The MOVIMOS Multimedia Search Engine

Thomas M. Breuel
DFKI & U. Kaiserslautern

MOVIMOS

Query

Results

moi=c01e9d4b0014aae85fe6f32206876083

Distance: 0.0345801411051

moi=f5aaf11a1ce29cf2b44710a2d55626b9

Distance: 0.0603425964865

moi=c7254efb2f5271d46c64cef6c76685b8

Distance: 0.0611235638565

moi=15050cd3274d6f925c96db0f38ac5ed1

Distance: 0.0611709013685

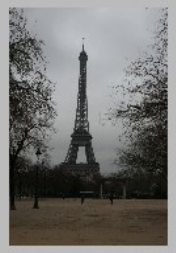
moi=2cf596e6e9e8668a82f5d58ceae4cb99

Distance: 0.0679455125112

moi=bbc7c285dd7a4ca40fcca85e072fa736

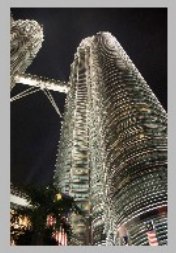
Distance: 0.071544180458

moi=90b1b35ea392250aa3e020790c2048b8



Distance: 0.0757749417915

moi=224fabec2f739022a16fefbfa1ea9591



Distance: 0.0793789953514

moi=fae8d138d066fddbfd6fc7addbe4a04b

Distance: 0.0803499452054

moi=ea745c716162576bafbcdd3b8c4b919d

Distance: 0.133241150515

User Interfaces

E.g.: Navidgator: similarity-based hierarchical navigation through image/video query results

MOVIMOS target applications

- **InViRe, TagMyDuck, OCRosearch, ...**

- **driven by IUPR applications**
  - camera-based search for mobile phones
  - mobile augmented reality
  - image/video search engines
  - digital forensics
  - automated pornography filtering
  - book and OCR search

Retrieval Systems

- First image DB system at MIT (1980)
- IBM QBIC (1995), multidim. indexing, closed source
- Viper/GIFT (1999), open source, large hw-requ.
- PicSOM (2002), SOM-based index, not available
- Cortina (2004), scales > 1 mio imgs, not available
- INRIA LEAR group (2004 ...)
- IUPR group (2004 ...)
- FIRE (2005), open source, monolithic architecture

- … and many more

history

- fast, parallel model-based indexing (1989)
- appearance-based 3D recognition (1992)
- QBIC (1995)
- personalized web search (1998)
- TagMyDuck, MOVIMOS ... (2004-)
- OCRopus (2006-)

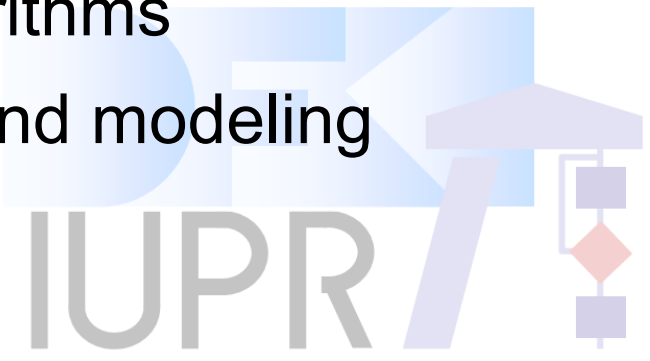
MOVIMOS technical goals

- **architectural**
  - full multimedia support
  - easy extensibility (new features, searches)
  - dynamic database updates
  - scalability (fast indexing + distribution)
- **functional**
  - standard CBIR functions
  - tagging and categorization
  - semi-supervised learning
  - context dependent search
  - personalized search

architectural

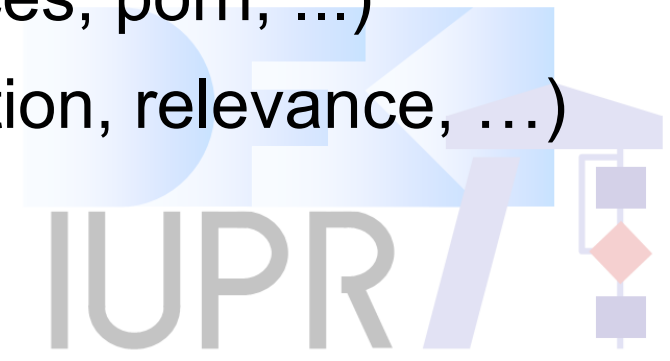
MOVIMOS multimedia support

- **content types**
  - images
  - video
  - audio
  - text
  - lattices
- **resulting application requirements**
  - open-ended set of format, features, algorithms
  - very data intensive: distributed storage and modeling
  - result integration, context modeling



MOVIMOS extensibility

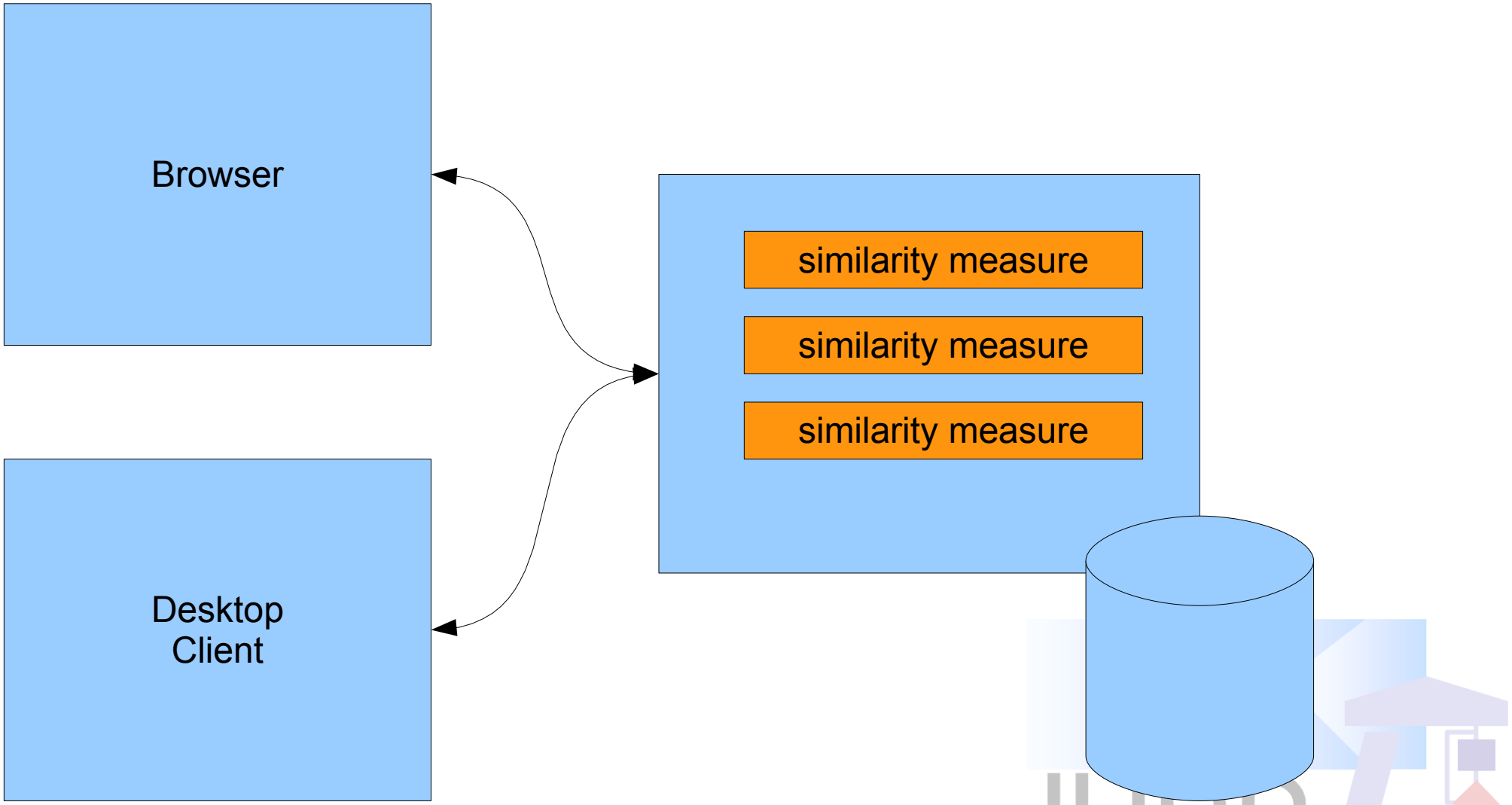
- **tools, architecture**
  - Python as glue code (“component architecture”)
  - prototyping in NumPy/SciPy (≈ Matlab)
  - easy access to native code for speed
  - CherryPy, REST for distribution / services
- **functionality**
  - standard CBIR primitives, operators (faces, porn, …)
  - new IUPR functionality (context, adaptation, relevance, …)
  - text, OCR plugins



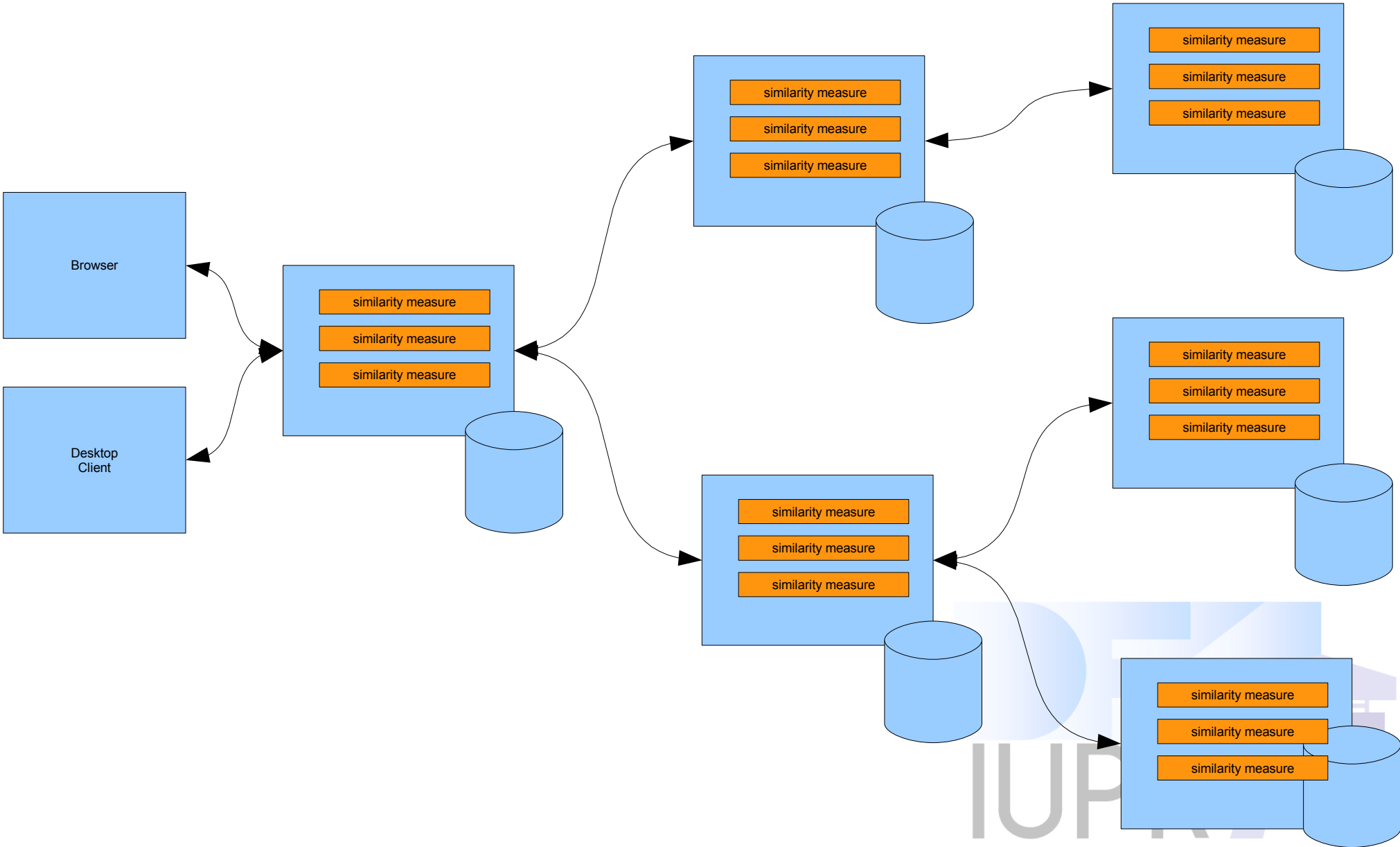
MOVIMOS scalability

- **fast indexing at each node**
  - index data structure + sublinear lookup
  - e.g.: bit vectors, inverted indexes
  - supported at MOVIMOS nodes
  - optional distributed index creation
- **distributed search**
  - motivation: some similarity measures hard to speed up
  - e.g.: geometric match verification, context-dependent simil.
  - supported between MOVIMOS nodes
  - support for multiple topologies
  - simple REST-based APIs

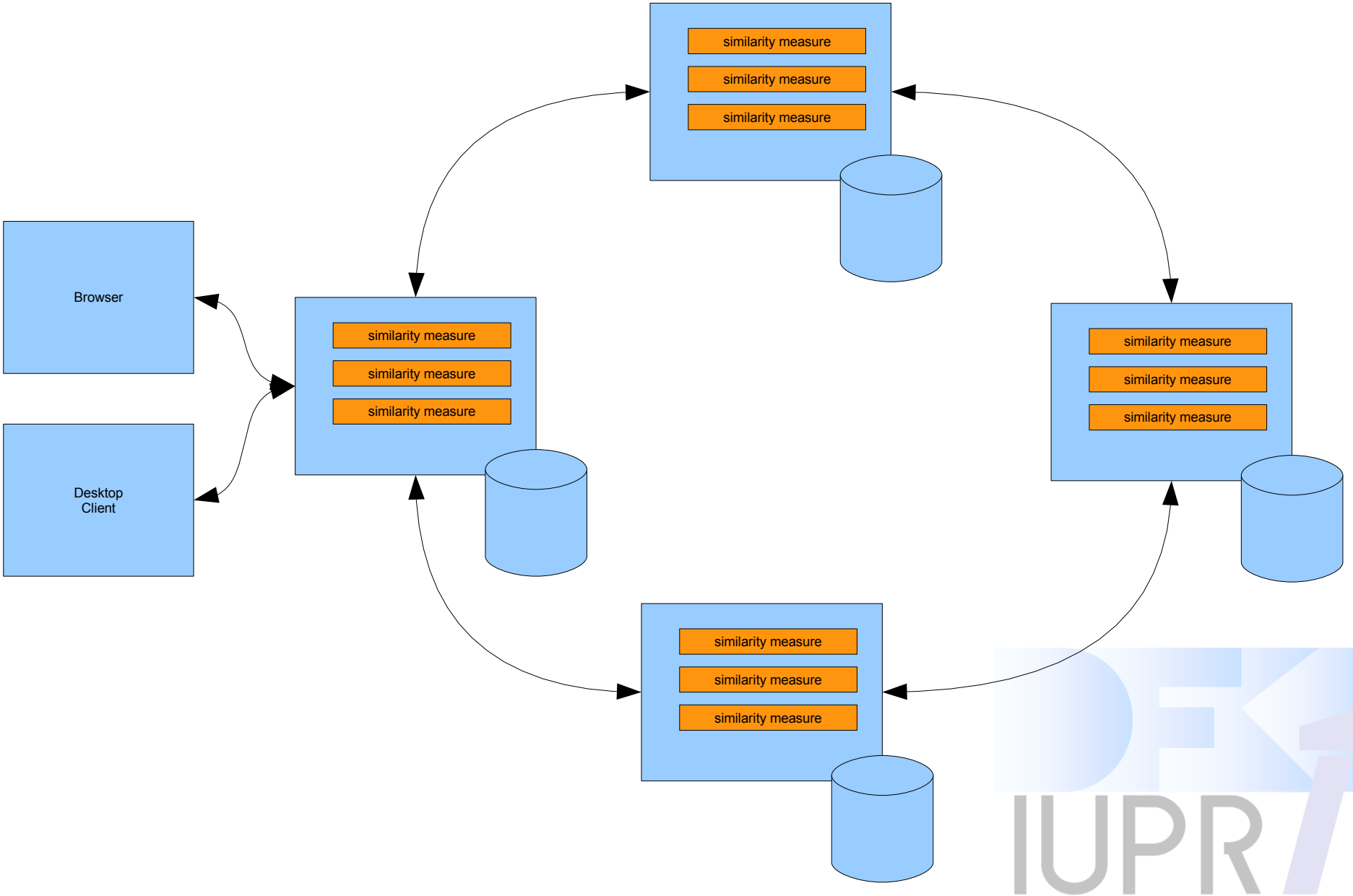
single node configuration



tree configuration



P2P configuration



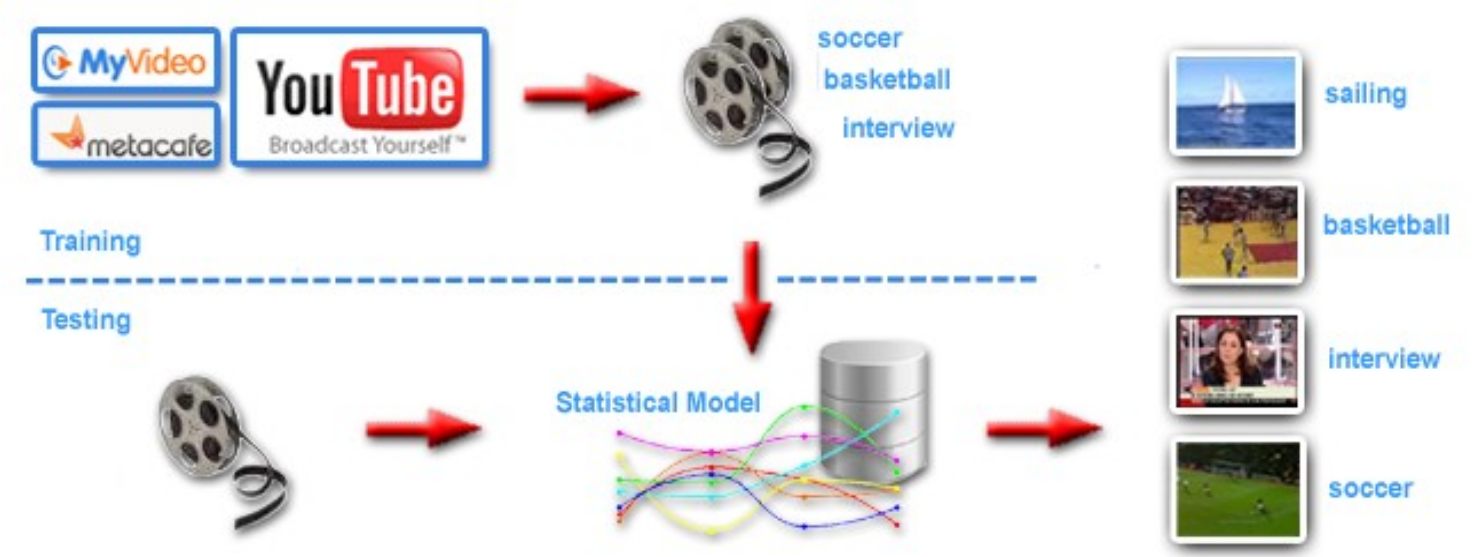
technology

automated tagging / categorization

- **goal**
  - assign descriptive tags to images / videos
- **applications**
  - search / categorization / personalized content delivery
- **challenges**
  - visual diversity of tags
  - many thousands of categories
  - lack of training data
  - context/user dependence

automated tagging / categorization

- **common approach**
  - build corpora, then train
- **our approach**
  - autonomous learning from the web (YouTube, Flickr)
  - using web tags as (noisy) ground truth



semi-supervised visual learning

- **challenges**
  - web tags are coarse, unreliable, and subjective
  - web datasets contain “non-relevant” parts (noise)
  - training automatic taggers on this material is difficult

semi-supervised learning II

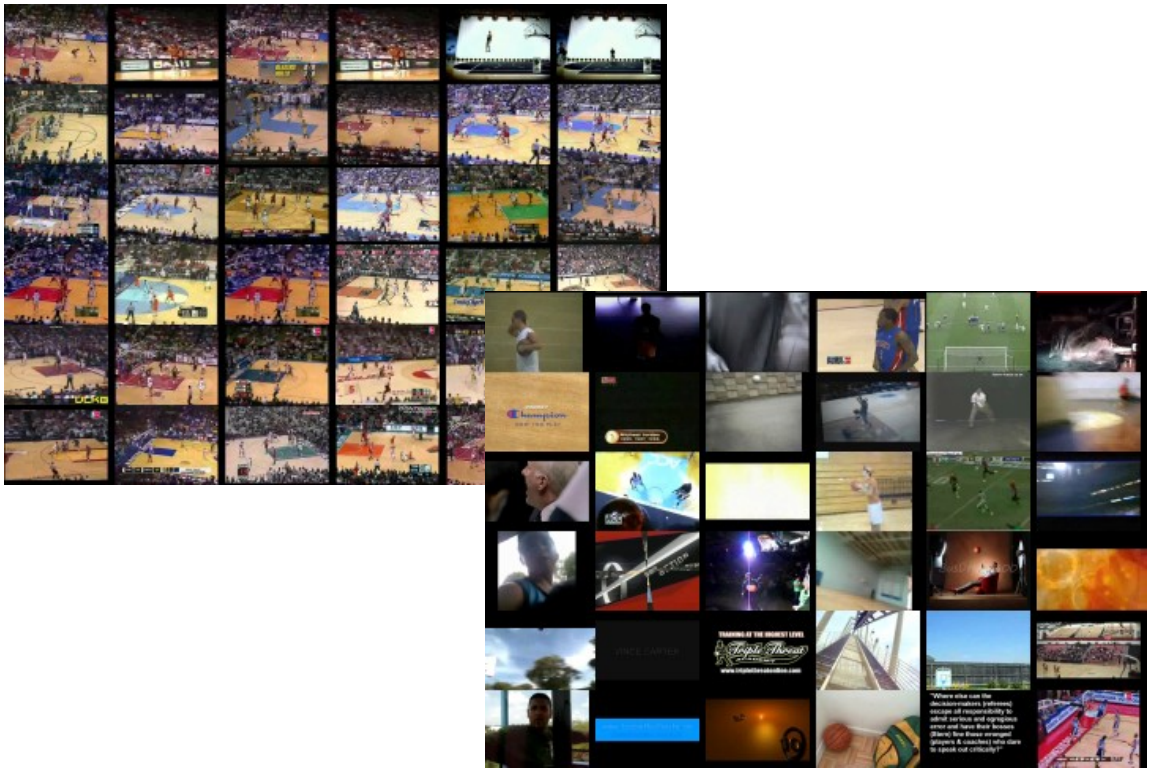
- **approach**
  - filter non-relevant content as outliers during training
  - model distributions of relevant and non-relevant content
  - parameterized kernel density estimators
  - β_i = feature i is relevant

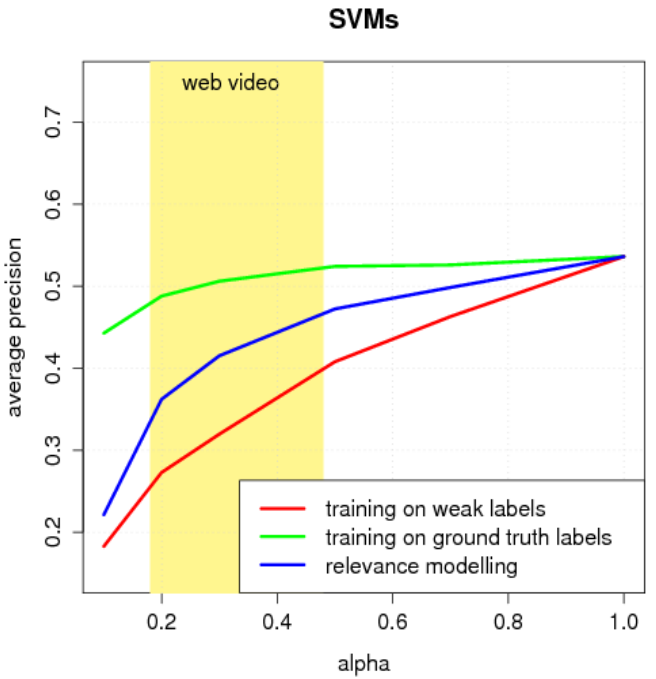
$$p_\beta^1(x) = \frac{1}{Z} \cdot \sum_{i=1}^{n} \beta_i \cdot K_h(x; x_i),$$

$$p_\beta^0(x) = \frac{1}{Z'} \cdot \sum_{i=1}^{n} (1 - \beta_i) \cdot K_h(x; x_i),$$

semi-supervised learning III

- automatically disregards irrelevant content
- improves tagging / categorization
- additional approach: motion segmentation





style / context / user adaptation

- **picture context / style**
  - pictures taken over the same trip / event
  - pictures taken by the same user
  - video frames from the same show / movie
  - users tag differently
  - queries have different objectives

- **solution**
  - adapt classifiers to context / style

style modeling

- **style modeling**
  - previously used in OCR / handwriting recognition
- **application to image tagging**
  - extend image annotation with a latent style variable
  - tags t, visual words v, style s
  - improves tagging significantly
  - best result to date on COREL-5K benchmark

$$P(t|d,s) = \sum_{z \in Z} P(t|z,s) \cdot P(z|d)$$

$$P(v|d,s) = \sum_{z \in Z} P(v|z,s) \cdot P(z|d)$$

summary

MOVIMOS

- **new, flexible, distributed platform**
  - images, video, text, lattices
  - standard CBIR, VQ, indexing, matching, verification
- **state-of-the-art technologies, e.g.**
  - categorization, tagging, rretrieval
  - context, style modeling
  - semi-supervised learning
- **research platform**
  - open standards (Python, NumPy, REST, etc.)
  - open source release planned for Fall 2009

papers, demos, links

www.iupr.com