Università della Svizzera italiana

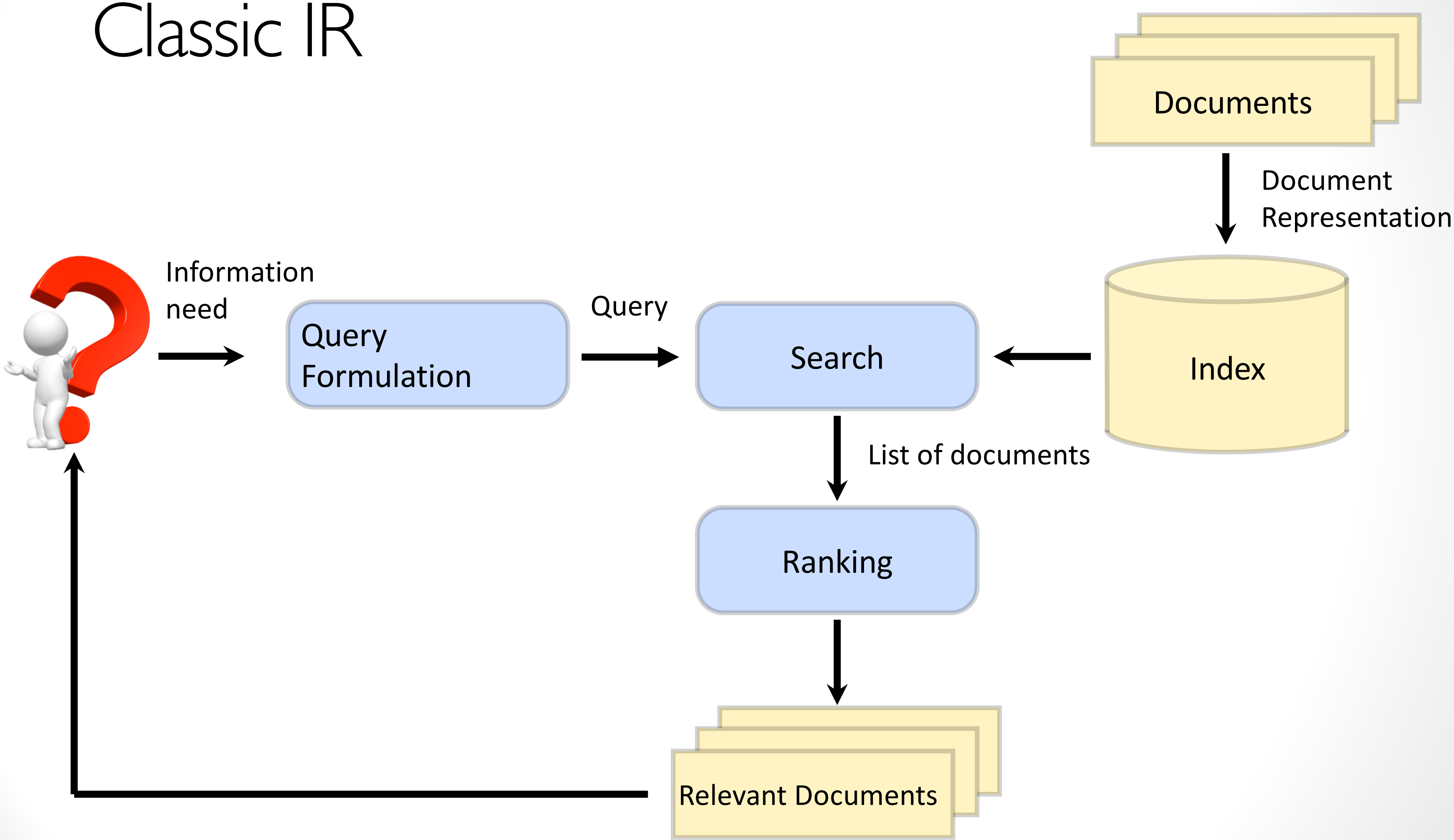
Distributed meets Mobile Information Retrieval

Prof. Fabio Crestani
Universitá della Svizzera Italiana (USI)
Switzerland

Outline

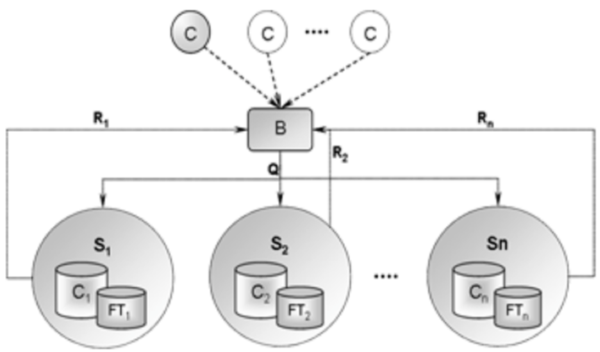
- Background
  - Distributed IR
  - Mobile IR
- Context as a driving element of search
  - Capturing context in Mobile IR
- Unified Mobile Search (UMS)
  - Evaluation results
- Conversational Mobile Search
  - Preliminary investigations
- Conclusions

Classic IR

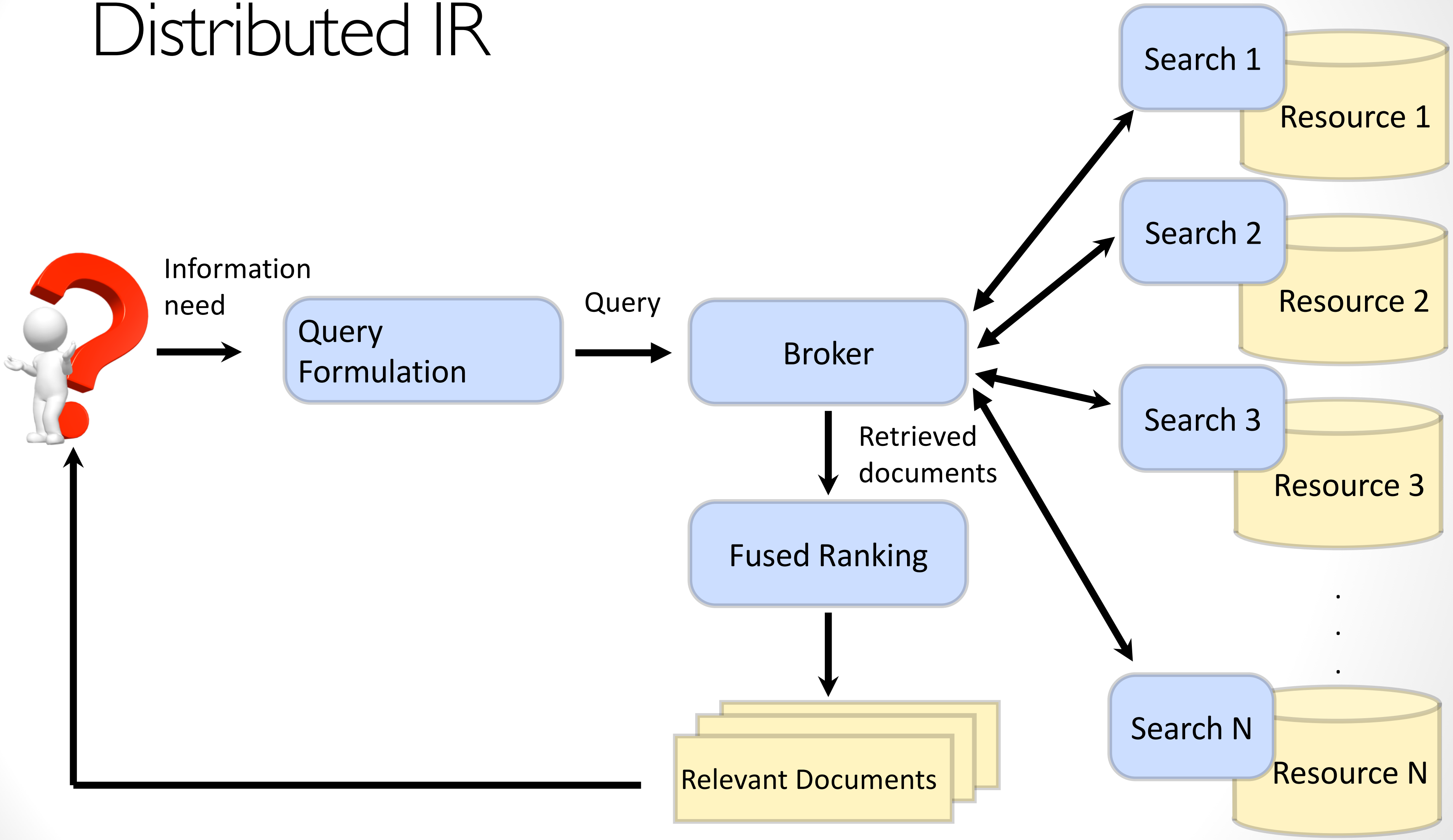

Documents
Document
Representation
Information
need
Query
Formulation
Query
Search
Index
List of documents
Ranking
Relevant Documents

Distributed Information Retrieval

- Distributed IR (or DIR) is concerned with enabling users to search for information that is distributed across different resources
  - Each resources is independent and has its own search engine
  - DIR can be deployed in fully distributed environments
  - DIR is also known as Federated Search, Metasearch, Aggregated Search. …
- Typical phases of a DIR systems are:
  1. Resource description
  2. Resource selection
  3. Results fusion



Distributed IR



Mobile Information Retrieval

- Mobile Information Retrieval (Mobile IR) is a more recent branch of IR that is concerned with enabling users to carry out IR on a mobile device
  - A mobile device is personal, full of sensors, …
  - Mobile information needs change more rapidly
  - Mobile IR is naturally **context sensitive!**

What is Context?

- Over 150 definitions of context, spread across several domains!
- In Information Retrieval:
  - *"A set of factors that have influence on the user's perception and acceptance of a particular item retrieved by a system"*
  - *"A set of environmental states and settings that either determines a system's behaviour or in which an system event occurs and is of interest to the user"*
- In practice …

The role of context

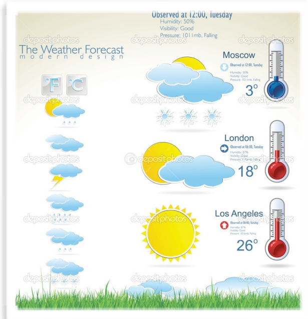
Time

Location

Social Situation

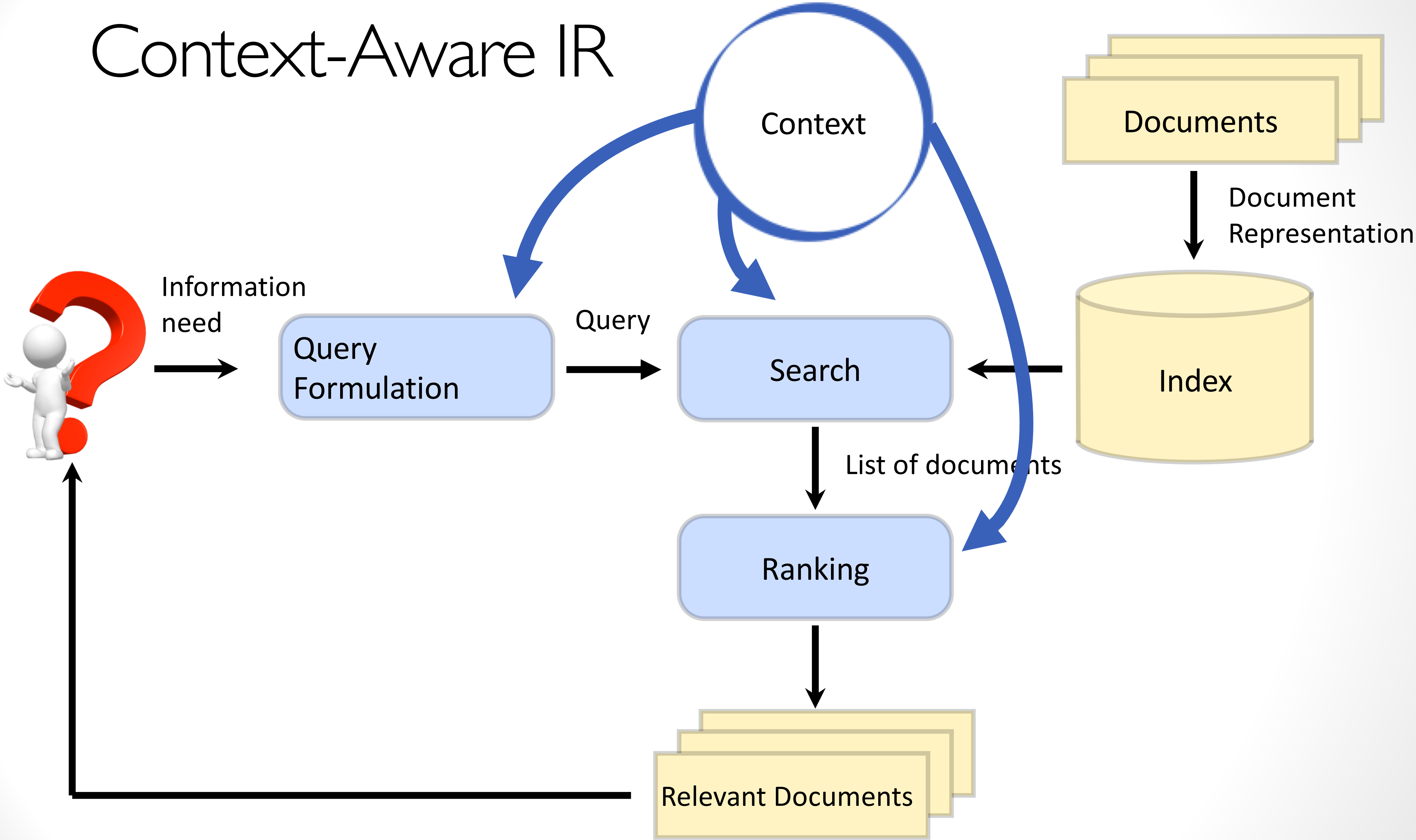
Context

Physical Environment



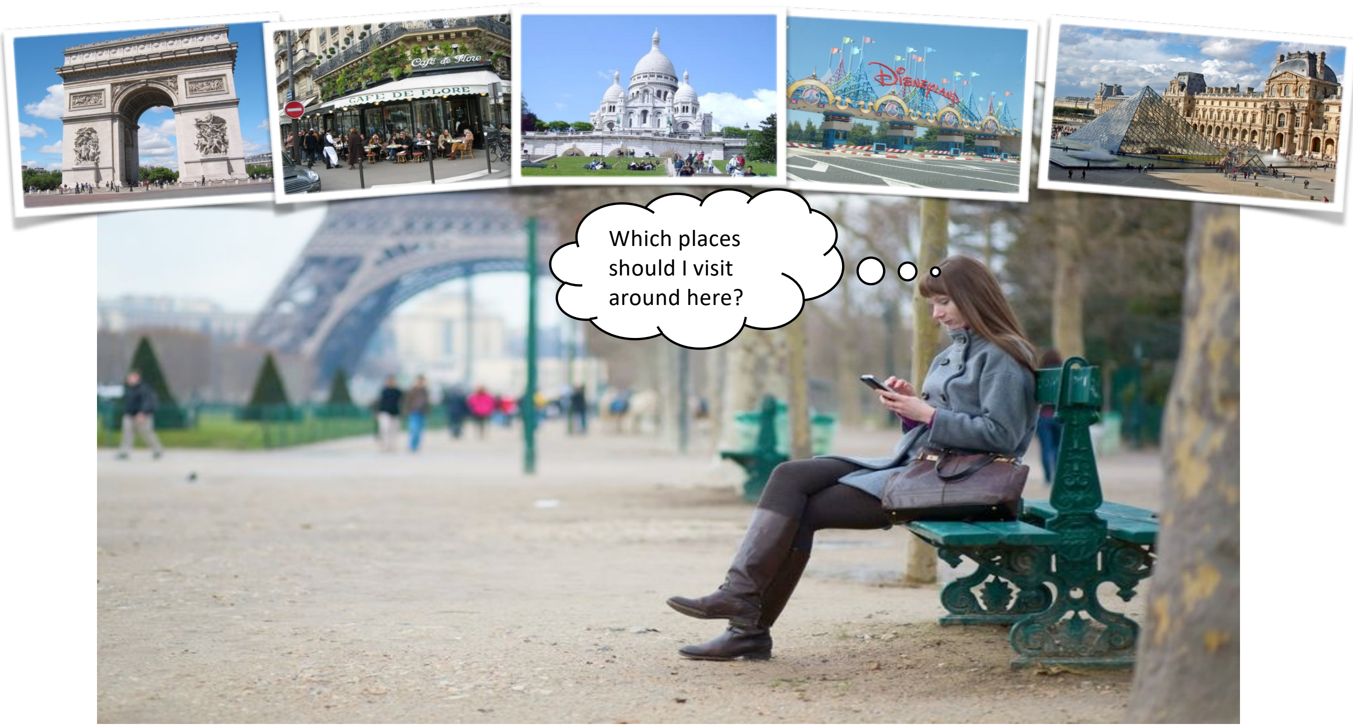
?

Context-Aware IR



Why is Context so important?

- Example: TREC Contextual Suggestion Track

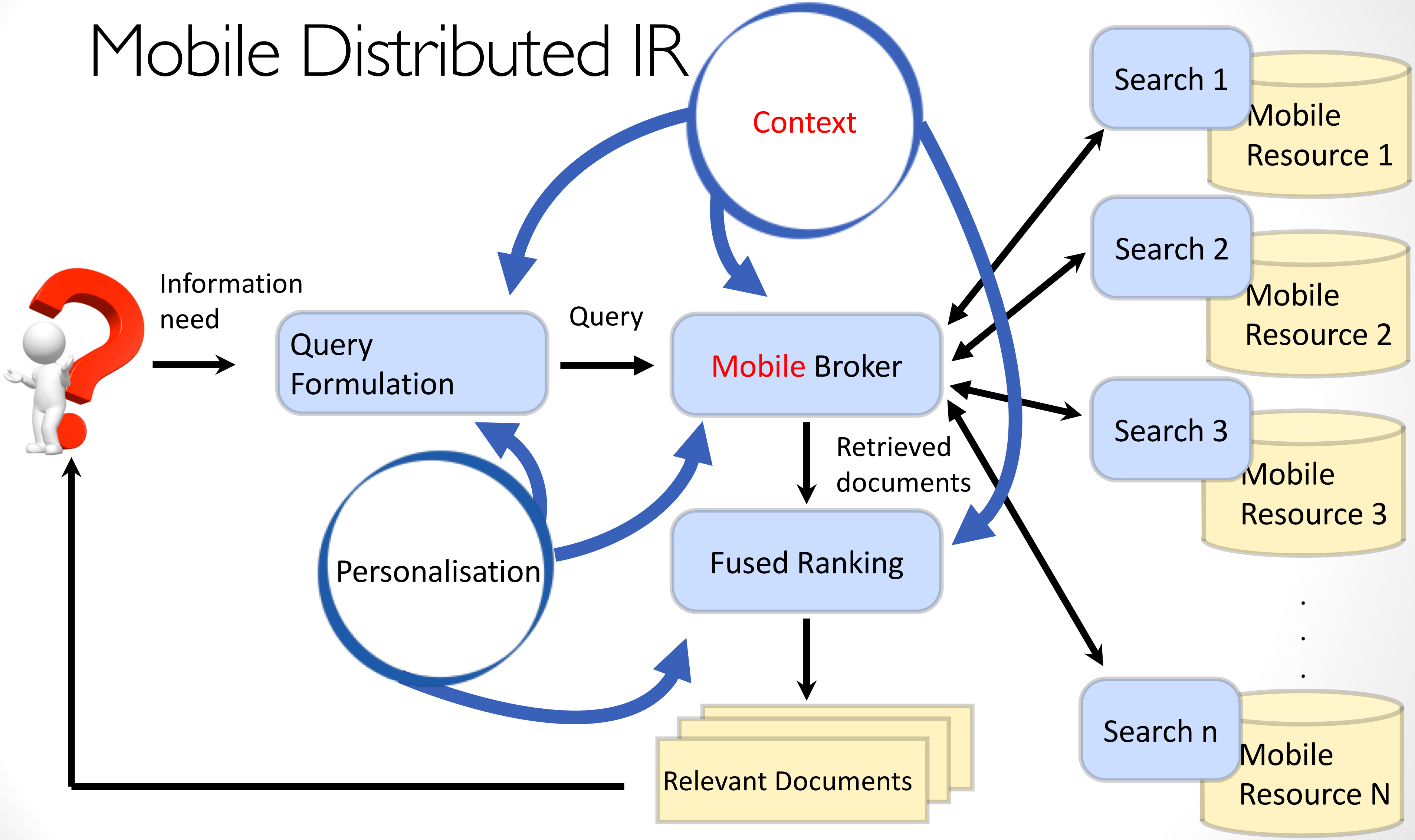


See http://trec.nist.gov/

Context in Mobile IR

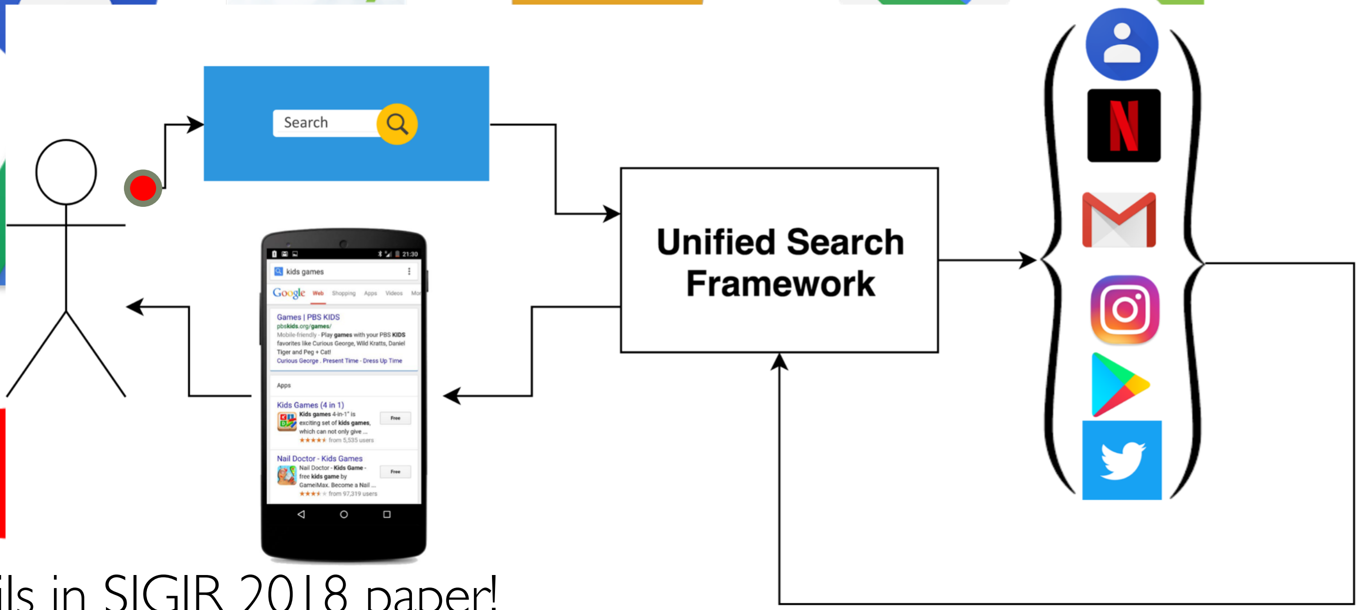
- Idea: why not use context combined with the user information need in a distributed information setting to drive the search process?
- Context can:
  - Help better understand the user information need
  - Direct the information need towards the most appropriate mobile applications
  - Organise results in a way that is context-aware
- Use context for Mobile IR in a distributed environment: Mobile Distributed IR

Mobile Distributed IR



Unified Mobile Search (UMS): the idea

A unified search framework that replaces all the search boxes in the apps



See details in SIGIR 2018 paper!

Feasibility study

- Does it all make sense?
- We organized a study using Crowdsourcing
  1. Study cross-app mobile search to define real-life mobile search tasks across different apps
  2. Collect query-app pairs for each task

1: Cross-App Mobile Search Tasks

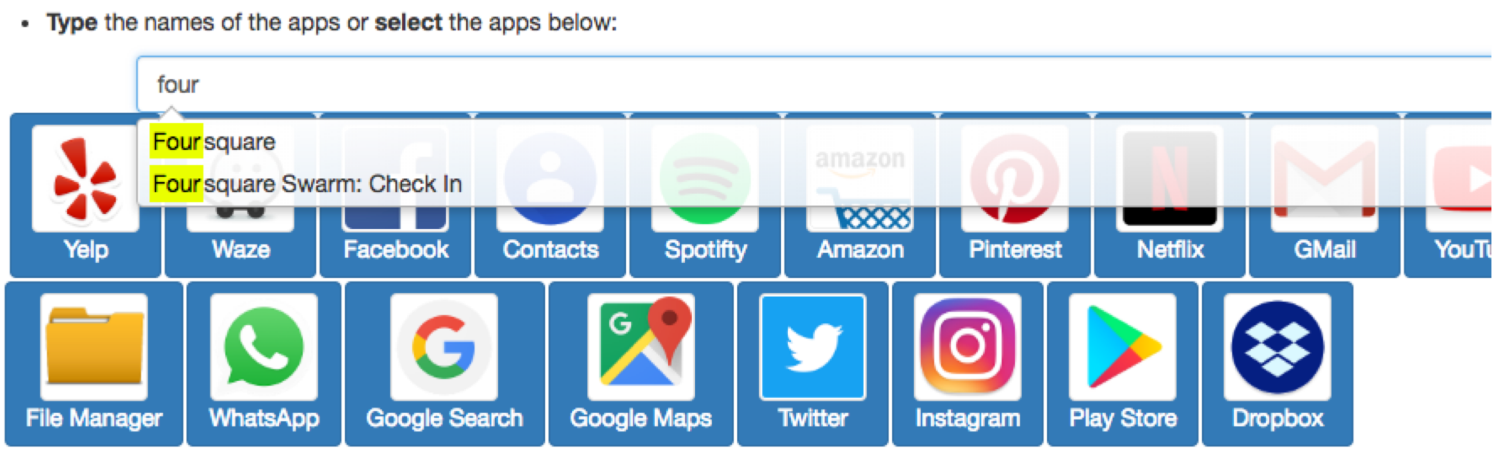
- Workers were asked to report their latest mobile search experience.
- The HIT was launched for various categories.
- We used the answers to generate more generic tasks.

"I was searching for a new refrigerator to buy. The first thing I did was search for the best refrigerators of 2017 and then narrow down my search for exactly the type of refrigerator that I was looking for..."

"Consider one of the oldest appliances in your home. You have been thinking of changing it for a while. Now, it's time to order it online."

2: Query and App Pairs

- For each search task workers had to:
  1. Type or select the name of the apps they would use to complete the task.
  2. Type the search query they would submit to the selected apps.
- The listed apps were among the most popular for Android
- The list was shuffled for each assignment



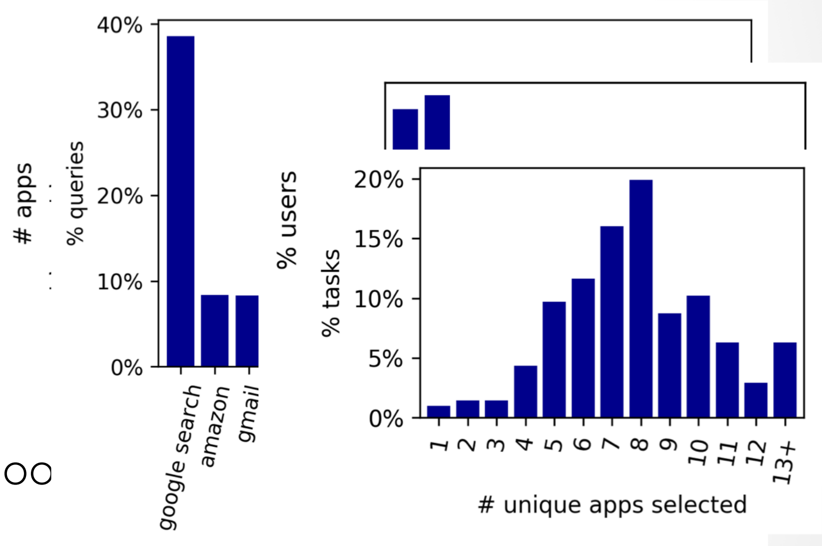
Results

- After defining 206 cross-app search tasks, we collected 5,812 query-app pairs
- The collection of search tasks and queries was called UniMobile
- Statistics of the collected data:
- Data publicly available at: http://bit.ly/UniMobile

# queries	5,812
# unique queries	5,567
# users	625
# search tasks	206
# unique apps	121
# unique first apps	70
# unique second apps	89
Mean unique apps per task	7.51 ± 10.57
Mean query per user	9.30 ± 20.30
Mean query per task	28.21 ± 12.72
Mean query terms	4.21 ± 2.45
Mean query characters	24.83 ± 12.88

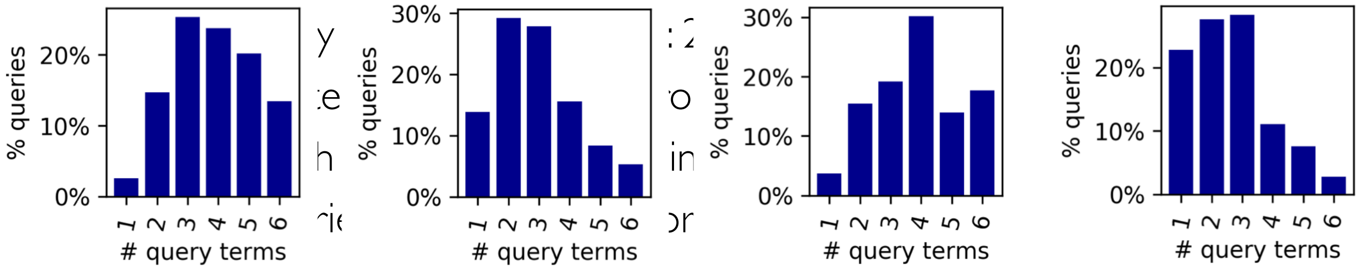
Data Analysis

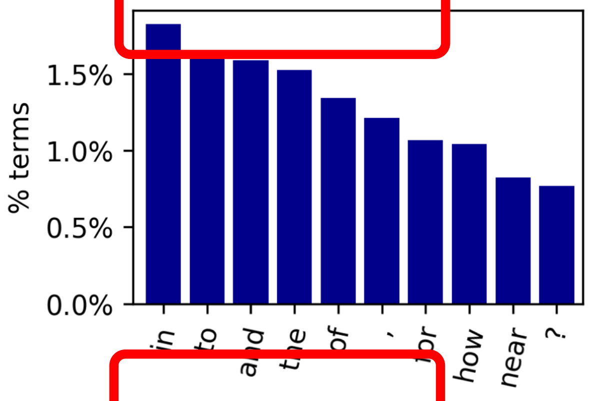
- Apps are distributed according to a power-law
- Apps are selected:
  - 39% of queries targeted to Google Search
  - 79% of workers selected more than one app
  - Users selected an average of 7.5 unique app
  - Google Search was the most selected app ir tasks
  - For 13% of the tasks, no workers selected Goo Search.



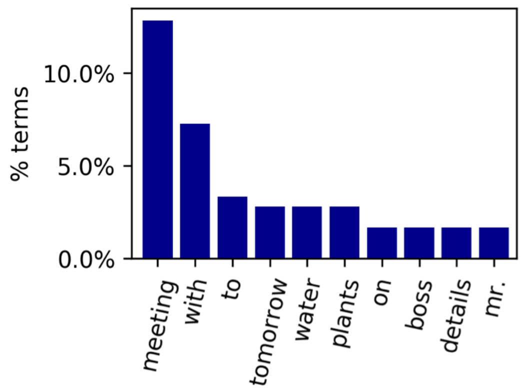
Data Analysis (cont.)

- Different app different query:
  - Different query term distributions are observed across apps.
    - Average query terms for Google Search: 4.82.

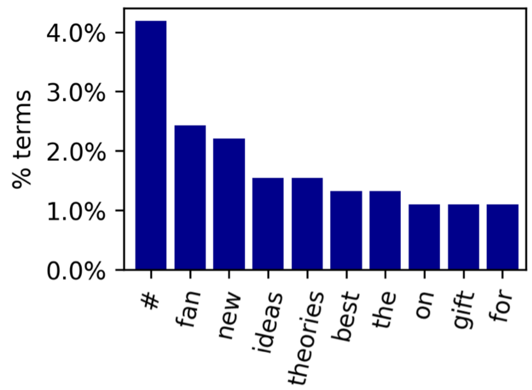




Google Search



Calendar



Twitter

Key Findings

- The majority of mobile queries are targeted to a few apps, each of which provides a different service
- Workers often chose two different apps for a single query, suggesting that many users submit the same query to multiple apps
- On average, workers selected more than 7 different apps to complete tasks
- Only 39% of queries were submitted to Google Search

Conclusion: mobile IR should really consider cross-app search

Mobile target apps selection

- Similar to resource selection in DIR: first step forward towards a unified mobile search
- User submits a query to the unified mobile search framework that ranks the target apps for the given query
- Assumptions:
  - Uncooperative resources: the mobile device cannot access apps' data
  - Heterogenous environment: different apps provide various types of information
  - Queries that can be passed to an app that do not need match with the content of the app's description
  - Apps' descriptions cannot be used for apps' representations

Neural Target Apps Selection

- We developed a Neural Target App Selection (NTAS) algorithm based on a deep learning architecture
- NTAS learns a high-dimensional representation for each app, based on their corresponding queries
- We propose two neural frameworks:
  1. NTAS1: App Scoring Model (i.e. provides a score for a given query-app pair)
  2. NTAS2: Query Classification Model (i.e. provides a a probability for each app targeted by query)

NTAS

- Given query q and a candidate app a:

$$\text{score} = \psi(\phi_Q(q), \phi_A(a))$$

que

$$\phi_Q(q) = \sum_{i=1}^{|q|} \widehat{\mathcal{W}}(w_i) \cdot \mathcal{E}(w_i)$$

ion

query terms

vocabulary

norm

weigh

$$\widehat{\mathcal{W}}(w_i) = \frac{\exp(\mathcal{W}(w_i))}{\sum_{j=1}^{|q|} \exp(\mathcal{W}(w_j))}$$

maps each vocabulary term to a real-valued number showing its global importance

embedding space

NTAS: learning

- Pointwise learning:

$$\mathcal{L}_{MSE}(b) = \frac{1}{|b|}\sum_{i=1}^{|b|}\left(y_i - \psi(\phi_Q(q_i), \phi_A(a_i))\right)^2$$

label in the i^{th} training instance

- Pairwise learning:

margin of hinge loss

$$\mathcal{L}_{Hinge}(b) = \frac{1}{|b|}\sum_{i=1}^{|b|}\max\Big\{0, \epsilon - \text{sign}(y_{i1} - y_{i2}) \\ \big(\psi(\phi_Q(q_i), \phi_A(a_{i1})) - \psi(\phi_Q(q_i), \phi_A(a_{i2}))\big)\Big\}$$

NTAS

- For a given query-app pair, it computes the probability of each app being targeted by a given query

$$\mathcal{L}_{ce}(b) = \frac{1}{|b|} \sum_{i=1}^{|b|} \sum_{j=1}^{N} (p(a_j|q_i) \log \underbrace{\gamma(\phi_Q(q_i))}_{\text{probability of the } k^{th} \text{ app being targeted}})$$

- γ is modeled using a fully-connected feed-forward network
- We use dropout to regularize the model

(full details in SIGIR2018 paper)

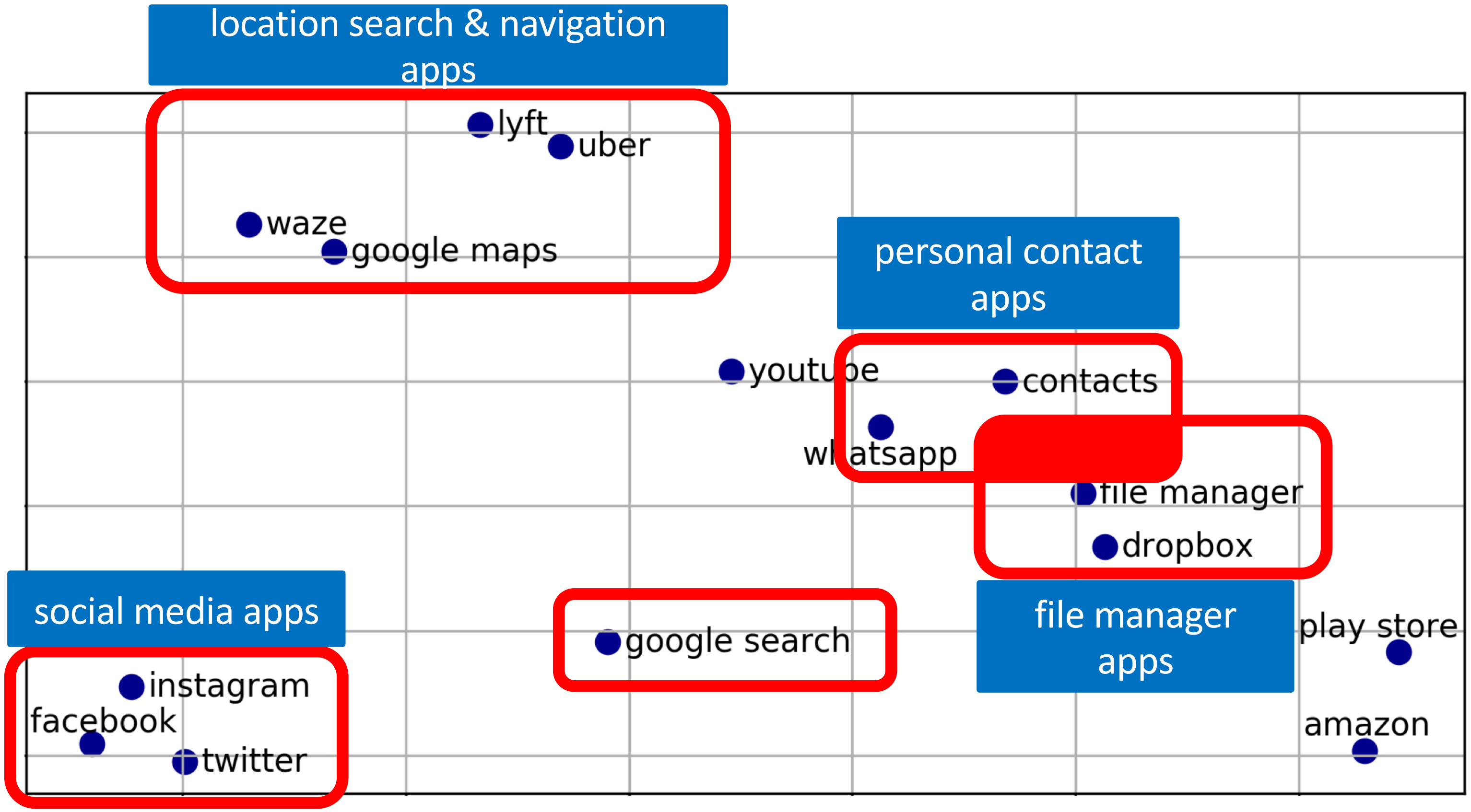
Experimental setup

- Data splits:
  - UniMobile-Q: random splits based on queries
  - UniMobile-T: random splits based on tasks
- Experiments repeated 5 times on different data splits.
- Evaluation metrics: MRR, P@1, nDCG@1, nDCG@3, nDCG@5
- Compared methods: StaticRanker (based on app popularity), QueryLM, BM25, BM25-QE, k-NN, k-NN-AWE, LambdaMART

Results

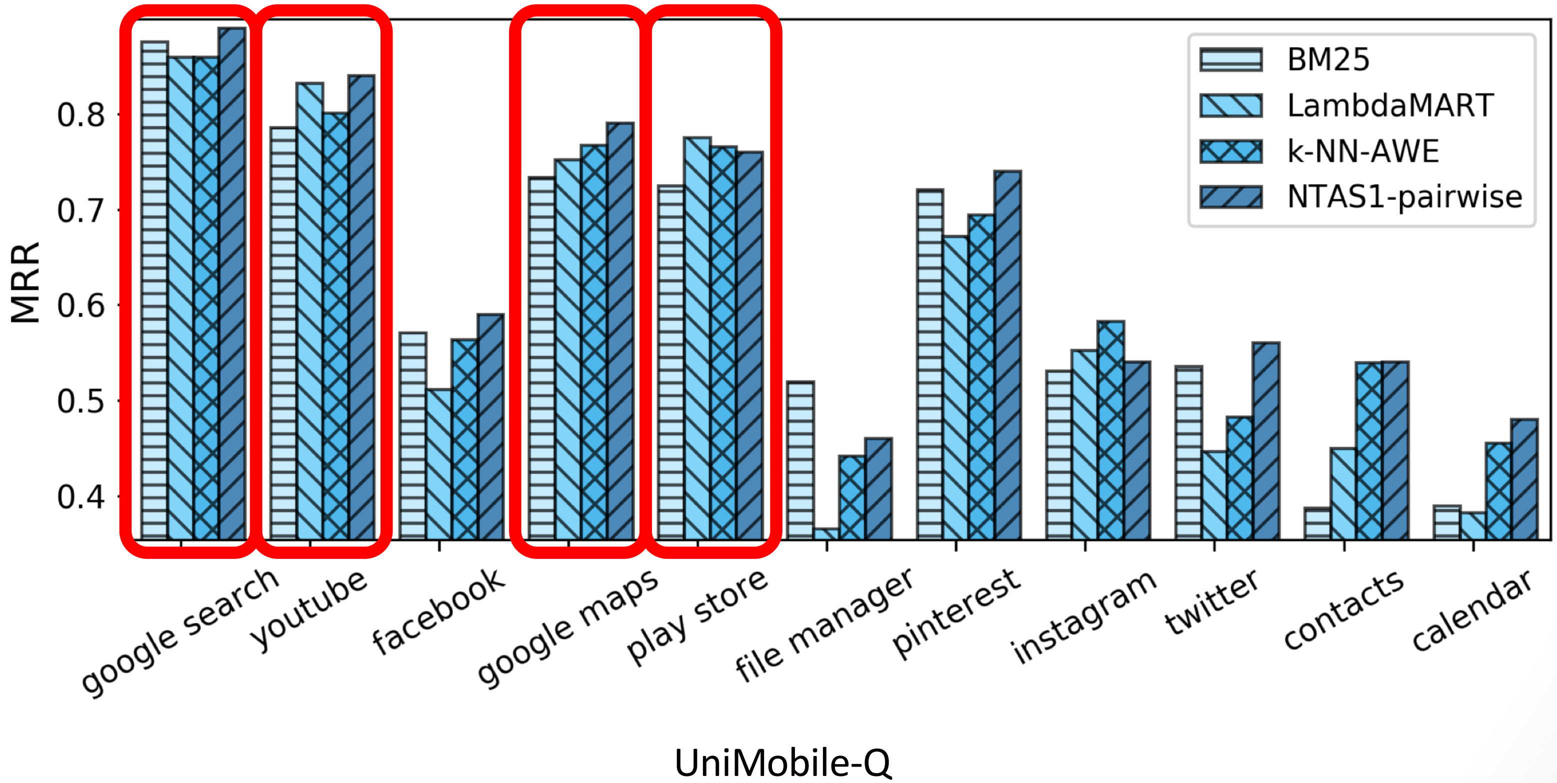
Method	UniMobile-Q Dataset					UniMobile-T Dataset				
	MRR	P@1	nDCG@1	nDCG@3	nDCG@5	MRR	P@1	nDCG@1	nDCG@3	nDCG@5
StaticRanker	0.6485	0.5293	0.4031	0.4501	0.5144	0.6718	0.5507	0.4247	0.4853	0.5446
QueryLM	0.5867	0.3803	0.3068	0.4676	0.5508	0.5178	0.3272	0.2619	0.3716	0.4503
BM25	0.7523	0.6233	0.4915	0.6298	0.6859	0.6780	0.5244	0.4101	0.5392	0.5992
BM25-QE	0.6948	0.5177	0.4116	0.5909	0.6498	0.6256	0.4276	0.3312	0.5015	0.5704
k-NN	0.7373	0.6031	0.4794	0.6091	0.6633	0.6879	0.5414	0.4287	0.5413	0.6003
k-NN-AWE	0.7420	0.6081	0.4842	0.6156	0.6682	0.6984	0.5551	0.4407	0.5560	0.6117
LambdaMART	0.7313	0.6127	0.4864	0.6110	0.6426	0.6749	0.5469	0.4323	0.5419	0.5704
NTAS1-pointwise	0.7591*	0.6214	0.4897	0.6328	0.6934*	0.7047*	0.5582*	0.4493*	0.5506*	0.6258*
NTAS1-pairwise	**0.7661***	**0.6285***	**0.5012***	**0.6364***	**0.7018***	**0.7192***	0.5661*	**0.4709***	**0.5941***	**0.6471**
NTAS2	0.7638*	0.6271*	0.4996*	0.6351*	0.6976*	0.7144*	**0.5723***	0.4608*	0.5689*	0.6334*

Results (cont.)



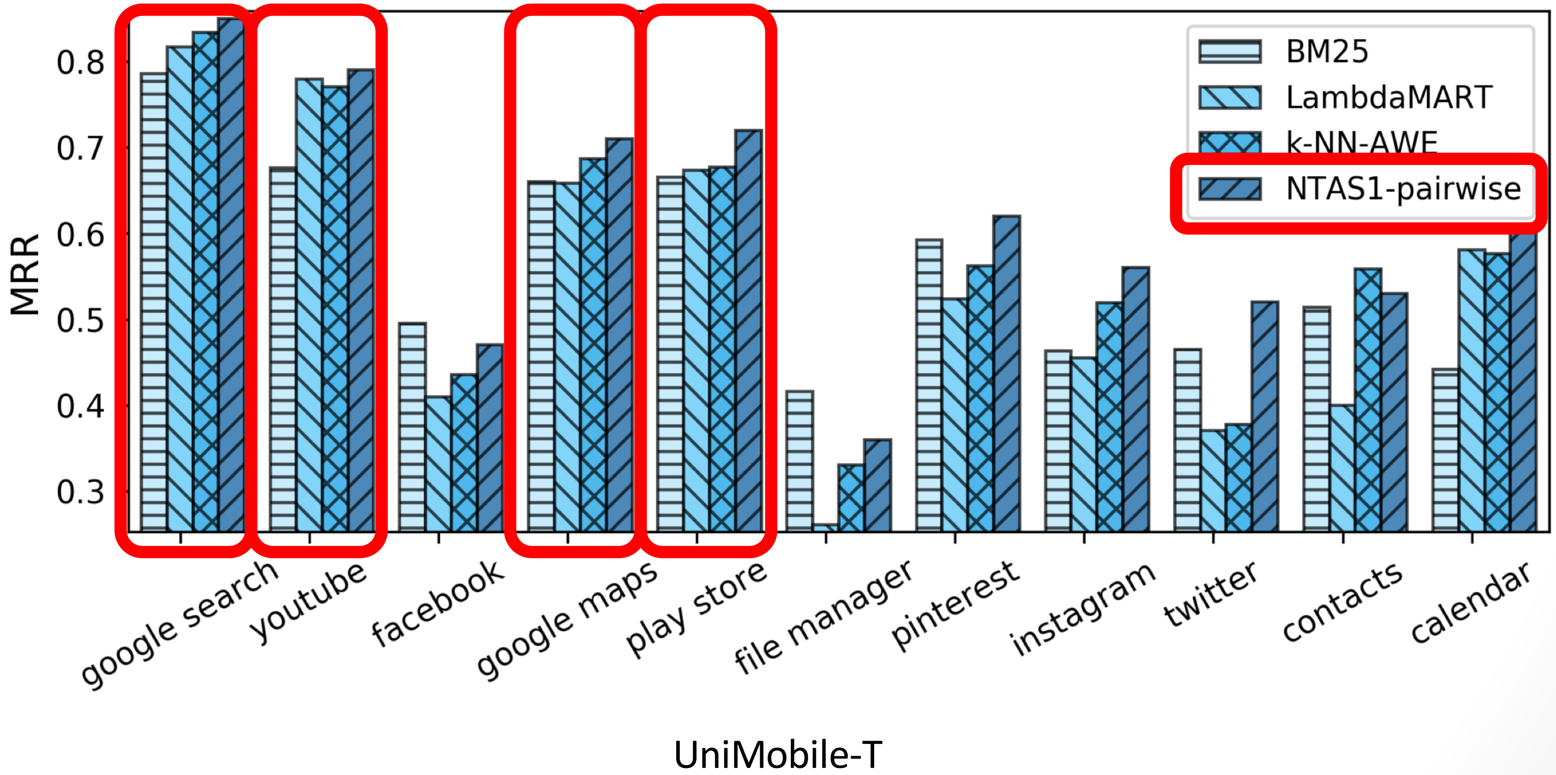
Results (cont.)

Personalisation effect in relation to queries



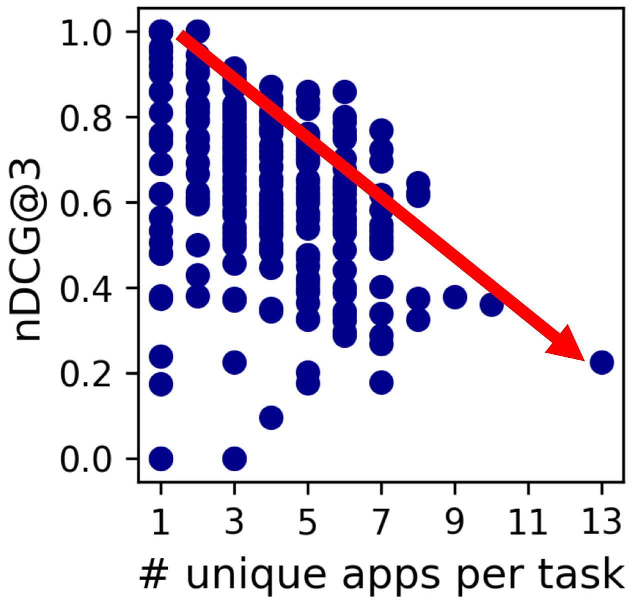
Results (cont.)

Personalisation effect in relation to tasks is even stronger!

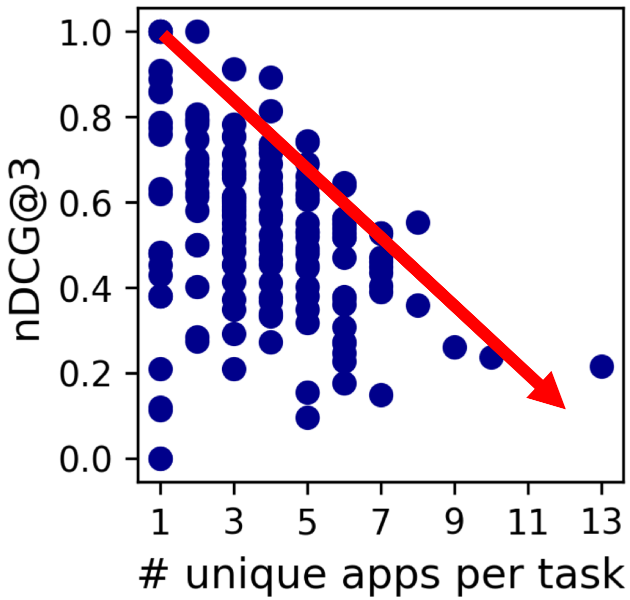


Results (cont.)

When the number of unique apps per task raises, performance drops (i.e. if a query can be done with different apps, then it is more difficult).



UniMobile-Q



UniMobile-T

Key Findings

- We introduced and studied the task of target apps selection
- We presented the first analysis of mobile cross-app search queries and user behaviors in terms of the apps chosen to complete different search tasks
- Unsurprisingly 39% of queries were submitted to Google Search
- Average unique apps per task: 7.5

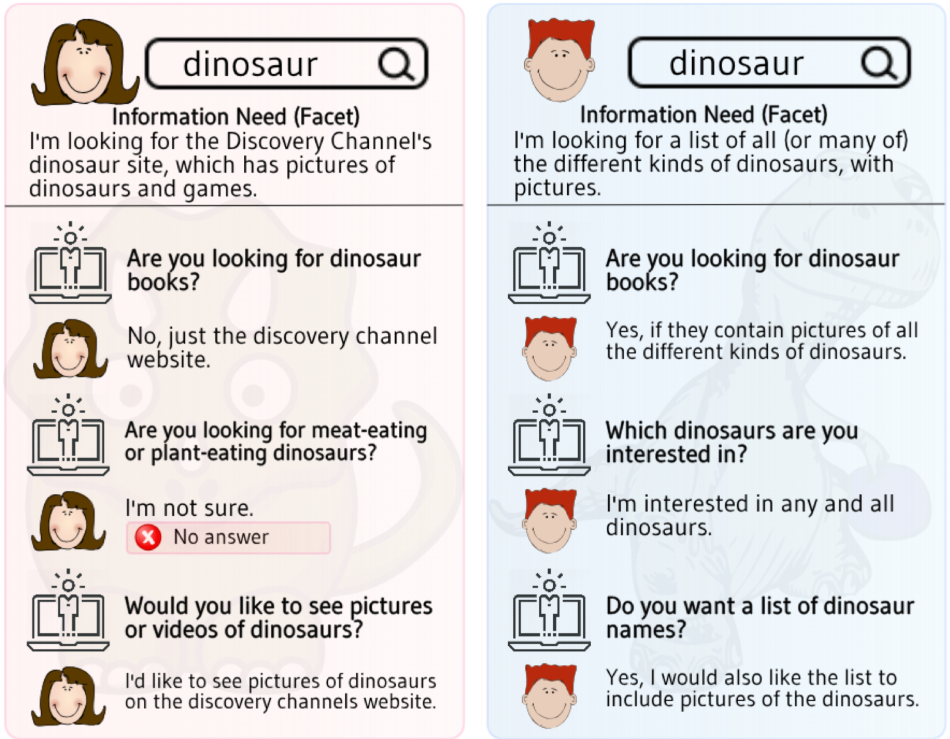
Conclusion: clear need for a unified mobile search framework and our first approach with UMS worked pretty well!

What next?

- Problem: users often fail to formulate their complex information needs in a single query
  - A long and complex interaction is needed to clarify the user's information need
  - Users need to scan multiple result pages and/or reformulate their queries
  - This is not easily done on a mobile terminal
- Solution: **voice-driven query interaction**!
  - Voice is a much easier and richer modality of interaction than text and would help the resource selection
  - It is the ideal modality for MobIR!
- We are currently working on engaging users with "clarifying questions" in open-domain information-seeking conversations

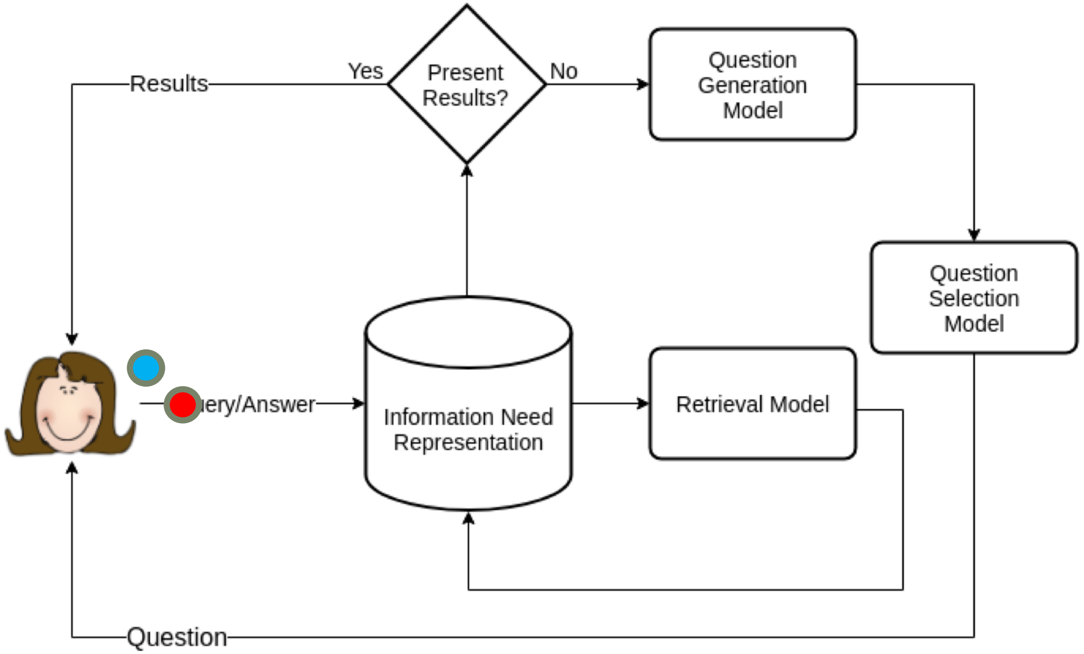
Why clarifying questions?

- To get a better understanding of what the user is looking for
- Here are a couple of examples of clarifying questions
  - From our Qulac collection
  - Clarifying questions are automatically generated from a large collection of clarifying questions collected using crowdsourcing
  - Very helpful for resource selection!



Voice-driven query interaction

- Here is how the interaction would take place:
  - Standard query
  - Clarifying query



Preliminary results: great potential impact

- Our experiments on an *oracle model* demonstrated that asking even only one good question lead to over 100% retrieval performance improvement
- However, sometimes you get it right … sometimes wrong:

Query	Facet Description	Selected Question	User's Answer
dog heat	What is the effect of excessive heat on dogs?	Would you like to know how to care for your dog during heat?	No, I want to know what happens when a dog is too hot.
sit and reach test	How is the sit and reach test properly done?	Do you want to know how to perform this test?	Yes, I do.
alexian brothers hospital	Find Alexian Brothers hospitals.	Are you looking for our schedule of classes or events?	No, I don't need that.
east ridge high school	Information about the sports program at East Ridge High School in Clermont, Florida	What information about East Ridge High School are you looking for?	I'm looking for information about their sports program.
euclid	Find information on the Greek mathematician Euclid.	Do you want a biography?	Yes.
rocky mountain news	Who are the sports reporters for the Rocky Mountain News?	Would you like to read recent news about the Rocky Mountain News?	No, I just want a list of the reporters who write the sports for the Rocky Mountain News.

(detai

Conclusions

- Personalised DIR is still a very powerful paradigm that can be effectively combined with Mobile IR and Voice Driven IR
- However, its deployment is difficult because of the many and different factors involved
- Considering *all available factors* and finding an *effective combination* of them is the best approach, but also the most difficult one!
- Success could result in a more effective and more user-friendly way to access distributed information
- But there are still many open challenges!

Future Work

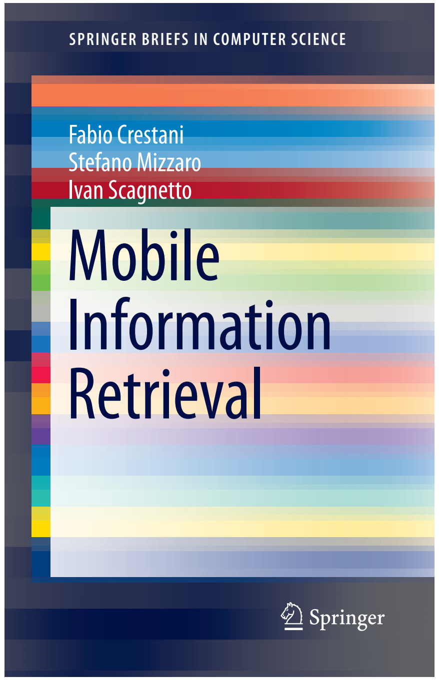
- Study result aggregation from multiple apps after submitting a query
- Add more information sources to the scoring algorithm and, possibly, use real data!
- Look into proper ways of using all contextual information acquired
- Study the best ways of presenting aggregated search results on the mobile phone's screen

Thanks to …

- The last part of this work was carried out by **Mohammad Alliannejadi** (now a final year PhD student), supported by the SNSF under the project *"Relevance Criteria Combination for Mobile Information Retrieval (RelMobIR)"*
- Full details of this work can be found the following papers:
  - Mohammad Aliannejadi, Fabio Crestani: Venue Appropriateness Prediction for Personalized Context-Aware Venue Suggestion. *SIGIR 2017*: 1177-1180.
  - Mohammad Aliannejadi, Fabio Crestani: Personalized Context-Aware Point of Interest Recommendation. *ACM Trans. Inf. Syst.* 36(4): 1-28, 2018.
  - Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, W. Bruce Croft: Target Apps Selection: Towards a Unified Search Framework for Mobile Devices. *SIGIR 2018:* 215-224.

Advertisement

Read:



It will do you <u>GOOD</u>!

<u>DISCLAIMER NOTICE!!!</u>

(also available at CoRR – arXiv)

Thank you very much for your attention!

Questions

also to fabio.crestani@usi.ch