Re-ranking Web Service Search Results Under Diverse User Preferences

Dimitrios Skoutas Mohammad Alrifai Wolfgang Nejdl

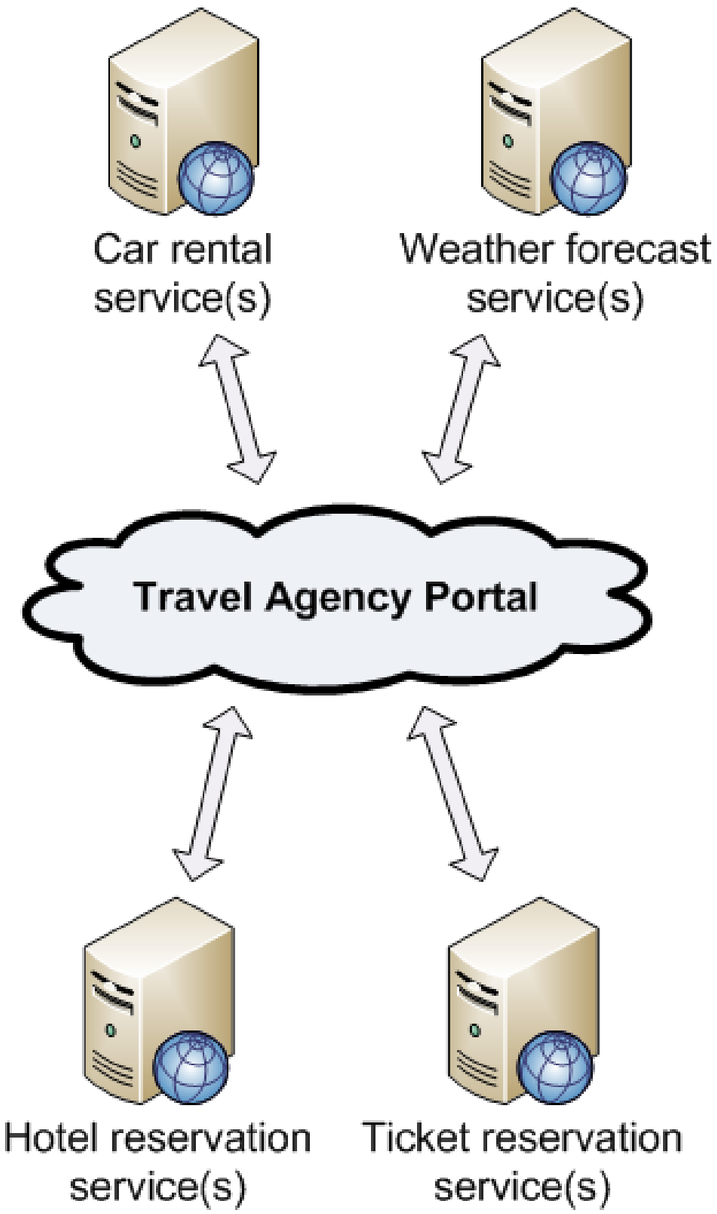
L3S Research Center
Hanover, Germany

{skoutas,alrifai,nejdl}@L3S.de

WEB OF SERVICES

Web services (in the traditional sense)

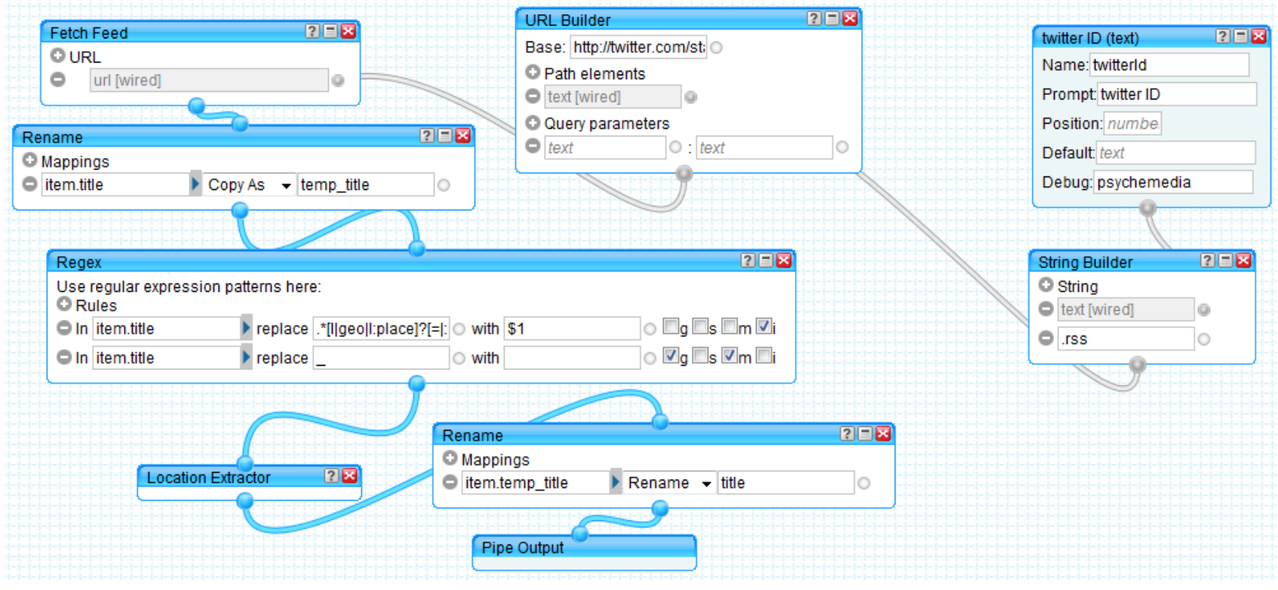
- network accessible software entities
- performing well-defined, self-contained operations
- concretely specified interfaces in machine-processable format
- loosely coupled, reusable, composable
- typically used for Enterprise Application Integration and B2B communication



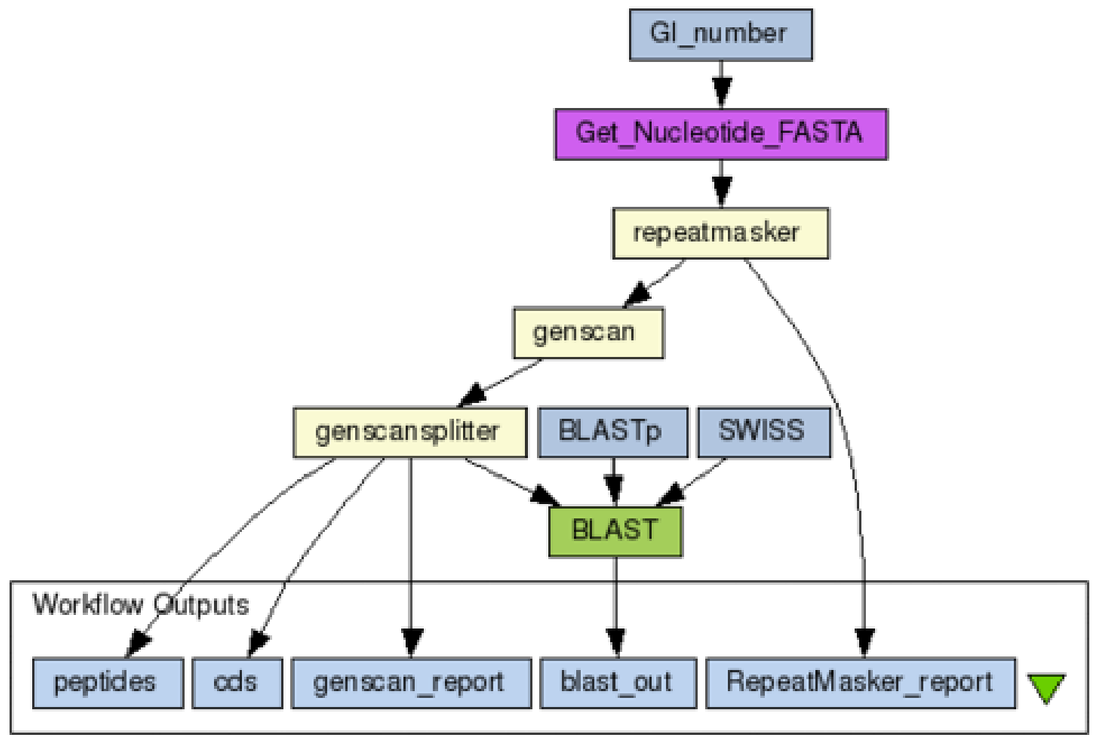
WEB OF SERVICES

Web services (in new trends)

- mashups
- e-Science
- search computing



Geo-Twitter Mashup



Genome annotation pipeline

upcoming EVENTS & THINGS TO DO

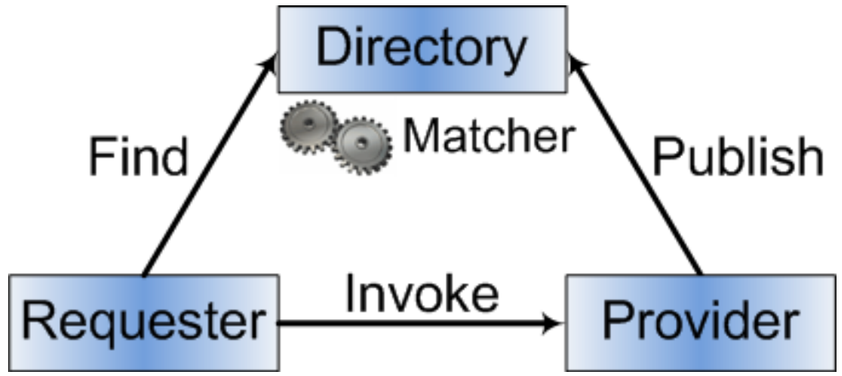
WIKIPEDIA The Free Encyclopedia

Google maps

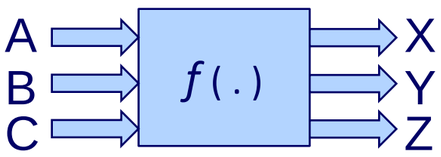
last.fm

find rock festivals nearby central European capital cities…

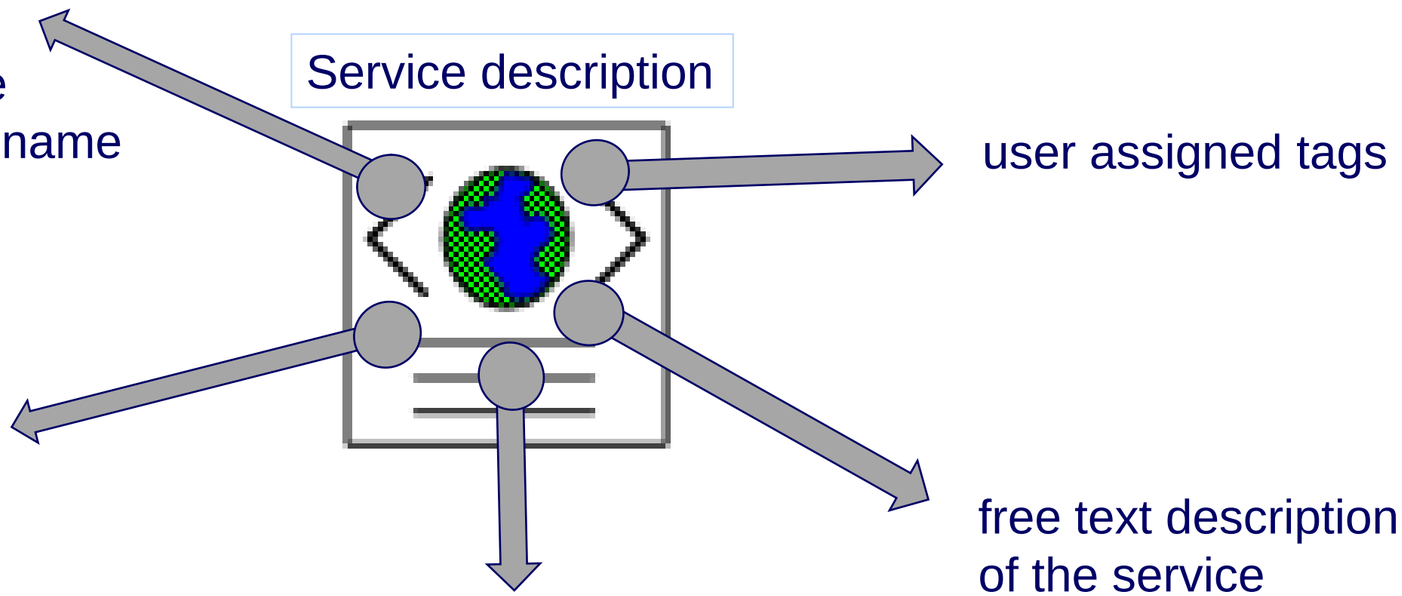
SERVICE DISCOVERY



Basic service discovery architecture



functional view of the service
e.g., *Input*: book title, author name
Output: book price



user assigned tags

free text description
of the service

QoS parameters
- price
- response time
- availability
- performance
- reliability
- …

other service attributes
- provider
- category
- protocols
- formats
-…

SERVICE DISCOVERY

Possible solutions

- maintain preference information in user profiles
- drawbacks
  - difficult to create and maintain
  - preferences often depend on the particular information need or point in time
- collect user preferences at query time
- drawbacks
  - tedious for the user
  - the user often does not know the available options

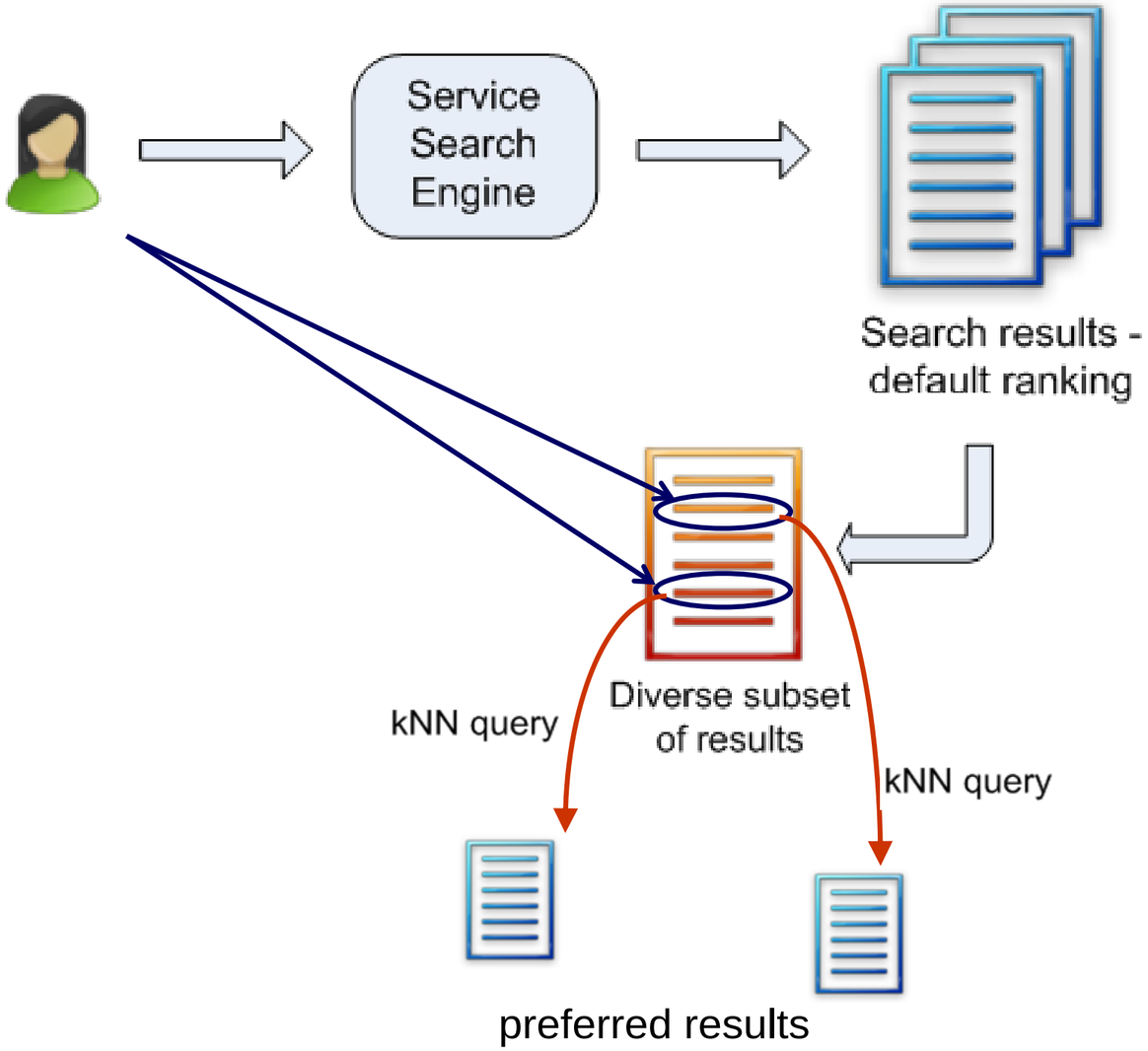
The problem:

- service descriptions are complex objects with different types of parameters
- often many similar services may partially match a user request
- different users have different preferences

SERVICE DISCOVERY

Proposed approach

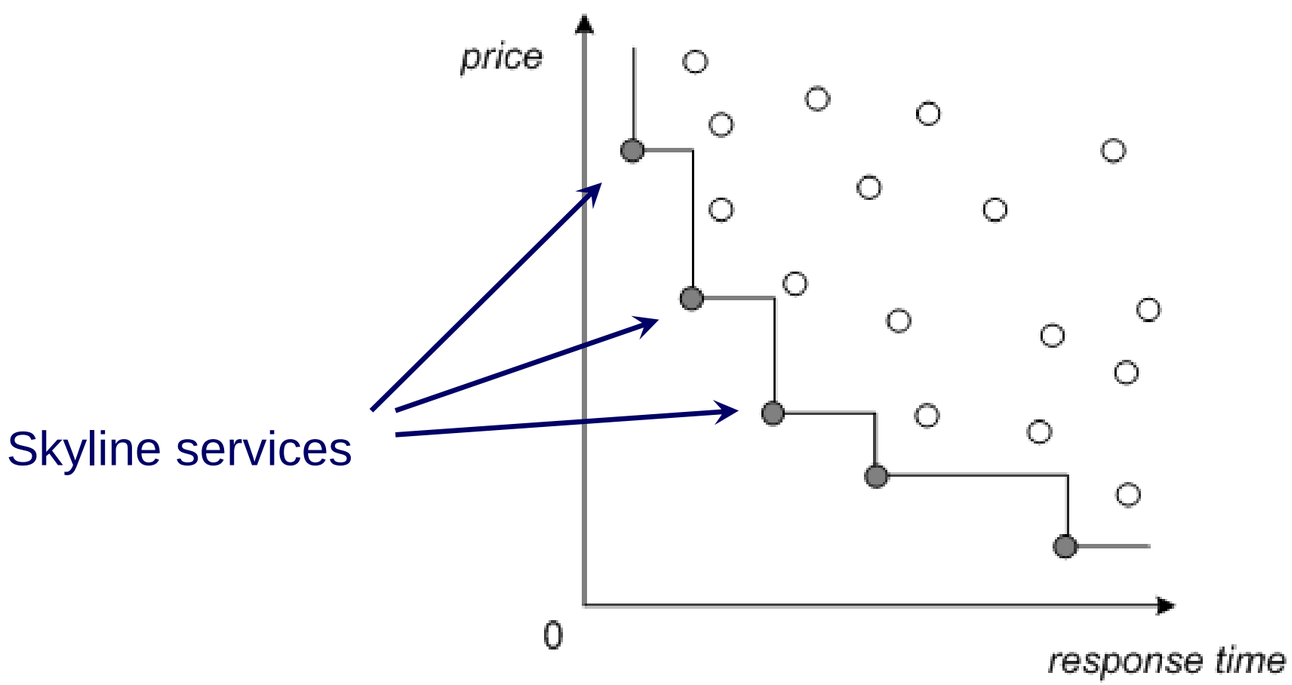
- implicitly identify user preferences at query time
- allow the user to select among different options
- *requirement*: the presented results are *diverse* and *representative*, but also *relevant*



SELECTING DIVERSE RESULTS

Deal with trade-off between different parameters

- a skyline-based approach works well for numeric attributes with total ordering [1]



[1] Skoutas et al. Top-k dominant web services under multi-criteria matching. EDBT 2009

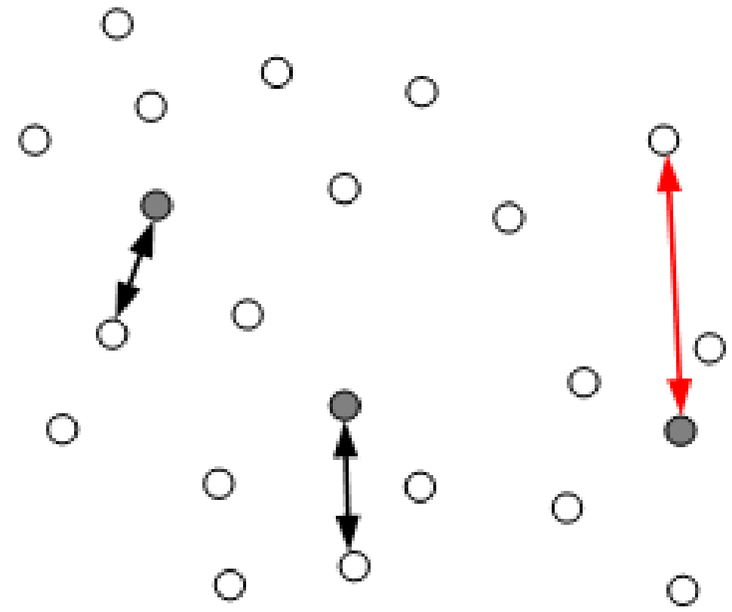
OBJECTIVE FUNCTION

Representation error

- select a subset of k results (S_k)
- for each non-selected result S, find the most similar one in the selected list

$$cerr(S, \mathcal{S}_k) = \min_{S' \in \mathcal{S}_k} dist(S, S')$$

- Objective: minimize the maximum representation error of the non-selected results



Objective Function

Combine representation error and scoring of individual services:

$$f(\mathcal{S}_k) = \max_{S \in \mathcal{S}} \{dom(R,S)^{\lambda} \times cerr(S, \mathcal{S}_k)\}$$

degree of match computed on input and output parameters

- ontology-based match (distance measure on the ontology graph)
- parameter name matching (string similarity)

utility score (aggregate QoS values)

(dis)similarity measure on other attributes (Jaccard coefficient)

$$sim(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Objective: minimize *f*

ALGORITHM

- Computing the optimal solution of the objective function *f* is NP-hard
- Greedy algorithm (2-approximation, based on the minimum *k*-center problem)
  - Initialize S_k with the service having the highest individual score
  - For steps 2 to *k*
    - find the service s with the maximum (weighted) representation error w.r.t. the current contents of S_k
    - add *s* to S_k

EVALUATION

Dataset: OWLS-TC v2[1]

- 1007 real-world web service descriptions
- 28 service requests
- semantically annotated in OWL-S using ontologies from 7 different domains
- Synthetically added nominal attributes
  - 4 attributes (message encoding, security protocol, transport binding protocol, transaction protocol) with 4, 12, 3 and 9 distinct values, respectively
  - values randomly assigned to each service for these attributes

[1] http://projects.semwebcentral.org/projects/owls-tc/

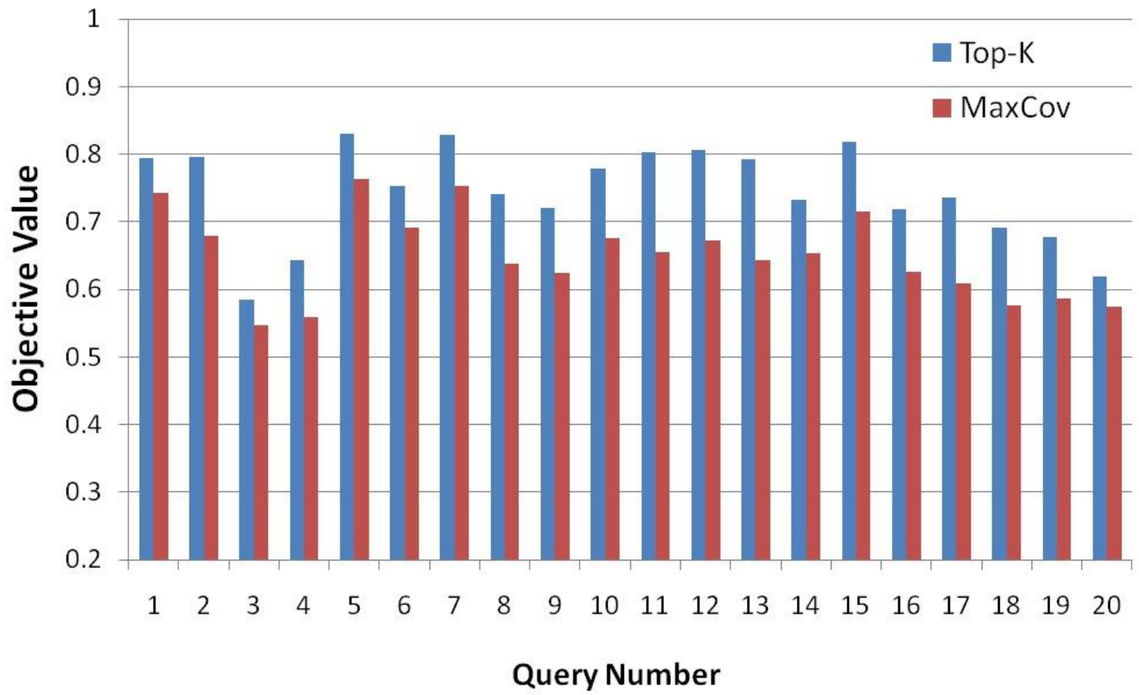
EVALUATION

Functional degree of match computed
- using an ontology reasoner to compare classes annotating service input/output parameters (quantified as the ratio of common ancestors)

Service distance computed using
- Jaccard coefficient on the added attributes

Comparison:

- *Top-k*: default ranking using only the functional degree of match
- *MaxCov*: re-ranked top-*k* results selected using the objective function *f*



Future Work

Evaluation

- user study to evaluate the quality of the results
- comparison with other diversification algorithms (e.g., [2])

Combine with user profiles

- user choices indicate implicit preferences
- query dependent and relative
- store in user profile and use to infer preferences in similar contexts

[2] S. Gollapudi, A. Sharma: An axiomatic approach for result diversification. WWW 2009

THANK YOU

Questions?