Automatic people tagging for expertise profiling in the enterprise

Pavel Serdyukov *

(Yandex, Moscow, Russia)

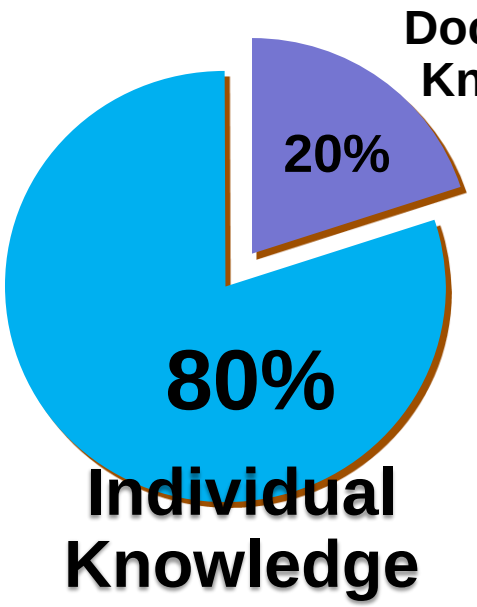
Mike Taylor, Vishwa Vinay, Matthew Richardson, Ryen White

(Microsoft Research, Cambridge / Redmond)

*** Work was done while visiting MSR Cambridge**

The need for experts

- Some knowledge is not easy to find
  - Not stored in documents
  - Not stored in databases
  - **It is stored in peoples' minds!**



Meet People!

What people do **without** special expert finding tools?

- Independent survey of 170 UK companies
- 50% want to be able to locate expertise
- Only 9% have tools for expert finding
- To find experts:
  - 71% “ask around”
  - 46% use the company directory
  - 34% use the company intranet
  - 30% send a company-wide email

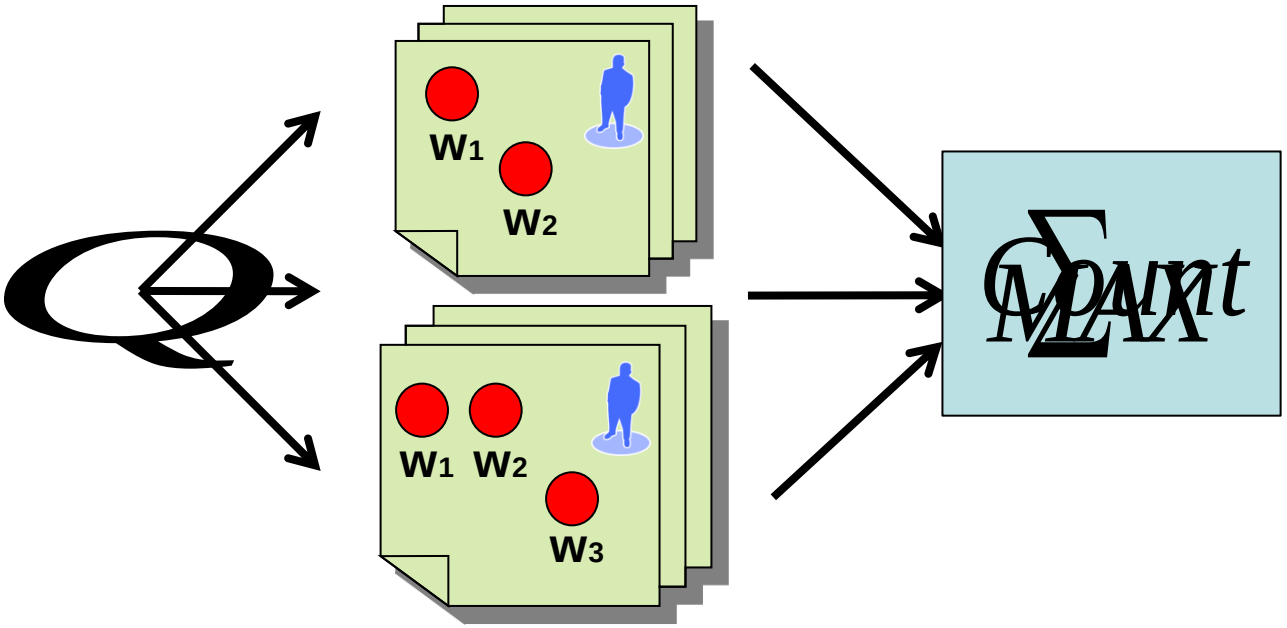
Marisa Peacock. The Search for Expert Knowledge continues. CMSWIRE. 12 May 2009.

Expert finding

- **Task definition**:
  - Given a short query
  - Rank employees **judged as experts higher than non-experts**
  - Very similar to document retrieval, but…
  - Finding **relevant people**, not **documents**
- Existed as a part of **TREC Enterprise track** for 4 years (2005 - 2008):
  - Community developed nice datasets
  - Lots of papers published
  - **And almost no industrial research!**

Traditional approach

- 1st step: Rank all documents with 
- 2nd step: Aggregate document scores



Typical expert finding output
Query: “csharp programming”

Hard to estimate relevance…
So, why should I click?

People Matches

Ali Pezeshk
SDET
US-WLX-SDET

Weiying Chen
Content Publishing
WIN Web Services & Content

Sunghwa Jin
SDET
US-Distributed Application Server

View more people »

Sharp C
Get the best deal. **Sharp C** at Shopzilla.
www.**shopzilla.co.uk**

Sharp Standard Televisions Sale
Save on **Sharp Standard** Televisions. Compare 100s of Sites
www.**shop.com**

Sharp C
Sharp C
www.**bizrate.co.uk**

C Standard at Amazon.co.uk
Millions of titles, new and used. Free UK Delivery on Amazon Orders
Amazon.co.uk/books

Compare to snippets/ads!
3 sources of evidence:
Title, URL, Description

Problem

- delicious e **do not trust** the plain list of names,

Andrew Kennedy (**Microsoft Research Cambridge**)

research.microsoft.com/en-us/um/people/akenn/

2

language ocaml science types uk

- So, we [illegible] build personal summaries:
  - Experi[illegible]ific
  - Concis[illegible]ontent-bearing
  - Sentence-free,
- Let's generate [illegible]

David Elsweiler

@delsweil Nuremberg

twitter

@arjenpdevries/ir-db-research

@giovannaroda/ir-al

Information Retrieval people and conferences (& NLProc, IE, ML, RecSys)

@leobard/pim-researchers

who is who of not forgetting that important tweet - Personal Information Management researchers

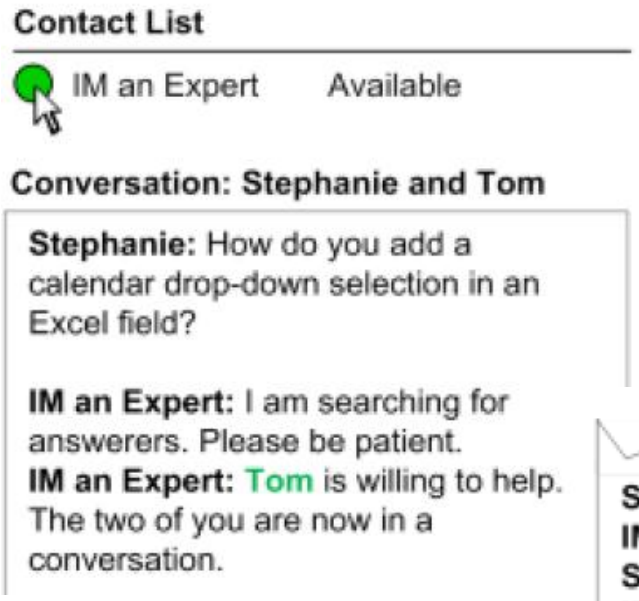
People like to tag each other

Farrell, S., Lau, T., Nusser, S., Wilcox, E., and Muller, M. 2007. **Socially augmenting employee profiles with people-tagging**. UIST '07.

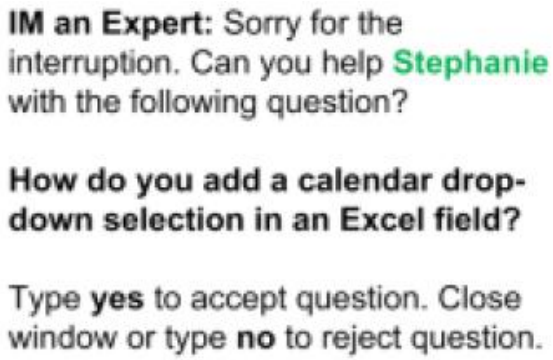
Microsoft IM-an-Expert

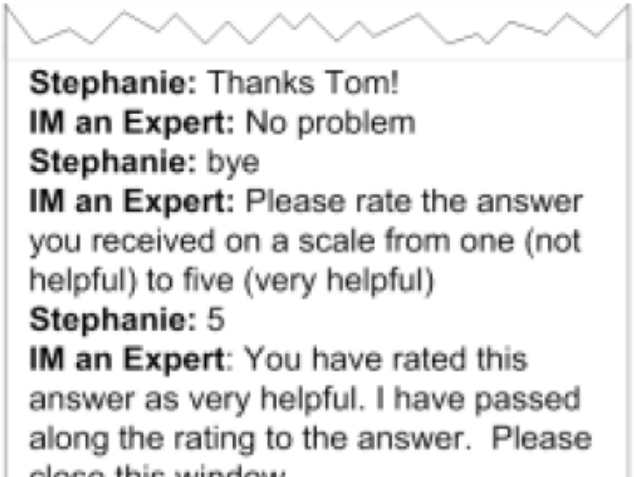
Q&A system that finds experts to answer specific questions and mediates the dialog between **an expert** and **the answer seeker**

Stephanie asks **IM-an-Expert** to find an expert



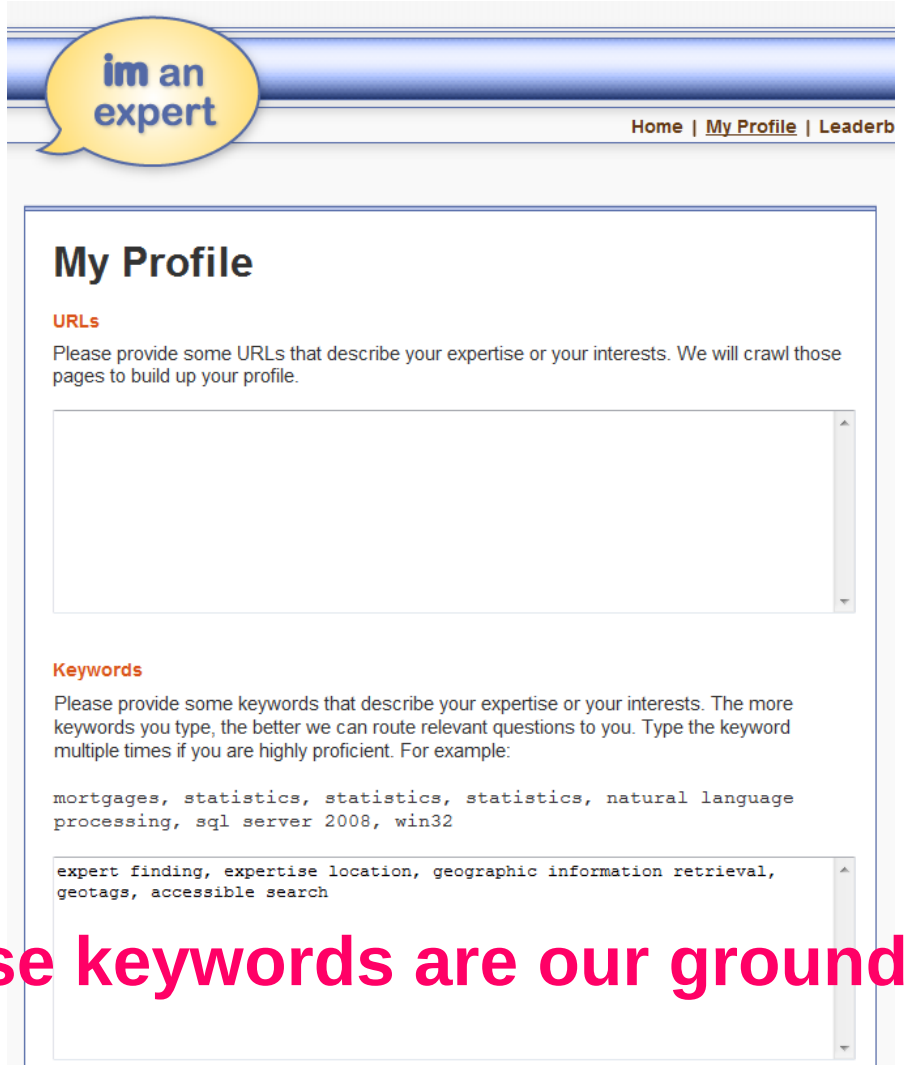
IM-an-Expert finds **Tom** and asks to help **Stephanie**





How to make yourself found?

Candidate experts in IM-an-Expert describe their expertise by keywords, so they **tag** themselves



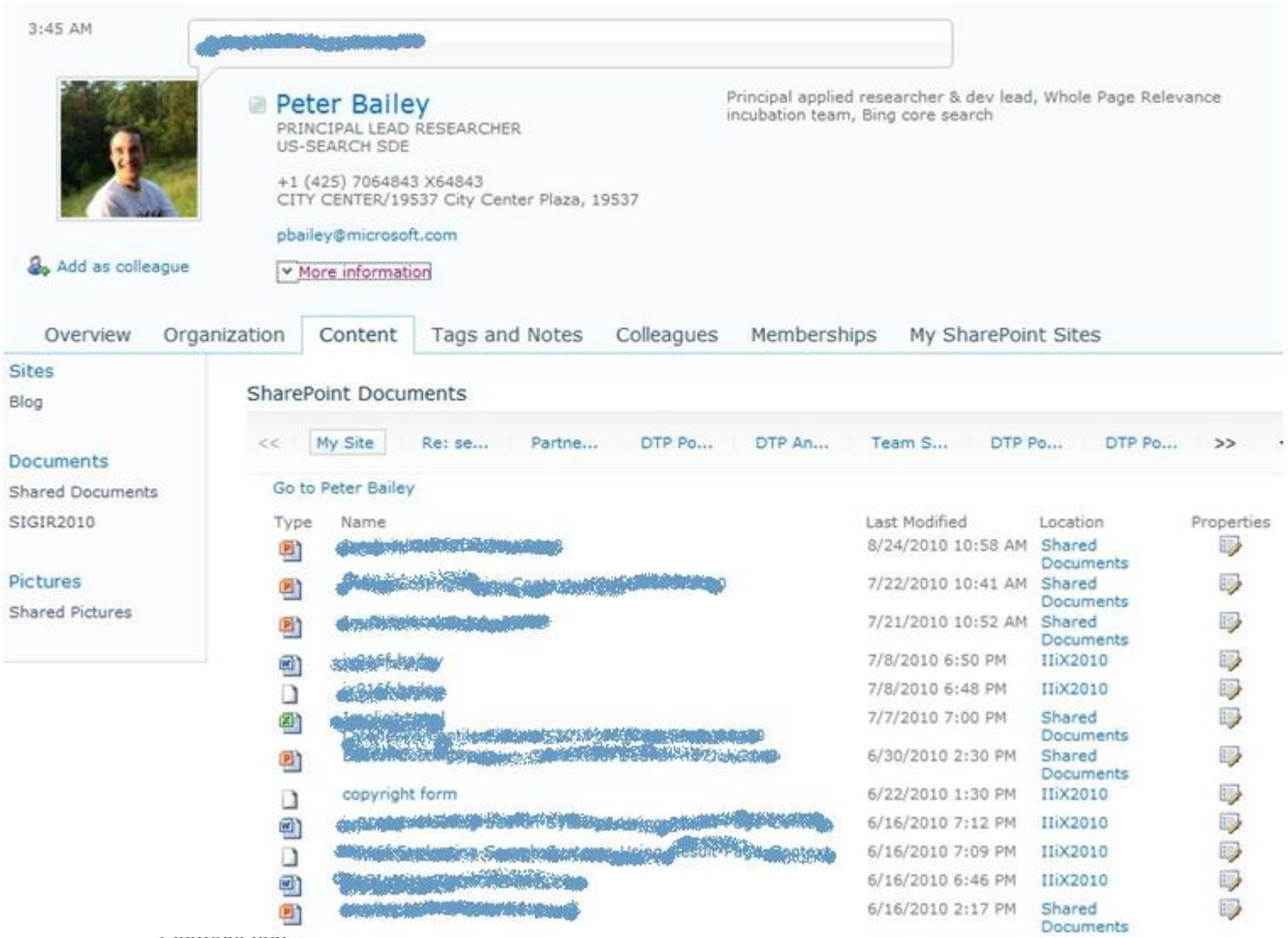
These keywords are our ground truth!

Our task

- **Predict** those **tags** that person specified in personal profile…
- … **using various expertise evidence sources** related to the person
- Non-unique tags from our training data are our **controlled vocabulary:**
  - So, the task is as well **to recommend tags** for newcomers
  - And actually for any person in Microsoft
- So, let's **rank known tags w.r.t. each person in the enterprise**

- **1167 profi** es
  - **Alias** + **L**
  - Gathered
- **Tag stats:**
  - **4450** unic
  - **1275** tags **employee**
  - **5.5** non-u
  - **1.47** worc

Expertise evidence sources:

- 

Expertise evidence streams: Click-through sources

- Personal queries to Sharepoint
  - 6 months of queries to Sharepoint (January 2010 – June 2010)
  - **67 unique queries** on average per person
- Clicked documents
  - **433 clicks** on average per person
  - **47 clicked documents** on average per person
- Queries with clicks on authored docs
  - **24 clicks** on average per person
  - **12 unique queries** on average

Streams and features

- Each source contributes streams
  - Authored/related/web/clicked docs:
    - Filenames, titles, snippets, body content
    - Body contents are crawled only for **authored** and **related**
  - Queries, lists:
    - Just query strings / names
- For each stream and each tag we calculate:
  - Binary (***1*** if stream contains tag, ***0*** - otherwise)
  - Language model based score:

$$P(tag \mid \theta) = \prod_{w \in tag} \left\{ (1-\lambda)\, p(w \mid \theta) + \lambda\, p(w \mid Global) \right\}$$

  - Sum of scores of all records (e.g. ***titles*** or ***queries***) in each stream **are our features**

Importance of deviation

- It's important not only to be "rich" in tag
- But "richer" on average!
- So, transformed features as:

$$X_{tag}^{employeeX} = X_{tag}^{employeeX} - \bar{X}_{tag}, \bar{X}_{tag} = \frac{1}{|training|} \sum_{employeeY \in training} X_{tag}^{employeeY}$$

Additional features

- **Popularity-based priors:**
  - Profile frequency
  - Frequency as query in Sharepoint
- **Quality of tag:**
  - Frequency in Enterprise data (***IDF***)
  - Probability of words in the tag based on Web corpus
    - Using Bing Web N-Gram service *****
- **Phrase length:**
  - In words
  - In characters

***** http://research.microsoft.com/en-us/collaboration/focus/cs/bingiton.aspx

Ranking

- 1167 profiles:
  - **700 (~60%)** as training set, **300** (**~25%**) as test set
  - **167** (**~15%**) as validation set (to tune parameters)
- In average**: ~ 5.8 tags** per person
  - **4098 positive examples**
  - **~1270 x 700 = ~900,000** negative examples?
    - **Too imbalanced…**
    - **Too slow to learn…**
  - Sampled negatives randomly, tested on validation set:
    - **~60,000 was enough to reach maximum AP**
- Learned **Logistic Regression** model

Measures

- We rank **tags** by their classification scores

- Measures:
  - **Precision** at ranks *1*, *5*, *10* **(P@1/5/10)**
  - Average Precision at rank *100* **(AP)**
  - **Success** at rank 5 **(S@5)**

Individual feature performance

	Feature p				erformance	
Stream	P@5					S@5
ProfileFrequency (baseline)	0.066	0.046	0.253	-	-	-
AuthoredContentLM (baseline)	0.044	0.030	0.180	0.081	0.057	0.310
ListNamesBinary	**0.122**	**0.086**	**0.437**	**0.125**	**0.087**	**0.437**
AuthoredFileNamesBinary	0.071	0.058	0.3	0.093	0.066	0.367
AuthoredTitlesLM	0.072	0.053	0.283	0.085	0.059	0.313
PersonalQueriesBinary	0.055	0.040	0.210	0.059	0.041	0.220
QueriesToAuthoredBinary	0.059	0.038	0.230	0.069	0.043	0.307
RelatedSummaryLM	0.035	0.024	0.163	0.059	0.041	0.257
ClickedTitlesBinary	0.021	0.018	0.087	0.033	0.025	0.133
WebTitlesLM	0.023	0.012	0.093	0.023	0.013	0.097
ALL features	**0.171**	**0.124**	**0.543**	-	-	-

"No expertise evidence" baseline

Feature group importance

Stream	P@1	P@5	P@10	AP	S@5
ALL	**0.266**	**0.171**	**0.122**	**0.124**	**0.543**
- *ProfileFrequency*	0.240	0.138	0.102	0.110	0.460
- *List*	0.146	0.096	0.073	0.074	0.370
- *Authored*	0.130	0.078	0.065	0.063	0.300
- *PersonalQueries*	0.090	0.057	0.046	0.047	0.250
- *Related*	0.060	0.053	0.044	0.030	0.220
- *Clicked*	0.033	0.046	0.039	0.025	0.180
- *Web*	0.010	0.034	0.032	0.019	0.143
- *QueriesToAuthored*	0	0.025	0.023	0.007	0.123

Removing features by feature groups (evidence sources)

Click-through evidence importance

Clickthrough = {PersonalQueries, QueriesToAuth, ClickedDocs}

Stream	P@1	P@5	P@10	AP	S@5
ALL	0.266	0.171	0.122	0.124	0.543
ALL - Click-through	0.266	0.160	0.112	0.117	0.513
Typical enterprise: *ALL - Lists - ProfileFrequency*	0.146	0.096	0.073	0.074	0.37
Typical enterprise - *Click-through*	0.093	0.078	0.056	0.050	0.310
Click-through (only)	0.09	0.061	0.047	0.048	0.247

Error analysis (I)

- Some tags **are not predictable** with Enterprise data:
  - Work non-related relevant tags:

    “ice cream”, “traveling”, “cooking”, “dancing”, “cricket”, “camping”, “judaism”
  - Tags which are not likely to be used in documents and/or too general:

    “design patterns”, “customer satisfaction”, “public speaking”, “best practices”

Error analysis (II)

- Alternative tags used:
  - Predicted: **csharp, e-learning, t-sql**
  - Relevant: **c#, elearning, transactsql**
- More or less general concept used:
  - Predicted: **sql server 2008**
  - Relevant: **sql server**
- Concept expressed differently:
  - Predicted: **machine learning, web search**
  - Relevant: **data mining, search engines**

Susan Dumais

Predicted

search
msr
bing
information retrieval
web search

Relevant, but not mentioned

Relevant

search
web search
enterprise search
desktop search
hci

Relevant, but named differently

Vsevolod Dmitriev

Predicted

russia
ocs
exchange
c#
.net

Relevant, but less general concept is mentioned

Relevant

exchange 2003
exchange 2007
exchange 2010
ocs 2007
outlook
exchange

Relevant, but not mentioned

Conclusions and Future work

- We’ve shown the way to solve a novel task of **automatic people tagging:**
  - Treated the problem as learning to combine evidences to rank areas of expertise
- Click-through evidence is important
  - But not decisive, at least, for Microsoft
- Future work should consider:
  - Diversity of recommended tagsets
  - Specificity of tags
  - Query dependent tagsets
  - Uncontrolled vocabulary