ImageCLEF 2013
Personal Photo Retrieval Subtask

David Zellhöfer

david.zellhoefer@tu-cottbus.de

Acknowledgments

b·tu

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Federal Office of
Consumer Protection
and Food Safety

Main Goal of the Task

- To assess the retrieval effectiveness of retrieval engines in a model of a life-spanning personal photo collection
  - Information needs include events and visual concepts
  - IN input is based on different search strategies, i.e., directed QBE and browsed image samples
  - Visual and metadata features can be used
  - Effectiveness is calculated for different user groups

Resources

- 5,555 images with GPS, Exif, and IPTC data
- 74 sample information needs based on directed search or browsing data
  - 49 events
  - 25 visual concepts
  - Classification was unknown to the participants
- Graded relevance assessments

- Baseline system (C++ for Win, Mac, and Linux)
  - Up to 17 features (with OpenCV)
  - Various distance and similarity measures
  - trec_eval compatible output

Ground Truth

- 42 assessors with different demographic background (ca. 50% female, mostly economics students, no (M)IR expertise)

- Graded relevance assessments, from irrelevant (0) to fully relevant (3)

- In average 2.69 topics were evaluated by an assessor

Personas

- 5 user groups based on the demographics and survey results accompanying the relevance assessments
  - Experts
    - A group of users that stated that they have an expertise with IR.
  - Non-Experts
    - The complement of the experts group.
  - Male/Female
    - The assessors divided by gender.
  - IT
    - This groups consists of assessors with an IT background.
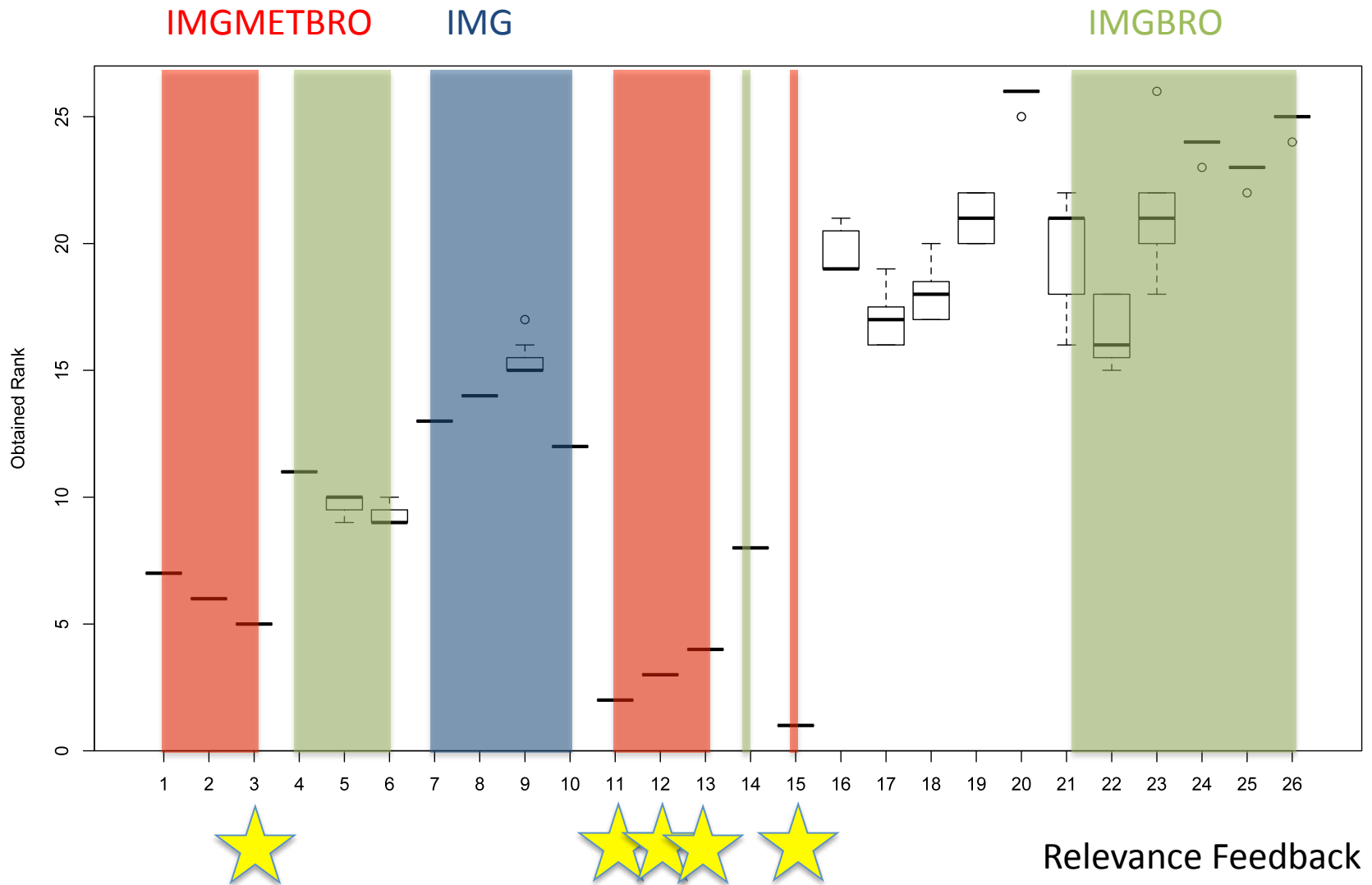 - Non-IT
    - The complement of the IT group.

Participation and Results

- 10 interested groups (at least)
- 7 participants, but only 2 publications

- Best run: **nDCG at 20 of 0.7427** for the average user group
- Worst run: **nDCG at 20 0.2302** for the average user group

Obtained Ranks over Personas
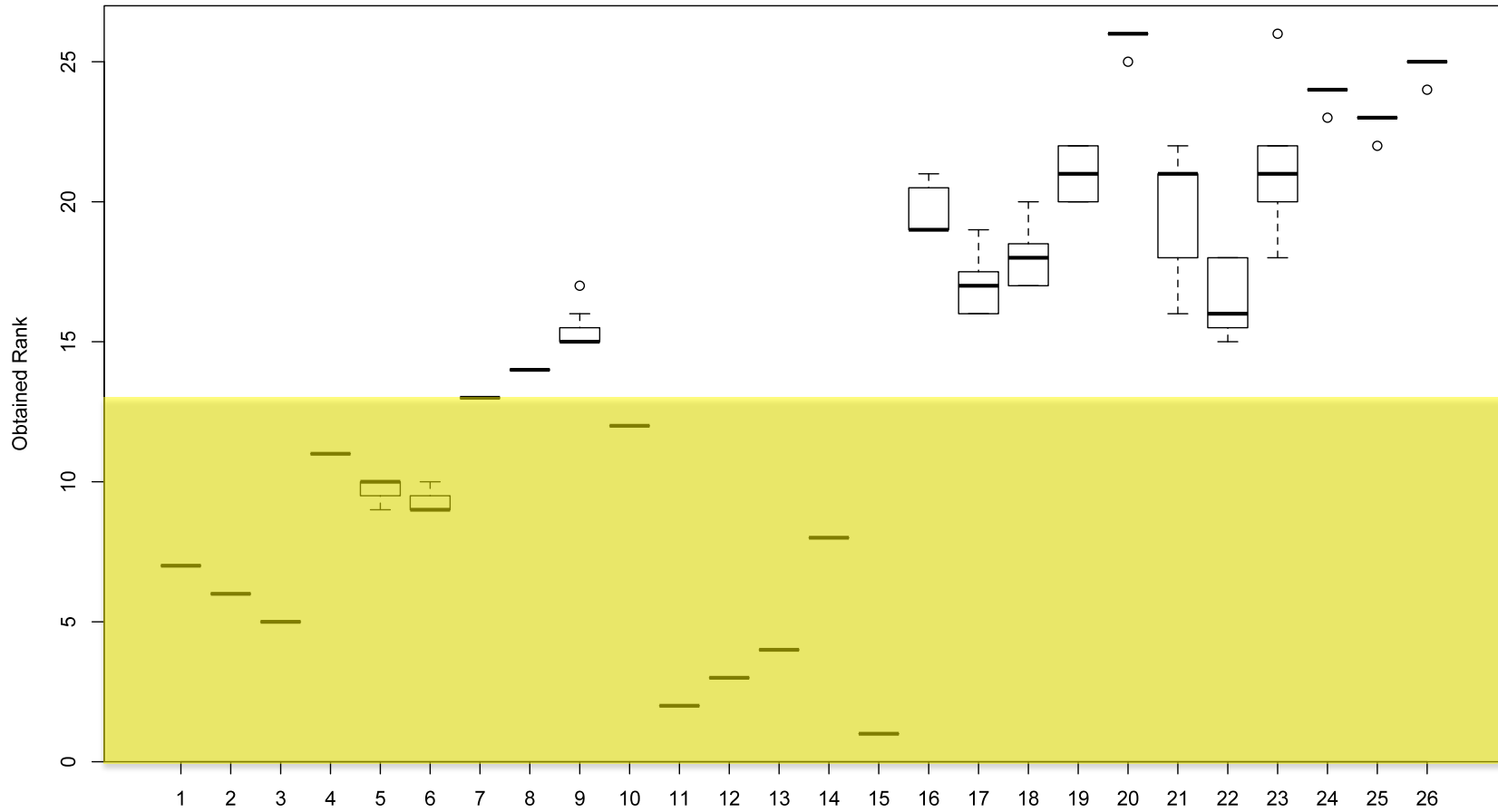
Obtained Ranks over Personas

Results

- Best performing groups used visual low-level features and metadata to solve the task, some relied on relevance feedback

- Ca. 50% of the runs perform equally well for all examined user groups

- The weak performing runs are relatively instable over different user groups

Key Trends and Lessons Learned

- In accordance with the findings of the last years' ImageCLEF tasks, there is evidence that the utilization of multiple modalities can increase the retrieval effectiveness

- Only 1 out of 7 participants used the provided baseline system