
# The Cross Language Image Retrieval Track: ImageCLEF 2006

**Paul Clough[1], Henning Müller[2], Thomas Deselaers[3], Michael Grubinger[4], Thomas M. Lehmann[5],Allan Hanbury[6] Bill Hersh[7]**

[1]Sheffield University, UK
[2]University and Hospitals of Geneva, Switzerland
[3]RWTH Aachen University, Computer Science Dep. Germany
[4]Victoria University, Australia
[5]RWTH Aachen University, Medical Informatics, Germany
[6]Vienna University, Austria
[7]Oregon Health Science University


# Overview

- General overview
  - Image retrieval and CLEF
- Tasks in 2006
  - Photographic image retrieval task
  - Medical image retrieval task
  - Medical image annotation and object classification tasks
- Summary
  - Overall achievements


# Image Retrieval and CLEF

- Cross-language image retrieval
  - Images often accompanied by text
    - can be annotated automatically
- Began in 2003 as pilot experiment
  - In 2006, 47 groups registered from 18 countries
- Aims of ImageCLEF
  - Promote & initiate research for cross language image retrieval
  - Further understanding of the relationships between multilingual texts and images for retrieval
  - Create useful resources for the research community

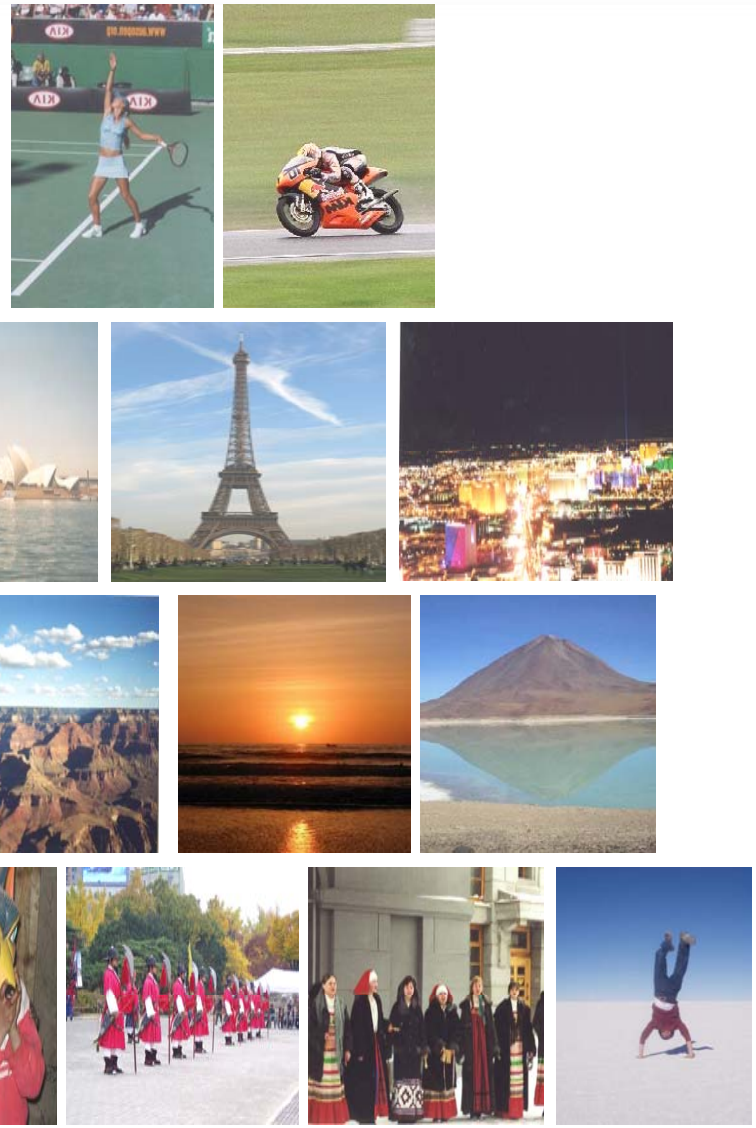

# Photographic Retrieval Task

# Photographic Retrieval task

- New collection of 20,000 photographs (mainly) from an independent German travel company (viventura)
  - IAPR-TC12 Benchmark
  - Replaces St Andrews collection of historic photographs
- Set of 60 topics for ad-hoc retrieval task
  - Based on realistic topics (log-file analysis and interviews)
  - Based on testing various types of linguistic and pictorial attributes, e.g. visual vs. semantic, specific vs. general objects and use of proper names
- Submissions categorised by
  - Manual or automatic
  - Source and target languages (English/German)
  - Use of visual and text features (modality)
  - Use of feedback (e.g. for query expansion)

# Image Collection

- 20,000 colour photographs
  - Wide variety
  - Global scope
- Accompanied by semi-structured captions
  - English and German
- Many images have similar visual content but varying
  - illumination
  - viewing angle
  - background



# Images and captions

- Created as a resource for evaluation

keywords describing image, location, and date as supplied by photographer

description of image supplied by benchmark authors

```
<DOC>
<DOCNO>annotations/16/16019.eng</DOCNO>
<TITLE>Flamingo Beach</TITLE>
<DESCRIPTION> a photo of a brown sandy beach; the dark blue sea with small
breaking waves behind it; a dark green palm tree in the foreground on the
left; a blue sky with clouds on the horizon in the background;
</DESCRIPTION>
<NOTES> Original name in Portuguese: "Praia do Flamengo"; Flamingo Beach is
considered as one of the most beautiful beaches of Brazil; </NOTES>
<LOCATION>Salvador, Brazil</LOCATION>
<DATE>2 October 2002</DATE>
<IMAGE>images/16/16019.jpg</IMAGE>
<THUMBNAIL>thumbnails/16/16019.jpg</THUMBNAIL>
</DOC>
```

# Example topic

- 60 topics created
  - Balance between realism and controlled parameters
- Varied according to
  - Whether derived directly from log file
  - Whether containing geographical constraint
  - Notions of “visualness” and linguistic complexity
  - Completeness of annotations
- Titles translated into 15 languages

```
<top>
<num> Number: 1 </num>
<title> accommodation with swimming
pool </title>
<narr> Relevant images will show the
building of an accommodation facility
(e.g. hotels, hostels, etc.) with a
swimming pool. Pictures without
swimming pools or without buildings
are not relevant. </narr>
<image> images/03/3793.jpg </image>
<image> images/06/6321.jpg </image>
<image> images/06/6395.jpg </image>
</top>
```

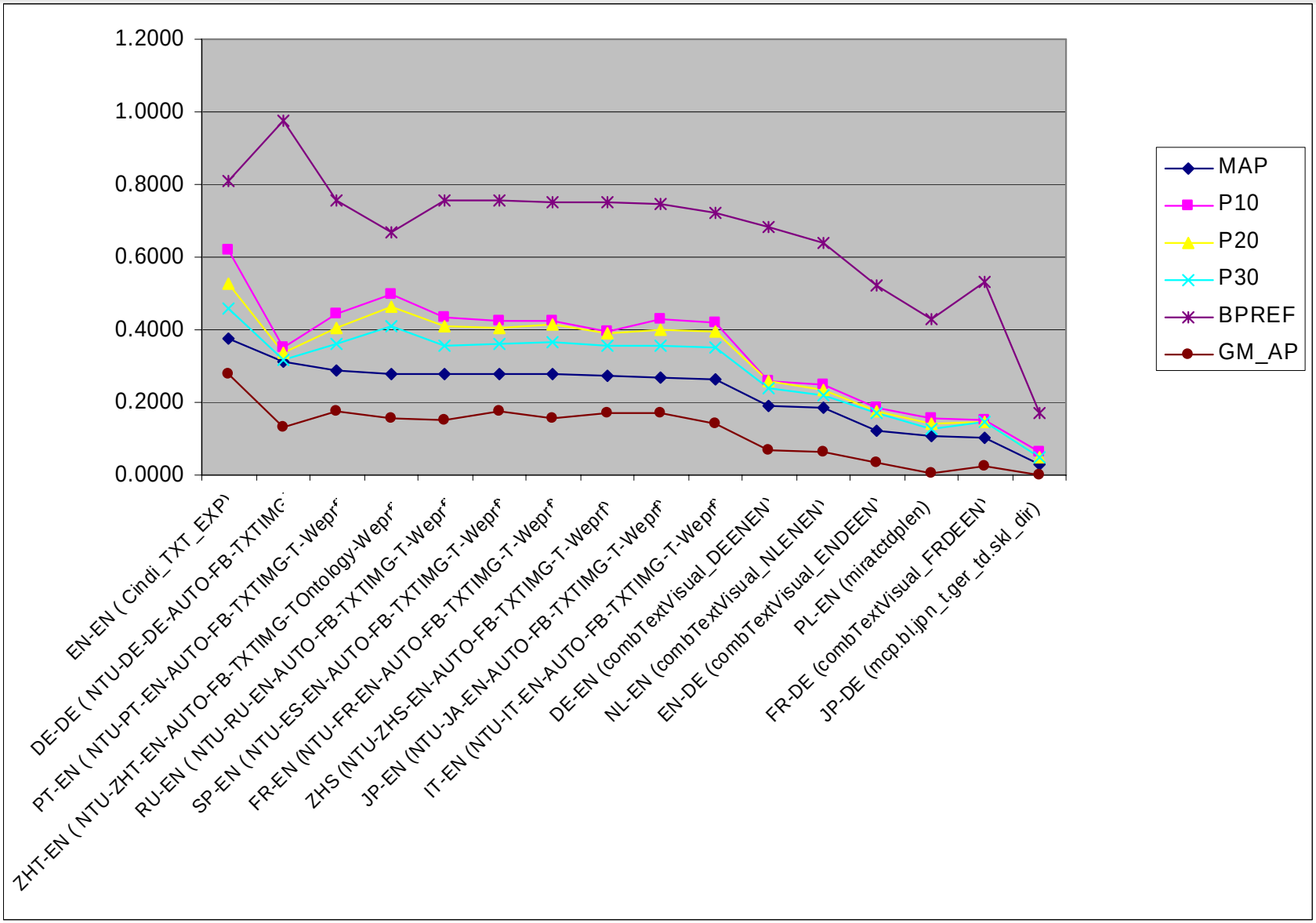


# Registration and Participation

- 36 groups registered
  - 12 participated (5 new)
- 157 runs submitted
  - variety of approaches
- 59% runs bilingual
  - 85% X-English
  - 15% X-German
- 31% runs involved use of image retrieval
  - 27% mixed text and visual
- 46% runs with feedback
- 1 manual run

| | | |
|---|---|---|
| NTU | National Taiwan U. - NLP | Taiwan |
| **Berkeley** | UC-Berkeley-Inf.Management | USA |
| **CINDI** | Concordia U. | Canada |
| Daedalus | Daedalus Consortium | Spain |
| Jaen | U.Jaen-Intell.Systems | Spain |
| **RWTH-Comp.Sci** | RWTH Aachen U. - comp-sci. | Germany |
| **CELI** | CELI- srl, Torini | Italy |
| DCU | Dublin City U.-Computing | Ireland |
| NII-1 | Nat.Inst.Informatics (Testbeds) | Japan |
| IPAL-I2R | Inst.for Infocomm Research | Singapore |
| **TU Chemnitz** | TU Chemnitz - Comp.Sci | Germany |
| LIC2M-CEA | Centre CEA Saclay | France |

# Results – automatic
## *highest MAP per language pair*



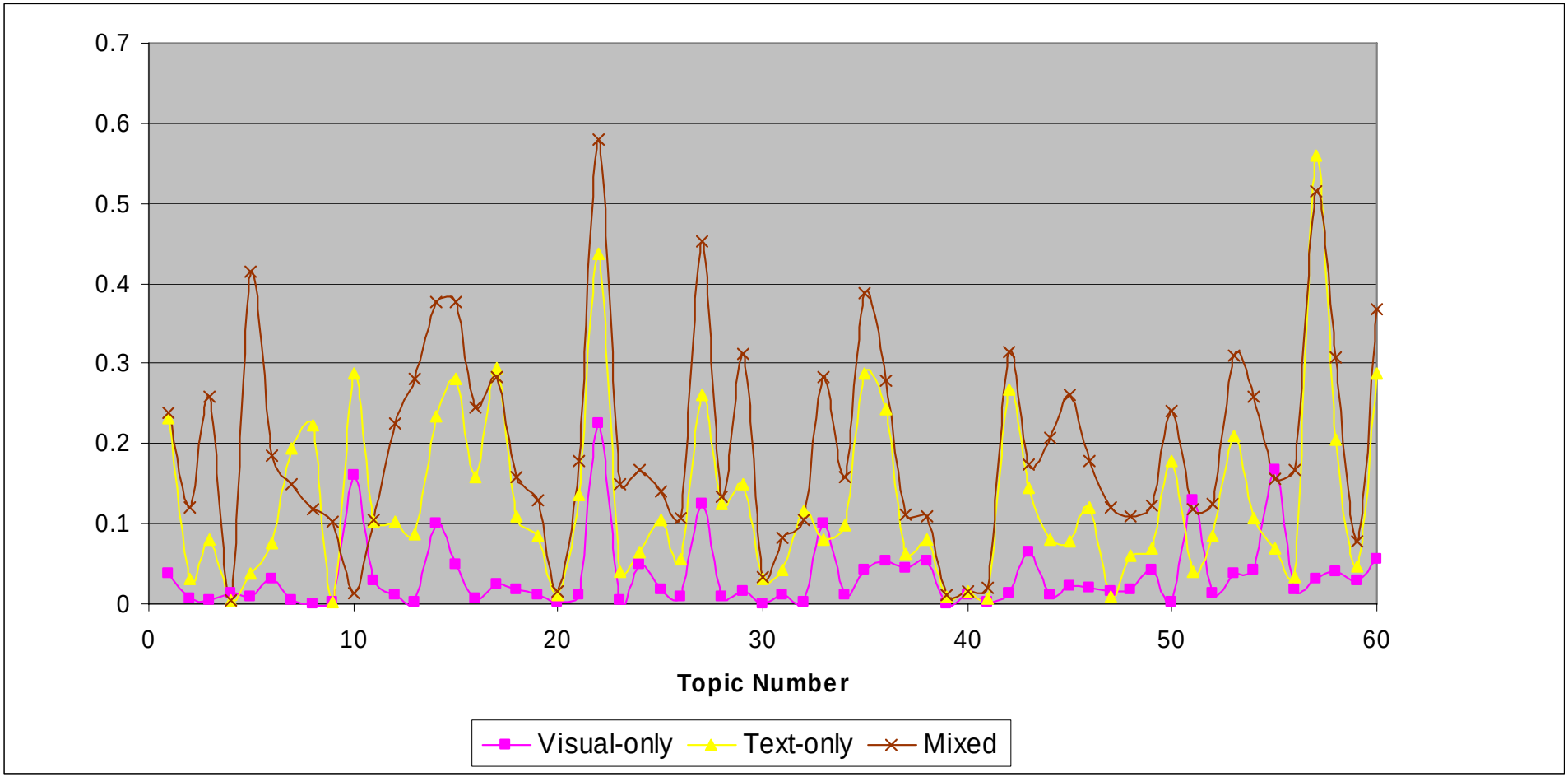
# Results – automatic

- For *highest* MAP per language pair
  - English-English 19% higher than German-German
    - particularly noticeable for languages with limited language resources (e.g. JP-EN)
  - Highest X-English 74% monolingual
    - PT-EN
  - Highest X-German 39% monolingual
    - EN-DE
  - 83% use feedback
  - 82% use mixed visual and textual features

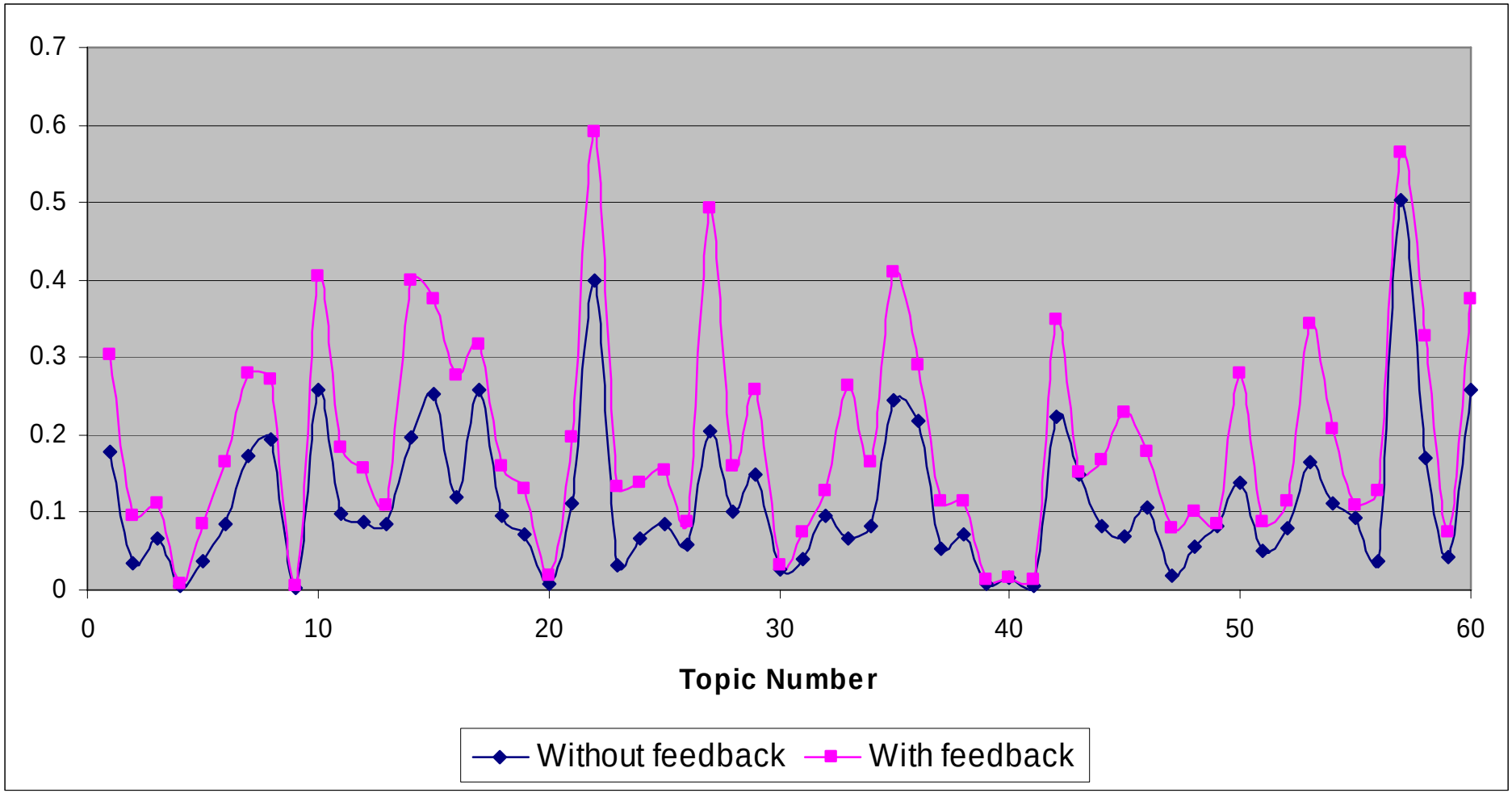

# Average MAP results

- Combining visual features from image and semantic information from text
  - on average 54% improvement over text alone
- Feedback (generally in form of PRF)
  - on average 39% improvement with
- Bilingual retrieval performs
  - 7% lower than monolingual
- Target language
  - Results for English annotations are 26% higher than German

# Average MAP by topic
## *Visual-only, text-only, mixed/combined*



# Average MAP by topic
## *Without and with feedback*



# Visual topics subtask

- Aimed to evaluate success of Content-Based IR techniques on IAPR collection
- 30 topics selected from ad hoc and made more “visual”
  - Topics consisted of 3 example images only
- Two groups participated in this task
  - IPAL and RWTH
- Results for 30 topics are low
  - Highest MAP 0.1010 (RWTH)
  - Results better at higher precision
- Results may reflect difficulty of domain (generic)


# Summary

- New collection in response to feedback
  - Real-life collection
  - English *and* German captions
  - Benchmark collection and very flexible for research and evaluation
- Challenging for both text and visual retrieval
  - Limited amount of text for retrieval
  - Heterogeneous content
  - German target language
- Retrieval varies widely according to topic
  - But generally runs with relevance feedback and combination of visual and text retrieval perform best


# Medical Retrieval Task

# Medical Retrieval Task

- Same databases as in 2005
  - 50’000 images, four sets of annotations, partly in French, German, English
  - Large variety, 2005 data for training
- Topics based on survey and log file analysis (medical web media search)
  - Categories for visual, semantic and mixed retrieval
- Submissions rated for two axes
  - Interaction (automatic, manual, interactive), media used for retrieval (visual, textual, mixed)


# Registration and Participation

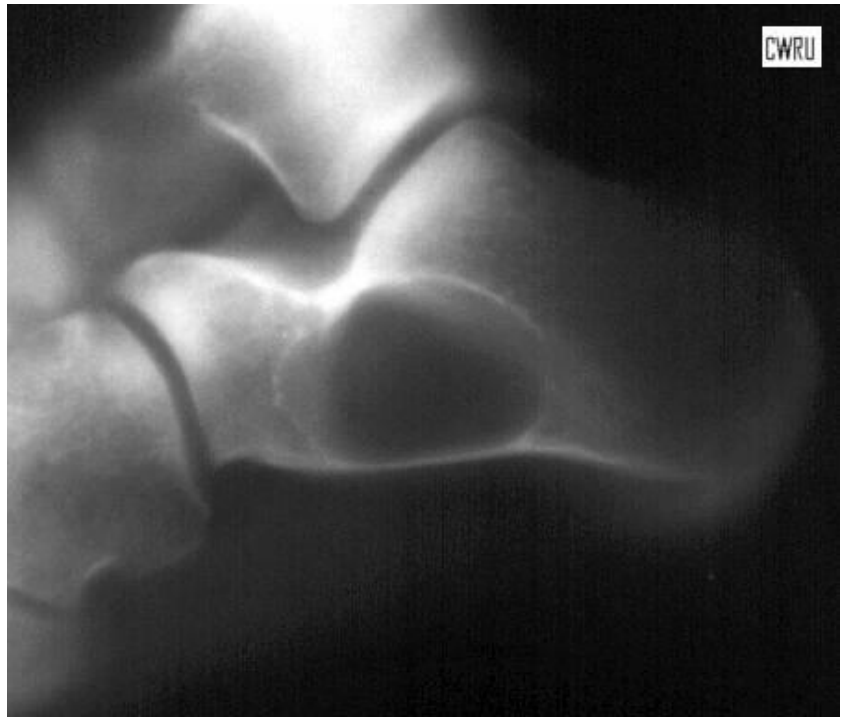
- 37 registered groups
  - 12 groups from 8 countries submitted results
  - Lack of time, too large databases, but useful!
- 101 runs submitted
- Large variety of techniques
  - Visual, textual, and combinations
- Most often automatic runs
- Textual and multimodal runs are most frequent, a few purely visual runs


# Participants

- CINDI group, Concordia University, Canada.
- Microsoft Research, China.  Microsoft research, China
- Institute for Infocomm Research I2R - IPAL, Singapore.
- University Hospitals of Freiburg, Germany
- Jaen University (SINAI), Spain
- Oregon Health and Science University (OHSU), USA.
- I2R Medical Analysis Lab, Singapore
- MedGIFT, University and Hospitals of Geneva, Switzerland
- RWTH Aachen University - Computer Science, Germany
- RWTH Aachen University - Medical Informatics, Germany
- State University New York, Buffalo, USA
- LITIS Lab, INSA Rouen, France.

# An example topic

3.6
Show me x-ray images of bone cysts.
Zeige mir Röntgenbilder von Knochenzysten.
Montre-moi des radiographies de kystes d'os.



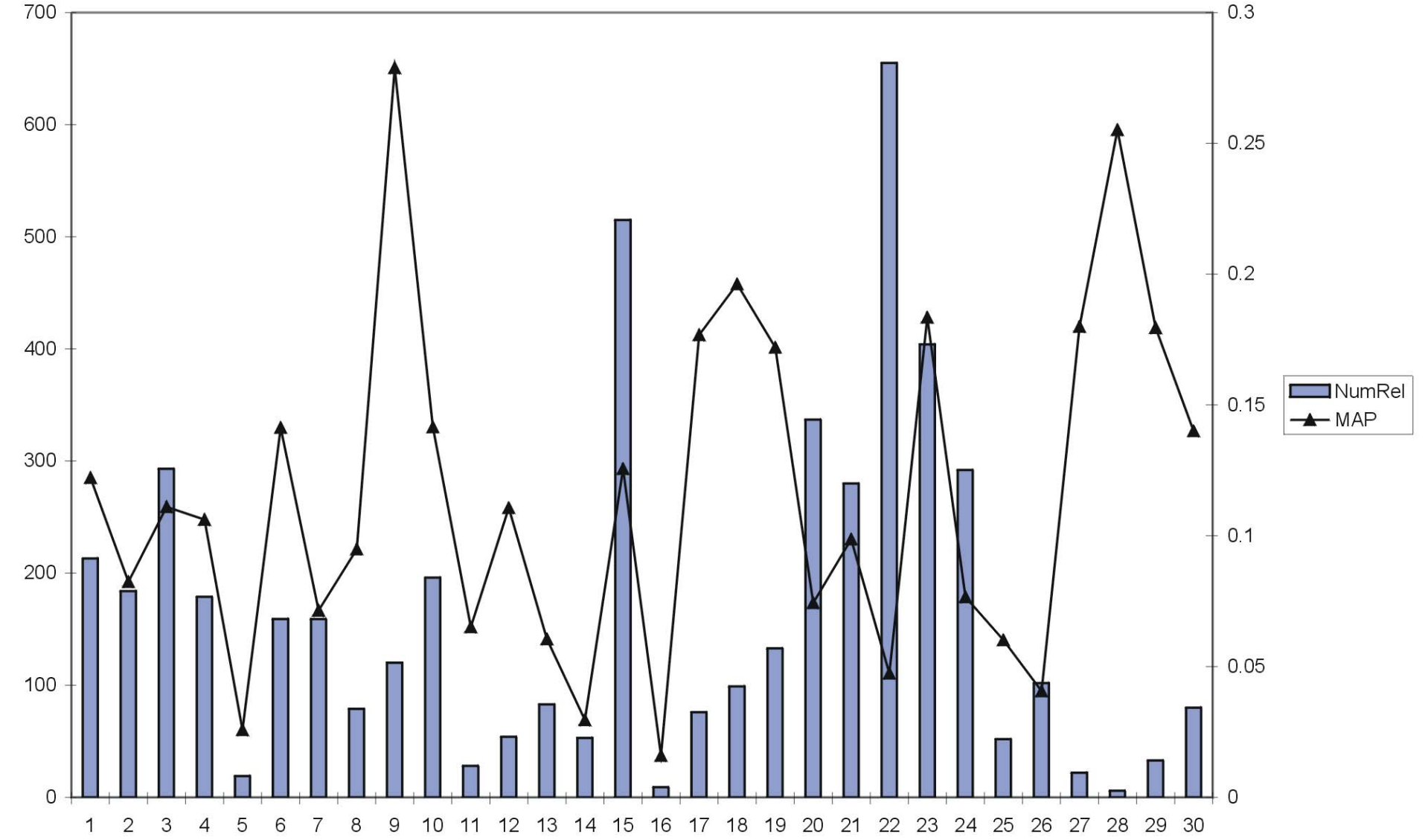

# Results - automatic

| Run | Topic | System | MAP | R-prec | B-pref |
|---|---|---|---|---|---|
| IPAL-IPAL_Cpt_Im.eval | automatic | mixed | 0.3095 | 0.3459 | 0.3922 |
| UB-UBmedVT2.eval | automatic | mixed | 0.2027 | 0.2225 | 0.2947 |
| RWTHi6-EnFrGePatches.eval | automatic | mixed | 0.1696 | 0.2078 | 0.2499 |
| GE_vt10.treceval.eval | automatic | mixed | 0.12 | 0.1703 | 0.1717 |
| SINAI-SinaiGiftT50L20.eval | automatic | mixed | 0.0467 | 0.095 | 0.1246 |
| UKLFR-UKLFR_mids_en_all_co.eval | automatic | mixed | 0.0167 | 0.0145 | 0.1568 |
| IPAL-IPAL_Textual_CDW.eval | automatic | textual | 0.2646 | 0.3093 | 0.354 |
| GE_8EN.treceval.eval | automatic | textual | 0.2255 | 0.2678 | 0.301 |
| UB-UBmedT1.eval | automatic | textual | 0.1965 | 0.2256 | 0.2881 |
| UKLFR-UKLFR_origmids_en_en.eval | automatic | textual | 0.1698 | 0.2127 | 0.2434 |
| RWTHi6-En.eval | automatic | textual | 0.1543 | 0.1911 | 0.2308 |
| OHSU_baseline_trans.eval | automatic | textual | 0.1264 | 0.1563 | 0.1827 |
| SINAI-SinaiOnlytL30.eval | automatic | textual | 0.1178 | 0.1534 | 0.2001 |
| cindi-CINDI_Fusion_Visual.eval | automatic | visual | 0.0753 | 0.1311 | 0.166 |
| MSRA_WSM-msra_wsm.eval | automatic | visual | 0.0681 | 0.1136 | 0.1551 |
| IPAL-IPAL_Visual_SPC+MC.eval | automatic | visual | 0.0634 | 0.1048 | 0.1398 |
| RWTHi6-SimpleUni.eval | automatic | visual | 0.0499 | 0.0849 | 0.1208 |
| GE-GE_gift.eval | automatic | visual | 0.0467 | 0.095 | 0.1246 |

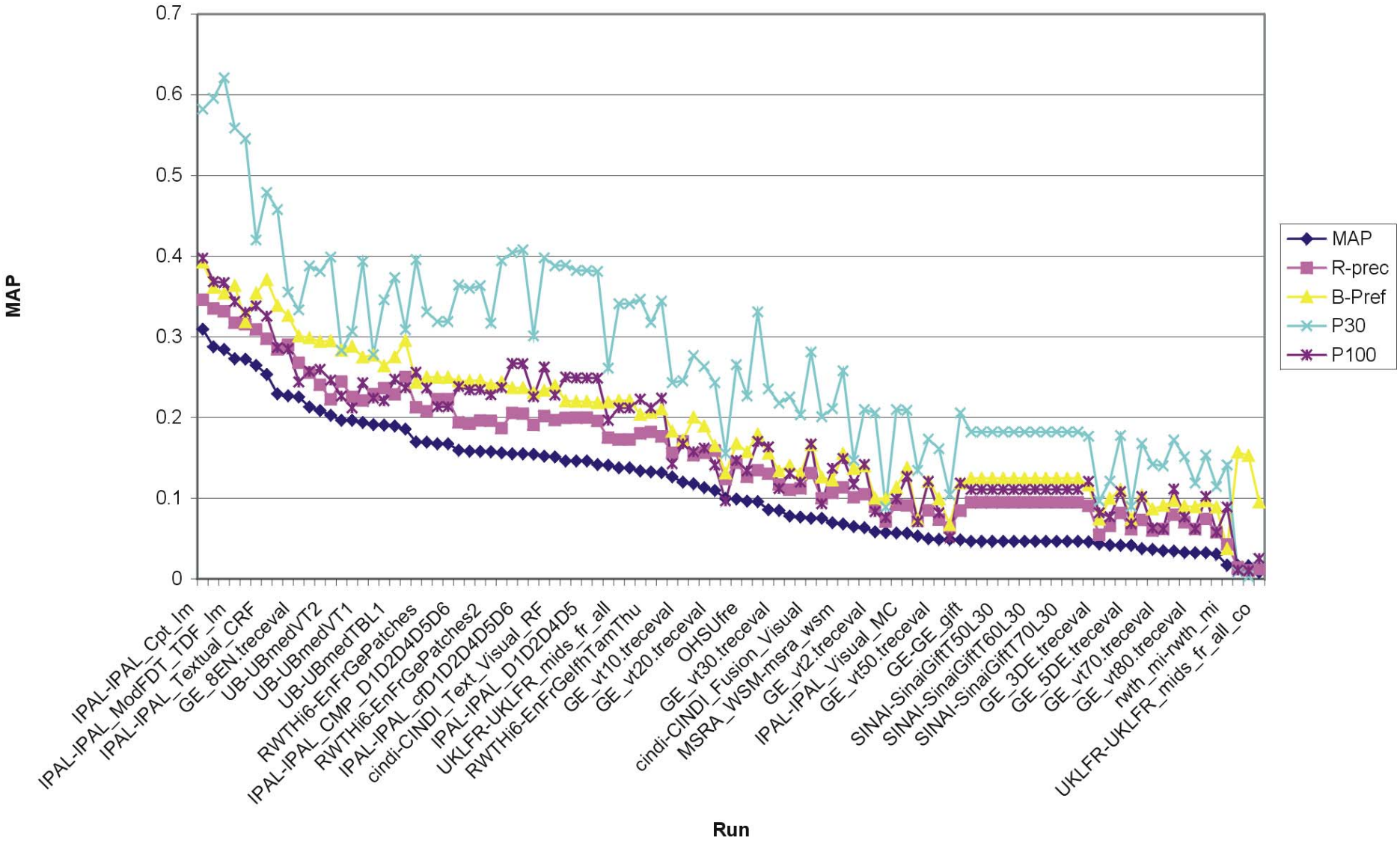
# Results (2) – manual, interaction

| | | | | | |
|---|---|---|---|---|---|
| OHSU-OHSU_m1.eval | feedback | mixed | 0.1563 | 0.187 | 0.2441 |
| cindi-CINDI_Text_Visual_RF.eval | feedback | mixed | 0.1513 | 0.1969 | 0.2397 |
| IPAL-IPAL_Textual_CRF.eval | feedback | textual | 0.2534 | 0.2976 | 0.3707 |
| cindi-CINDI_Visual_RF.eval | feedback | visual | 0.0957 | 0.1347 | 0.1796 |
| INSA-CISMef.eval | manual | mixed | 0.0531 | 0.0719 | 0.0731 |
| OHSUeng.eval | manual | textual | 0.2132 | 0.2554 | 0.2987 |
| IPAL-IPAL_CMP_D1D2D4D5D6.eval | manual | visual | 0.1596 | 0.1939 | 0.2452 |

# Results (3)


700
600
500
400
300
200
100
0
0.3
0.25
0.2
0.15
0.1
0.05
0
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
NumRel
MAP

# Results (4)


MAP
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
MAP
R-prec
B-Pref
P30
P100
IPAL-IPAL_Cpt_Im
IPAL-IPAL_ModFDT_TDF_Im
IPAL-IPAL_Textual_CRF
GE_8EN.treceval
UB-UBmedVT2
UB-UBmedVT1
UB-UBmedTBL1
UB-EnFrGePatches
RWTHi6-EnFrGePatches
IPAL-IPAL_CMP_D1D2D4D5D6
RWTHi6-EnFrGePatches2
IPAL-IPAL_cfD1D2D4D5D6
cindi-CINDI_Text_Visual_RF
IPAL-IPAL_D1D2D4D5
UKLFR-UKLFR_mids_fr_all
RWTHi6-EnFrGelfhTamThu
GE_vt10.treceval
GE_vt20.treceval
OHSUfre
GE_vt30.treceval
cindi-CINDI_Fusion_Visual
MSRA_WSM-msra_wsm
GE_vt2.treceval
IPAL-IPAL_Visual_MC
GE_vt50.treceval
GE-GE_gift
SINAI-SinaiGiftT50L30
SINAI-SinaiGiftT60L30
SINAI-SinaiGiftT70L30
GE_3DE.treceval
GE_5DE.treceval
GE_vt70.treceval
GE_vt80.treceval
rwth_mi-rwth_mi
UKLFR-UKLFR_mids_fr_all_co
Run


# Interpretation

- Visual retrieval works only really well on visual topics
- Multimodal retrieval is hard but can result in very good results
  - Fine tuning is needed
- Groups prefer automatic retrieval and work less on interaction
  - Automatic results are generally best


# Automatic Annotation Tasks


# Automatic Image Annotation

- 2 automatic image annotation tasks
- Purely visual tasks
- Explore the state of the art of image annotation techniques
- Aiming to be used as a first step for multi-modal retrieval


# Medical Image Annotation

- 10,000 training and 1,000 test images
- 116 classes identifying modality, body orientation, body region and biological system (IRMA code)
    - *e.g. 01: plain radiography, coronal, cranuim, musculosceletal system*
- Classes in English and German; unevenly distributed

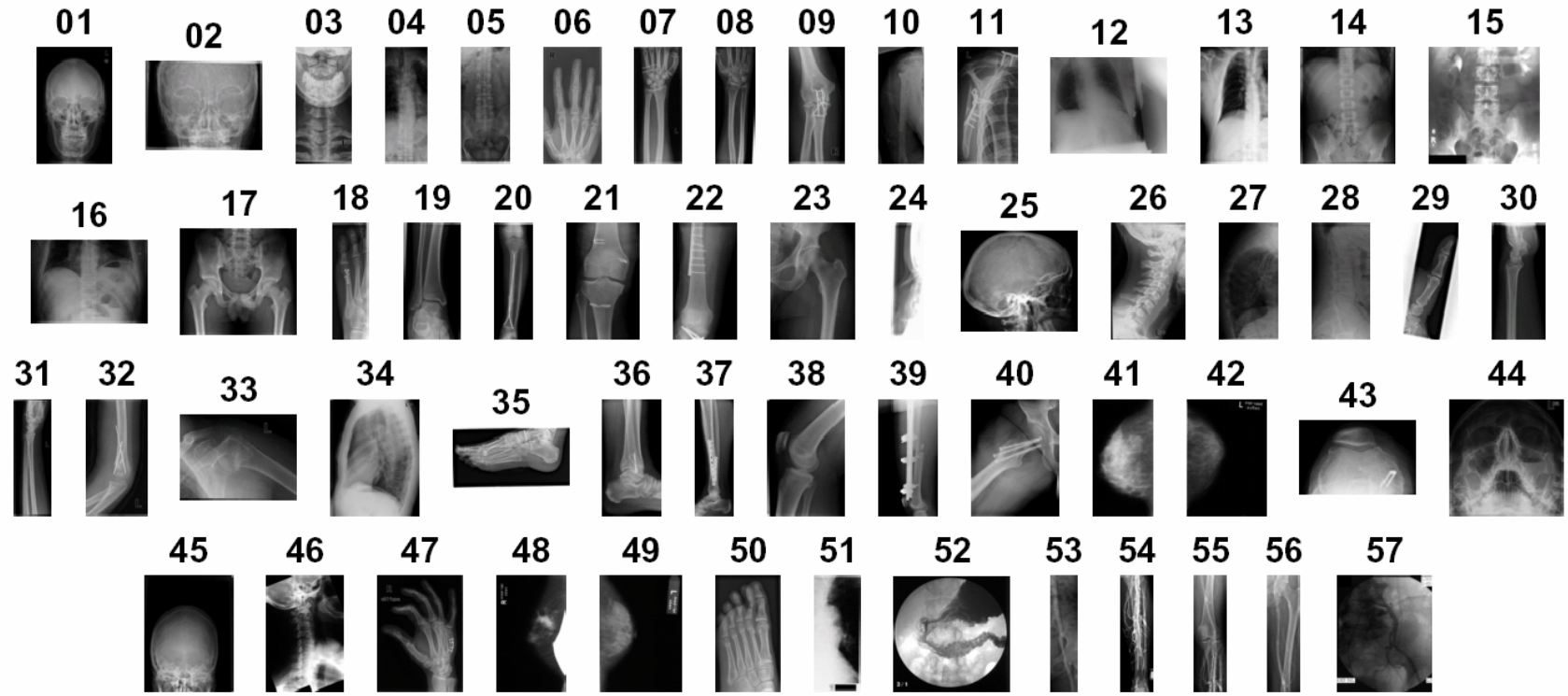
Organized by Thomas Deselaers and Thomas Lehmann, RWTH Aachen University

# Example of IRMA code

- Example: 1121-127-720-500
  - radiography, plain, analog, overview
  - coronal, AP, supine
  - abdomen, middle
  - uropoetic system
- This year: each unique code is a possible label

# Example Images

- 10,000 training images
- 1,000 test images
- 116 classes




# Participants

- Groups
  - 29 registered
  - 13 participated
- Runs:
  - In total 30 submitted
- Several groups participating the second time

- CINDI (Canada)
- DEU (Turkey)
- MedIC-CISMeF (France)
- MSRA (China)
- MU I2R (Singapore)
- NCTU DBLAB (Taiwan)
- OHSU (Oregon, US)
- RWTHi6 (Germany)
- RWTHmi (Germany)
- UFR (Germany)
- ULG (Belgium)
- UTD (Texas, US)
- MedGIFT (Switzerland)


# Results

| Rank | Group | Run | ER |
|---|---|---|---|
| 1 | RWTHi6 | SHME | 16.2 |
| 2 | UFR | UFR-ns-1000-20x20x10 | 16.7 |
| 4 | MedIC-CISMeF | local+global-PCA335 | 17.2 |
| 6 | MSRA | WSM-msra-wsm-gray | 17.6 |
| 12 | RWTHmi | opt | 21.5 |
| 14 | CINDI | cindi-svm-sum | 24.1 |
| 19 | OHSU | OHSU-iconGLCM2-tr | 26.3 |
| 21 | NCTU | dblab-nctu-dblab2 | 26.7 |
| 22 | MU | I2R-refine-SVM | 28.0 |
| 24 | ULG | SYSMOD-RANDOM-SUBWINDOWS-E | 29.0 |
| 25 | DEU | DEU-3NN-EDGE | 29.5 |
| - | medGIFT | combination | 29.7 |


# Analysis of the Results

- Performance of systems strongly improved since last year:
  - the system that performed best last year is rank 11 this year
- large variety in submitted methods
  - image retrieval approaches
  - discriminative classification approaches
- large variety in used features
  - local features
  - global features

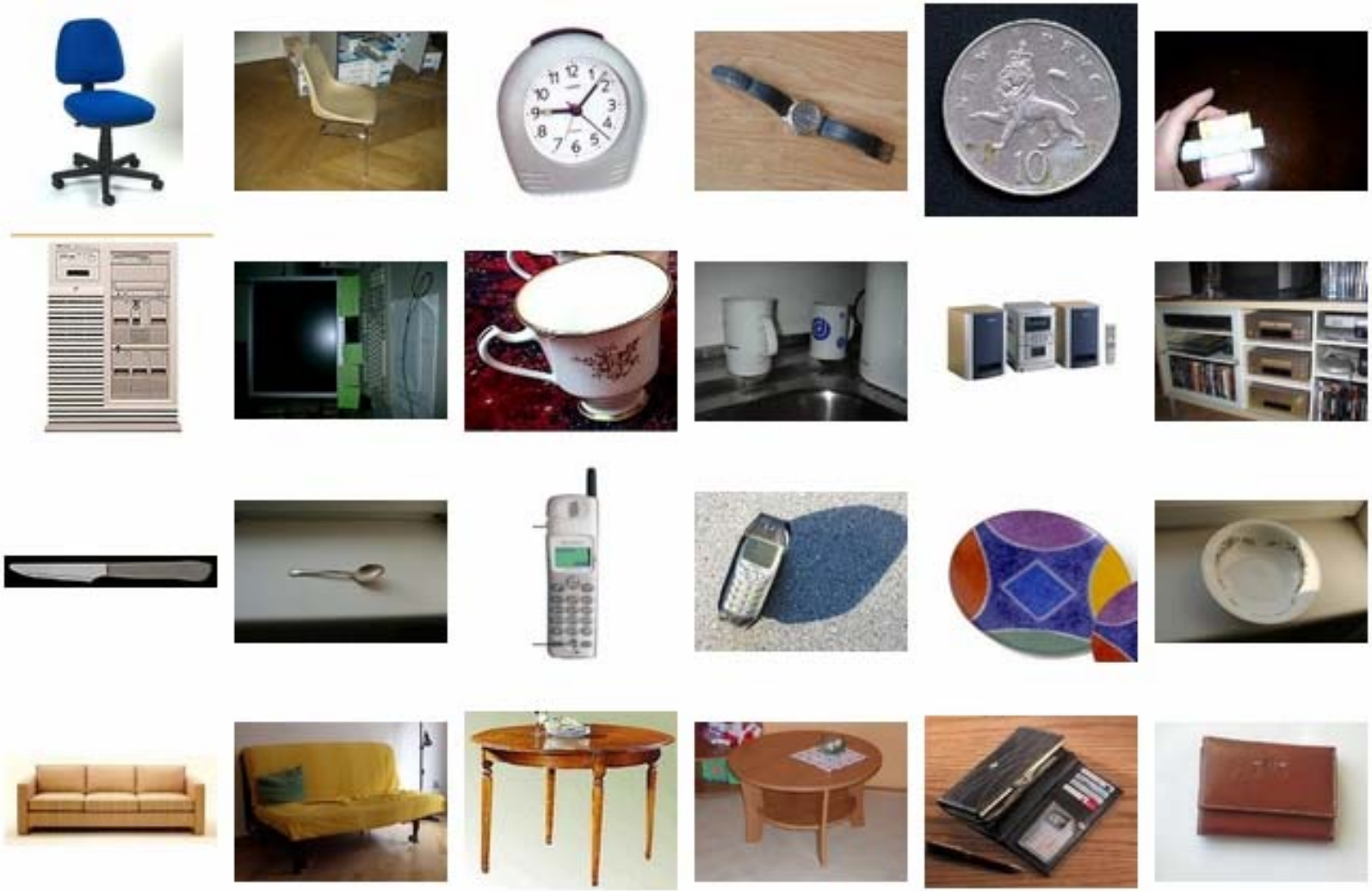
⇨ combination of good classifiers leads to even better results


# Object Annotation Task

- Database provided by LTUtech consisting of >80,000 images from 267 classes
- To ease participation, reduce dataset
  - 13,963 training images
  - 100 optimization images
  - 1,000 test images
  - 21 classes:
    - Ashtrays, Backpacks, Balls, Banknotes, Bench, Books, Bottles, Calculators, Cans, Chairs, Clocks, Coins, Computing equipment, Cups, HiFi, Cutlery, Mobile Phones, Plates, Sofas, Tables, Wallets

Organized by Thomas Deselaers,RWTH Aachen University and Allan Hanbury, TU Wien

# Example Images




# Participants & Results

- CINDI (Canada), DEU (Turkey), RWTH (Germany), MedGIFT (Switzerland)

| Rank | Group | Run | ER |
|---|---|---|---|
| 1 | RWTHi6 | SHME | 77.3 |
| 2 | RWTHi6 | PatchHisto | 80.2 |
| 3 | cindi | Cindi-SVM-Product | 83.2 |
| 4 | cindi | Cindi-SVM-EHD | 85.0 |
| 5 | cindi | Cindi-SVM-SUM | 85.2 |
| 6 | cindi | Cindi-Fusion-knn | 87.1 |
| 7 | DEU-CS | edgehistogr-centroid | 88.2 |
| 8 | DEU-CS | colorlayout-centroid | 93.2 |


# Conclusion and Discussion

- Two purely visual tasks
  - medical: clear progress visible
  - objects: very difficult task
- High variety in competing methods
- Future plans:
  - use image annotation as first step for multi-modal retrieval
  - allow for more complex annotations


# Conclusions

- Continued global participation from variety of research communities
- ImageCLEF has continued to improve
  - Medical retrieval task
    - realistic topics and larger medical image collection
  - General photographic retrieval task
    - new collection (IAPR) and representative topics
  - Medical annotation task
    - more training data and larger number of classes
  - Introduction of general annotation task
- Overall combining text and visual approaches works well for ad-hoc tasks