Entity track 2011

Planning Session

Krisztian Balog (University of Amsterdam)
Pavel Serdyukov (Delft University of Technology)
Arjen P. de Vries (CWI)

Agenda

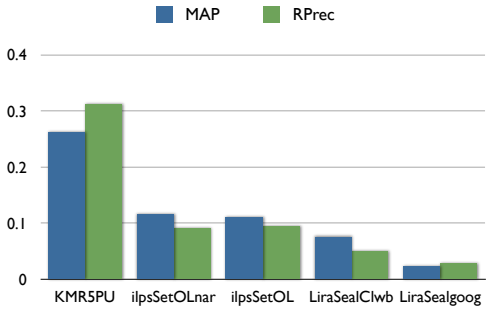
- Entity List Completion results
- Collecting resources
- Plans, wishes, desires for 2011

Entity List Completion

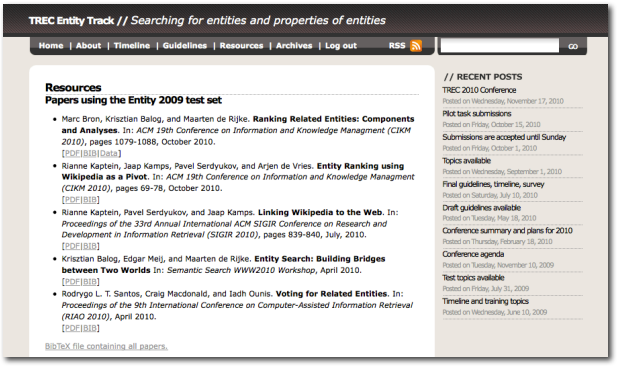
qID	Topic	#Examples	#Relevant
4	Professional sports teams in Philadelphia.	6	5
5	Products of Medimmune, Inc.	2	1
7	Airlines that currently use Boeing 747 planes.	23	25
11	Donors to the Home Depot Foundation.	6	8
12	Airlines that Air Canada has code share flights with.	13	17
15	Universities that are members of the SEC conference for football.	10	3
17	Chefs with a show on the Food Network.	22	21
20	Scotch whisky distilleries on the island of Islay.	7	1

Results

(3 groups, 5 runs)



http://ilps.science.uva.nl/trec-entity/



Plan for Entity 2011

- Repeat the REF task
- Turn the ELC pilot to an official task

Related Entity Finding

Return a ranked list of entity homepages, given

Input entity

Name + Homepage

Target type

- Person
- Organization
- Product
- Location

"Shops in Amsterdam that sell the Samsonite X'Blade laptop backpack"

Narrative

Description of the relation in free text

Related Entity Finding

- 50 new topics
- More fine grained type definition using the DBPedia ontology
- Drop "relevant" only "primary"
- Supporting document needed
- How should we go about Wikipedia?
  - Wikipedia page doesn't count as primary

REF 2010 qrels

`qID ClueWebID name relDoc class relName`

- `relDoc`: Judgment for the document
  - 0 = not relevant
  - 1 = relevant
  - 2 = primary
- `class`: Class number for the document
- `relName`: Judgment for the name for the page returned. Only if the page is primary!
  - 0 = incorrect
  - 1 = inexact
  - 2 = correct

Getting names for correct answers

```
grep " 2 .* 2$" ent-ref10.qrels | cut -d" " -f1,5,3
```

```
29 hewlett_packard 9
29 hewlettpackard 9
29 hewlettpackard_co 9
48 air_line_pilots_association 9
48 air_line_pilots_association_(alpa) 9
48 air_line_pilots_association_28alpa29_ 9
48 air_line_pilots_association__alpa__ 9
48 air_line_pilots_association_alpa 9
48 air_line_pilots_association_international 9
```

Evaluation based on names is strongly discouraged!
There are less names than primary homepages!

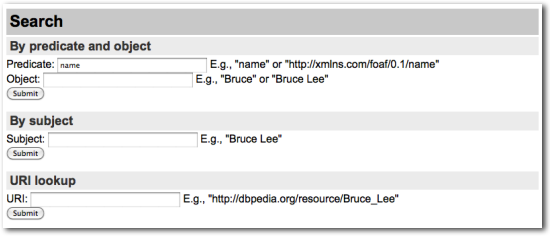
Entity List Completion

- Same as the REF task
- Entities are represented by a unique URI in the Linked Open Data Cloud
- Topic definition includes example entities

Example topic

```
<query>
<num>7</num>
<entity_name>Boeing 747</entity_name>
<entity_URL>clueweb09-en0005-75-02292</entity_URL>
<target_entity>organization</target_entity>
<narrative>Airlines that currently use Boeing 747 planes.
</narrative>

<!-- Available only for the ELC task -->
<entity_URIs>
 <URI>http://dbpedia.org/resource/Boeing_747</URI>
</entity_URIs>
<target_type_dbpedia>dbpedia-owl:Airline</target_type_dbpedia>
<examples>
 <entity>
  <URI>http://dbpedia.org/resource/Northwest_Airlines</URI>
  <URI>http://www.daml.org/2002/08/nasdaq/nasdaq#NWAC</URI>
 </entity>
 <entity>
  <URI>http://dbpedia.org/resource/British_Airways</URI>
  <URI>http://twitter.com/British_Airways</URI>
 </entity>
 ...
</examples>
</query>
```

Interface for exploring the collection

http://zookst1.science.uva.nl:8888/btc-webapp-0.2/btc2009