

# Cumulative Citation Recommendation: Classification vs. Ranking

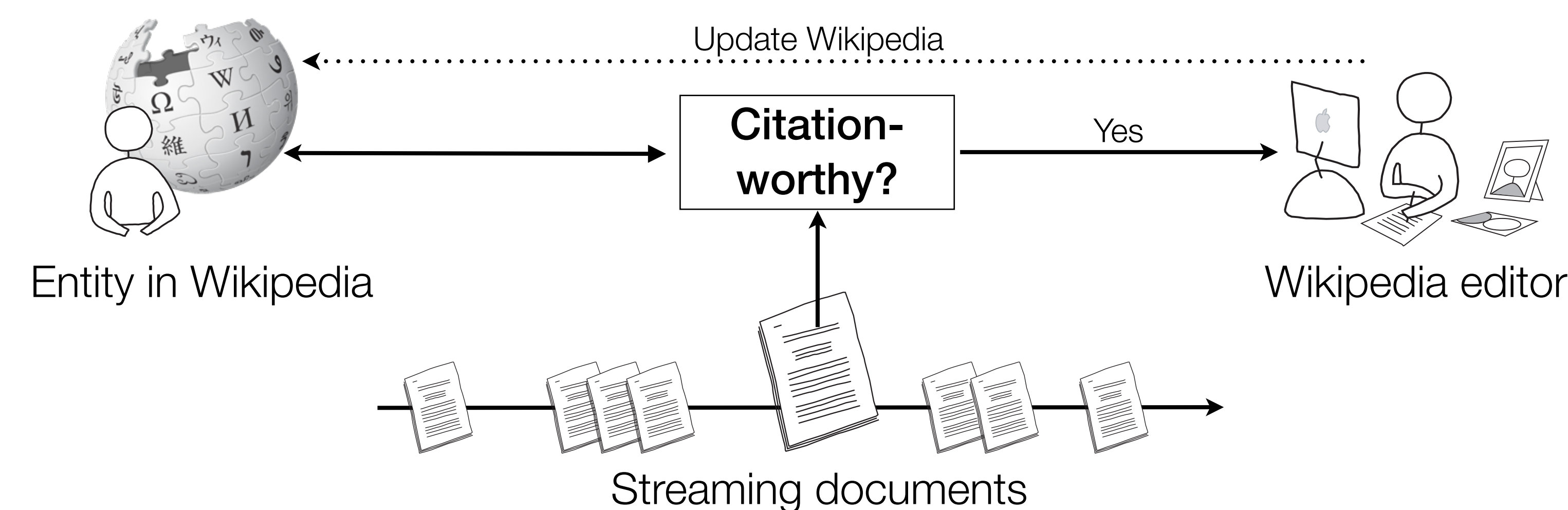


<http://bit.ly/16RraPB>

Krisztian Balog, University of Stavanger, Norway  
Heri Ramampiaro, Norwegian University of Science and Technology

## Cumulative citation recommendation (CCR)

Filtering a time-ordered corpus for documents that are highly relevant to a predefined set of entities.



## CCR @TREC KBA

- Process the collection in hourly batches
- Assign a score to each document based on its citation-worthiness
- Annotations along two dimensions: mentions and relevance
- Set-based evaluation metrics using a confidence cutoff

## Features

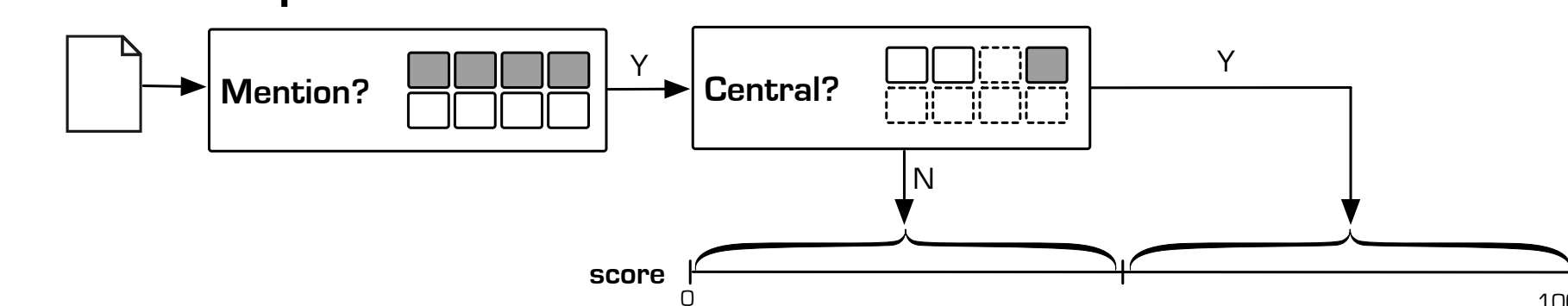
- **Document** (type, length of various document fields, language)
- **Entity** (number of related entities)
- **Document-entity** (entity occurrences and mentions of related entities in the document, similarity between document and entity)
- **Temporal** (average volume, changes, and bursts in stream volume and Wikipedia pageviews)

\* K. Balog, H. Ramampiaro, N. Takhirov, and K. Nørnvåg. **Multi-step classification approaches to cumulative citation recommendation**. In: *OAIR'13*, 2013.

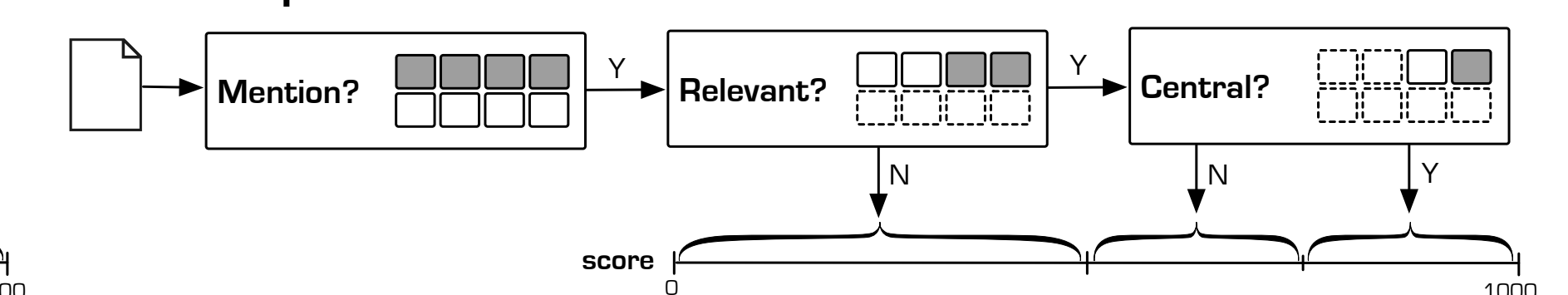
## Approaches

- **Classification**. Multiple steps, each is a binary classification task
  - Step 1: entity detection, based on a set of entity surface forms
  - Steps 2-3: supervised classification (J48, Random Forest)

### 2-step classification



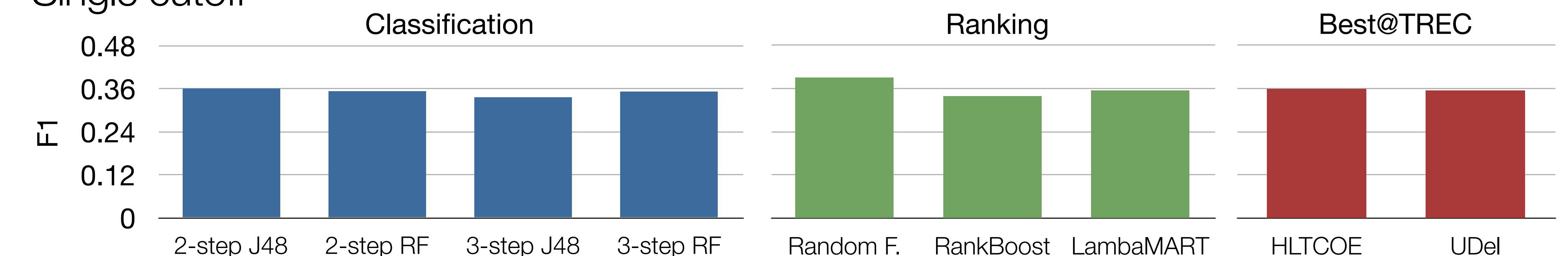
### 3-step classification



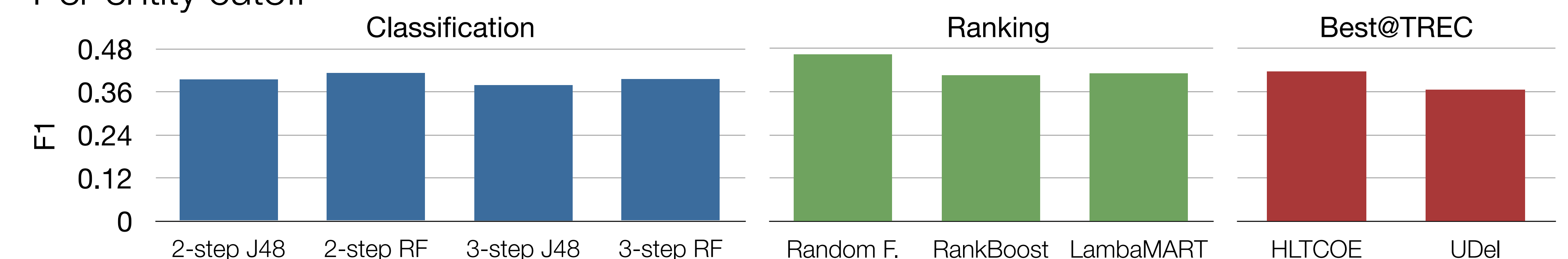
- **Ranking**. Learning-to-rank
  - Pointwise (Random Forests), Pairwise (RankBoost), Listwise (LambdaMART)

## Results

### Single cutoff



### Per-entity cutoff



## Upshot

- Ranking methods are a better fit for the CCR task (more precisely: the combination of the task with this particular evaluation methodology)