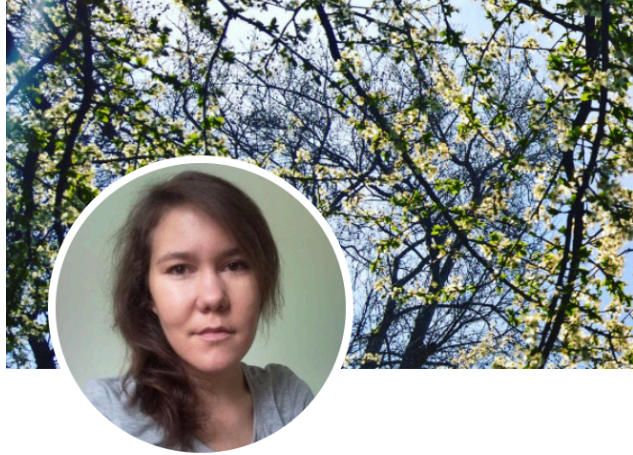


An Expertise Recommender System based on Data from an Institutional Repository (DiVA)

Milena Angelova, Vishnu Manasa Devagiri, Veselka Boeva, Peter Linde, Niklas Lavesson
mangelova@mail.com



About me



Milena Angelova

Python developer at Audience Platform

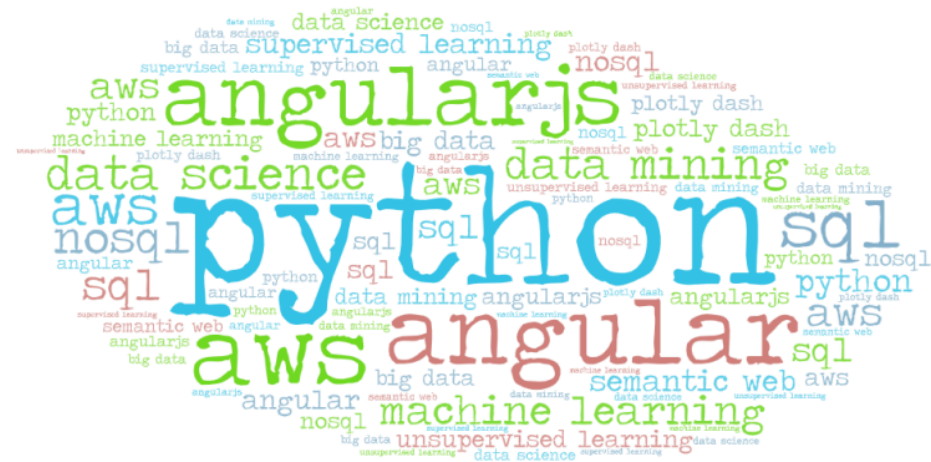
Contacts

LinkedIn:

<https://www.linkedin.com/in/milena-angelova-7254966a/>

Email:

mangelova@mail.com



Jobs in Audience Platform

<https://dev.bg/company/audience-platform/#jobs>


Google Scholar Citations

<https://scholar.google.com/citations?hl=bg&user=xHzxlzQAAAAJ>

Agenda

- ▶ Expertise retrieval
- ▶ Possible applications
- ▶ The architecture of the Expert Finding System
- ▶ The ontology model
- ▶ Ambiguity
- ▶ Results
- ▶ Demo

Publications



An Expertise Recommender System based on Data from an Institutional Repository (DiVA)

Milena Angelova, Vishnu Manasa Devagiri, Veselka Boeva, Peter Linde et Niklas Lavesson

p. 135-149

RÉSUMÉ INDEX TEXTE BIBLIOGRAPHIE NOTES AUTEURS ILLUSTRATIONS

RÉSUMÉ

Finding experts in academics is an important practical problem, e.g., recruiting reviewers for reviewing conference, journal or project submissions, partner matching for research proposals, finding relevant MSc or PhD supervisors, etc. In this work, we discuss an expertise recommender system that is built on data extracted from the Blekinge Institute of Technology (BTH) instance of the institutional repository system DiVA. The developed prototype system is evaluated and validated on information extracted from the BTH DiVA installation, concerning thesis supervision of researchers affiliated with BTH. The extracted DiVA classification terms are used to build an ontology that conceptualizes the thesis domain supported by the university. The supervisor profiles of the tutors affiliated with the BTH are constructed based on the extracted DiVA data. These profiles can further be used to identify and recommend relevant subject thesis supervisors.

ENTRÉES D'INDEX

Keywords : data mining, DiVA, expertise retrieval, knowledge management, natural language processing

RECHERCHER DANS LE LIVRE

TABLE DES MATIÈRES

CITER PARTAGER

CITÉ PAR

AJOUTER À ORCID

**CONNECTING THE KNOWLEDGE COMMONS —
FROM PROJECTS TO SUSTAINABLE INFRASTRUCTURE**
Leslie Chan, Pierre Mounier
<https://books.openedition.org/oep/9078>

An Expertise Recommender System Based on Data from an Institutional Repository (DiVA)

Milena Angelova ¹, Vishnu Manasa Devagiri ², Veselka Boeva ², Peter Linde ², Niklas Lavesson ² [Details](#)

1 TU-Sofia - Technical University of Sofia [Bulgaria]
2 Blekinge Institute of Technology

Abstract : Finding experts in academics is an important practical problem, e.g. recruiting reviewers for reviewing conference, journal or project submissions, partner matching for research proposals, finding relevant M. Sc. or Ph. D. supervisors etc. In this work, we discuss an expertise recommender system that is built on data extracted from the Blekinge Institute of Technology (BTH) instance of the institutional repository system DiVA. The developed prototype system is evaluated and validated on information extracted from the BTH DiVA installation, concerning thesis supervision of researchers affiliated with BTH. The extracted DiVA classification terms are used to build an ontology that conceptualizes the thesis domain supported by the university. The supervisor profiles of the tutors affiliated with the BTH are constructed based on the extracted DiVA data. These profiles can further be used to identify and recommend relevant subject thesis supervisors.

Keywords : natural language processing data mining DiVA expertise retrieval knowledge management

Document type : **Conference papers**

Domain : **Humanities and Social Sciences / Library and information sciences**

ELPUB 2018, Jun 2018, Toronto, Canada.
<https://hal.archives-ouvertes.fr/hal-01816680v1>

Expertise Retrieval

- ❑ Finding the right person with the appropriate skills and knowledge.
- ❑ Organizations search for new employees not only relying on their internal information sources, but they also use data available on the Internet.
- ❑ The data is very dispersed and of distributed nature.
- ❑ The expertise retrieval process is supported by IT-based solutions - *expert finder*, *expert seeker*, *expert miner*.

Possible applications



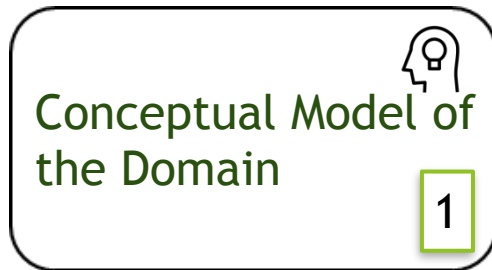
- ▶ Identification of experts in a particular technological domain, e.g. for the purpose of *technology scouting*.
- ▶ Visualization of research activities and experts within geographical regions, e.g. in the context of *technology brokerage*.
- ▶ *Partner matching* for research proposals.
- ▶ *Reviewer assignment* to conference/journal papers.
- ▶ Finding relevant *M. Sc. or Ph. D. supervisors*.
- ▶ ...

Expertise Retrieval Tasks

Expert finding is the task of finding experts given a topic describing the required expertise.

Expert profiling is the task of returning a list of topics that a person is knowledgeable about.

Building Blocks of Expert Finding Systems



Expert Finding

□ A list completion task

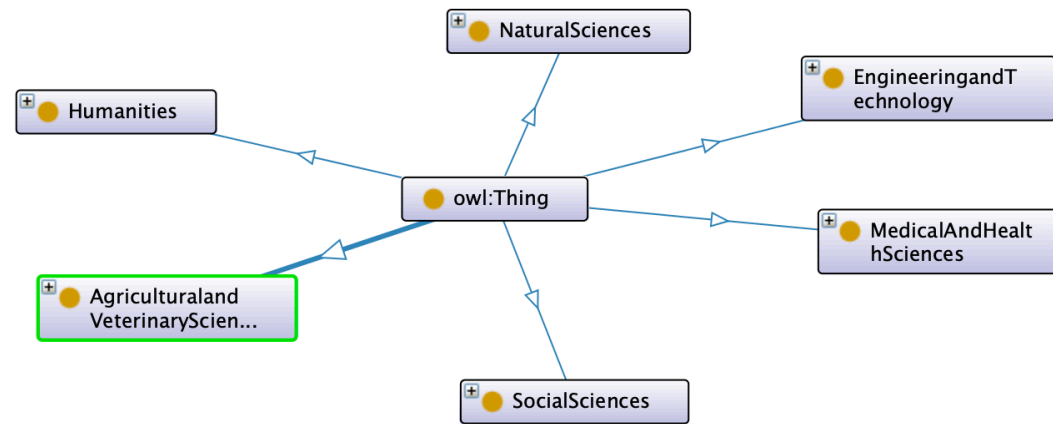
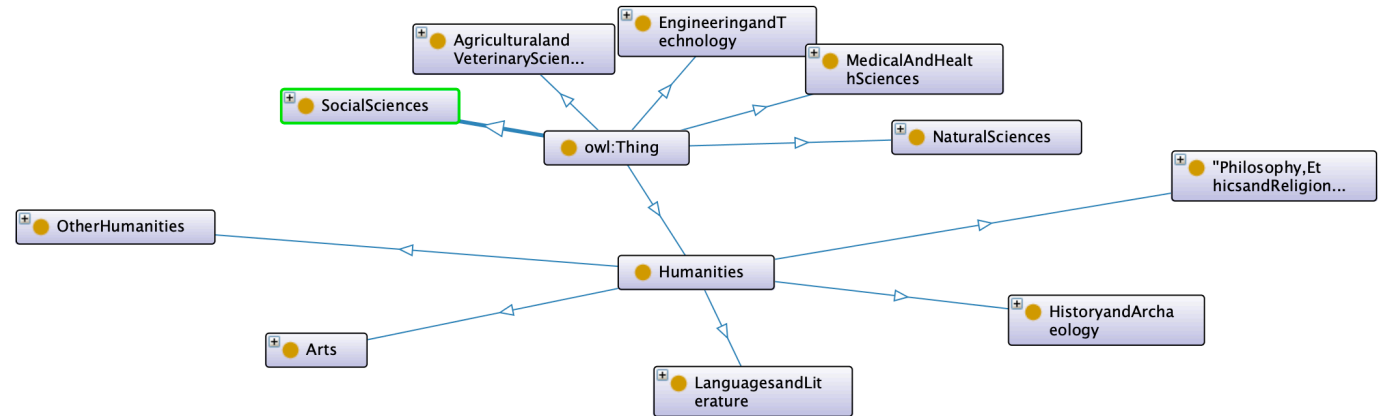
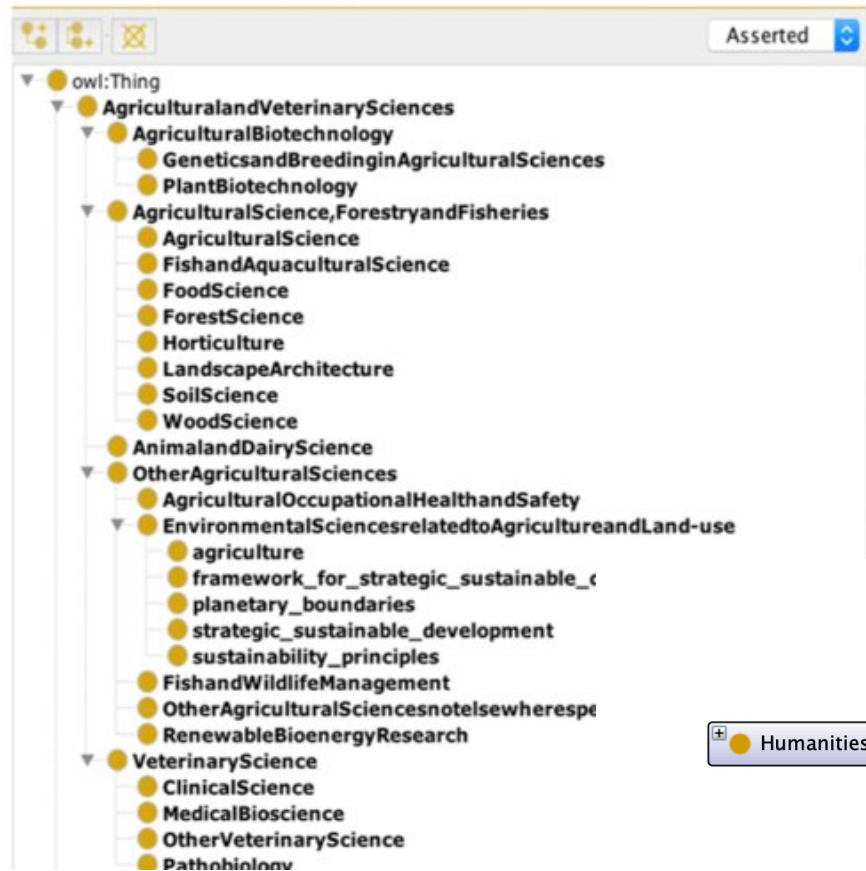
- The user provides an example expert who has been used to work on similar problems in the past, and the system returns a ranked list of experts.

□ A clustering task

- The experts are grouped into a number of disjoint expert areas (clusters) by using some clustering algorithm.
- The user restricts her/his considerations only to those experts who are within the cluster that is more close to the task's subject.



Conceptual Model of the Domain



Expert Profiling

- ▶ An expert profile may be quite complex and can, for example, be associated with information that includes: email address, affiliation, a list of publications, co-authors, but it may also include or be associated with: educational and (or) employment history.
- ▶ This information can be separated into two parts: expert's personal data and information that describes the area of competence.
- ▶ Personal data is used to resolve the problem with ambiguity.



Ambiguity Resolving Algorithms

- ▶ The problem refers to the fact that multiple profiles may represent one and the same person and therefore must be merged into a single generalized expert profile.
- ▶ The process of merging expert profiles is driven by the calculation of the similarity scores between different entities composing the profile, e.g. expert name, affiliation, email address and etc.

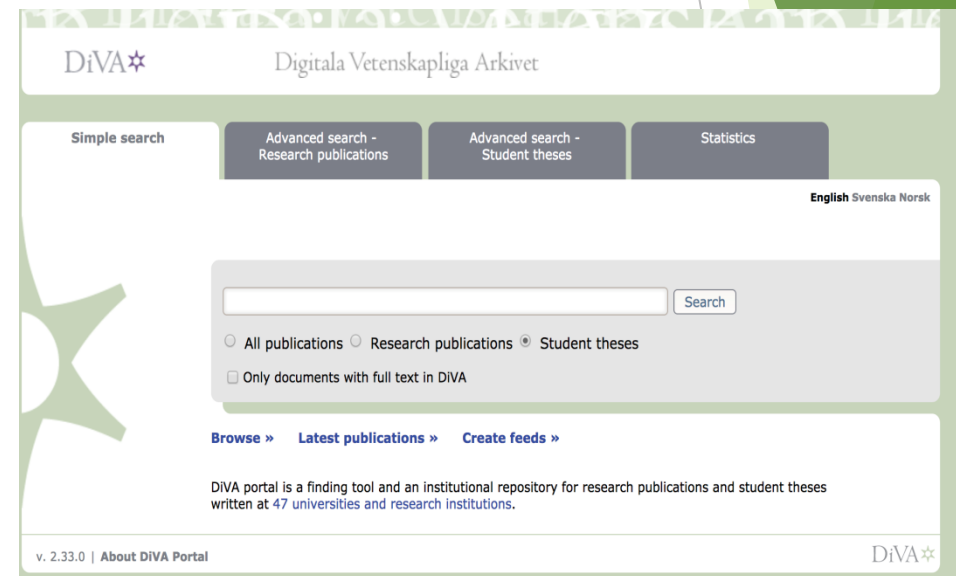


Source: <http://yhwu.me/publications/vast16-slides.pdf>

DATA: Institutional Repository DiVA



- ❑ Data is extracted from the BTH DiVA installation, concerning thesis supervision of researchers affiliated with BTH.
- ❑ DiVA is a publication and archiving platform for research publications and student essays used by 46 universities and authorities in the Nordic countries
- ❑ The data set consists of 2216 records of student theses published between 2010-2017.



First Experiment Scenario

The example expert expertise: *database, performance, usability, web server, cloud computing and Amazon web services.*

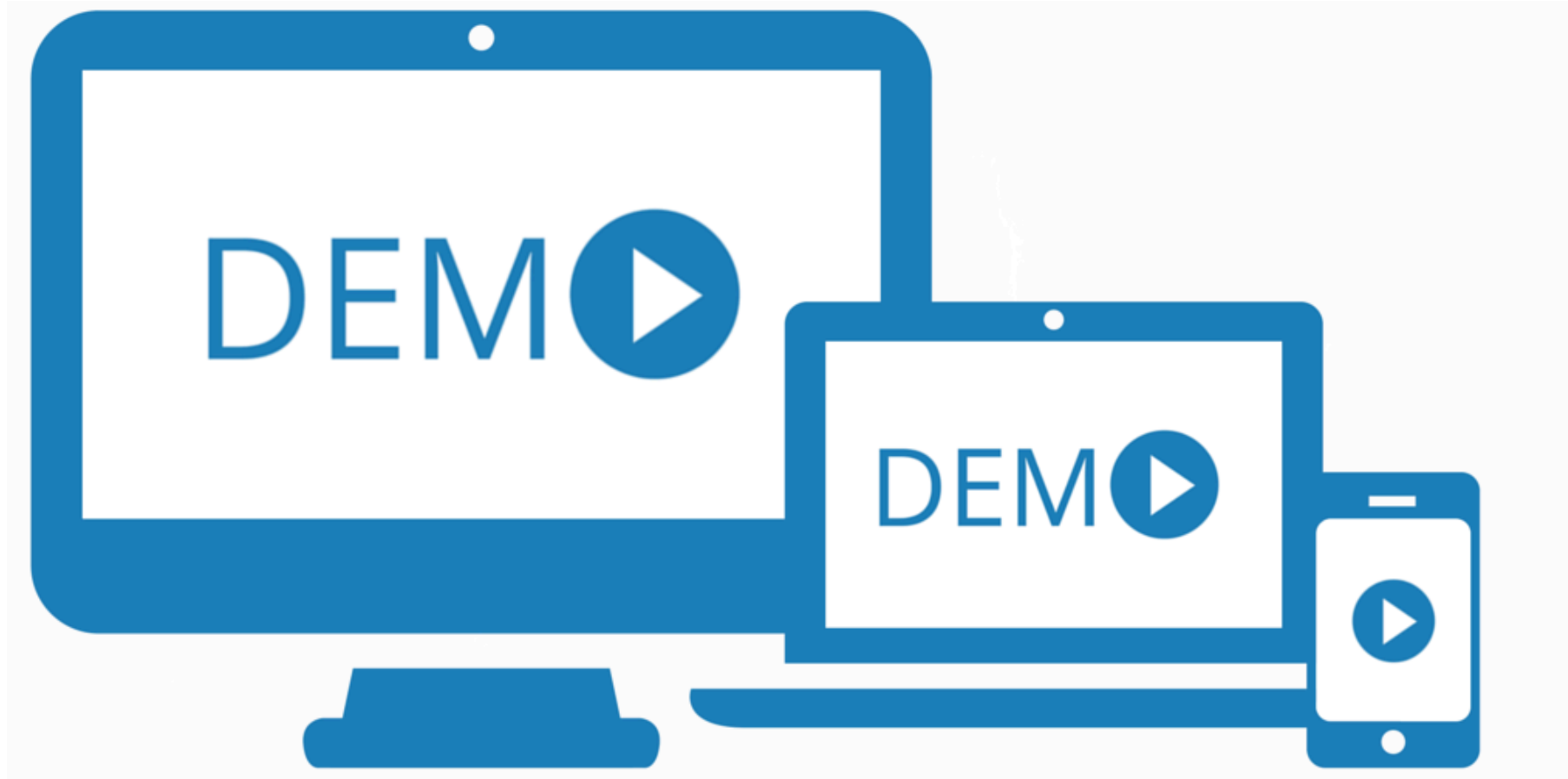
Expert	Keywords	Expertise similarity score
1	user experience, usability	0.770
2	privacy, security, cloud computing	0.763
3	cloud computing , security metrics, security threats, security measurement frameworks	0.760
4	procedural city generation, perlin noise, performance , game content	0.760
5	machine learning, parallel computing, multiprocessor, performance	0.760
6	mobile, power, consumption, android, native, web , enterprise service bus, performance , framework	0.754
7	mongodb, couchdb, python, pymongo, couchdb-python, nosql, document database , json, dbms, database	0.754
8	compression, sms, arithmetic, lambda, huffman, lzw, lz77, lz78, fristående kurs, voltdb, mysql , databases , main-memory database , primary memory database , performance	0.751
9	non-functional search-based software testing, non-functional system properties, search-based software testing, meta-heuristic techniques, performance testing , load testing, load patterns	0.75
10	digital multimedia broadcasting, mpeg-2 standard, mpeg-4 standard, video transport stream	0.75

Second Experiment Scenario

The example expert expertise: *database, performance, usability, web server, cloud computing and Amazon web services.*

Experts	Cluster	Description of clusters
3, 4, 7, 8	0	usability ; tessellation; android; security threats; main-memory database ; database ; distributed databases ; parallel computing; security; data mining and etc.
2, 5, 6, 10	3	usability ; data mining; performance monitoring ; systematic review; video streaming; parallel computing; mpeg-2 standard; mpeg-2 standard; nosql database ; machine learning; cloud computing and etc.
1, 9	15	usability ; quality of experience; urban design; systematic literature review and etc.

Demo Time



Questions?



Thank you for your attention! ;)