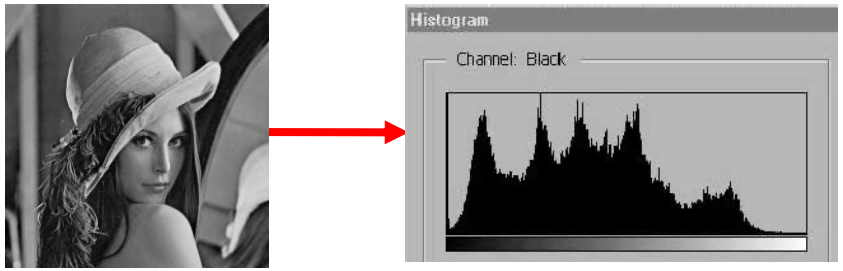
universität koblenz
weiter:denken

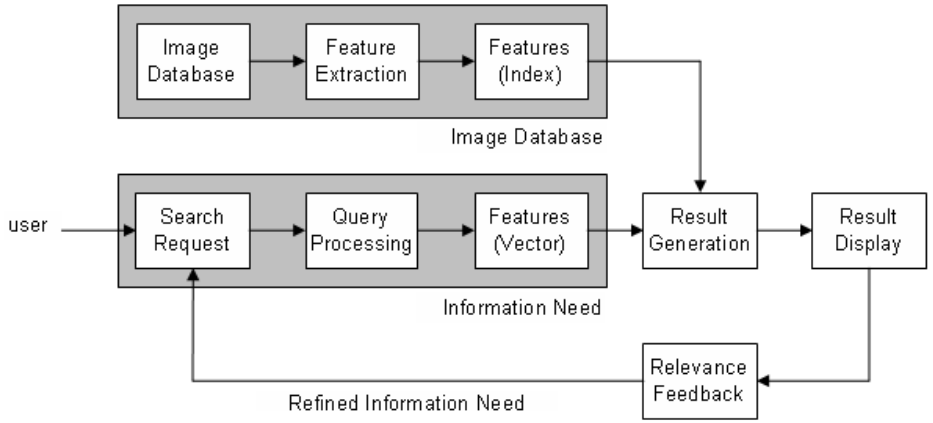
- Manual annotation is expensive (time)
- Meaning of image difficult to interpret/express in written form
- Indexed text might not relate to image content
- Often short texts to index with (e.g. captions)
- Manual annotation is subjective and suffers from low agreement between individuals (and groups)
- Vocabulary mismatch between indexer and user
- Difficult to express more abstract needs (e.g. “visual” or emotive queries)
- Notions of relevance (i.e. based on factors other than topic/theme of image)

Content-based image retrieval

universität koblenz
weiter:denken

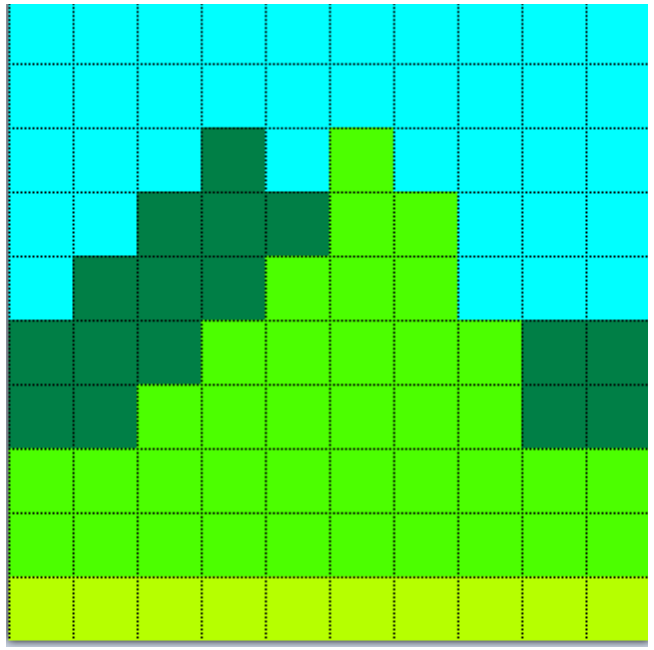
- Images are represented using "low-level" features which are automatically generated
  - E.g. based on colour, shape, texture, etc.
  - Image represented as "feature vector"
- System compares the "feature vector" of the query and to the "feature vector" of all stored images
  - Uses a similarity/distance measure
  - This generates a score (a number)
- A ranking is then produced by sorting this score, e.g. from largest to smallest





Examples of a very simple global “colour” feature

universität koblenz
weiter:denken

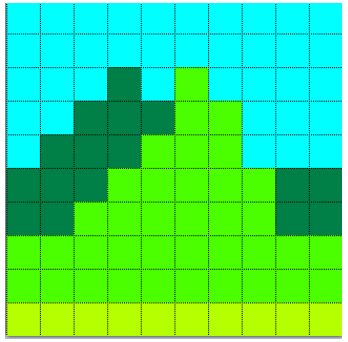


Frequency 37 37 16 10

Feature vector: (0.37, 0.37, 0.16, 0.10)

(scaled by the image size of 10 by 10 pixels,
i.e. 37/100, 37/100, 16/100, 10/100)

Query

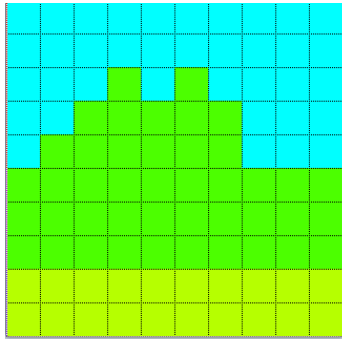


Feature: 0.37, 0.37, 0.16, 0.10

Match (similarity measure, e.g. cosine)

$$\frac{0.37x0.37 + 0.37x0 + 0.16x0.43 + 0.10x0.20}{\sqrt{(0.37^2 + 0.37^2 + 0.16^2 + 0.10^2)} + \sqrt{(0.37^2 + 0.00^2 + 0.43^2 + 0.20^2)}}$$

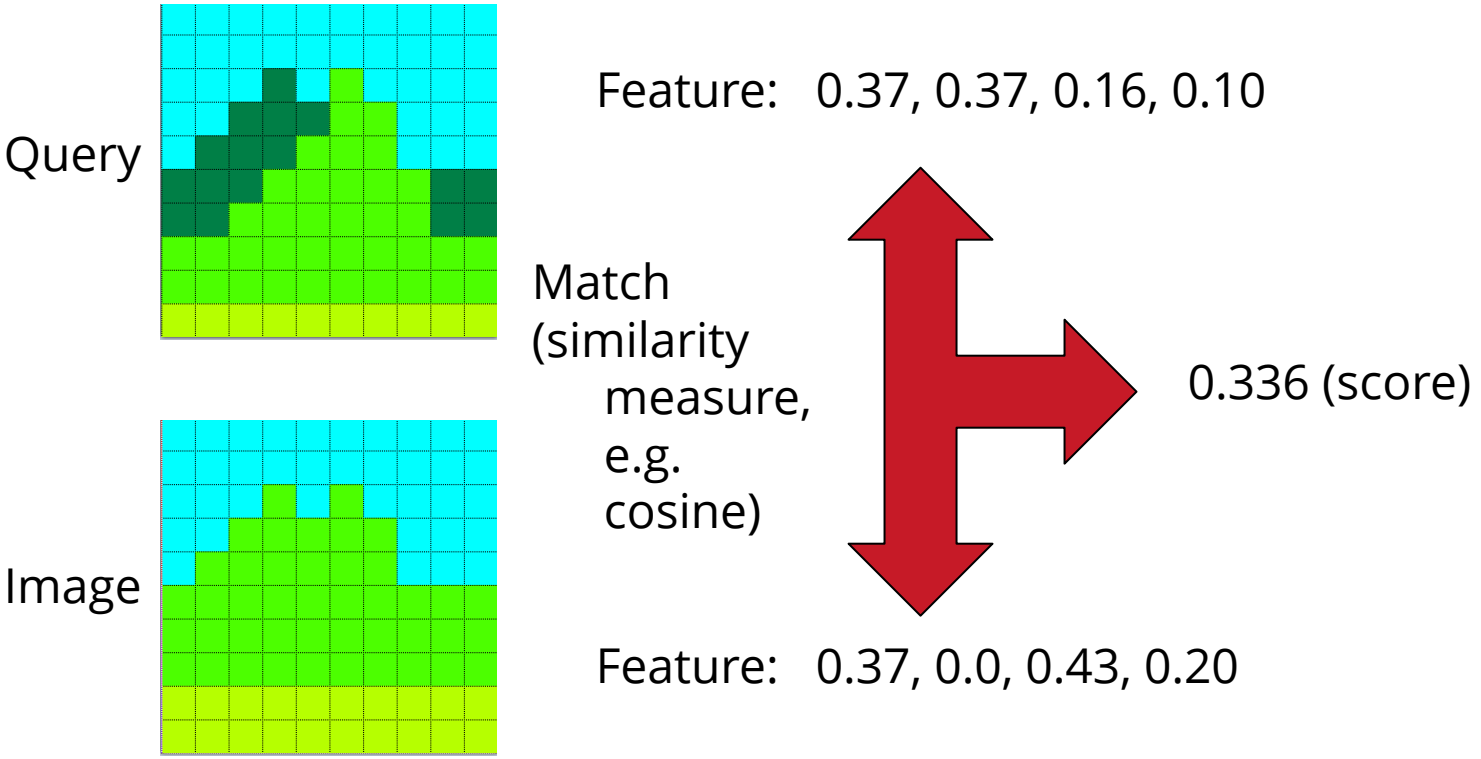
Image



Feature: 0.37, 0.0, 0.43, 0.20

Cosine similarity: page 78 of Baeza-Yates et. al. (2011)

Simple matching



Cosine similarity: page 78 of Baeza-Yates et. al. (2011)

Score is calculated against all images in the collection, images then ranked by score

Content-based image retrieval

universität koblenz
weiter:denken

- Examples of "global features" which can be extracted from images:
- Colour describes pixels (primitive feature)
  - Models include HSV (Hue, Saturation, Value) and RGB (Red, Green, Blue)
  - Define distribution of colour pixels in an image (histogram)
- Texture describes (small-scale) regions
  - Smoothness, contrast, regularity, directionality etc.
  - Often of limited value on its own
- Shape (image segmentation)
  - Represent the size, shape and orientation of objects (blocks & regions)
  - Represent relative position of objects

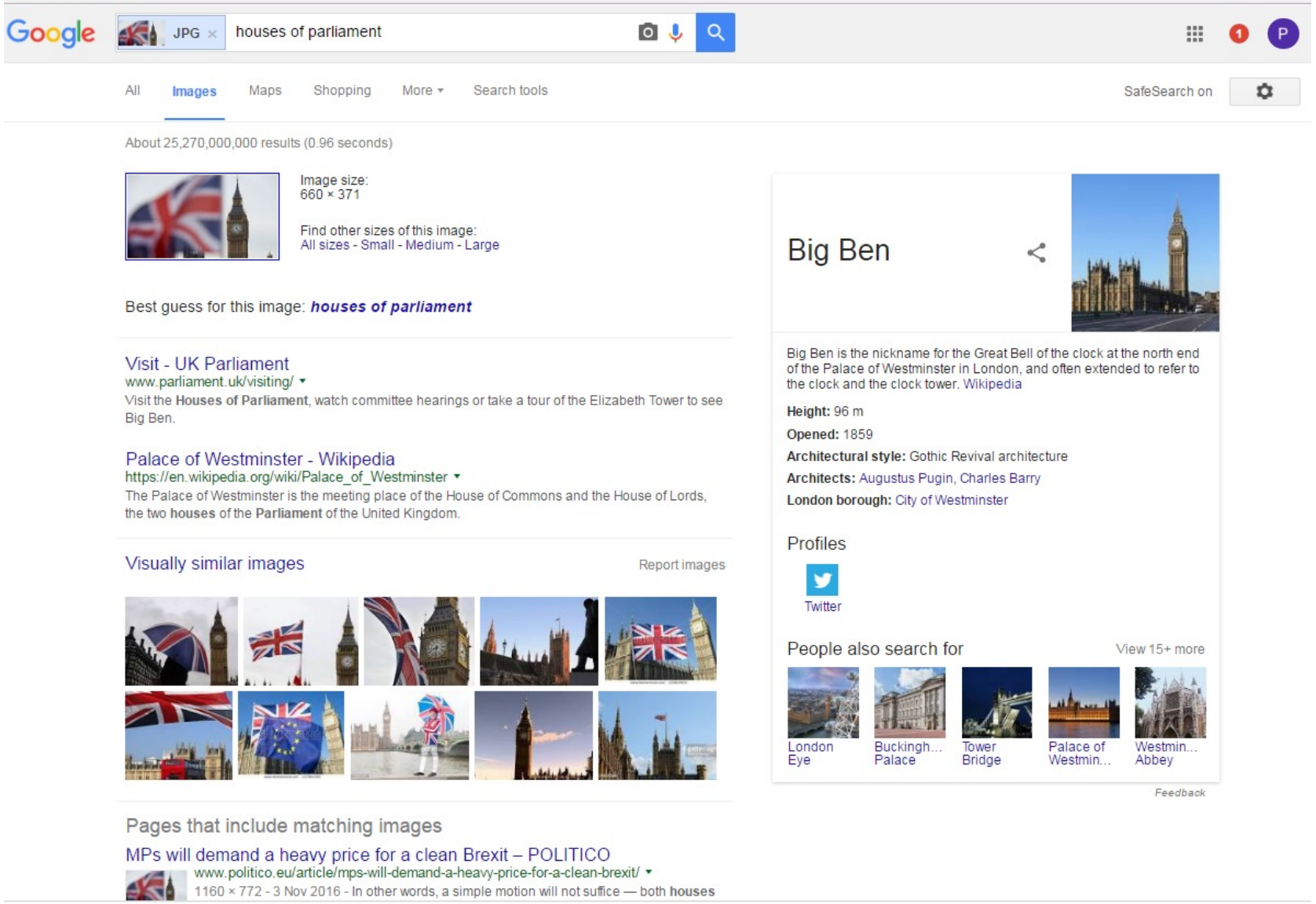
Querying

universität koblenz
weiter:denken

- Query by example
  - One or more example images
- Query by sketch
  - Allow the user to “sketch” an image, which is then used for search
- Relevance feedback
  - Can browse a collection by selecting one or more existing results, which will then be used as the query

Query by example

universität koblenz
weiter:denken



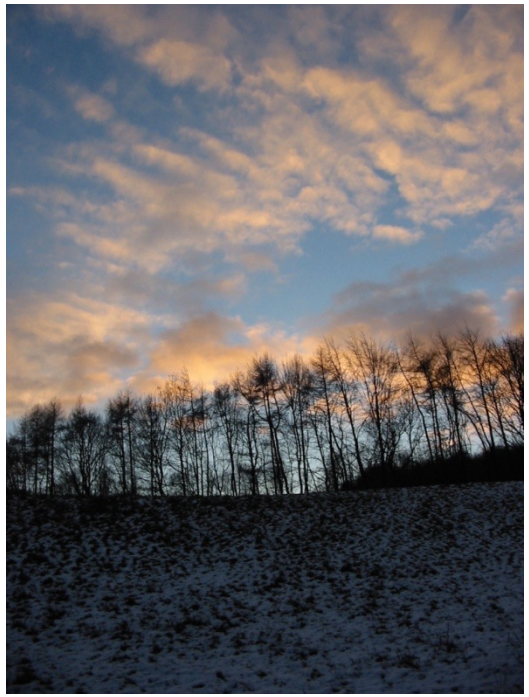
Query by sketch (imgseek)

universität koblenz
weiter:denken

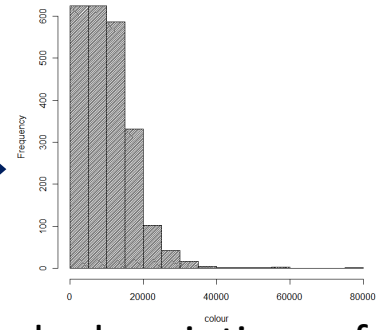
From http://www.imgseek.net/

Big problem: The Semantic Gap

Probably the biggest challenge in multimedia retrieval



It’s difficult to make this jump



Simple descriptions of the image, such as how frequently a colour occurs

“Low level”

“Landscape”
“Evening”
“Forest”
“Snow”
etc.

Semantic descriptions of the image

“High level”

Other problems with content-based image retrieval

universität koblenz
weiter:denken

- Difficult to express information needs visually
  - May not always have an image at hand
  - The same objects may be visually very different
  - Drawing or sketching is difficult for most people
  - Difficult to express more abstract needs
- May require training
  - e.g. to find a picture of “Justin Bieber”, need lots of examples of his image to train a classifier
- Generally poorer performance!

3. Audio retrieval

Audio information retrieval

universität koblenz
weiter:denken

- Two main types:
  - Techniques which assume speech in the audio
    - Use Automatic Speech Recognition (ASR) to generate text transcripts
  - Techniques which do not assume speech!
    - E.g. Music retrieval

Jonathan Foote. 1999. An overview of audio information retrieval. *Multimedia Syst.* 7, 1 (January 1999), 2-10.

What is audio?

universität koblenz
weiter:denken

- Analogue audio signal which is sampled and digitised
  - E.g. MP3 music file, most digital audio
- Symbolic music
  - E.g. MIDI (Musical Instrument Digital Interface)
- [concentrate here on the former]

Speech recognition

universität koblenz
weiter:denken

- Much audio is speech based
  - Interviews, lectures, radio plays, etc.
- Can use Automatic Speech Recognition (ASR) systems to transcribe the speech
  - Feed out audio files to an ASR system
  - Index the resulting transcripts using text retrieval techniques
  - Query the text transcripts, which can include timestamps back into the audio stream

Problems and issues

universität koblenz
weiter:denken

- ASR systems have improved considerably, but their performance can still be less than ideal
  - The audio may contain a wide range of different accents, in varying audio quality
  - ASR systems work best with *training*
    - Ideally on individuals, but also on types of accents
    - An ASR system trained on American accents is unlikely to work well with Yorkshire accents
  - Audio often involves multiple speakers
    - E.g. consider a TV or radio interview, such as on news programs

Other associated technologies

- Speech discrimination
  - Aims to discriminate speech from non-vocal music or other sounds
    - Important to disable ASR system when there is no speech
- Speaker identification
  - Aims to identify who is speaking
    - Works from a set of known voices (“closed set identification”)
  - Related area of speaker verification aims to determine whether a speaker is who he/she claims to be

Non-speech approaches

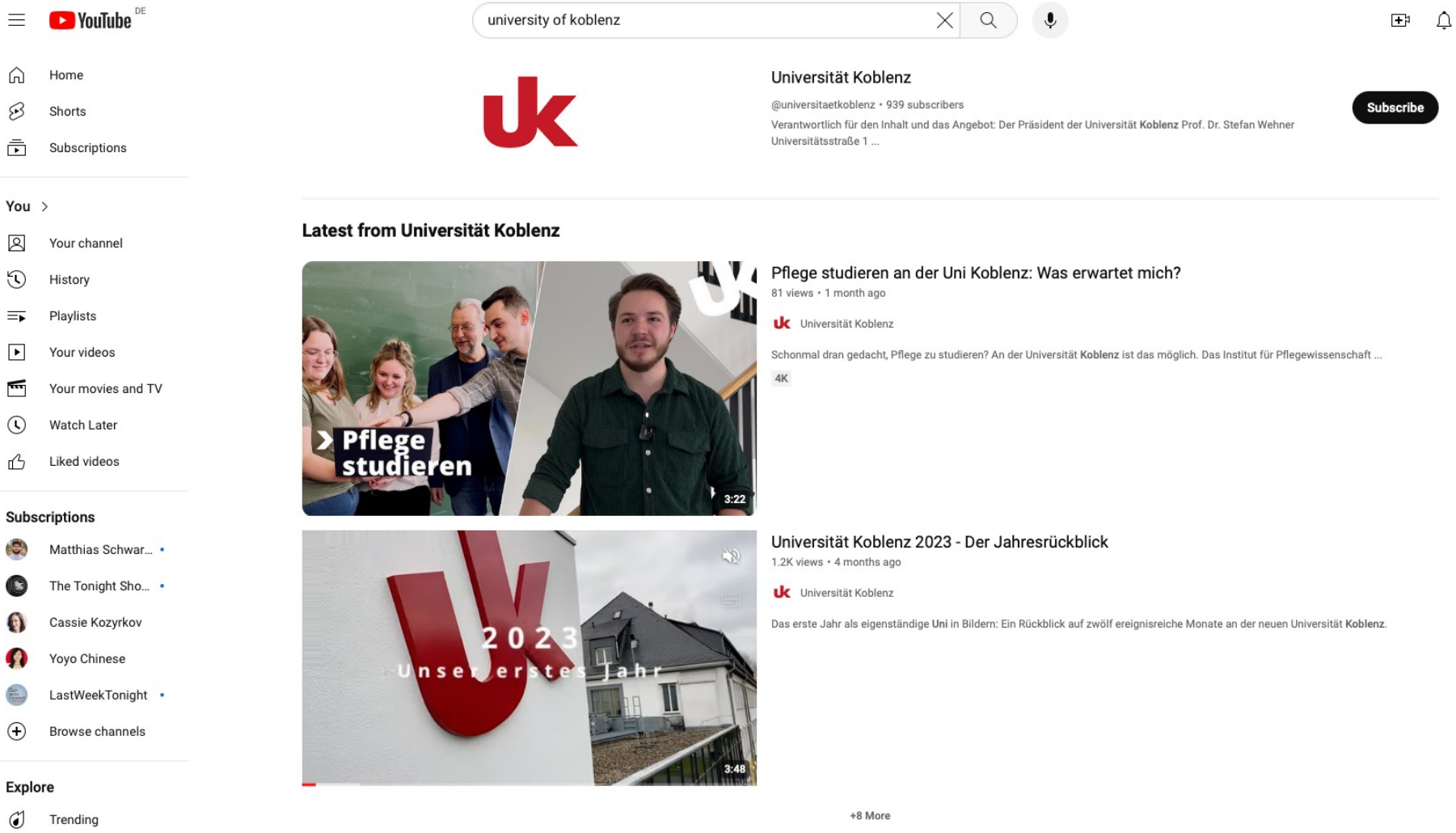
universität koblenz
weiter:denken

- Some audio does not consist of speech
  - E.g. Music, special effects, etc.
- Many approaches, including:
  - Audio classification
  - Query by Singing/Humming
  - Music fingerprinting

3. Video retrieval

Text-based video retrieval

universität koblenz
weiter:denken



Youtube.com

Text-based video retrieval

universität koblenz
weiter:denken

- Entire video is retrieved using assigned metadata
  - Popular for online video sharing systems (e.g. Youtube)
  - Suffice for short videos, but offer limited interactivity with the video itself
- Problems:
  - What if your videos are very long?
  - Looking for part of a video, buried somewhere in the middle?
  - Who assigns the metadata?
    - On YouTube, it's the user who uploads the video

Content-based video retrieval

universität koblenz
weiter:denken

- Aims to work automatically, without human effort
  - No need for a human to assign metadata
- Indexes and retrieves sections of videos, rather than whole videos
  - Tries to give direct access to the most relevant parts of a video
- Provides ways in which a video can be visualised without watching it
  - Watching video takes time (even if it’s enjoyable)

What is a video?

universität koblenz
weiter:denken

1. A sequence of image frames

2. An associated audio track, which can contain speech, music, effects, etc.

"The quick brown fox jumped over the lazy dog . . . "

Content-based video retrieval typically uses the visual frames, plus the speech, to index and search videos

Shot segmentation

- Splits the video into sequences of visually related frames, called *shots*
  - Pairs of frames are compared, large differences between frames indicate shot boundaries
  - Some shot boundaries are more difficult to detect, such as gradual fades from one shot to the next

Keyframes: Representing shots

- To visually represent each shot, one or more frames can be extracted
  - Often take the “iframes” from the compressed mpeg video stream
  - Called “keyframes”
- A single image can then be used to represent the shot when searching or visualising the video

Audio stream: text transcripts

universität koblenz
weiter:denken

- A text transcript of the audio stream is often an important source of information for retrieval
  - Automatic Speech Recognition (ASR) systems can be used
    - Can be very unreliable without training (as previously mentioned)
  - Closed captions (subtitles)
    - Much more reliable, but not always available
    - e.g. TV rushes, pre-production video
- Not all videos have speech
  - E.g. some cartoons for pre-school children, the films of “Charlie Chaplain”, etc.

Representing shots using a text transcript

universität koblenz
weiter:denken

A text transcript, such as generated by an ASR system, has timestamps for each word which can be aligned with the shots

Summary: What do we have?

universität koblenz
weiter:denken

- A video has been segmented into a sequence of shots
- Each shot can be represented by:
  - One or more *keyframe* images
  - Associated text from a transcript

Video surrogates

universität koblenz
weiter:denken

- Summary of a video created to help users make sense of a video without watching it
  - Keyframe
  - Storyboards
  - Collage of image
  - Fast-forward/skimming
  - Keywords/descriptions

What can we do now?

universität koblenz
weiter:denken

- Search:
  - Can now index this information to allow it to be searched
  - The transcripts can be searched by text
  - The keyframes can be visually searched or browsed using image retrieval methods
    - E.g. “find me another shot visually similar to *this* one”
- Presentation/visualisation
  - Keyframes and transcripts can be used visualise videos and parts of videos

Keyframes, title, description, tags

universität koblenz
weiter:denken

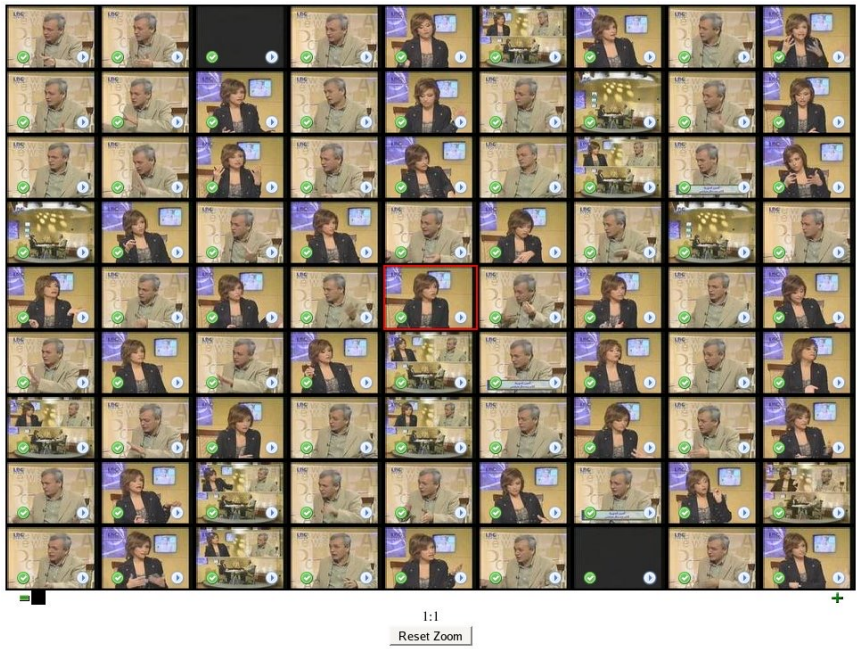
Mediating Awareness and Communication through Digital but Physical Surrogates (1996)

Video demonstration from the 1999 CHI conference. Digital but physical surrogates are tangible representations of remote people positioned within an office and under digital control. Surrogates selectively collect and...

Genre: Educational
Keywords: CHI; UID; Surrogates; Remote Sensing
Duration: 00:06:53
Popularity (downloads): 377

From the Open Video project (http://www.open-video.org)

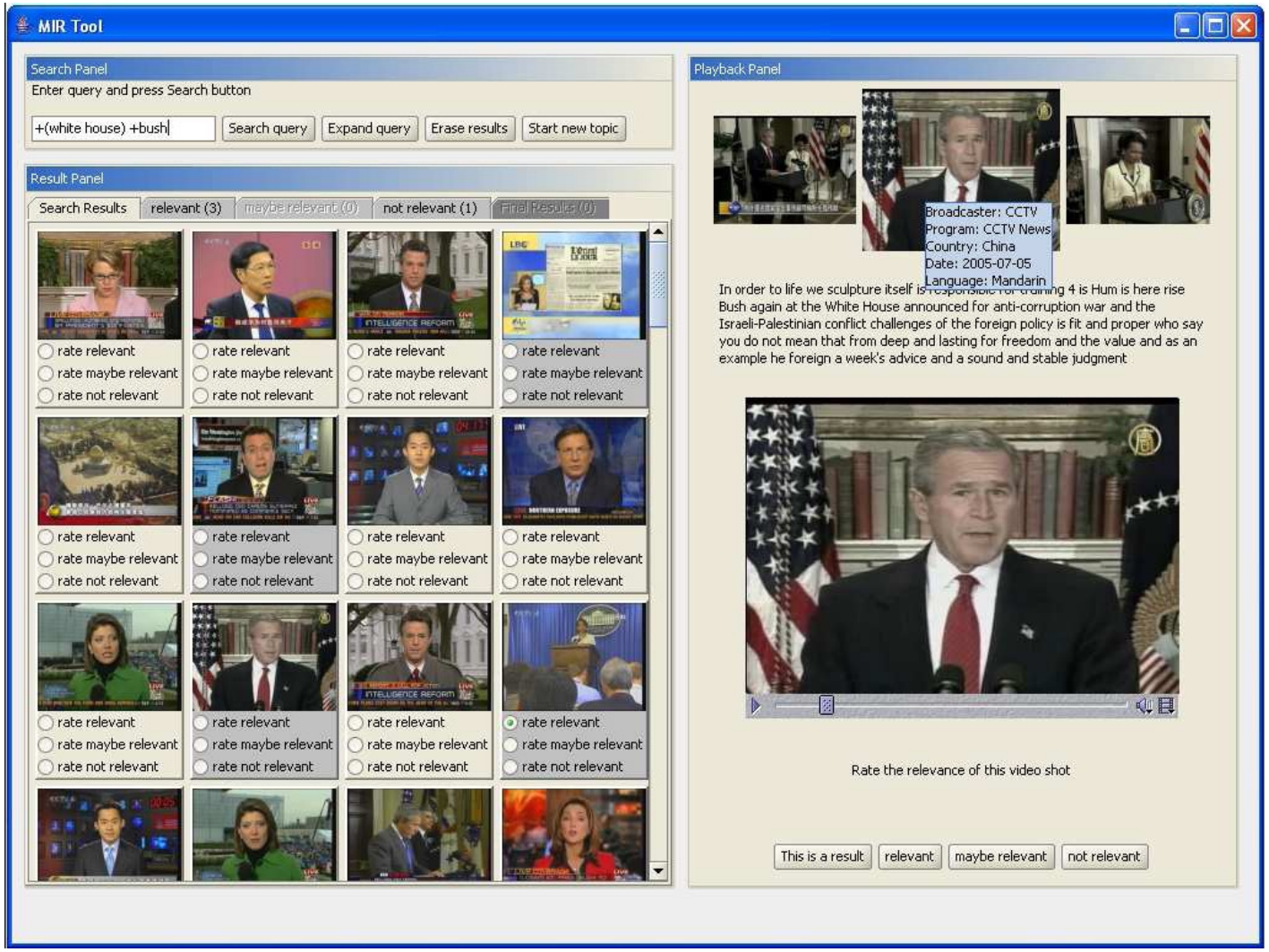
Storyboards



Peter Wilkins, Raphael Troncy, Martin Halvey, Darragh Byrne, Alia Amin, P. Punitha, Alan Smeaton, Robert Villa. User Variance and its Impact on Video Retrieval Benchmarking ACM International Conference on Image and Video Retrieval (CIVR 2009), July 2009

Examples search interfaces

- Screenshot of the Multimatch Video Interface (with selected video document already loaded)
  - Carmichael, J., Larson, M., Marlow, J., Newman, E., Clough, P., Oomen, O., and Sav, S. (2008), Multimodal Indexing of Digital Audio-visual Documents: A Case Study for Cultural Heritage Data, In Proceedings of the Sixth International Workshop on Content-Based Multimedia Indexing (CBMI2008), London, UK, 18-20th June, pp. 93-100.



J. S. Urban, X. Hilaire, F. Hopfgartner (2006) “Glasgow University at TRECVid 2006”, TREC Video Retrieval Evaluation Notebook Papers, www-nlpir.nist.gov/projects/tvpubs/tv.pubs.org.html

Challenges

universität koblenz
weiter:denken

- Similar to those for image retrieval
- Robust shot boundary detection
  - Combing audiovisual cues (over low-level features alone)
- Story segmentation
  - Context and application-dependent
- High-level feature and semantic concept extraction
  - e.g. using the 1,000 concepts from LSCOM (Large Scale Concept Ontology for Multimedia)
- Intuitive search and browse interfaces
  - Personalising search

4. Summary

Summary

universität koblenz
weiter:denken

- Multimedia is everywhere, it's ubiquitous
  - People want access to it
- Multimedia retrieval is typically based on
  - Indexing associated metadata (often simpler)
  - Processing the content of the media (harder)
- There are many challenges in multimedia search
- Online applications provide limited interactivity
- Future work is moving towards:
  - Integrating text and visual features and improving user's experience
  - Improving automatic annotation of concept with high-level concepts