

A Personal Photo Collection for Graded Relevance Assessments

David Zellhöfer

Brandenburgische Universität Cottbus

Institut für Informatik

Lehrstuhl für Datenbank- und Informationssysteme



Overview

- Current collections
 - Are biased
 - Not necessarily reflecting personal photo collections
 - Do not contain typical images (duplicates, „tainted“ images)
 - Do only provide binary relevance assessments
 - Do not provide demographic information
 - User simulation...
 - Do not care about events
 - Are not really extensible



The Bias

- Images are often collected from the web
 - Who is uploading ugly or raw images to Flickr?
 - Who is really searching for hundreds of *airbrush angels*?



Contents of Personal Photo Collections

- Motif duplicates
- Blurred images
- Low contrast
- Back lit etc...



Binary Relevance Assessments

- Is there something like clearly relevant?
- Is MAP a useful metric?
- How can we measure improvements by relevance feedback?

- Doesn't NDCG reflect the user need slightly better?



Demographic Information

- Age, gender, usage behaviour etc. for both assessors and photographers
- Useful for:
 - User simulation (RF experiments etc.)
 - Usability engineering (interaction...)
- More IIR than system-centric IR (out of ImageCLEF's focus?)

Event Information

- WordNet-based tagging as groundtruth
- GPS data
- Metadata (EXIF/IPTC)



Extension

- Topic mappings between different collections
 - Caltech 101/256
 - Wang
 - UCID v2
 - MS RAMM
 - Our own



What can we provide?

- Personal Photo Collection Sample
 - 16 contributors (aging from ca. 25 to 56 yrs., female and male)
 - 4000-5000 documents with manual ground truth (i.e. 50% of 2011's photo annotation task)
 - Ca. 15 topics
 - Event groundtruth
 - GPS, EXIF, IPTC and low-level features

