The LICHEN Framework: A new toolbox for the exploitation of corpora

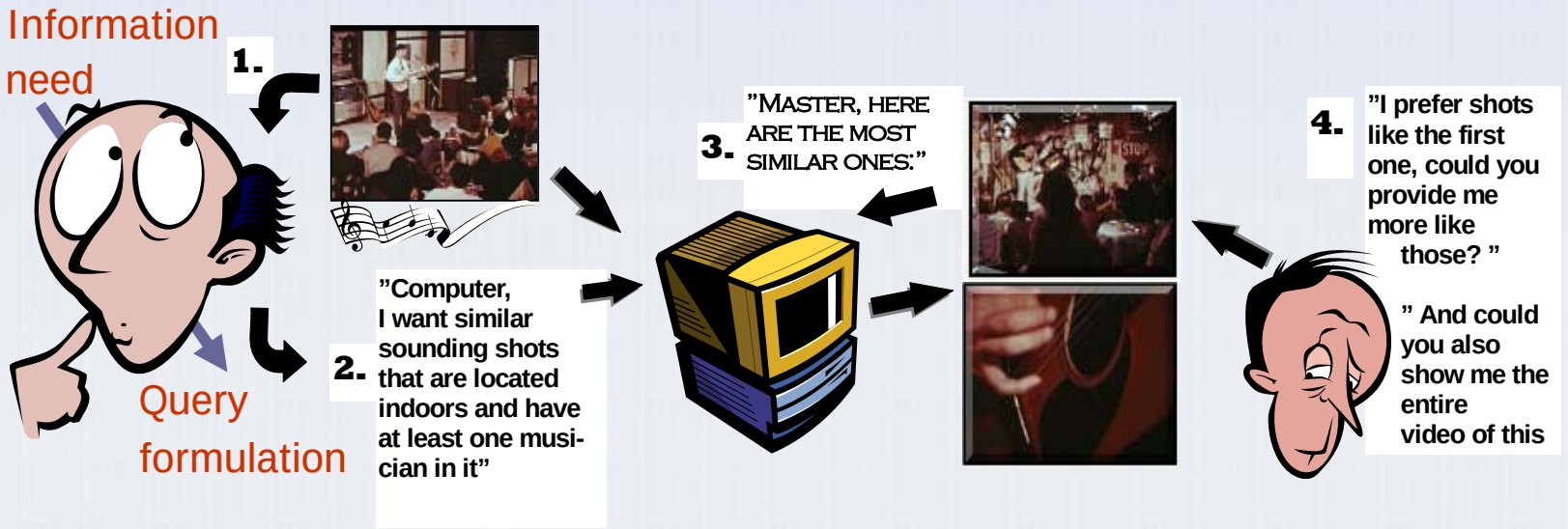
Lisa Lena Opas-Hänninen,
Tapio Seppänen, Ilkka Juuso and
Matti Hosio (University of Oulu, Finland)

Background

- **cultural inheritance is increasingly preserved in multiple media: text, images, speech, audio, graphics, animation etc.**
- **digitalization of content creation and storage devices produces an increasing amount of digital data for databases (digital convergence)**
- **need for database tools for accessing and analyzing the data**

Content-based information retrieval

- **Basic principle**
  - The user has a need for information and an illustrating example (text, image, video shot, sound)
  - The user formulates a query from the properties of the example
  - The retrieval system provides the user with hits that are supposed to be relevant
  - The user checks the hits and refines his query to get better hits

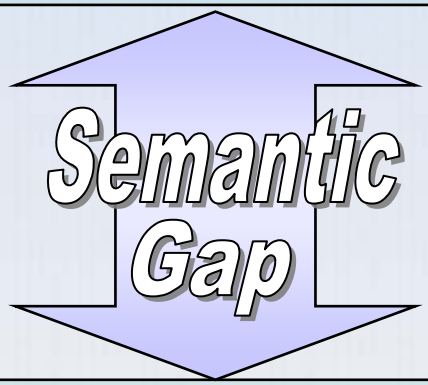


The semantic gap

Information need

High-level semantic concepts

objects, scenes, persons, actions, events, feelings



Retrieval engine

Tries to map data-driven low-level features to user-driven concepts.

- *feature computation (representation)*
- *inference, classifiers*
- *machine learning (modeling, feature fusion)*

Data: text, texture, shape, color, layout, motion

Low-level features

(Automatically computed) metadata

The problem of annotation

- **(Semi)automatic extraction of useful information from video content for the purpose of retrieval, browsing and indexing**
- **Required for training and testing search engines**
- **Why (semi)automatic instead of manual annotation?**
  - Sheer volume of data may render manual methods impractical
  - Manual methods are subject to personal interpretations
  - Manual methods are subject to human errors
- **Still, manual annotation is very important**
  - (Semi)automatic methods may not be robust enough

Document image retrieval

- Paper documents have been scanned and stored in DBs
- Search for documents with specific layout structures
- Search for documents containing specific text or markings
- Books, emails, poems, articles, etc.

Various document types

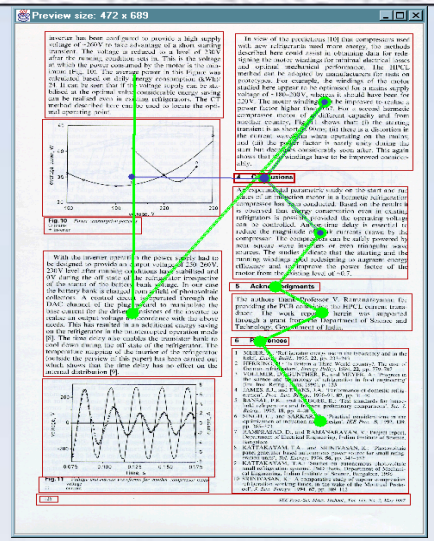
(a) Advertisement (b) Advertisement (c) Newsletter

(d) Article (e) Article (f) Book page

(g) WWW page (h) WWW page

Document image retrieval

- Image retrieval is based on:
  - subject
  - color
  - texture
  - date



- **Document retrieval is conventionally based on:**
  - text
  - subject
  - attributes

+

- **In IDIR documents can be queried <u>also</u> by:**
  - layout (position and size of objects in documents)

Effect of OCR errors on document retrieval

- **Optical Character Recognition (OCR)**
  - Conversion of document images to text
- **The TREC community experiments**
  - For OCR accuracy of <80%, not useful
  - For OCR accuracy of 80-95%, use enhanced IR
    - *Filtering of noise, approximate string matching, fuzzy methods, OCR confusion statistics, n-gram*
  - For OCR accuracy of 95-100%, most IR work fine

Arbitrary image retrieval

- **digital images of ...**
  - cultural content: shamans, Lapps, ceilidh, herd of reindeer, whisky stillpots, santa claus, fishing, sauna, etc.
  - other tourism-related images
  - ads, illustrations, etc.

Image

- color
- shape
- texture
- objects
- layout

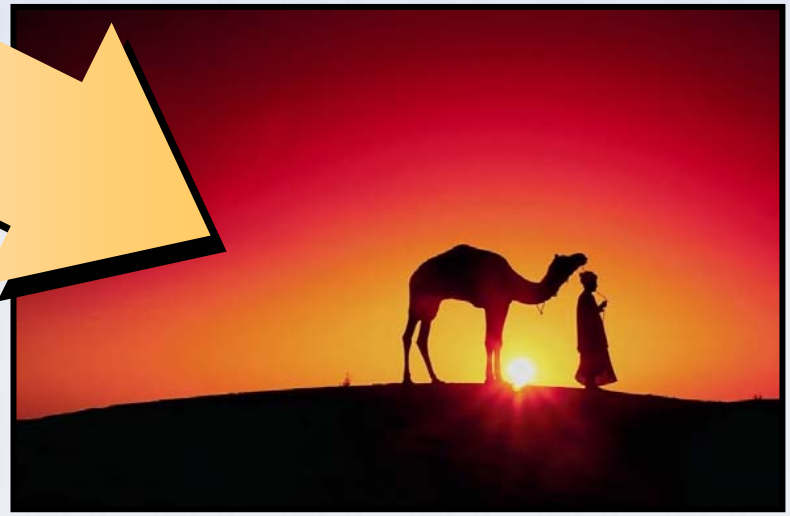
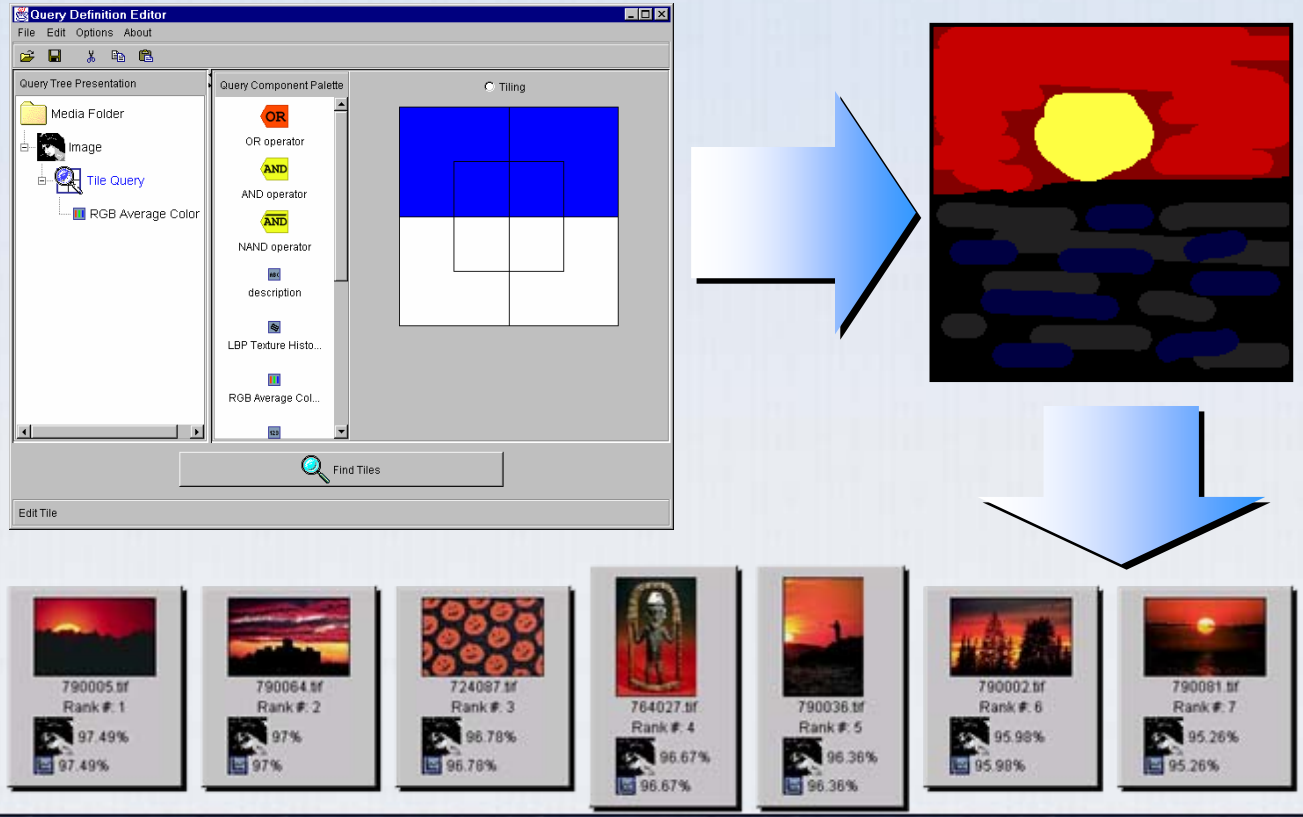


Image Retrieval System

- **Searching with content-based search interfaces, flexible search trees, sketch based retrieval, example-based search, fast indexing, and similarity metrics**



Audio and speech retrieval

- **interviews, conversations, TV broadcasts, speeches, etc.**
- **search for instances of words, utterances, expressions,...**
  - *“wonderful”*, “*yeah”*
- **play the sounds while displaying the accompanying textual transciption or images/video**
  - do smiles always indicate happiness?
  - does a knotted brow always indicate puzzlement?
- **samples of environmental sounds, such as from nature or animals**

Audio and speech

Audio

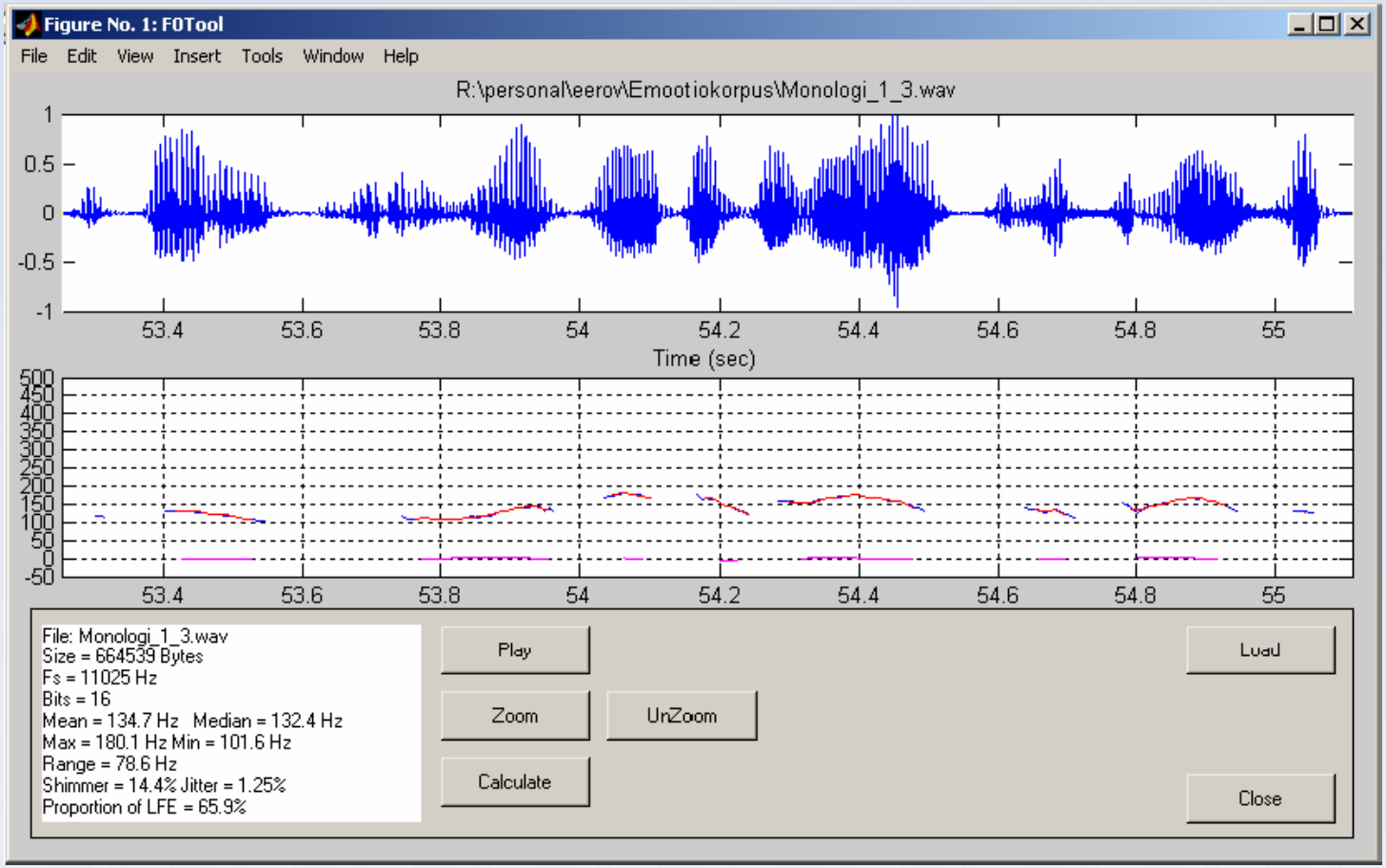
- speech
- music
- other sounds

- speech-driven UI
- simultaneous interpretation
- audio material indexing

- music classification
- music storage and search
- voice-sample based queries

- voice-effect libraries
- effect-based video categorization

Prosodic analysis tools



Video retrieval

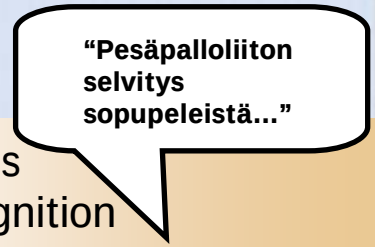
- **TV broadcasts, political speeches, videoed events, etc.**
- **Search for specific videos or video shots**
- **Display hits and their metadata or interpretations**
  - gestures, facial expressions (eg. political speeches)

Video and movies

Video

- auditory information
- static and dynamic visual information
- spoken information

- voice analysis
- speech recognition



- image analysis
- text analysis
- fusion techniques

- key frames
- shots
- scenes
- time-dependency

- video analysis
- media asset management
- activity recognition

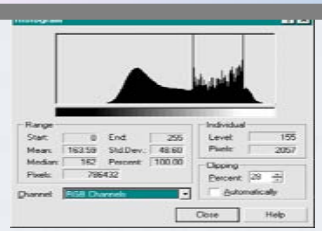
Data abstraction levels

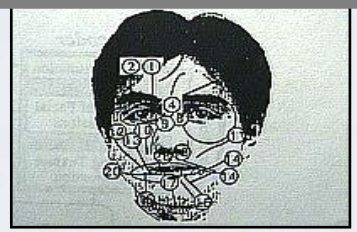
Raw data sequence

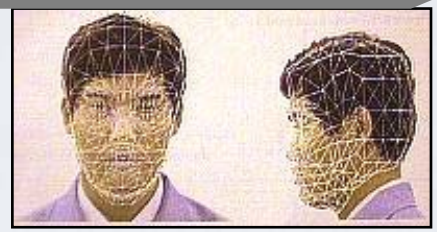
Low-level components	Intermediate-level components	High-level components
• atomic • pixel colors (image) • frames (video) • spectrum (speech)	• regional • temporal • segmented (image) • shots (video) • phonemes (speech) • audio types: speech/music	• meaning • objects (image) • scenes (video) • words, clauses (speech)

- **Semantics**
- **Objects**

Automatic | **Semi-automatic** | **Semi-automatic or manual**





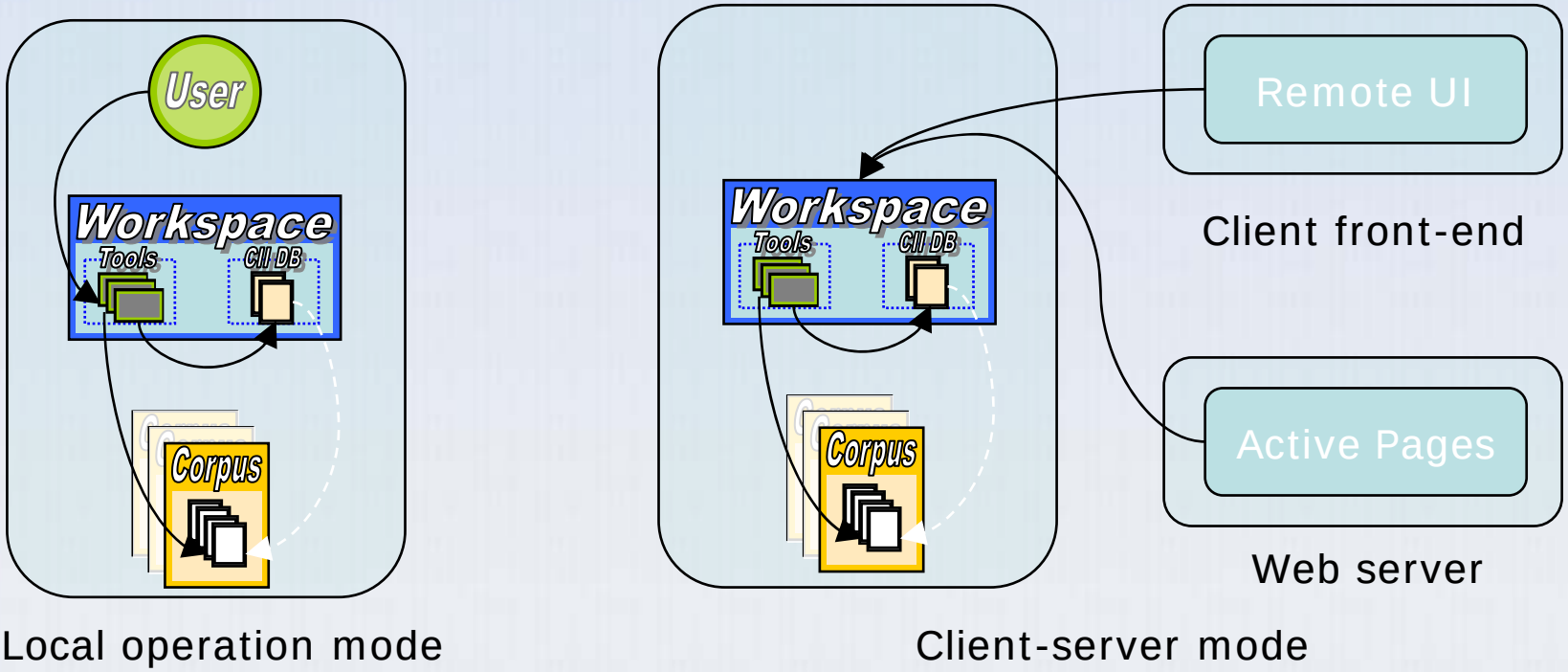


Query examples

- **Find shots of Condoleeza Rice**
- **Find shots of Iyad Allawi, the former prime minister of Iraq**
- **Find shots of Omar Karami, the former prime minister of Lebannon**
- **Find shots of Hu Jintao, president of the People's Republic of China**
- **Find shots of Tony Blair**
- **Find shots of Mahmoud Abbas, also known as Abu Mazen, prime minister of the Palestinian Authority**
- **Find shots of a graphic map of Iraq, location of Bagdhad marked - not a weather map**
- **Find shots of tennis players on the court – both players visible at same time**
- **Find shots of people shaking hands**
- **Find shots of a helicopter in flight**
- **Find shots of George Bush entering or leaving a vehicle, e.g., car, van, airplane, helicopter, etc - he and the vehicle both visible at the same time.**
- **Find shots of something (e.g., vehicle, aircraft, building, etc) on fire with flames and smoke visible**
- **Find shots of people with banners or signs**
- **Find shots of one or more people entering or leaving a building**
- **Find shots of a meeting with a large table and more than two people**
- **Find shots of a ship or boat**
- **Find shots of basketball players on the court**
- **Find shots of one or more palm trees**
- **Find shots of an airplane taking off**
- **Find shots of a road with one or more cars**
- **Find shots of one or more tanks or other military vehicles**
- **Find shots of a tall building (with more than 5 floors above the ground)**
- **Find shots of a goal being made in a soccer match**
- **Find shots of an office setting, i.e., one or more desks/tables and one or more computers and one or more people**

The LICHEN Framework

- **Architecture supports both stand-alone (local) and client-server (over the network) operation modes**



Digital rights management

- **License servers**
- **Cryptography-based protection of data**
- **Digital watermarking of images, audio, speech, videos**