

---

# Personalized Information Retrieval

Elena Holobiuc  
Iulia Pasov  
Alexandru Agape  
Octavian Sima  
Bogdan Cap-Bun



# Content

---

- Overview
- Enhancing Personalized Web Search
- Intent and interest in personalized search
- Online Advertising
- Opinion Mining
- Trends



# Interest-based Personalized Search

---

- Search results are based only on the query, not on the user interests or search context
- Results are usually so many that they are partitioned into several pages
- Individual differences in information needs, polysemy and synonymy pose many problems
- Solution:
  - A personalized search approach for extending a conventional search engine on the client side
  - Results that "look" different for each user



# Interest-based Personalized Search (2)

---

- User basic information is known (skills, interests,...)
- Identification of categories associated to each defined user interest
- URLs are used as training examples
- Help user focus on results of interest decreasing the time spent in searching
  
- The personalized categorization system:
  - outperforms non-personalized categorization systems for searches with free-form queries
  - helps users find relevant pages with less effort, even if they cannot issue relevant queries
  - is not universally better than any another system
- What if a user searches for something not defined as his interest?



# Content

---

- Overview
- Enhancing Personalized Web Search
- Intent and interest in personalized search
- Online Advertising
- Opinion Mining
- Trends



# Enhancing Personalized Web Search

---

- Re-ranking query results returned by search engines locally using personal information;
  - bandwidth intensive

or

- Sending personal information and queries together to the search engine
  - most used by search engines - tailor results on server
  - privacy issues due to exposing personal information to public
  - It requires the user's permission



# Enhancing Personalized Web Search (2)

---

- **Hierarchical user profile:**
  - It's not realistic to require that every user to specify their personal interests explicitly and clearly.
  - Offers an easy way to protect and measure privacy.
- **Construction of the User Profile**
  - based on frequent terms in the user documents
  - general terms with higher frequency are placed at higher levels
  - relationships between the frequent terms :
    - Similar terms : two terms that cover the document sets with heavy overlaps might indicate the same interest.
    - Parent-Child terms : Specific terms often appear together with general terms, but the reverse is not true



# Enhancing Personalized Web Search (3)

---

- Top-down approach for building the profile
- Tree structure - each node (labeled as term  $t$ ) is associated with a set of supporting documents  $S(t)$
- The root node is created without a label and attached with  $D$ , which represent all personal documents.
- Starting from the root, nodes are recursively split until no frequent terms exist on any leaf nodes.





# Content

---

- Overview
- Enhancing Personalized Web Search
- Intent and interest in personalized search
- Online Advertising
- Opinion Mining
- Trends



# Addressing User Needs in Search Results

---

- Do the queries have a question-answering intent ?
- Log and interpret user interactions with search results
  - A query is considered abandoned if no results are clicked
- The absence of interaction behavior can be useful in understanding the information's value.
- The query logs include:
  - query text, userID, timestamp
  - list of results and their positions
  - whether or not each result was clicked
- A search engine tries to meet an information need within the context of the search result page.



# Detecting User Goals from Interacting Data

---

- Same user might expect different results for the same query at different times
  - identification of the user's intent is needed
- User needs are not always revealed through clicks
  - eye-tracking - devices are uncommon to most users
  - mouse movement and scrolling might also reflect user attention
- Builds a user behavior model that captures queries, clicks and fine-grained interaction with the search results
  - predicts the searcher's current goal and future behavior



# Exploring mouse movements for inferring query intent

---

- Navigational queries
  - users often go directly to the interested result (spending little time on reading)
  - simple mouse trajectories
- Informational queries
  - users spend more time reading the result page
  - complex mouse trajectories.
- Problems :
  - the mouse is not always used to mark user's interests



# Variation in User Intent

---

- Similar queries from different people might target different results
- Identify queries that show the most variability across individuals
  - measure explicit relevance judgements and large-scale log analysis of user interaction patterns
- Identify queries that benefit most from personalized ranking
  - features of the query, the results of the query and people's interaction history with the query
  - click-based measures indicate when different people find different results relevant to the same query



# User interest

---

- Reasonable *user model* that captures relation between user's click history and his interest
  - $P(\text{visit}) = \text{Sum} (P(\text{topic}) * P(\text{visit} | \text{topic}))$ ,  $P(\text{topic})=?$
- Learning method for finding parameters of the model to use for predicting interest
  - linear regression - poor results due to sparsity
  - maximum likelihood - maximize probability of history
- Ranking mechanism to consider user's interest in generating search result
  - Adapt Topic Sensitive Page Rank method
    - Rank given by  $P(\text{topic}|\text{query})$
    - Bayes rule, learned  $P(\text{topic})$  and  $P(\text{query}|\text{topic})$  estimated using Open Directory



# No results found

---

- What does the user do when no results are found?
  - Typical for eCommerce sites, with difference between seller and buyer vocabularies and changing inventory
  - Power users differ from novice ones
- Building a dataset
  - Assign logged pages to a set of classes:
    - Homepage, search, product details, purchase page
  - Build browser *trails* from click stream history of browsing
    - mark trails containing zero recall searches
- Characterize zero-recall searches
- Study user behavior
  - Power users refine search, resulting in better conversion rate



# Content

---

- Overview
- Enhancing Personalized Web Search
- Intent and interest in personalized search
- Online Advertising
- Opinion Mining
- Trends





# Online advertisement

---

- Financial engine of search
- How to measure effectiveness?
  - click rate, conversion rate
  - improved statistics if considering user search history
    - correlations in ads seen by users and user's actions in the near future
- Adapt powerful statistical models to mine user-level advertising data and specialized IR algorithms for advertisement evaluation
  - Markov models, graphical models
  - MapReduce
  - PageRank



# Online advertisement (2)

---

- Nice reports to have (for advertisers)
  - top  $k$  ads with largest impact
  - ads with significant long-term effects
    - missed by non user-based methods
  - top  $k$  adds with largest marginal increase
    - where should I put more money?
- Use generative model
  - build graph with events (impressions, conversions)
  - Assign weights also based on higher-order interactions
    - LastAd (default) - direct conversion rate
    - PageRank contribution
    - Eventual Conversion
    - Removal effect



# Content

---

- Overview
- Enhancing Personalized Web Search
- Intent and interest in personalized search
- Online Advertising
- Opinion Mining & Sentiment Analysis
- Trends



# Opinion Mining & Sentiment Analysis

---

- "What other people think ?"
- "Positive vs. Negative"
  
- 81% of Internet users have done online research on a product at least once;
- Consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item
  
- Companies realize the importance of consumer voices
- Opinion-rich resources (reviews, forums) are constantly growing - 75,000 new blogs and 1.2 million posts daily
  
- New technologies are needed for retrieving and tracking this information



# Opinion/Review Search Engine

---

## 1. Determine whether the user is looking for subjective material.

- indicator terms, checkbox
- query classification is a difficult problem (2005 KDD Cup)

## 2. Determine which documents or parts of documents contain reviews or opinions

- easy on review-aggregation sites (Epinions.com, Amazon)
  - stereotyped format is used
- on blogs, the subjective content vary quite widely in content, style, presentation, level of grammaticality.



# Opinion/Review Search Engine (2)

---

3. Identify the overall sentiment expressed by these documents regarding the item/topic in question

- easier when the user must specify grades for pre-defined sets of characteristics ( Yahoo! Movies )
- a lot of processing is needed for free-form text

4. Present the sentiment information to the user

- aggregate “votes” registered on different scales (e.g. one reviewer uses a star system, another uses letter grades).
- selective highlighting of some opinions
- visuals are better than a textual summary



# Applications

---

- **Review-related websites**
  - review-oriented search engine can serve as the basis for the creation and automated upkeep of opinion-aggregation websites
  - summarize and automatically fix wrong ratings
- **Recommendation systems**
  - avoid recommending items with negative feedback
  - bring up product ads when relevant positive sentiments are detected
  - improve IR by discarding info found in subjective sentences
- **Business intelligence**
  - “Why aren’t consumers buying our laptop?”



# Challenges

---

- Classify an opinionated text as either positive or negative

*“If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut.”* (Givenchy perfume review)

- It's hard for humans to come up with the best set of keywords for expressing a sentiment

	Proposed word lists	Accuracy (%)	Ties (%)
Human 1	positive: <i>dazzling, brilliant, phenomenal, excellent, fantastic</i> negative: <i>suck, terrible, awful, unwatchable, hideous</i>	58	75
Human 2	positive: <i>gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting</i> negative: <i>bad, cliched, sucks, boring, stupid, slow</i>	64	39
Statistics-based	positive: <i>love, wonderful, best, great, superb, still, beautiful</i> negative: <i>bad, worst, stupid, waste, boring, ?, !</i>	69	16





# Challenges (2)

---

- Hard to recognize the quality of a review.

## It sucks.

by allsimlover ([movies profile](#))  
(Nov 22, 2004)  
3 of 12 people found this review helpful

I loved this movie when I was little.  
Now I hate it! I've grown out of it. I  
might like it again, eventually.  
Burton rules!

Overall Grade: F

## GREAT MOVIE

by cheshiremusic ([movies profile](#))  
(Oct 19, 2003)  
1 of 2 people found this review helpful

i loved this movie. it was relli  
good. everyone shoud see  
it...<<<<that was relli all i had to  
say but apperently i have to write  
30 words so lemme tell u...

[Full Review](#)

Overall Grade: A+

- Objective or subjective text ?
  - “How mad are you?”



# Summarization

---

- Aggregate & Represent opinions
- Single-document summarization
  - author's opinion
  - the most positive and negative phrases
- Multi-document summarization
  - textual summaries
  - non-textual summaries - based on pre-defined polarity
- Document polarity defined by:
  - thumbs up / thumbs down
  - number
  - grade

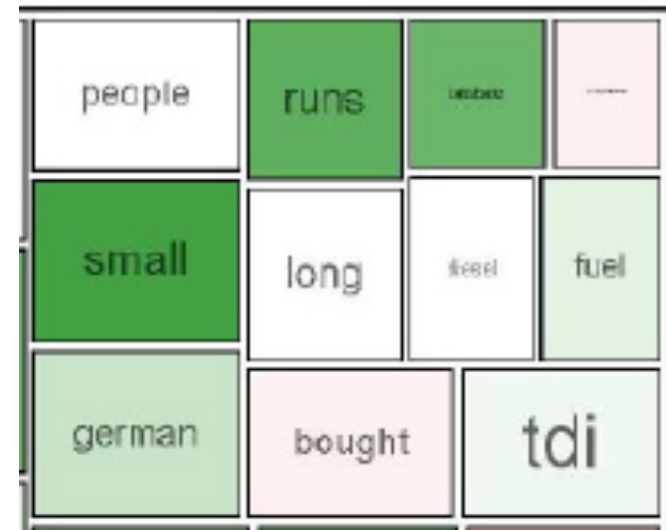


# Non-Textual Summarization

- "Bounded" summary statistics
  - "Thermometer"-type images



- Color shading representation
  - determines topics from a review
  - size - no of occurrences
  - color - average sentiment
  - extracts the most extreme opinions



# Non-Textual Summarization (2)

- "Unbounded" summary statistics
  - actual number of opinions
  - number of positive and negative reviews
  - average rating
  - details on average rating (how many people gave 7 out 10)



## Average Customer Review

★★★★★ (1,247 customer reviews)

Share your thoughts with other customers

Create your own review

Amazon ratings

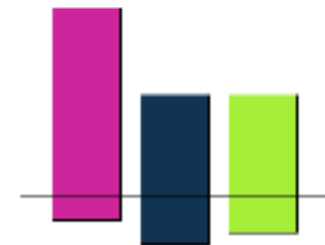


**Your rating:** ★★★★★★★★ -/10

Ratings: 7.0/10 from 8,050 users Metascore: 65/100

Reviews: 63 user | 140 critic | 24 from Metacritic.com

IMDB ratings



"Horizon" line representation

# Summarization (2)

---

- **Opinion timelines**
  - Order reviews in reverse chronological order
  - Track opinion changes over time
- **Review(er) quality**
  - Is the review helpful or useful?
  - Recent and low-ranked reviews have few utility votes
  - "Rich get richer" phenomena
  - Reviewer credibility

**The most helpful favourable review**

2,731 of 2,773 people found the following review helpful:



# Ranking

---

- Utility evaluation
- Review score
- Number of stars assigned
- Similarity between the review and the product specification
  
- Readability score (review length in characters divided by number of sentences)
  
- Is the review a spam?
  - duplicate reviews
  - insertion of brands unrelated to the product
  - reviews without opinion



# Sentiment analysis implications

---

- Privacy violations
- Manipulation
  - spam reviews
  - "game the system" - suppress negative publicity
- Economic impact
  - reviews seem to be influential for expensive products
  - negatives ratings have an effect, while positives ones do not
  - "word of mouth"- the amount of feedback, and not the polarity matters
- User impact
  - review text vs score



# Implementation

---

- Data sets
  - WordNet, SentiWordNet
- Approaches
  - Naive Bayes, SVM
  - Different weights for POS - higher for nouns, verbs, adjectives
- Libraries
  - Rapid Miner, NLTK (Python), LingPipe (Java)
- Examples
  - [twitrratr.com](http://twitrratr.com) - Twitter
  - [tweetpsych.com](http://tweetpsych.com) - psychological profile for Twitter accounts
  - [tweetfeel.com](http://tweetfeel.com)
  - [ubervu.com](http://ubervu.com) - social media connotation analysis





# Content

---

- Overview
- Enhancing Personalized Web Search
- Intent and interest in personalized search
- Online Advertising
- Opinion Mining
- Trends



# Trends

---

- Computer vision-based personalization
  - Identify user facial expression
  - Serve different content according to user's emotions
- Avatars for:
  - interacting with users and recommending items
  - building profiles for users from online conversations
- Dynamic page content based on user profile and mouse/eye tracking
- Social filtering
  - personalize results taking into account what users with a similar profile considered useful
  - recommend items based on friends preferences, search history, "likes", "shares" etc.



# Conclusions

---

- Personalized search can improve user search experience
- Need of user intent and user interest prediction
- New evaluation methods to make use of personalized search data
- Ensure privacy of user data
- Opinion Mining & Sentiment Analysis are "hot" topics with many possible applications
- Information can be extracting from various environments, not only text



# References

---

- B. Pang, L. Lee, *Opinion Mining and Sentiment Analysis*
- F. Qiu, J. Cho, *Automatic Identification of User Interest For Personalized Search*
- G. Jeh, J. Widom, *Scaling Personalized Search*
- G. Singh, N. Parikh, N. Sundaresan, *User Behavior in Zero-Recall eCommerce Queries*
- J. Teevan, S. T. Dumais, D. J. Liebling, *To personalize or not to personalize. modeling queries with variation in user intent*
- L.A. Granka, T. Joachims, G. Gay, *Eye-tracking Analysis of User Behaviour in WWW Search*
- L. B. Chilton, J. Teevan, *Microsoft Research Addressing people's information needs directly in a web search result page*
- N. Archak, V. S. Mirrokni, S. Muthukrishnan, *Mining advertiser-specific user behavior using adfactors*
- Q. Guo, E. Agichtein, *Exploring Mouse Movements for Inferring Query Intent*
- Q. Guo, E. Agichtein, *Ready to buy or just browsing. detecting web searcher goals from interaction data*
- S. Xu, H. Jiang, F.C.M. Lau, *Personalized Online Document, Image and Video Recommendation via Comodity Eye-Tracking*
- Z. Ma, G. Pant, O.R. Liu Sheng, *Interest-Based Personalized Search*
- Y. Xu, B. Zang, Z. Chen, *Privacy-Enhancing Personalized Web Search*
- Y. Zhu, L. Xiong, C. Verdery, *Anonymizing User Profiles for Personalized Web Search*



# Q & A



Q & A



Căutare

Aproximativ 2.870.000.000 (de) rezultate (0,15 secunde)

Toate

Sfat: [Căutați numai rezultate în Română](#). Puteți specifica limba dvs. de căutare în [Preferințe](#).

Imagini

Videoclipuri

Știri

Cărți

Discuții

Mai multe

[Q&A - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Q%26A](https://en.wikipedia.org/wiki/Q%26A) - Traducerea acestei pagini

**Q&A**. From Wikipedia, the free encyclopedia. Jump to: navigation, search. Look up **q&a** in Wiktionary, the free dictionary. **Q & A**, generally meaning "Questions ...

[Q&A \(Australian talk show\)](#) - [Q&A \(software\)](#) - [Q&A \(film\)](#) - [Q&A website](#)

[Q & A \(novel\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Q\\_%26\\_A\\_\(novel\)](https://en.wikipedia.org/wiki/Q_%26_A_(novel)) - Traducerea acestei pagini

**Q & A** is a novel by Vikas Swarup, an Indian diplomat. Published in 2005, it was the author's first novel. Set in India, it tells the story of Ram Mohammad Thomas, ...

