

Finding Experts and their Details in E-mail Corpora

Krisztian Balog and Maarten de Rijke

Background

Enterprise search: finding valuable information in an organization's intranet

- **Mixture of document types:** web pages, news/email archives, document archives, etc.
- **Searching for:** services, objects, **people**, ...

Tasks

- **Finding experts:** given a topic, who are experts on the topic in the organization?
- **Expert details:** once we've found an expert, how do we get in touch with them?

Approach

I. Finding Experts

- What is the probability of a candidate ca being expert on the query topic q ? -- $p(ca|q)$
- Use Bayes: rank candidates in proportion to $p(q|ca)$
- Use documents as a latent variable between the query and the candidates
- Rank documents according to the query; then determine how likely it is that a candidate is an expert by considering the set of associations
- Formally:

$$p(q|ca) \propto \sum_d p(q|d)p(ca|d)$$

- Form candidate-document associations by exploiting the fielded structure of e-mail messages

$$p(ca|d) = \sum_{d_i \in D} a(d_i, ca) \quad a(d, ca) = \sum_{i=0} \pi_i A_i(d, ca)$$

From:	David Hawking	[A ₀] EMAIL_FROM
To:	Dan Connolly	[A ₁] EMAIL_TO
Cc:	Maria Fernandez	[A ₂] EMAIL_CC
Subject:	Conferences	
Date	21 Oct 2003 12:07	
I'd like to send Smith to ADC2004. She's entitled under section whatever on p.27 of the corporate manual. Jones wants to go but she already went on that junket to Maui.		
--		
Dan Connolly, W3C http://www.w3.org/People/Connolly/ tel: +1-512-310-2971 (office, mobile) mailto:connolly.pager@w3.org (put your tel# in the Subject:)		[A ₃] EMAIL_CONTENT

II. Mining Contact Details

- Retrieving contact details of experts is a natural component of an operational expert finder
- Signatures contain reliable information
- Signature detection using heuristics (precision oriented)
- Pattern-based extraction of details from signatures

Experimental evaluation

- **Lists** part of the W3C corpus (198.000 docs)
- **TREC 2005 expert search data**
 - list of W3C people (candidates)
 - topics, relevance assessments

Results

I. Finding Experts

- 65% of relevant experts found
- MAP = 0.24 (mean average precision)
- RRI = 0.45 (reciprocal rank of the first relevant result)

II. Mining Contact Details

- Signatures extracted (w3c people): 15.514
- Unique signatures: 3.447
- Personal data found in signatures for 75% of (w3c) people

Example

Query: semantic web coordination

Experts found	Documents found
Rating: ★★★★★ Name: Steve Bratt no: candidate-0598 E-mail: steve@w3.org Search expert on: Citeseer Portal.acm.org	Rating: ★★★★★ Title: Web services at W3C - slide "Semantic Web" no: www-007-9315926
Rating: ★★★★★ Name: Hugo Haas no: candidate-0003 E-mail: hugo@w3.org Search expert on: Citeseer Portal.acm.org	Rating: ★★★★★ Title: Specifics of "Aligning" Web Services and the Semantic Web - slide "Why 'align with the semantic web'?" no: www-007-9502816
Rating: ★★★★★ Name: Eric Miller no: candidate-0007 E-mail: em@w3.org Search expert on: Citeseer Portal.acm.org	Rating: ★★★★★ Title: Are we done yet? - slide "Web Services and the Semantic Web" no: www-005-7691978
Rating: ★★★★★ Name: Tim Berners-Lee no: candidate-0161 E-mail: timbl@w3.org Search expert on: Citeseer Portal.acm.org	Rating: ★★★★★ Title: The W3C Semantic Web Activity - slide "W3C Semantic Web Activity - Structure" no: www-021-14331430
Rating: ★★★★★ Name: Michael Wilson no: candidate-0959 E-mail: M.D.Wilson@rl.ac.uk Search expert on: Citeseer Portal.acm.org	Rating: ★★★★★ Title: Circles and arrows diagrams using stylesheet rules no: www-002-2879356
Rating: ★★★★★ Name: Orly Rapaport no: candidate-0892 E-mail: orly.rapaport@comverse.com Search expert on: Citeseer Portal.acm.org	Rating: ★★★★★ Title: W3C: Leading the Web -- and Web Services -- to their Full Potential - slide "Web Services Coordination Group" no: www-007-10565502
	Rating: ★★★★★ Title: W3C Semantic Web Coordination Group Charter no: www-010-1008848

Further work

- Identify more expertise indicators
- Use external (threaded) structure of e-mail
- Extract additional personal details (eg. address, affiliation)

Formal Models for Expert Finding in Enterprise Corpora, K. Balog, L. Azzopardi, and M. de Rijke. In: *29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval*, 2006.



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