Image Annotation with Relevance Feedback

Inspirations, ideas & plans

Motivation

- Ideal situation: general-purpose image annotation with unlimited vocabulary

Flower, yellow, dandelion, detail, close-up, nature, plant, beautiful

- Reality:
  - Classifiers with limited vocabulary and dependency on labeled training data
  - Search-based solutions with low precision

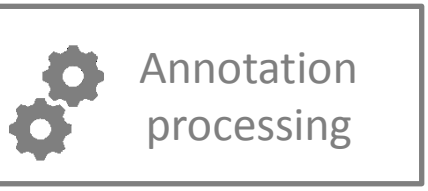
Keywords provided by MUFIN image annotation

car, show, vehicle, travel, transport, sports, motor, automobile, speed, person, luxury, coupe, new, museum, road, indoors, concept, color, view, manufacturers, front, three, automotive, horizontal, expensive, nobody, convertible, business, photography, roadster, industry, european, study, transportation, fast, photo, silver, modern, salon, make, street, white, showpiece, cars, black, republic, city, studio, district, state

Motivation (cont.)

- Possible solution: iterative annotation with user cooperation
  - Iterative refinement of annotation result
  - Takes into account user's individual needs and preferences

Vehicle, ~~person~~, ~~scenery~~



car, vehicle, transport, motor, automobile, luxury, coupe, new, expensive, convertible, silver, modern, salon, make, showpiece

Outline

- Relevance feedback
  - Principles
  - Issues to consider
- Image annotation with RF
  - Search-based image annotation overview
  - Annotation with RF: possibilities and challenges
- Inspiration from existing approaches
  - RF for text search
  - RF for image search
  - Cross-modality RF
  - RF for annotations
  - RF for graph ranking
- MUFIN IA with RF: solution outline

Relevance feedback – basic principles

1. The user issues a (short, simple) query
2. The system returns an initial set of retrieval results
3. The user marks some returned documents as relevant or nonrelevant
4. The system computes a better representation of the information need based on the user feedback
5. The system displays a revised set of retrieval results
6. Steps 3-5 are repeated until the user is satisfied

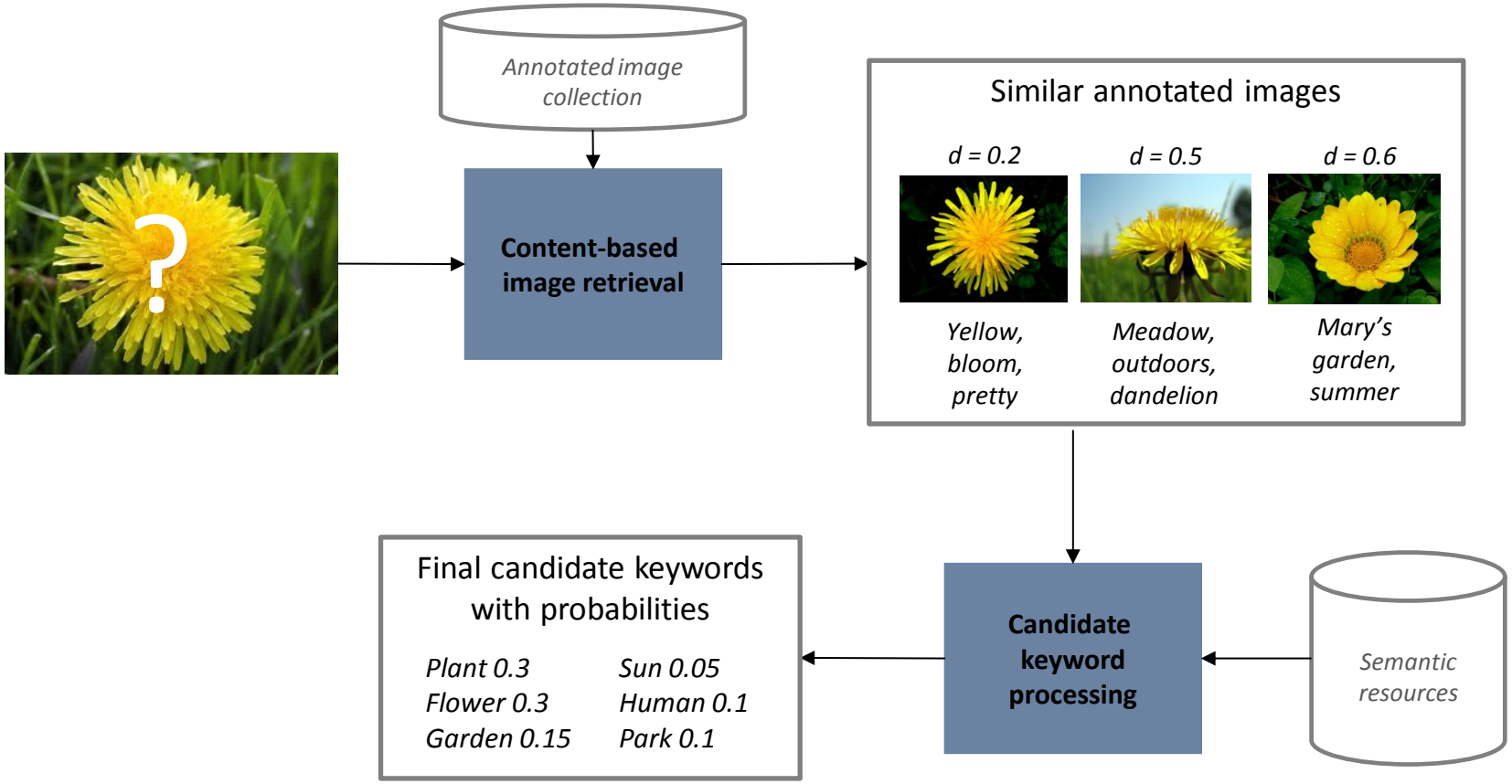
- Types of feedback:
  - Explicit / implicit / blind or pseudo-RF
  - Short-term / long-term

RF: issues to consider

- Getting explicit feedback
  - What to show the user
    - Greedy and impatient user – best known result in each step
    - Cooperative user – results that will provide the most information for the next step
  - What is realistic to expect from users?
    - How many results they will evaluate
    - Type of feedback: positive only / positive and negative / binary / multivalued / something more complex – e.g. organize images in 2D space, provide labels, etc.
- Utilizing feedback
  - How shall we utilize the information gained?
    - …
- Evaluating feedback effects on result quality
  - The information provided by the user automatically improves some quality metrics
    - Select evaluation methodology such that this “cheating” is eliminated

RF for search-based annotation – Part I: Understanding the task

Search-based annotation: Overview



RF for search-based annotation

- Annotation processing – first iteration
  - Input: image
  - Output: descriptive keywords
- Annotation processing – RF iteration
  - Original input: image
  - User feedback: positive/negative keywords
  - Output: descriptive keywords

- The problem is special in the following
  - Input modality is different from output/feedback modality
  - There are two distinct phases that may accommodate the feedback
    - CBIR for candidate keyword retrieval
    - Candidate keyword ranking
  - Existing works mostly focus on pseudo-RF in the first phase
    - There is more to be studied!

RF for search-based annotation (cont.)

- Phase 1: Content-based image retrieval with RF – retrieval task
  - Input: query image
  - User feedback: positive/negative keywords
    - Cross-media feedback!
  - Output: visually similar images / initial candidate keywords

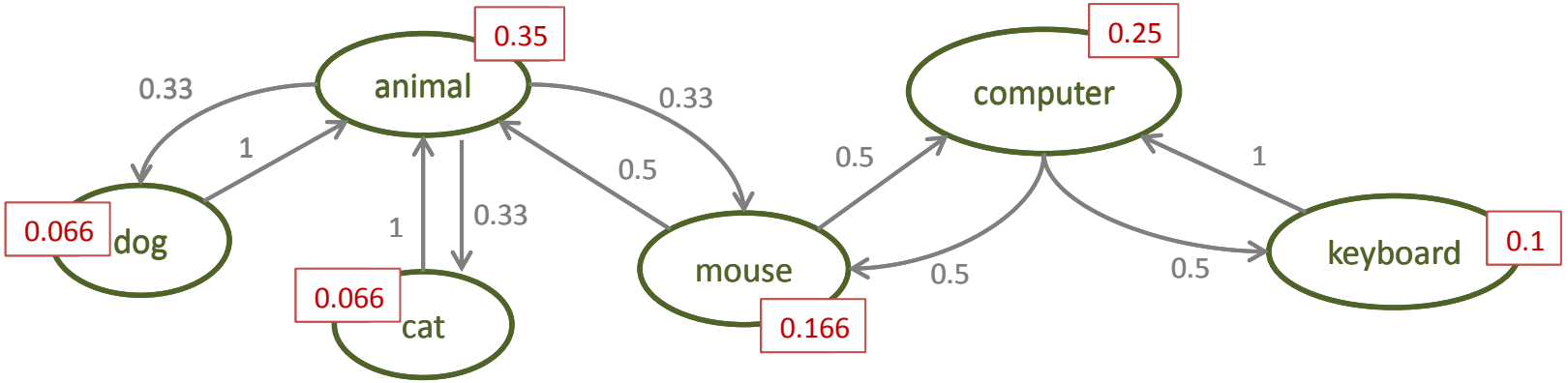
- Phase 2: Candidate keyword processing with RF – ranking task
  - Input: candidate keywords collected from similar images
  - User feedback: positive/negative keywords
  - Output: relevance scores for candidate keywords

MUFIN IA with RF: challenges

- Content-based image retrieval in MUFIN IA
  - Standard similarity search (we have a query)
  - CBIR with RF has been studied, however
    - We have cross-modality feedback
    - We want to consider negative feedback

MUFIN IA with RF: challenges (cont.)

- Candidate keyword ranking in MUFIN IA
  - ConceptRank algorithm: biased random walk over semantic graph of candidate keywords, inspired by PageRank



  - ConceptRank with RF: New problem!
    - Feedback for ranking not as well studied as for retrieval
    - PageRank is not used with ad-hoc feedback
    - Negative feedback is going to be particularly challenging, since negative and positive information should spread differently
      - Is a dog -> definitely is an animal
      - Is not a dog -> still may be animal and even very similar to dog, e.g. wolf

Looking for inspiration: RF in related areas

RF for text retrieval: Rocchio algorithm

- RF for text retrieval
  - Input: query keywords = short, sparse document
  - Collection: text documents
  - Search result: text documents
  - Feedback: positive/negative documents

- Rocchio algorithm
  - Classic implementation of RF in vector space model (1970)
  - Idea: adjust the query vector to maximize similarity with relevant documents and minimize similarity with nonrelevant documents

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

  - Empirical observations:
    - Positive feedback turns out to be much more valuable than negative feedback, so most IR systems set $\gamma < \beta$. Reasonable values might be $\alpha = 1$, $\beta = 0.75$, and $\gamma = 0.15$.

RF for image retrieval

- RF for image retrieval
  - Input: query image
  - Collection: images
  - Search result: images
  - Feedback: positive/negative images

- Some observations:
  - More ambiguities arise when interpreting images than words
    - user interaction more desirable
  - Judging a document takes time, while an image reveals its content almost instantly to a human observer
    - feedback process can be faster and more sensible for the end user
  - Efficient implementation is often a challenge

RF for image retrieval – early approaches

- First approaches were heavily influenced by the Rocchio algorithm
  - Query point movement
    - From the positive/negative feedback, compute the position of an “ideal query point”
    - The most direct application of the Rocchio algorithm
    - Easy evaluation – can reuse existing indexes
    - Problems: not possible in general metric space; assumes there exist the ideal query
  - Distance function adjustment
    - RF used for tuning of weights of individual descriptors/dimensions
    - Problems: querying with the new distance function may not be possible over existing index structures
      - Possible solution: use the new distance function only for reranking

RF for image retrieval – early approaches (cont.)

- Query expansion – multiple queries
  - Wu, Faloutsos, Sycara, Payne: FALCON: Feedback Adaptive Loop for Content-Based Retrieval. VLDB 2000
    - metric approach: a set G of good objects (the query is the first), aggregate dissimilarity function

$$(D_{\mathcal{G}}(x))^{\alpha} = \frac{1}{\sum_{i=1}^{k} w_i} \cdot \sum_{i=1}^{k} w_i (d(x, g_i))^{\alpha}$$

    - Implementation by multiple range queries
    - Applicable also to disjoint queries (all American presidents)

RF for image retrieval – later approaches

- Later works treat RF processing as an optimization / learning / classification problem
  - Main approaches: SVMs, probabilistic modeling, graph modeling
  - A lot of papers exist, new are still being published
  - No comparison available across all existing approaches
  - Mostly, the efficiency of RF processing over large collections is not discussed
    - Small test datasets, focus on answer quality improvement
    - The only possible implementation for large-scale retrieval is to apply the RF processing only on the top-N objects retrieved by initial similarity search

RF for image retrieval – later approaches (cont.)

- Very recent: CNN retraining
  - Tzelepi, Tefas: Relevance Feedback in Deep Convolutional Neural Networks for Content Based Image Retrieval. SETN 2016: 27:1-27:7
    - The proposed idea is to use the ability of a deep CNN to modify its internal structure in order to produce better image representations used for the retrieval based on the feedback of the user. To this end, we adapt the deepest neural layers of the CNN model employed for the feature extraction, so that the feature representations of the images that qualified as relevant by the user come closer to the query representation, while the irrelevant ones move away from the query.
    - Instead of modifying the query, the proposed method modifies the image representation in the seventh neural layer, FC7.
    - Two applications: single-session learning, long-term learning from multiple users
    - Efficiency never discussed

Cross-modality RF for image retrieval

- Multi-modal database: typically images accompanied by text metadata
- Query can be defined by
  - All modalities
    - Generalization of one-modality RF
  - A subset of available modalities – e.g. visual only or text only
    - Cross-modality RF: the feedback provides a new modality that was not present in the original query

- Cross-modality RF for image retrieval
  - Input: query image without text metadata
  - Collection: images + text metadata
  - Search result: images + text metadata
  - Feedback: positive/negative images + associated metadata

Cross-modality RF for image retrieval (cont.)

- Let us assume visual and text modalities
  - Much more frequent are text queries and pseudo-RF with visual modality
    - Text search for images with visual ranking of results
  - However, there also exist a few solutions where visual modality is the primary
    - CBIR with pseudo-RF text reranking
    - CBIR for annotations with user/pseudo RF

Pseudo-RF for improving text-based image search

- Ranking by pseudo-RF is frequently used to overcome the semantic gap problem
  - try to extract some useful information from the initial result
    - Initial result should contain a substantial ratio of relevant objects
- There are two information sources contained in the initial result set:
  - the properties of the candidate objects: try to discover some important dimension or descriptor that shows low variance for many of the result set objects
    - position in the search space (in case of the vector space model)
    - distance from the query (overall object distance/partial distances for individual modalities)
  - mutual relationships between candidates: relevant objects should be similar to each other while the less relevant ones will more probably be outliers in a similarity graph
    - similarity graph processing, typically by random walk
    - clustering, giving higher ranks to large clusters or to clusters which have their centroid near to the query object
    - reverse-kNN queries

RF for multi-modal image retrieval and annotation

- Example of graph-based approach:
  - J. Li, Q. Ma, Y. Asano, and M. Yoshikawa. Re-ranking by multi-modal relevance feedback for content-based social image retrieval. In *14th Asia-Pacific Web Conference on Web Technologies and Applications (APWeb 2012)*, pages 399–410, 2012
    - Graph model, both images and tags are nodes, there are image-image, image-tag and tag-tag edges
    - Users select relevance feedback instances among both images and tags!
    - Basic mutual reinforcement process: in each iteration, compute the score of a given image/node using scores of neighbors; distances provide weights. Basically the same as RW iteration.
    - Re-ranking with RF: at the beginning of each RF iteration, set scores of positive/negative RF instances to current maximum/minimum score in the candidate set. Propagate these scores through the graph edges to other nodes.

Pseudo-RF for improving visual-based image search

- Mensink, T., Verbeek, J., & Csurka, G. (2011). Weighted Transmedia Relevance Feedback for Image Retrieval and Auto-annotation, (RT-0415).
  - Transmedia Pseudo-RF: rank similar images by visual similarity to the query and text similarity to the visually most similar images
  - Basic formula

$$s_{ab}(q,d) = \sum_{i=1}^{k} s_a(q,d_i)\; s_b(d_i,d)$$

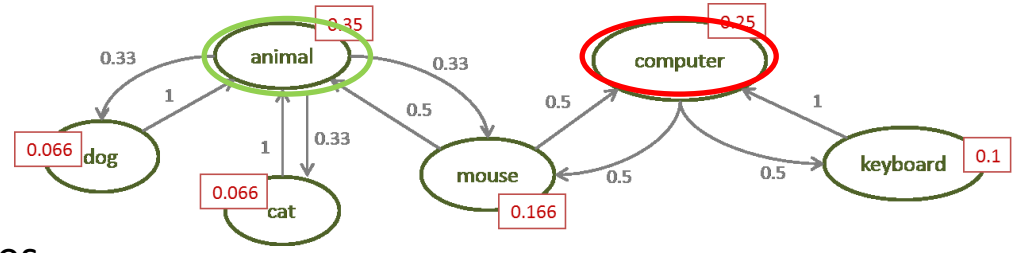
    - Extensions: parameters for importance of images based on rank;
  - Improvements of annotation precision not so big: 1-2 %.

RF for annotations

- Not many works exist
- Most solutions use pseudo-RF for CBIR phase
  - Techniques discussed on previous slides
- Alternative direction: assistive tagging
  - M. Wang, B. B. Ni, X.-S. Hua, T.-S. Chua. 2012. Assistive Tagging: A Survey of Multimedia Tagging with Human-Computer Joint Exploration, ACM Computing Surveys, 2012, 44(4):25.
  - Provide support for easy tagging of image collections:
    - (1) Tagging with data selection and organization: cluster data, require manual tagging only for several representative samples
    - (2) Tag recommendation: suggests candidate labels – possibly using information about the user
    - (3) Tag processing: refining human-provided tags or adding more information to them

RF for graph ranking problems

- Graph node ranking problem:
  - Input: graph
  - Ranking result: node scores
  - Feedback: positive/negative nodes



- Best known graph ranking algorithm: PageRank
  - TrustRank enhancement: some pages are more reliable sources of information – a-priori relevance information
    - Utilization: biased restart vector for the PageRank computation – information from reliable pages gets more weight during score propagation
  - However, PageRank is query independent
  - Query-dependent RF solved by re-ranking the top pages determined by PageRank
    - Google patent exists for this

Query-dependent random walk with feedback

- Rota Bulò, S., Rabbi, M., & Pelillo, M. (2011). Content-based image retrieval with relevance feedback using random walks. *Pattern Recognition*, *44*(9), 2109–2122.
  - RF for CBIR: Looking for image ranking such that images with RF=1 are on the top, images with RF=0 are at the bottom and the rank of visually similar images is similar
  - The resulting rank vector *x* has the following property: for each node *i,* the rank x_i expresses the probability that a random walker starting from node *i* will reach a relevant node sooner than an irrelevant node
- Lee, S. (2015). Explicit Graphical Relevance Feedback for Scholarly Information Retrieval.
  - Recommending research papers
  - The probability that a paper *p* is relevant the given query *q* and feedback *F* equals the probability that a random walk from node *p* will reach a positive node minus the probability of a random walk to the negative nodes

RF for search-based annotation – Part II: Solution outline

RF model

- Modeling the user input:
  - Both positive and negative feedback
  - Multivalued relevance from interval [0;1]

- The model is too general for most real applications, but it allows us to study the influence of different input characteristics on the RF effectiveness
  - Experiments with positive-only RF, 1/0 RF, etc.

Search-based annotation with RF - recap

- Phase I: CBIR search with cross-modality RF
  - What have we learned from related work?
    - Most solutions use (pseudo)-feedback in the form of positive/negative images
      - It is necessary to estimate the relevance of associated keywords, which is not our case
    - Main ideas: basic rank by text similarity; optimizing pair-wise ranking of images w.r.t. similarity of their descriptions

- Phase II: graph node ranking with RF
  - What have we learned from related work?
    - Option 1: fix scores of positive/negative nodes, compute the rest
      - The negative information is suppressed, but not exploited
    - Option 2: compute the probability that a positive node is reached before negative

CBIR with keyword RF

- Multiple possible solutions will be examined

- Solution 1: Standard CBIR with RF-based text-ranking
  - As opposed to systems that consider pseudo-RF, we have reliable feedback, therefore its utilization can be more straightforward
    - We do not have to consider probability of guessing the feedback correctly
  - Principle:
    - get N visually most similar images
    - rank the N images w.r.t. text similarity to positive keywords
    - rank the N images w.r.t. text similarity to negative keywords
    - combine the two ranked lists, return K<<N best images
  - Issues to deal with:
    - Optimal size of ranking lists (efficiency vs. effectiveness)
    - Possible gap between annotation vocabulary and dataset vocabulary
      - Possible solution: feedback expansion e.g. by WordNet synonyms
  - Pros: simple, efficient, can utilize both positive and negative feedback
    - From preliminary results, it works; however, we need to study the conditions
  - Cons: maybe too simple? Not new

CBIR with keyword RF (cont.)

- Solution 2: Transforming keywords from feedback to visual descriptor
  - Inspired by Carrara et al.: Picture It In Your Mind: Generating High Level Visual Representations From Textual Descriptions. CoRR abs/1606.07287 (2016)
  - Different possible ways to take:
    - Use only positive keywords to construct the descriptor of a new, “artificial” positive query image
      - Combine with original image descriptor to form a new one
        - We have doubts whether the result will make any sense, but will try
      - Use the original and the new descriptor for multi-object query
    - Use both positive and negative keywords -> positive and negative artificial images
      - Combine with original image descriptor: probably not feasible
      - Use positive artificial image for multi-object query, re-rank result with respect to negative example
  - Issues to deal with: effectiveness vs. efficiency
    - Multi-objects queries will likely be too expensive; will re-ranking give satisfactory results?
  - Pros: innovative, utilizes fashionable state-of-the-art approach – CNNs ☺
  - Cons: may not return better results

ConceptRank with RF

- Option 1: spreading only positive information
  - Principle:
    - Boost initial probabilities of positive keywords
    - Remove negative keywords from the network
  - Issues to deal with: reasonable setting of initial probabilities with respect to all available information
    - Initial keyword probabilities from CBIR phase
    - RF information
  - Pros: easy to implement, will result in smaller network -> fast processing
  - Cons: does not fully exploit negative information

ConceptRank with RF

- Option 2: spreading both positive and negative information
  - Principle:
    - Build two networks – for positive information spreading and negative information spreading
    - Compute ConceptRank on top of each network – this will give us a “positive score” and a “negative score” of each node; combine these
  - Issues to solve:
    - Building the negative network: is there anything we can derive from negative feedback apart from removing the respective part of the network?
    - Initial probabilities of nodes
    - Combining the positive and negative node scores
  - Pros: positive and negative information more fully exploited
    - Hopefully better results?
  - Cons: more computations; more parameters that need to be correctly tuned

More open questions

- How many RF iterations we will consider?
  - Try 1 and more, observe usefulness of each new iteration
- How shall we deal with RF history?
- What to feed back?
  - We simulate the user for experiments
  - How many assessed keywords? Include also partially relevant? From how many top results?
- All iterations the same, showing best current results, or the first more like active learning, showing possible categories?
- Efficiency vs. effectiveness!

Summary

- Search-based annotation is not sufficiently precise
  - Currently used as tag-hinting, user has to choose correct keywords
  - User relevance judgement could be exploited in a new iteration of the annotation process
- RF for image annotations has not been thoroughly studied yet
- We want to examine the possibilities of exploiting RF in the two main phases of annotation process
  - RF for cross-modality CBIR
    - We have two possible solutions, ready for implementation and testing
  - RF for graph node ranking
    - More thinking to be done yet