

## Now that you've bought into Model 2, we'll tell you why to get Model 1

Krisztian Balog and Maarten de Rijke

ISLA, University of Amsterdam  
<http://ilps.science.uva.nl>

## Two principal approaches to Expert Finding

- **Model 1:** Candidate models
  - Create a textual model ("profile" document) of candidates' knowledge according to the documents with which they are associated
  - Rank profiles using document retrieval techniques
- **Model 2:** Document models
  - Find documents that are relevant to the topic
  - Find out who is most strongly associated with these documents

## Results

	TREC 2005		TREC 2006		TREC 2007	
	MAP	MRR	MAP	MRR	MAP	MRR
Model 1	.1883	.4692	.3206	.7264	.3700	.5303
Model 2	.2053	.6088	.4660	.9354	.4137	.5666

K. Balog, *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

- Baseline settings
  - Automatic smoothing parameter estimation
  - Baseline document-candidate associations
  - No query expansion

## Components / extensions

- Proximity-based versions
- Smoothing parameter estimation
- Document-candidate associations
- Query models

## Proximity-based versions

	TREC 2005		TREC 2006		TREC 2007	
	MAP	MRR	MAP	MRR	MAP	MRR
Model 1	.1883	.4692	.3206	.7264	.3700	.5303
Model 1B	.2020	.5928	.4254	.9048	.3608	.5003
Model 2	.2053	.6088	.4660	.9354	.4137	.5666
Model 2B	.2194	.6096	.4544	.9235	.4303	.5656

K. Balog, *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## Proximity-based versions

	TREC 2005		TREC 2006		TREC 2007	
	MAP	MRR	MAP	MRR	MAP	MRR
Model 1	.1883	.4692	.3206	.7264	.3700	.5303
Model 1B	.2020	.5928	.4254	.9048	.3608	.5003
Model 2	.2053	.6088	.4660	.9354	.4137	.5666
Model 2B	.2194	.6096	.4544	.9235	.4303	.5656

K. Balog, *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## Smoothing parameter estimation

		TREC 2005		TREC 2006		TREC 2007	
		MAP	MRR	MAP	MRR	MAP	MRR
M1	auto	.1883	.4692	.3206	.7264	.3700	.5303
	opt.	.1912	.5747	.3834	.8647	.3801	.5571
M2	auto	.2053	.6088	.4660	.9354	.4137	.5666
	opt.	.2211	.6302	.4697	.9558	.4142	.5671

K. Balog. *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## Smoothing parameter estimation

		TREC 2005		TREC 2006		TREC 2007	
		MAP	MRR	MAP	MRR	MAP	MRR
M1	auto	.1883	.4692	.3206	.7264	.3700	.5303
	opt.	.1912	.5747	.3834	.8647	.3801	.5571
M2	auto	.2053	.6088	.4660	.9354	.4137	.5666
	opt.	.2211	.6302	.4697	.9558	.4142	.5671

K. Balog. *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## Smoothing parameter estimation

		TREC 2005		TREC 2006		TREC 2007	
		MAP	MRR	MAP	MRR	MAP	MRR
M1	auto	.1883	.4692	.3206	.7264	.3700	.5303
	opt.	.1912	.5747	.3834	.8647	.3801	.5571
M2	auto	.2053	.6088	.4660	.9354	.4137	.5666
	opt.	.2211	.6302	.4697	.9558	.4142	.5671

K. Balog. *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## Smoothing parameter estimation

		TREC 2005		TREC 2006		TREC 2007	
		MAP	MRR	MAP	MRR	MAP	MRR
M1	auto	.1883	.4692	.3206	.7264	.3700	.5303
	opt.	.1912	.5747	.3834	.8647	.3801	.5571
M2	auto	.2053	.6088	.4660	.9354	.4137	.5666
	opt.	.2211	.6302	.4697	.9558	.4142	.5671

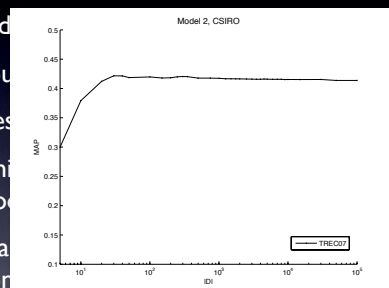
K. Balog. *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## The preferred approach is...

- Model 2, because
  - outperforms Model 1 in nearly all conditions
  - less sensitive to smoothing
  - mimics the process how real users would look for experts
  - can be implemented with very limited effort on top of a document search engine
  - effective (enough to examine the top ranked documents)

## The preferred approach is...

- Model 2, because
  - outperforms Model 1 in nearly all conditions
  - less sensitive to smoothing
  - mimics the process how real users would look for experts
  - can be implemented with very limited effort on top of a document search engine
  - effective (enough to examine the top ranked documents)



## But...

- Better document retrieval  $\Rightarrow$  better expert finding?
- No, the relation is not that simple

	Baseline	QM_BFB	QM_EX
Document search	.3576	.3797	.4449
Expert search (M2)	.4137	.3795	.4652

MAP scores on the TRECENT 2007 document and expert search tasks.  
K. Balog, *People Search in the Enterprise*. PhD thesis, University of Amsterdam, 2008.

## Moreover ...

- So far: changing one component at a time
- What if we start combining and stacking extensions?
- How to integrate external (e.g., Web) evidence?
- How to visualize?

## Stacking things

	estimated		optimal	
	MAP	MRR	MAP	MRR
Model 1	.3700	.5303	.3801	.5571
Model 1B	.3608	.5003	.4633	.6236
Model 2	.4137	.5666	.4142	.5671
Model 2B	.4303	.5656	.4323	.5790

## Stacking things

	estimated		optimal	
	MAP	MRR	MAP	MRR
Model 1	.3700	.5303	.3801	.5571
Model 1B	.3608	.5003	.4633	.6236
Model 2	.4137	.5666	.4142	.5671
Model 2B	.4303	.5656	.4323	.5790

## Stacking things

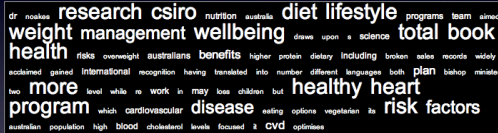
	estimated		optimal	
	MAP	MRR	MAP	MRR
Model 1	.3700	.5303	.3801	.5571
Model 1B	.3608	.5003	.4633	.6236
Model 2	.4137	.5666	.4142	.5671
Model 2B	.4303	.5656	.4323	.5790

## Stacking things

	estimated		optimal	
	MAP	MRR	MAP	MRR
Model 1	.3700	.5303	.3801	.5571
Model 1B	.3608	.5003	.4633	.6236
Model 2	.4137	.5666	.4142	.5671
Model 2B	.4303	.5656	.4323	.5790

## Visualizing the model

- Model 1



- Model 2

- List of documents? Snippets?

ILPS@TRECENT 2008

## Approach

- Proximity-based version of Model 1
- Use global evidence to estimate the importance of a text snippet [Balog & de Rijke, CIKM 2008]
- Expansion
  - Profile side: evidence from the Web
  - Query side: terms from candidate profiles

## Setup

- Minimal corpus preprocessing (stopword removal, no stemming)
- Use candidate list from [Balog & de Rijke, CIKM 2008]
  - Made available for TREC participants <http://es.csiro.au/cerc/data/balog/>
- Science communicators are filtered out
- Results reported on 2007 topics

## Query expansion using candidate profiles

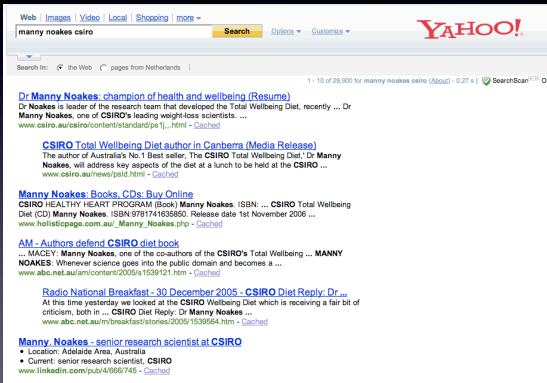
- Add top K (=10) terms from the profiles of top M (=5) experts

	Query model	MAP	MRR
Model 1B	QM_BL	.4838	.6280
	QM_PROFILES	.4933	.6422
Model 2	QM_BL	.4559	.6172
	QM_PROFILES	.4799	.6268

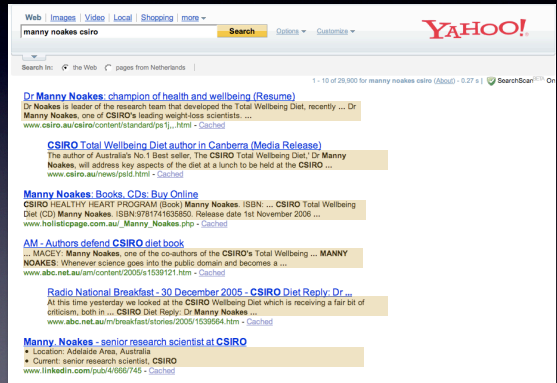
## Profile expansion using the Web

- Use candidate name + CSIRO as a query
- Use Yahoo! API to get documents
- Build candidate profile from result snippets

## Profile expansion using the Web



## Profile expansion using the Web



## Profile expansion using the Web

	Query model	MAP	MRR
Model IB	QM_BL	.4838	.6280
	QM_PROFILES	.4933	.6422
Web	QM_BL	.3840	.5217
	QM_PROFILES	.4153	.5478
Model IB (1) + Web (4)		<b>.5323</b>	<b>.6371</b>

## Now what?

Model 1



Model 2

## What about...

Model 1



Model 2

## Model 1 + Model 2

	MAP	MRR
Model 1B	.4838	.6334
Model 2	.4799	.6268
M1B + M2	.5267	<b>.6828</b>
M1B + M2 + Web	<b>.5405</b>	.6468

## Wrap-up

- Model 1: better candidate representations result in better performance
- Model 2: better document retrieval doesn't necessarily result in better ranking of experts
- Which of Model 1 or Model 2 should be preferred then?
  - It depends on...

## Conclusions

Now that you've bought into Model 2, we'll tell you why to get Model 1 **too**

## Future?

## Questions?

**Krisztian Balog**

[K.Balog@uva.nl](mailto:K.Balog@uva.nl)  
<http://www.science.uva.nl/~kbalog>