**Resource selection**

**ODP**
- Similarity between ODP’s query and resource categories
- Jaccard

**Retrieval model**
- Build pseudo-document and retrieve best matching resources
- Title: Results’ title
- Desc: Results’ snippet

**Hybrid run**
- Aggregates rankings from the other methods using Borda voting
- TF-IDF
- BM25

**FedWeb 2012**
- 108 resources
- Top 10 results (snippets + pages)
- Dec 2011 – Jan 2012

**FedWeb 2013**
- 156 + 1 resources (specific + BigWeb)
- Top 10 results (snippets + pages)
- Apr – May 2013

**Results merging**

**Relevance**
- Documents ranked with respect to the query likelihood model:
  \[
  p(d|q) = \prod_{w \in q} p(w|d)
  \]
  - Run: IndriQL

**Cluster**
1. Rank resources (previous task)
2. Documents within a resource are ranked with IndriQL
   - Run: cITODPJ. Not submitted

**Diversity**
- IA-select diversification of IndriQL ranking using query relevance with respect to the resources
  \[
  p(S_q) = \prod_{r \in S_q} \prod_{d \in r} (1 - V(d, z))
  \]
  - Run: iaTODPJ

**Boost**
- Use directly the relevance with respect to the resources to boost the documents
  \[
  p(d(q, z) \propto p(d(q))p(q|z)
  \]
  - Run: bstTODPJ

**Discussion**

Results merging can be solved with simple IR techniques

Query likelihood obtains very good results

How to define a training set for an evolving test environment?

The rankings of the resources change
The content of the websites change
The type of queries is important:
  - are they tailored to be answered by a specific resource?