

# Category-based Query Modeling for Entity Search

Krisztian Balog, Marc Bron, and Maarten de Rijke  
ISLA, University of Amsterdam

## Summary

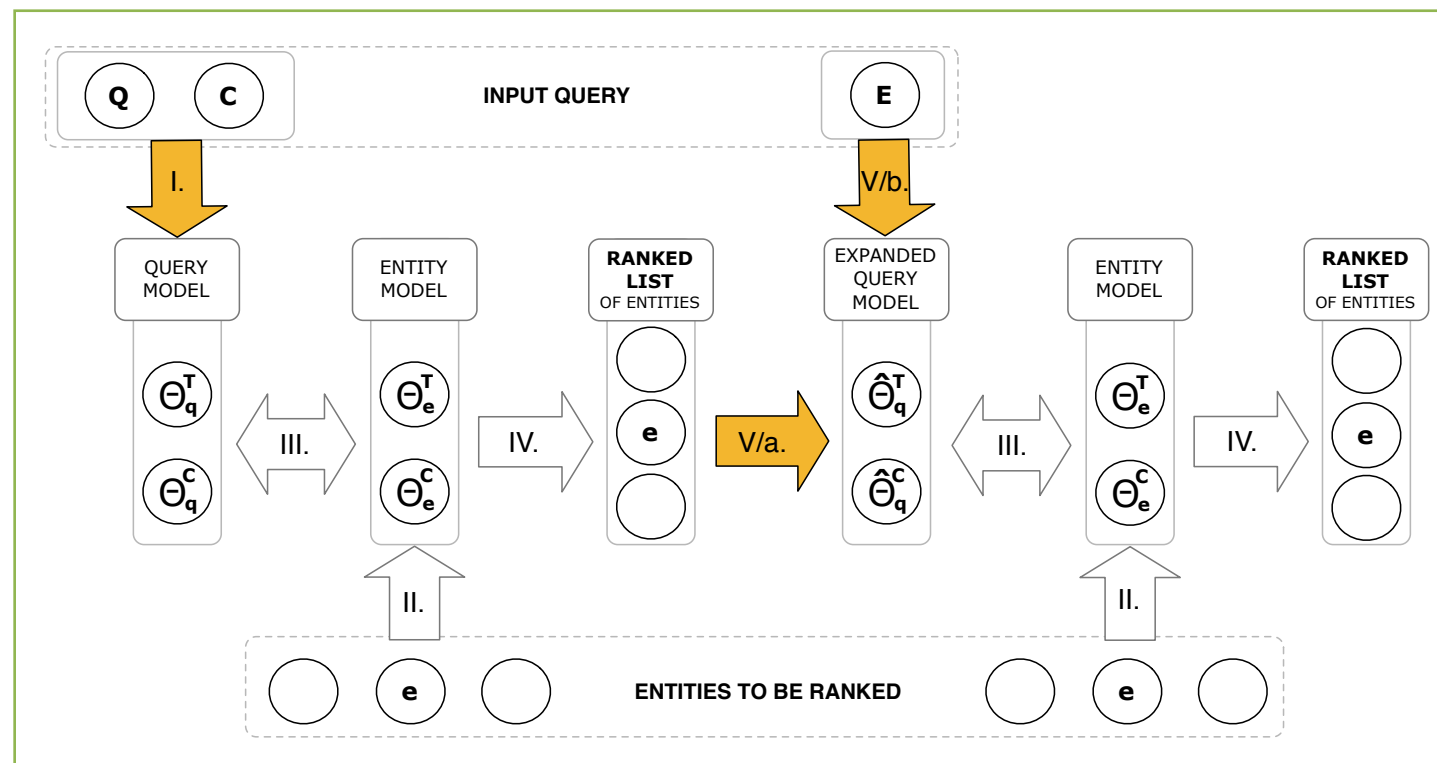
- Users often look for specific entities instead of documents mentioning them
- Entities represented by their Wikipedia page
- Introduction a **general probabilistic framework for entity retrieval**
- Focus on the **use of category information** in a theoretically sound way
- Demonstrate the advantage of a category-based representation over a term-based one

## Two tasks

- **Entity ranking:** topic consists of a keyword query (Q) and target categories (C)
- **List completion:** the topic also specifies example entities (E)

## Approach

- Information needs and entities both represented as a tuple: a **term-based** plus a **category-based** model, both characterized by probability distributions.



A general scheme for entity ranking

## Query models

Results on the **entity ranking** task (MAP for XER2007, xinfAP for XER2008).

<b>XER2007</b>	0.1798	0.1706	0.2410	0.2162	<b>0.2554</b>	0.1881	0.2255
<b>XER2008</b>	0.1348	0.1259	0.1977	0.3099	<b>0.3124</b>	0.2911	0.2950

## Expanded query models

Results on the **list completion** task (MAP for XER2007, xinfAP for XER2008).

	Baseline [I.]	Blind feedback [V/a.]			Example entities [V/b.]		
<b>XER2007</b>	0.2202	0.2138	0.2258	0.2197	0.2376	0.3141	<b>0.3267</b>
<b>XER2008</b>	0.2729	0.2814	0.2968	0.3017	0.2886	0.3873	<b>0.3926</b>

## Findings

- Category-based representation is very effective
- Category-based feedback outperforms term-based feedback