CLEF 2005: Multilingual Retrieval by Combining Multiple Multilingual Ranked Lists

Luo Si & Jamie Callan
Language Technology Institute
School of Computer Science
Carnegie Mellon University

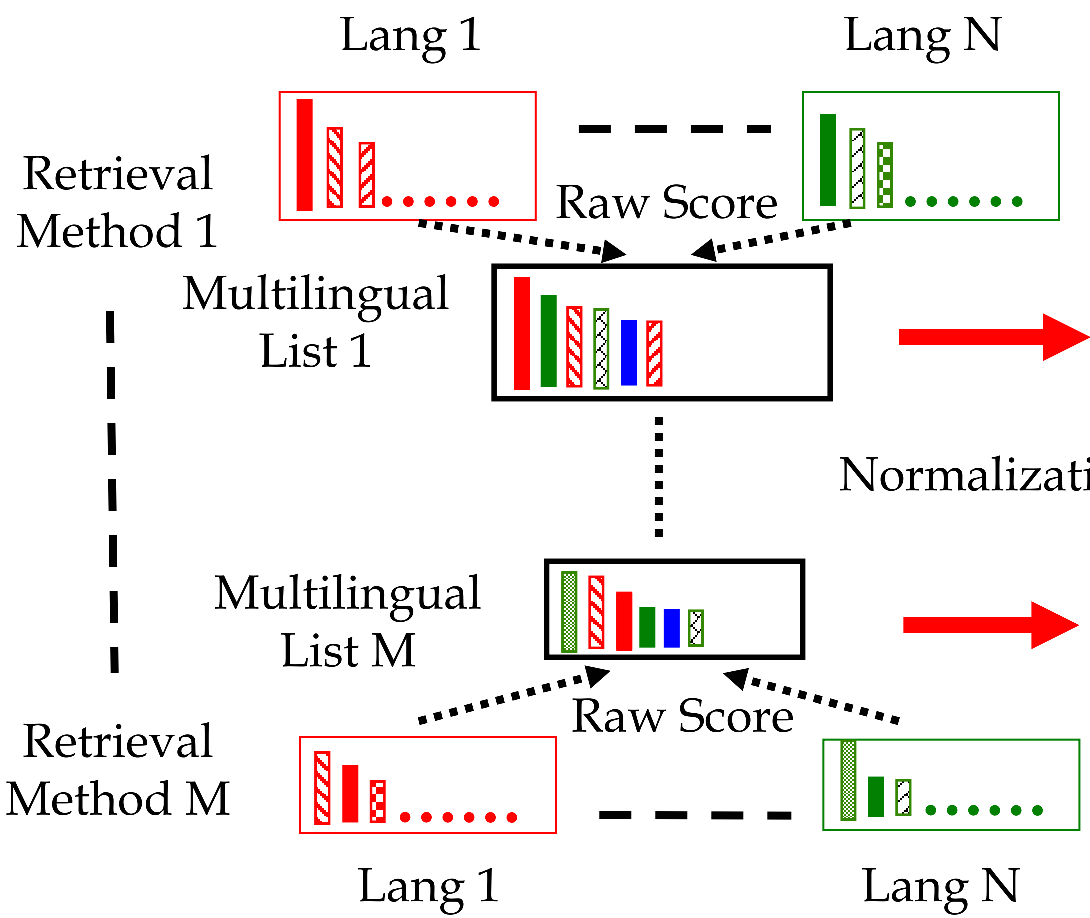
CMU, Two-Years On Task: Basic Idea

How \ What	Q	D
|| Corpus	✔	✔
MT S/W	✔	

- **Combination of different bilingual CLIR methods is helpful**
  - **What to translate:** Queries, documents
  - **How to translate:** Parallel corpus, MT software
- **Two ways to combine results of different methods and languages:**
  - Combine all results for language *l* into a language-specific result, then combine results from all languages into a final result
    - » Starts by combining scores from multiple methods, which **may be incompatible**
  - Combine all results from method *r* into a multilingual result, then combine results from all methods into a final result ✔
    - » **Easier**, because it starts by combining scores from a single method, which **may be more compatible**

CMU, Two-Years On Task: Combination Method

Our Strategy: Combining Multiple Simple Multilingual Ranked Lists



(a) Equal Weight Combination

(b) Learned Weight Combination by maximizing regularized MAP

CMU, Two-Years On Task: Evaluation Results

Mean average precision of each multilingual retrieval method

	Qry_fb (1)	Doc_nofb (2)	Qry_nofb (3)	Doc_fb (4)	UniNE (5)
MAP	0.353	0.360	0.335	0.332	0.330

Qry/Doc: what was translated

Fb/Nofb: with/without pseudo relevance back.

UniNE: UniNE system.

Mean average precision by combining multiple multilingual retrieval results

	M4_W1	M4_Trn	M5_W1	M5_Trn
MAP	0.432	0.434	0.446	0.449

M4/M5: Combine models (1)-(4) / (1)-(5); W1/Trn: Equal or learned weights

- Combined results of multiple methods provides large improvement

CMU, Result Merging Task: Basic Idea

- **We treat the task as a multilingual federated search problem**
  - All documents in language *l* are stored in resource (search engine) *s*
    - » So there are several independent search engines
  - Downloading and indexing documents takes time
    - » Do as little of this as possible
    - » Do as rarely as possible at retrieval (query) time
- **Goals:**
  - A high-quality merged list
  - Low communication and computation costs

CMU, Result Merging Task: Basic Idea

- **Offline:** Sample documents from each resource
  - To estimate corpus statistics (e.g., IDF)
- **Online:**
  - Calculate comparable scores for <u>top ranked</u> documents in each language
    - » Combine scores of query-based and doc-based translation methods
    - » Build language-specific query-specific logistic models to transform language-specific scores to comparable scores
      - [Si & Callan, SIGIR 02]
  - Estimate comparable scores for <u>all retrieved documents</u> in each language
    - » Combine them with exact comparable scores if available
  - Use comparable scores to create a merged multilingual result list

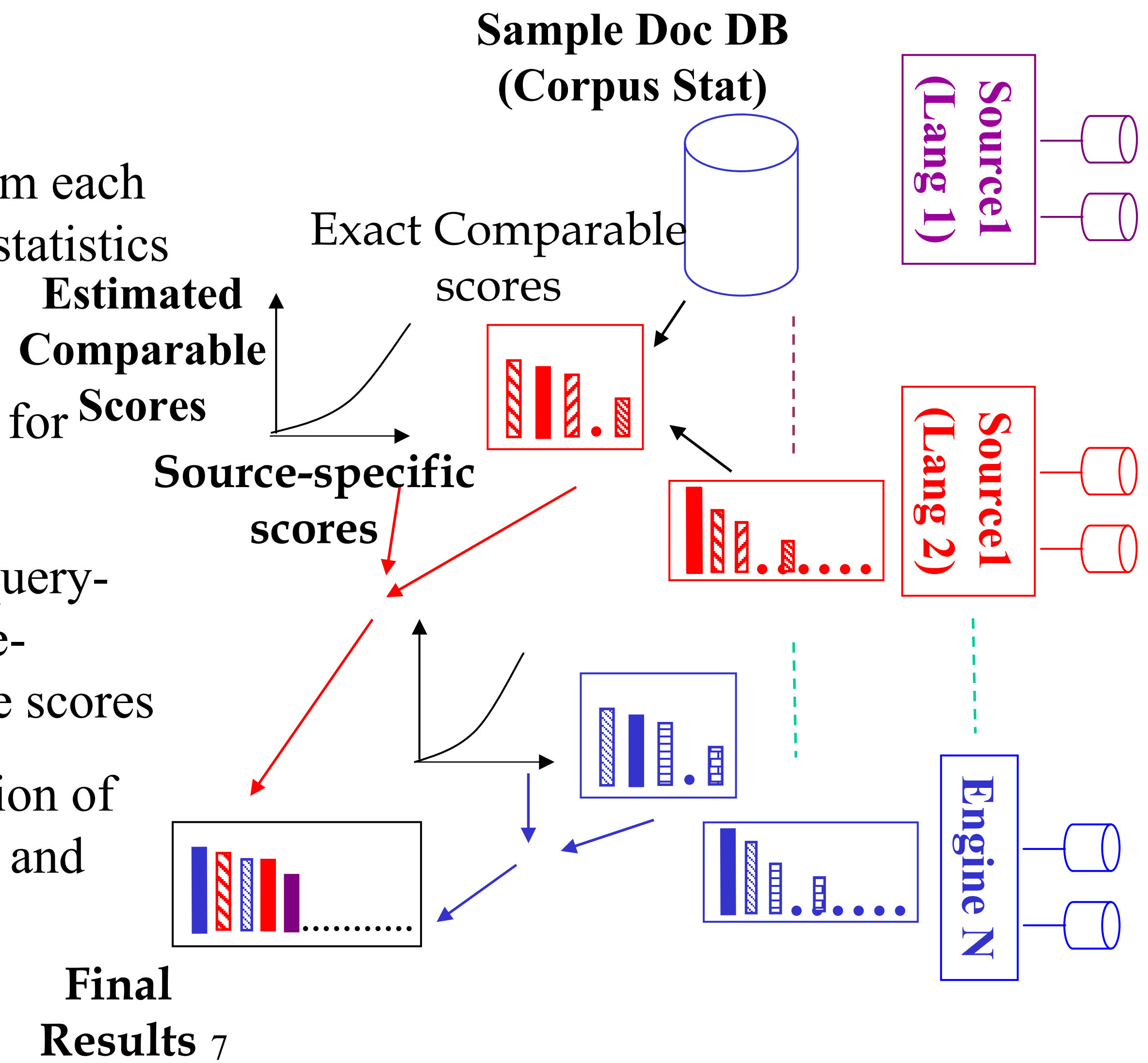
CMU, Result Merging Task: Language-Specific Query-Specific Model

Offline:

- Sample 3,000 documents from each language to estimate corpus statistics

Retrieval time:

- Calculate comparable scores for top ranked documents
- Estimate language-specific query-specific model to map source-specific scores to comparable scores
- Merge docs by the combination of estimated comparable scores and accurate scores



CMU, Result Merging Task: Evaluation Results

Language-specific logistic models are used to map resource-specific scores to comparable (resource-independent) scores

- **Should the models be query-specific or query-independent?**

Mean average precision of language-specific query-independent models (UniNE)

- TrainLog_MLE (logistic model by maximizing MLE): 0.301
- TrainLog_MAP (logistic model by maximizing MAP): 0.330

Mean average precision of language-specific query-specific models (UniNE)

	C_1000	C_500	Top_150_C05	Top_10_C05	Top_5_C05
MAP	0.382	0.384	0.412	0.393	0.372

C_X: top X docs from each list merged by exact comparable scores.

Top_X_0.5: top X docs from each list downloaded for logistic model to estimate comparable scores and combine them with exact scores by equal weight

Language-specific query-specific merging methods have a big advantage