

# Expertise Search in Heterogeneous Information Sources

---

*Luo Si, Yi Fang, and Aditya Mathur*  
*Department of Computer Science, Purdue University*

*Supported by State of Indiana and Purdue University*





# Goals

---

- In many applications, the goal of search is to identify **Experts** instead of specific documents
- Motivating Examples:
  - Government agencies and industry sponsors to locate potential faculty with desired expertise
  - Researchers to look for collaborators with complementary research expertise
  - Students to look for academic advisors with matched research interests



# INDURE System

---

- Indiana University Research Expertise Database (**INDURE**):
  - A scalable expertise search system analyzing heterogeneous sources (e.g., homepages, Ph.D. dissertations, NSF/NIH awards, manual inputs)
  - Over 12,000 faculty across four major universities in Indian (Purdue, Indiana Univ, Ball State and Notre Dame) are currently searchable
  - Evolving everyday to include additional information about faculty



# Discriminative Probabilistic Framework for Expertise Search

---

- Existing expertise search algorithms do not consider heterogeneous sources or simply treat different sources equally
- Our novel probabilistic algorithms combine evidence from different sources by analyzing different queries and different experts
  - Expert-Specific (ES) algorithm: Some senior faculty do not have home pages, while junior faculty do not have supervised Ph.D. dissertation.
  - Query-Specific algorithm (QS): For Query “Cancer”, more weights should be on NIH instead of NSF.
  - Expert and Query Specific algorithm (EQS): Consider different queries and experts simultaneously.



# Empirical Studies

---

- Comparing proposed algorithms (**ES**, **QS**, **EQS**) against state-of-art algorithms (Concatenation, ExpCombSUM, ExpCombMNZ) for precision/accuracy of top-ranked experts

	$P@5$	$P@10$
Concatenation	0.326	0.296
expCombSUM	0.342	0.313
expCombMNZ	0.332	0.310
<b>ES</b>	0.386	0.345†
<b>QS</b>	0.381	0.339†
<b>EQS</b>	0.408†	0.369†