

iRDQL - Imprecise RDQL Queries Using Similarity Joins¹

Abraham Bernstein and Christoph Kiefer



Overview

Typical problem when *querying*: not all relevant results are found or you find yourself buried in results

➔ Use similarity measures to improve the reach and ranking of the results

- *iRDQL* enables the user to query for similar resources in ontologies by the use of *Similarity Joins*
- Ranking is done by similarity measures which determine the degree of similarity between resources
- All similarity measures are implemented in *SimPack* (see below)
- Evaluation results show the usefulness of our approach to **improve recall** without overly sacrificing precision

SimPack Generic Java Library of Similarity Measures for the Use in Ontologies

Feature Vector Similarity

Object Features:

$$x' = \begin{pmatrix} 0 \\ name \\ type \end{pmatrix} \Rightarrow x = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, y' = \begin{pmatrix} age \\ 0 \\ type \end{pmatrix} \Rightarrow y = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Vector Measures:

$$simDice(x, y) = \frac{2 \cdot x \cdot y}{\|x\|_2^2 + \|y\|_2^2}$$

$$simJac(x, y) = \frac{x \cdot y}{\|x\|_2^2 + \|y\|_2^2 - x \cdot y}$$

$$simCos(x, y) = \frac{x \cdot y}{\|x\|_2 \cdot \|y\|_2}$$

String Similarity

Levenshtein string edit distance:

$$simLev(Rx, Ry) = \frac{form(x, y)}{form_wc(x, y)}$$

TFIDF (term frequency – inverse document frequency):

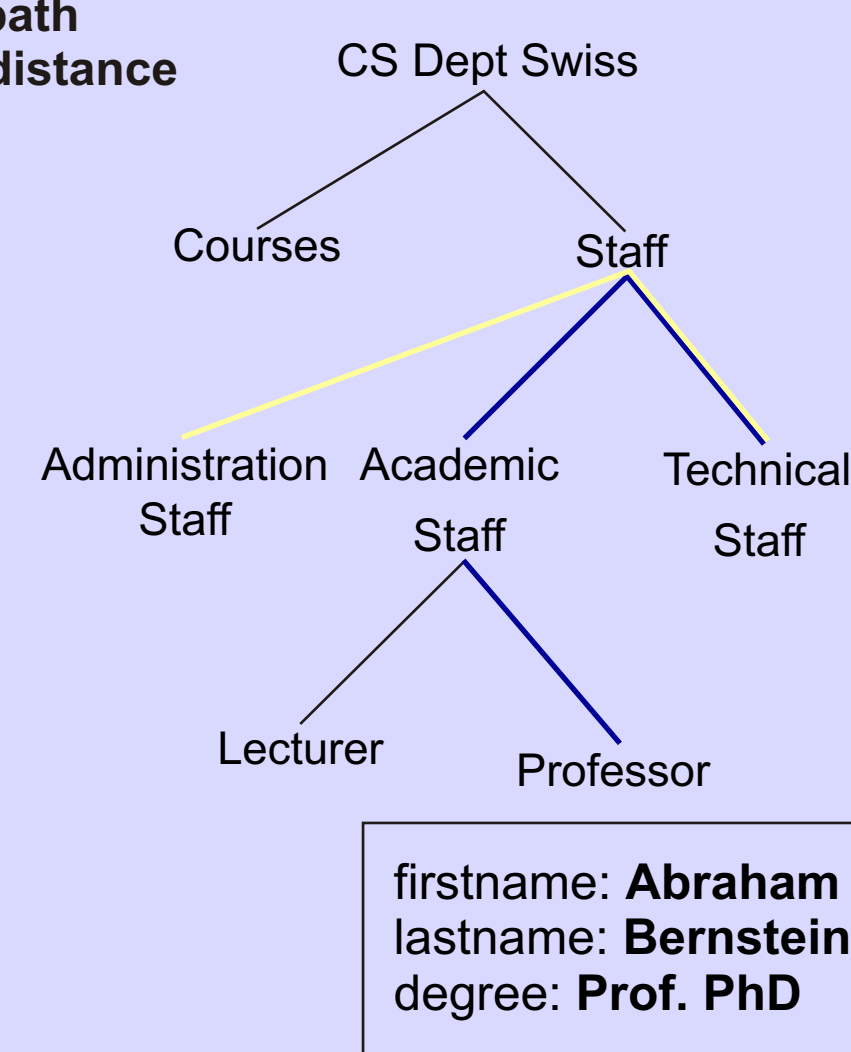
- term weights w_{ij} of term i in document j
- tf_{ij} : number of occurrence of i in j
- df_i : number of documents containing i

$$w_{ij} = tf_{ij} \times \log\left(\frac{N}{df_i}\right), tf_{ij} = \frac{n_i}{\sum n_k}$$

Tree/Graph Similarity

Tree/Graph comparison:

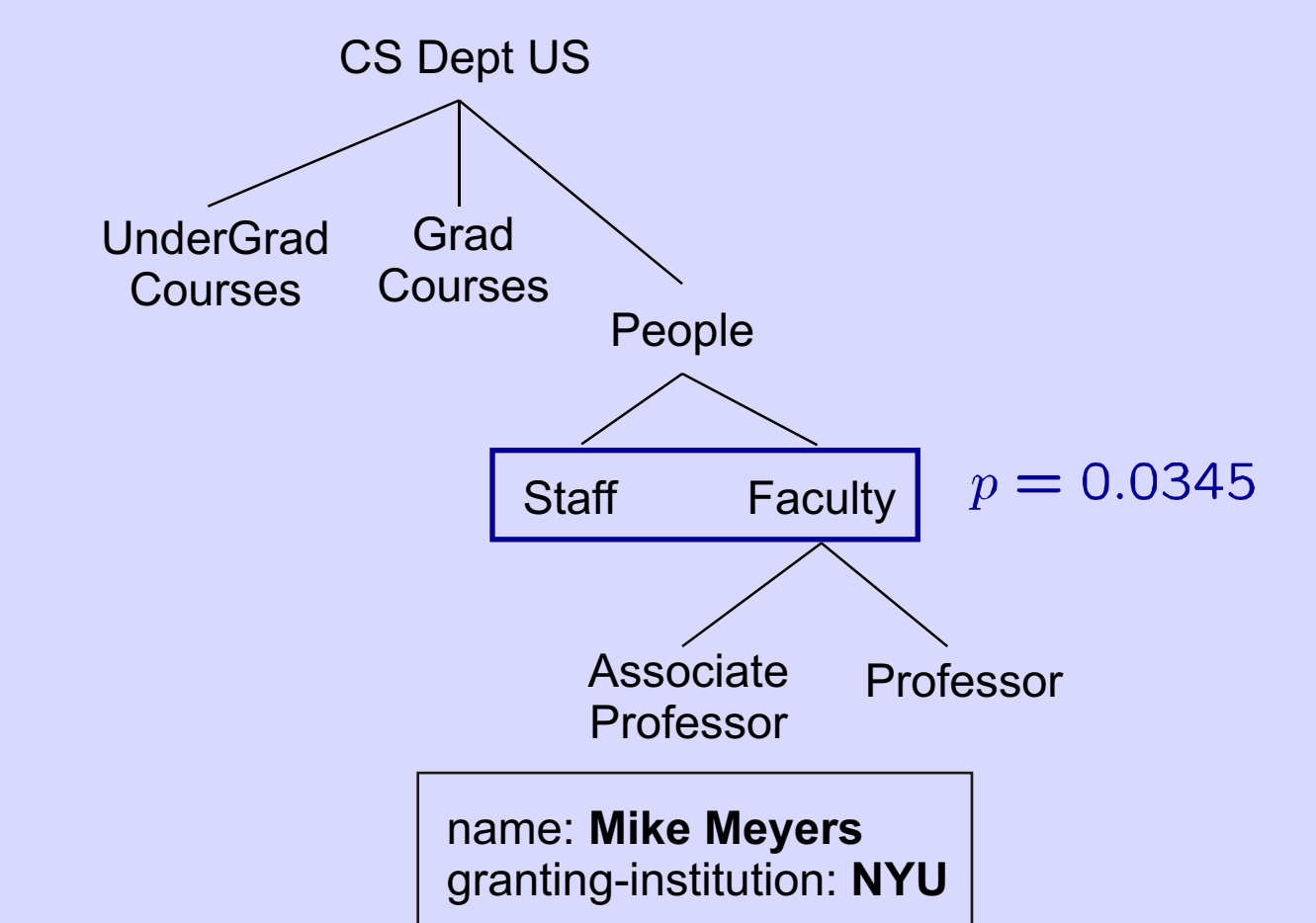
- Shortest path
- Tree-edit distance



Object Similarity

Amount of information contained in objects:

$$sim(Staff, Faculty) = -\log 0.0345$$



iRDQL Extending RDQL With Similarity Joins

3 additional language constructs

➔ IMPRECISE, SIMMEASURE, OPTIONS

```
SELECT ?s1, ?p1, ?p2
WHERE (?s1 presents ?p1)
      (?p2 serviceName "beach surfing")
IMPRECISE ?p1, ?p2
SIMMEASURE Levenshtein
OPTIONS IGNORECASE false THRESHOLD 0.7
```

Similarity Join

?s1	?p1	?p2	sim
Beach Surfing Service	Beach Surfing Profile	Beach Surfing Profile	1.0
Beach Broker Service	Beach Broker Profile	Beach Surfing Profile	0.85
Abstract Broker Service	Abstract Broker Profile	Beach Surfing Profile	0.7

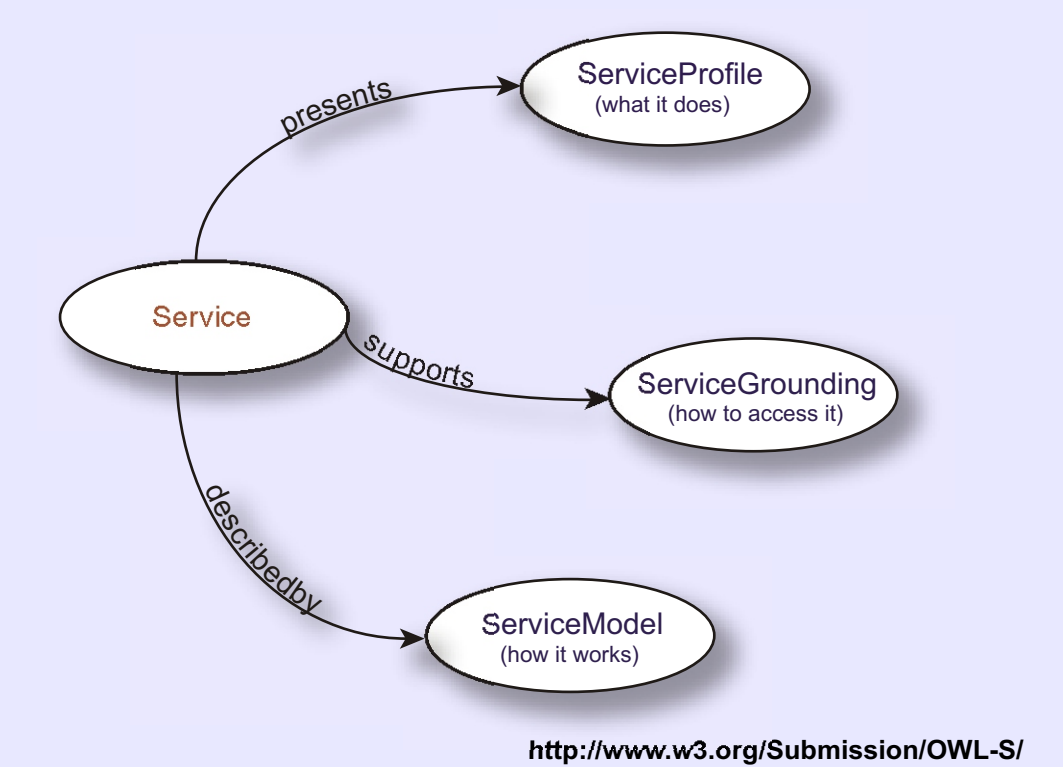
OWL-S Test Collection² Queries and Answers

406 OWL-S services of 6 different domains

Query service

```
<?xml version="1.0" encoding="UTF-8"?>
<service:Service rdf:ID="CITY_BROKER_SERVICE">
  <service:presents rdf:resource="#CITY_BROKER_PROFILE"/>
  <service:describedBy rdf:resource="#CITY_BROKER_PROCESS_MODEL"/>
  <service:supports rdf:resource="#CITY_BROKER_GROUNDING"/>
</service:Service>

<Profile:Profile rdf:ID="CITY_BROKER_PROFILE">
  <service:isPresentedBy rdf:resource="#CITY_BROKER_SERVICE"/>
  <profile:serviceName xml:lang="en">
    hotel reservation booking service
  </profile:serviceName>
  <profile:textDescription xml:lang="en">
    Provide the best hotel reservation system in a given city.
  </profile:textDescription>
  <profile:hasInput rdf:resource="#_CITY"/>
  <profile:hasOutput rdf:resource="#_BROKER"/>
  <profile:has_process rdf:resource="CITY_BROKER_PROCESS" />
</profile:Profile>
```

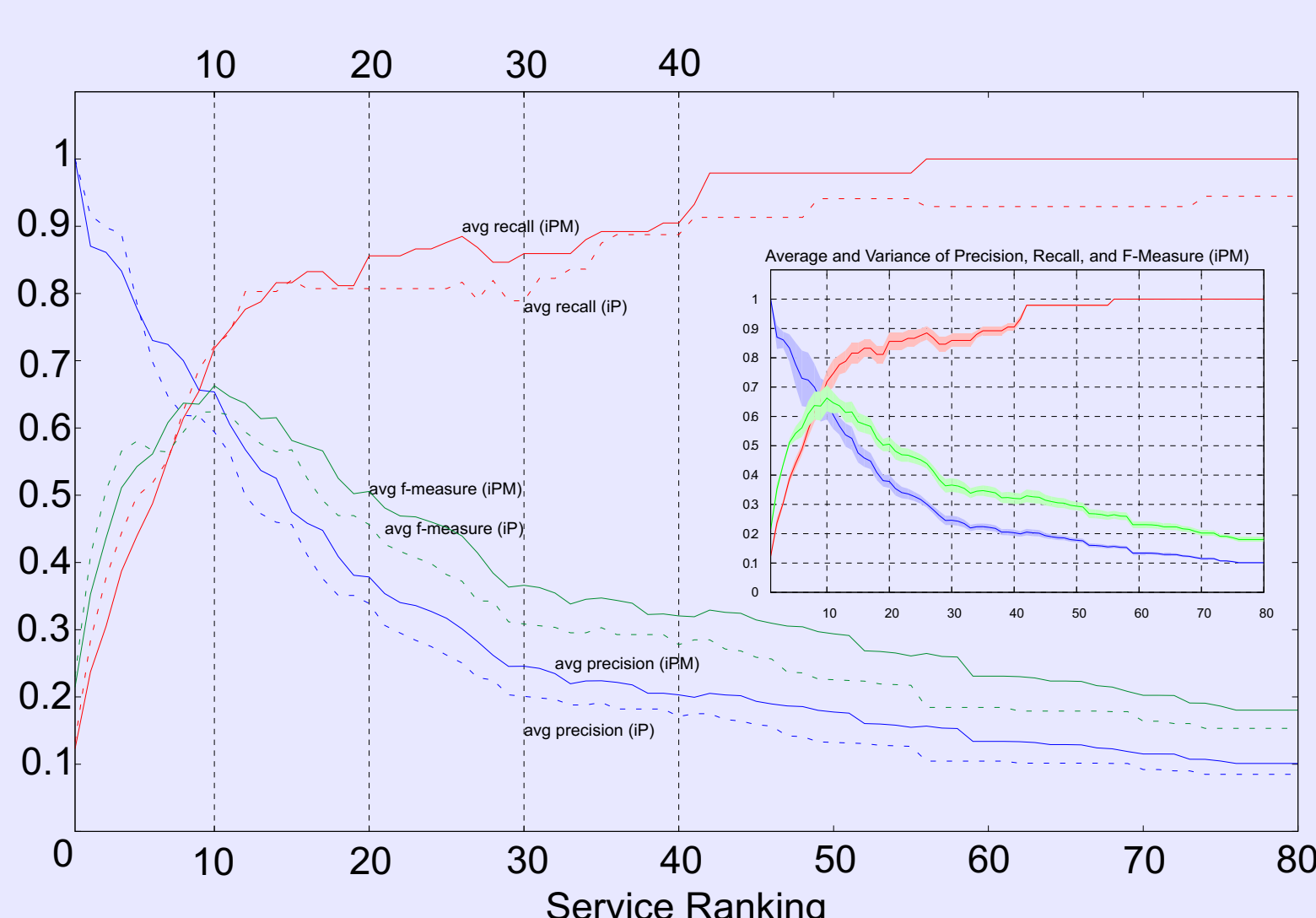


Query relevance set

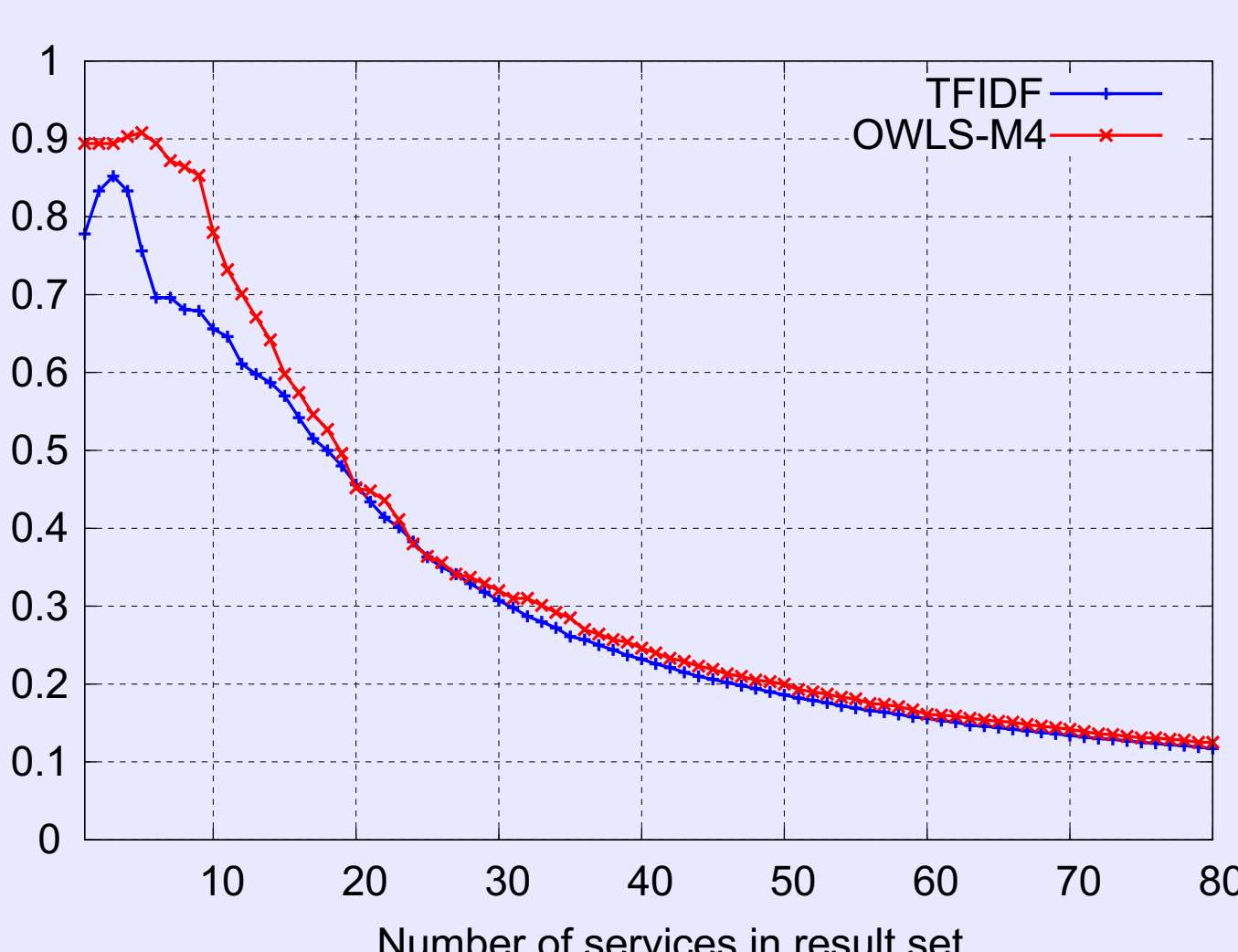
```
city_broker_service.owl
city_broker_service2.owl
city_financial_agent_service.owl
city_financial_agent_service1.owl
urbanarea_financial_agent_service.owl
city_organization_service.owl
...
```

Experiments Different Types of Similarity Joins & iRDQL vs. OWLS-M4²

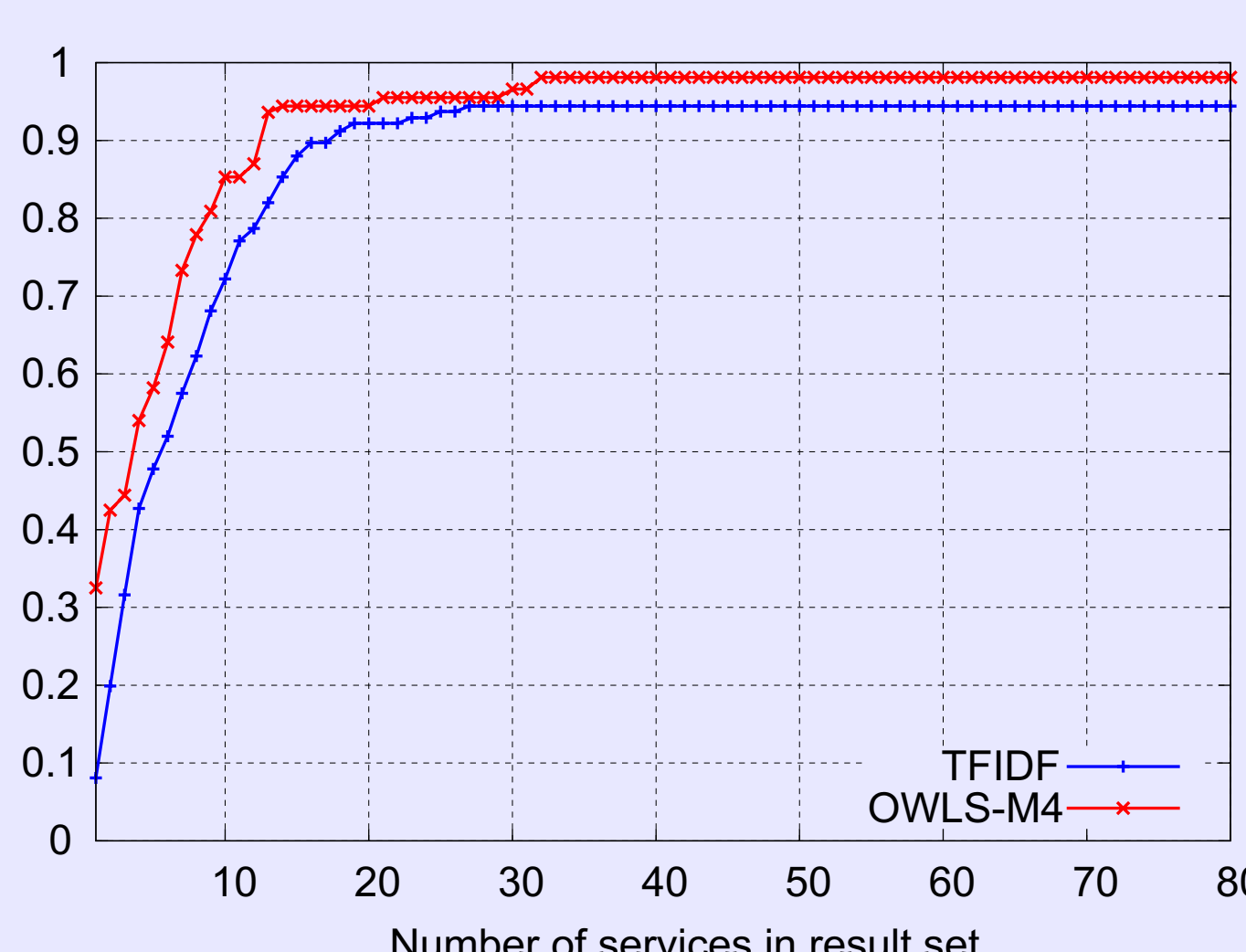
Precision, Recall and F-Measure (iP vs iPM)
Levenshtein



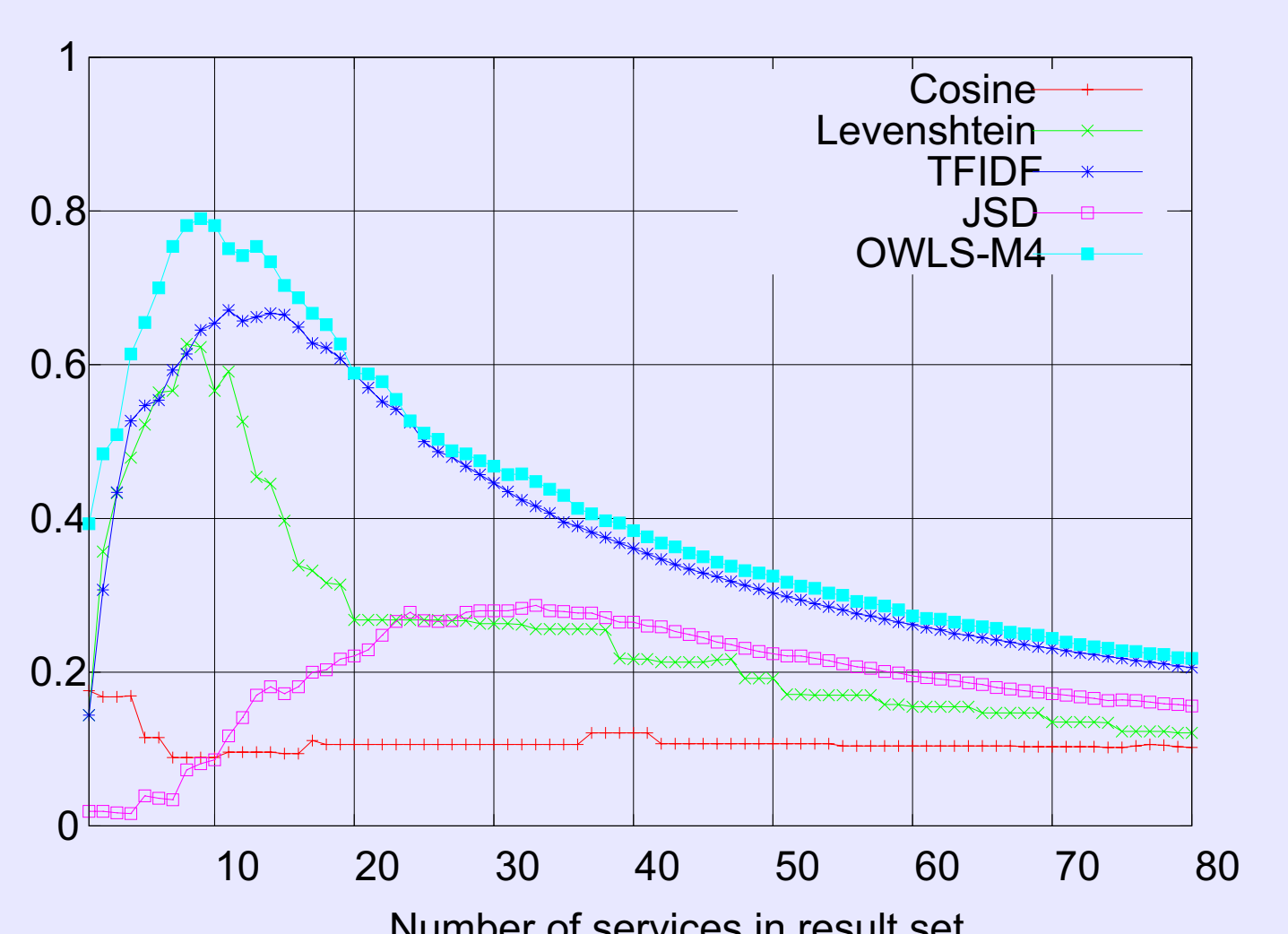
Precision



Recall



F-Measure



[1]iRDQL - Imprecise RDQL Queries Using Similarity Joins. Abraham Bernstein and Christoph Kiefer. Third International Conference on Knowledge Capture (K-CAP). Banff, Alberta, Canada. October, 2005.

[2]OWLS-MX: Hybrid Semantic Web Service Retrieval. Matthias Klusch, Benedikt Fries, Mahboob Khalid and Katia Sycara. 1st International AAAI Fall Symposium on Agents and the Semantic Web. 2005.