

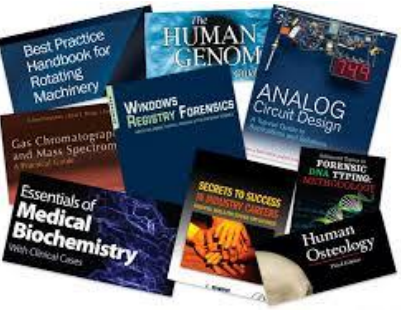


ELSEVIER

# Smart Data and Algorithms for the Publishing Industry

Antonio Gulli

# Elsevier – A data driven company



- Founded over 130 years ago
- Over the last 50 years the majority of Noble Laureates have published with Elsevier
- Employ over 7,000 employees in 24 countries



- Published over 330,000 articles
- Received over 1 million submissions
- Work with over 30 million Scientists, students, health & information professionals
- Over 58 million items indexed by Scopus
- Over 400 millions articles known by Mendeley
- Usage data on ScienceDirect, Scopus, Mendeley, EV, Scival, Pure and the submission platforms

# Smart Data is a central element for the Publishing Industry: content, usage, transactions



AUTHOR



READER



REVIEWER



RESEARCHER



COLLABORATOR



EDITOR



TEACHER

One login, one cookie

online   
ELSEVIER  
EDITORIAL  
SYSTEM

[www.elsevier.com](http://www.elsevier.com)  
[journals.elsevier.com](http://journals.elsevier.com)

Pure



ScienceDirect

EVISE\*

SciVal

Health Advance

Scopus

Search & Discovery

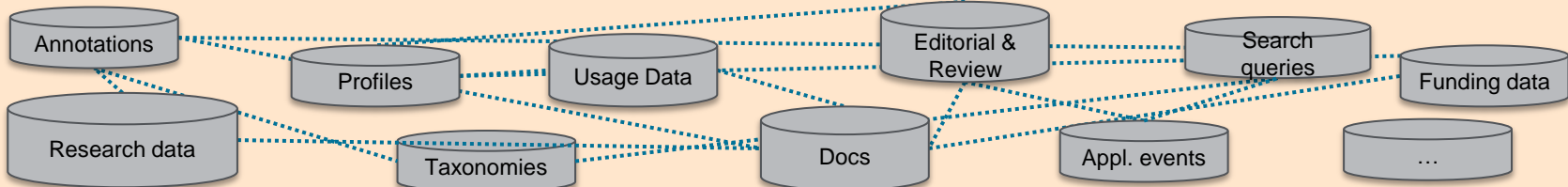
Data creation & Query

Identity & Authentication

A/B

Event notification

Content creation



Researchers

Products

Capabilities

Mining the signals

Data:



ELSEVIER



# Platforms





Advanced search

### ScienceDirect

is a leading full-text scientific database offering journal articles and book chapters from nearly 2,500 journals and more than 30,000 books.

12,609,638  
articles on ScienceDirect

#### Browse publications by subject

##### Physical Sciences and Engineering

- Chemical Engineering
- Chemistry
- Computer Science
- Earth and Planetary Sciences
- Energy
- Engineering
- Materials Science
- Mathematics
- Physics and Astronomy

##### Life Sciences

- Agricultural and Biological Sciences
- Biochemistry, Genetics and Molecular Biology
- Environmental Science
- Immunology and Microbiology
- Neuroscience

##### Health Sciences

- Medicine and Dentistry
- Nursing and Health Professions
- Pharmacology, Toxicology and Pharmaceutical Science
- Veterinary Science and Veterinary Medicine

##### Social Sciences and Humanities

- Arts and Humanities
- Business, Management and Accounting
- Decision Sciences
- Economics, Econometrics and Finance
- Psychology
- Social Sciences

#### Browse publications by title

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	0-9			

#### Browse all titles

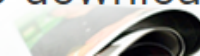
#### Open Access

Articles published in our Open Access journals are made permanently free for everyone to access immediately upon publication.

[View the Open Access journal directory](#)

#### ScienceDirect's Top 25

View the top downloaded articles.



#### Latest articles

Probing the structure of a liquid metal during vitrification  
*Acta Materialia*  
Atomic-scale intergranular crack-tip plasticity in tilt grain boundaries acting...  
*Acta Materialia*

Scopus Scopus SciVal | Library catalogue | Elizabeth Dyas ▾ Logout Help ▾ Brought to you by Scopus Team

Search Alerts My list My Scopus

Analyze search results

[Export](#) | [Print](#) | [E-mail](#)

TITLE-ABS-KEY ( heart attack ) [Back to your search results](#)

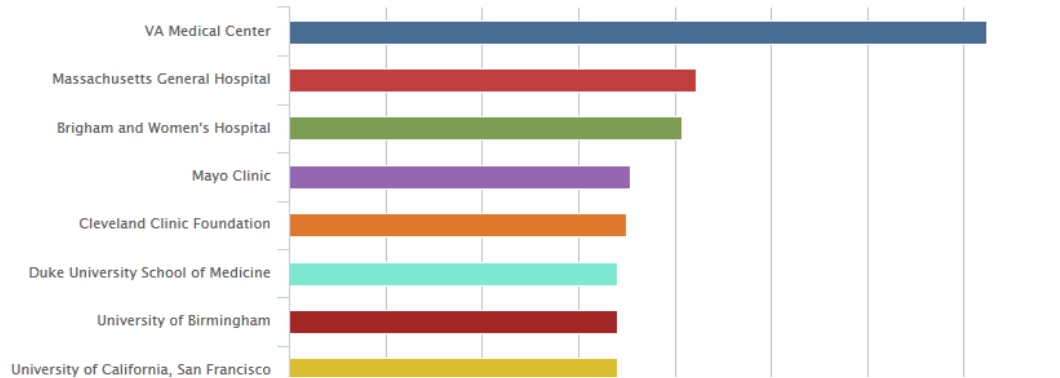
26105 document results Choose date range to analyze: 1885 to 2015 [Analyze](#)

Year	Source	Author	Affiliation	Country	Document type	Subject area
------	--------	--------	-------------	---------	---------------	--------------

Affiliation	Documents
<input checked="" type="checkbox"/> VA Medical Center	362
<input checked="" type="checkbox"/> Massachusetts General Hospital	211
<input checked="" type="checkbox"/> Brigham and Women's Hospital	204
<input checked="" type="checkbox"/> Mayo Clinic	177
<input checked="" type="checkbox"/> Cleveland Clinic Foundation	175
<input checked="" type="checkbox"/> Duke University School of Medi...	170
<input checked="" type="checkbox"/> University of Birmingham	170
<input checked="" type="checkbox"/> University of California, San Fr...	170
<input checked="" type="checkbox"/> Erasmus University Medical C...	162
<input checked="" type="checkbox"/> University of Toronto	152

## Documents by affiliation

Compare the document counts for up to 15 affiliations



The largest abstract and citation database of peer-reviewed research literature. Its used by academics, government researchers and corporate R&D professionals who need to search, discover and analyze research.

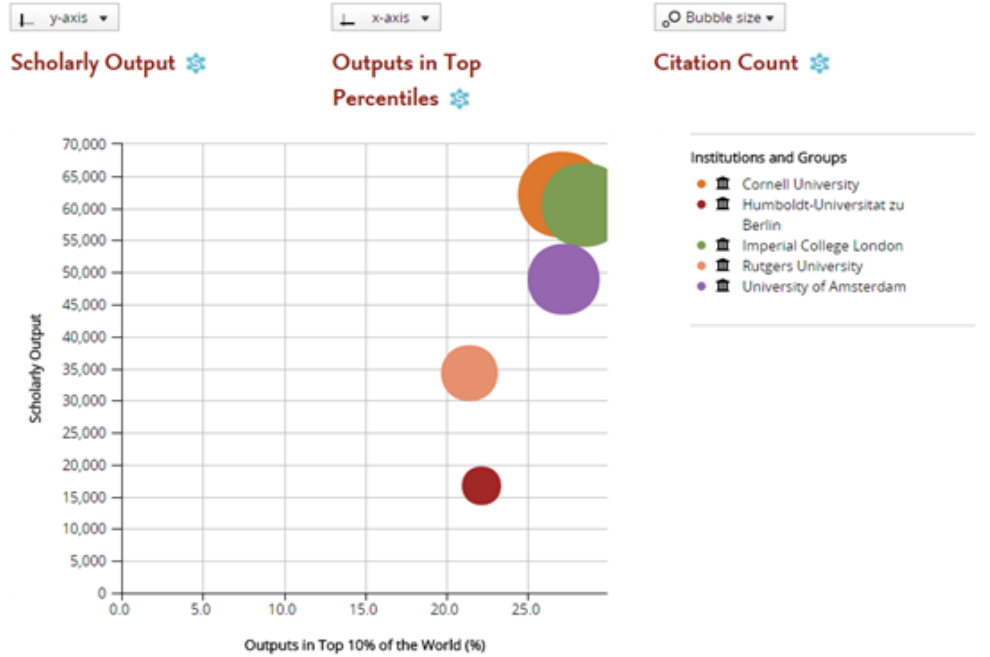
**21,900+ journals | 5,000 publishers | 58 million items | 3,000+ customers**

- Hide tags
- Institutions and Groups
- Cornell University
  - Humboldt-Universität zu Berlin
  - Imperial College London
  - Rutgers University
  - University of Amsterdam
  - Harvard University
  - King Abdullah University of Science and Technology
  - University of Cambridge
  - University of Groningen
  - University of Oxford
  - Wageningen University and Research Center
  - Yale University
- + Add Institutions and Groups
- x Remove all entities from this section
- Researchers and Groups
- Publication Sets
- Countries and Groups
- Research Areas

## Benchmarking

Source: Scopus data up to 16 Dec 2014 2006 to 2014 no filter selected

Chart Table Export



- Every 4 weeks the entire SciVal dataset is updated
- Combinations 34.4M Scopus articles and their 400M citations
- roughly 90,000B metric combinations



Welcome back antonio gulli

My Account

Install Mendeley Desktop

We're hiring! / Support

Get Mendeley

Dashboard

My Library

Papers

Groups

People

Papers graphene



## Papers

graphene



eg: scientific impact measures

Advanced search

Results 1 - 20 of 39,566

Prev 1 2 3 ... 1979 Next

Refine search

Open Access articles only

### The rise of graphene.

A K Geim, K S Novoselov in *Nature materials* (2007)

Graphene is a rapidly rising star on the horizon of materials science and condensed-matter physics

Save reference to library · Related research

11,650 readers

### Graphene transistors.

Frank Schwierz in *Nature nanotechnology* (2010)

Graphene has changed from being the exclusive domain of condensed-matter physicists to being explored by those in the electron-device community. In particular, graphene-based transistors have

Save reference to library · Related research

984 readers

### Graphene: status and prospects.

A K Geim in *Science (New York, N.Y.)* (2009)

Graphene is a wonder material with many superlatives to its name. It is the thinnest known material

Save reference to library · Related research

1,742 readers

### The chemistry of graphene oxide.

Daniel R Dreyer, Sungjin Park, Christopher W Bielawski, Rodney S Ruoff in *Chemical Society reviews* (2010)



Feedback





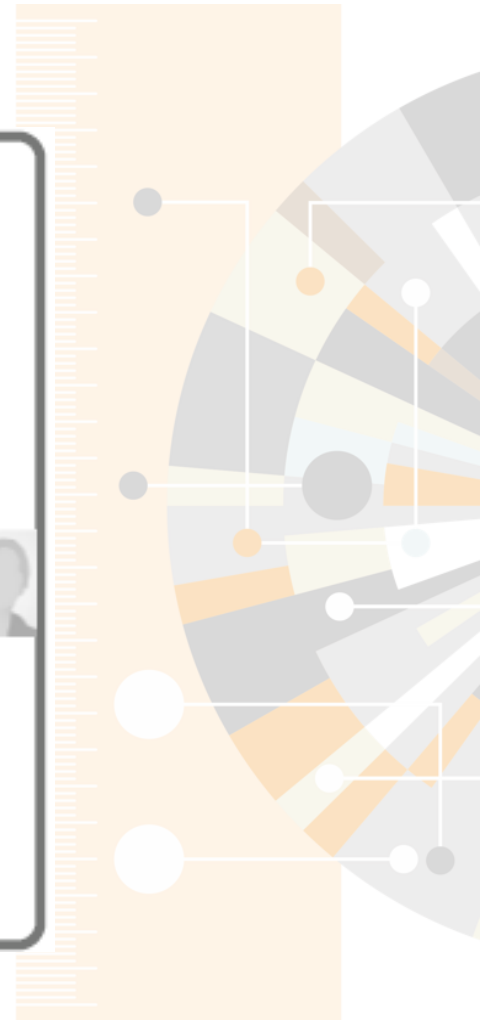
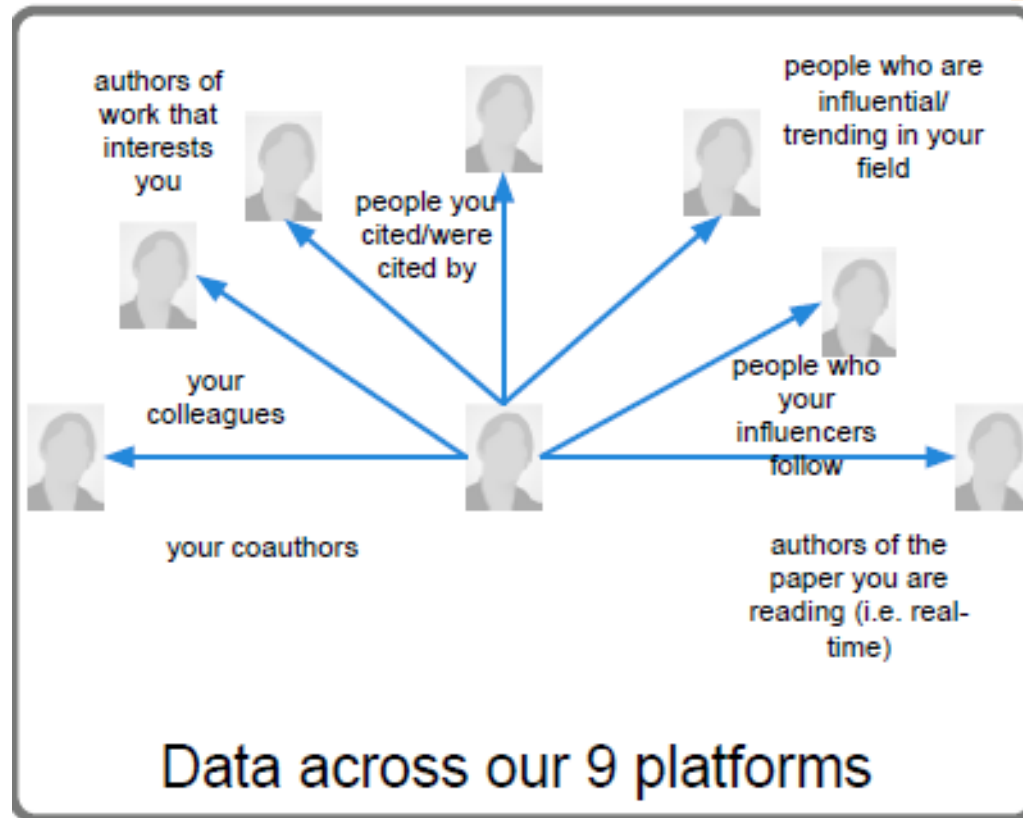
ELSEVIER



# Application: Recommenders



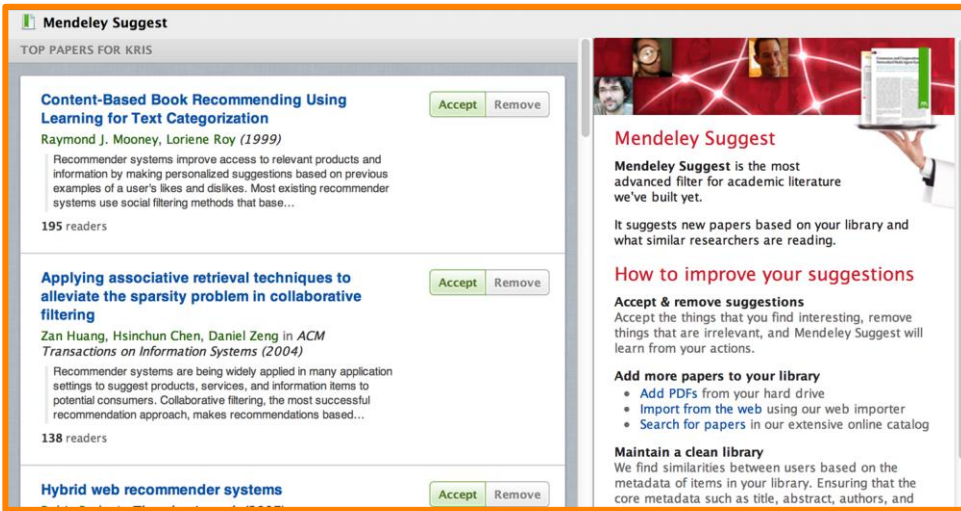
# “Related and relevant” topics for multidisciplinary areas



# Recommenders @ Mendeley

## Experiments & Results

- Offline Evaluation
- Online A/B testing
  - 2 email tests
  - ~40-50k users
  - 6-8 algorithms
  - 40-50% Open Rate
  - 8-12% CTR



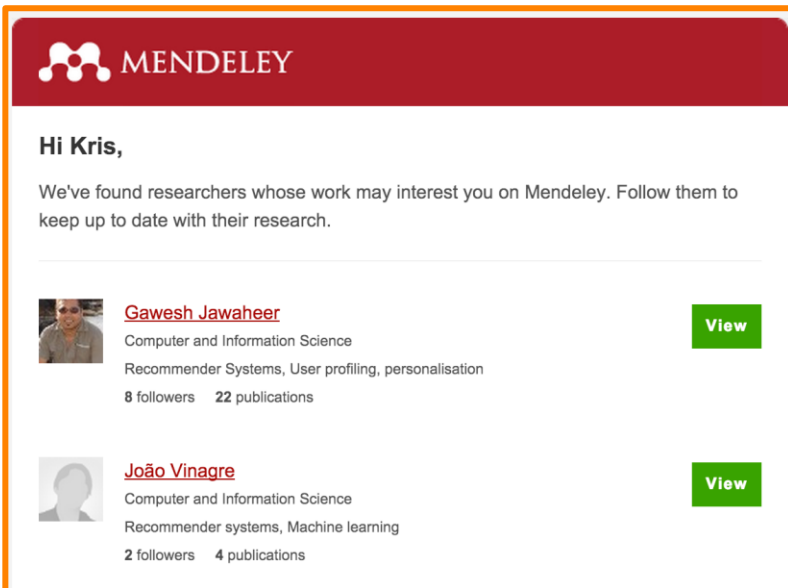
**Mendeley Suggest**  
TOP PAPERS FOR KRIS

**Content-Based Book Recommending Using Learning for Text Categorization**  
Raymond J. Mooney, Lorie Roy (1999)  
195 readers

**Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering**  
Zan Huang, Hsinchun Chen, Daniel Zeng in *ACM Transactions on Information Systems* (2004)  
138 readers

**Hybrid web recommender systems**

**Mendeley Suggest**  
Mendeley Suggest is the most advanced filter for academic literature we've built yet.  
It suggests new papers based on your library and what similar researchers are reading.  
**How to improve your suggestions**  
**Accept & remove suggestions**  
Accept the things that you find interesting, remove things that are irrelevant, and Mendeley Suggest will learn from your actions.  
**Add more papers to your library**  
• Add PDFs from your hard drive  
• Import from the web using our web importer  
• Search for papers in our extensive online catalog  
**Maintain a clean library**  
We find similarities between users based on the metadata of items in your library. Ensuring that the core metadata such as title, abstract, authors, and



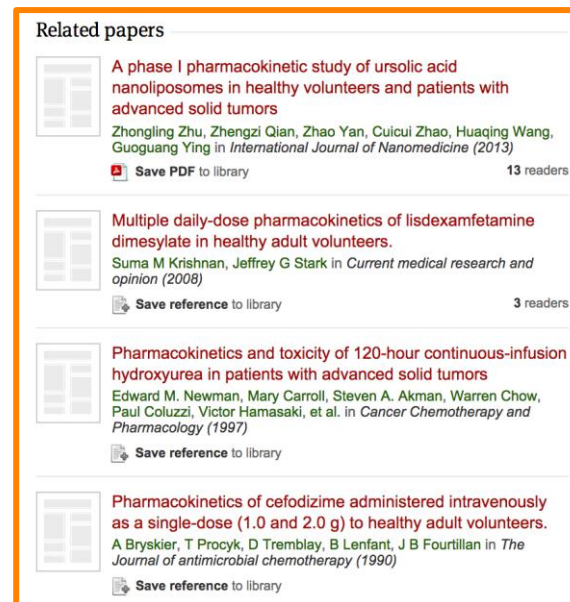
**MENDELEY**

Hi Kris,

We've found researchers whose work may interest you on Mendeley. Follow them to keep up to date with their research.

**Gawesh Jawaheer**  
Computer and Information Science  
Recommender Systems, User profiling, personalisation  
8 followers 22 publications

**João Vinagre**  
Computer and Information Science  
Recommender systems, Machine learning  
2 followers 4 publications



**Related papers**

A phase I pharmacokinetic study of ursolic acid nanoliposomes in healthy volunteers and patients with advanced solid tumors  
Zhongling Zhu, Zhengzi Qian, Zhao Yan, Cuicui Zhao, Huaqing Wang, Guoguang Ying in *International Journal of Nanomedicine* (2013)  
Save PDF to library 13 readers

Multiple daily-dose pharmacokinetics of lisdexamfetamine dimesylate in healthy adult volunteers.  
Suma M Krishnan, Jeffrey G Stark in *Current medical research and opinion* (2008)  
Save reference to library 3 readers

Pharmacokinetics and toxicity of 120-hour continuous-infusion hydroxyurea in patients with advanced solid tumors  
Edward M. Newman, Mary Carroll, Steven A. Akman, Warren Chow, Paul Coluzzi, Victor Hamasaki, et al. in *Cancer Chemotherapy and Pharmacology* (1997)  
Save reference to library

Pharmacokinetics of cefodizime administered intravenously as a single-dose (1.0 and 2.0 g) to healthy adult volunteers.  
A Bryskier, T Procyk, D Tremblay, B Lenfant, J B Fournillan in *The Journal of antimicrobial chemotherapy* (1990)  
Save reference to library

# EPFL/Recommend – Increase Click Through Rate and Article Accesses



- When customers look at articles on ScienceDirect, they're also provided with links to other articles of interest



The screenshot shows the ScienceDirect interface for an article. At the top, the ScienceDirect logo is on the left, and the user name 'Jeff Honious' with a 'Logout' link is on the right. A navigation bar contains links for 'Home + Recent Actions', 'Publications', 'Search', 'My settings', 'My alerts', and 'Help'. Below this is a search bar with the text 'Search ScienceDirect' and a search icon. The main content area features the Elsevier logo and the journal title 'Information Processing Letters', Volume 83, Issue 4, 31 August 2002, Pages 211–220. The article title is 'Mining association rules using inverted hashing and pruning'. The authors are John D. Holt and Soon M. Chung. The abstract is visible on the left. On the right, there is a 'Bibliographic information' section with a DOI link and a 'Get rights and content' button. Below this is a 'Recommendations' section with a list of suggested articles, each with a 'Show more information' link. A yellow box highlights the 'Article page view in ScienceDirect' text, and another yellow box highlights the 'Recommendations' section.

**Article page view in ScienceDirect**

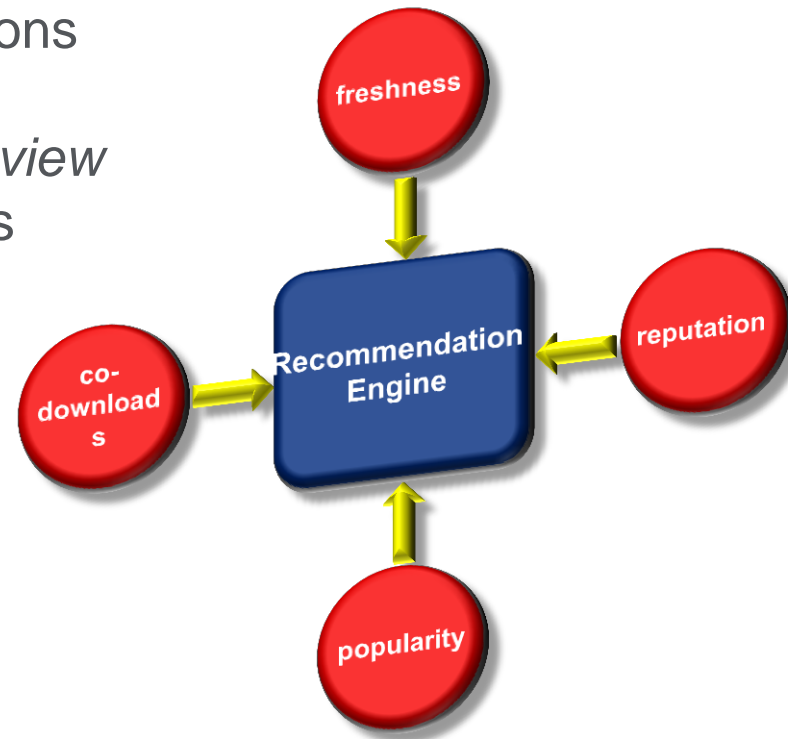
**Recommendations**

# EPFL / Recommend

EPFL: École Polytechnique Fédérale de Lausanne  
(Switzerland)

## Approach:

- Analyze usage data & articles on 4 dimensions
- Pre-compute recommendations on HPCC
- Offer recommendations during *article page view*
- A/B test variants for scoring / ranking results



# Recommendation Engine development



During the pilot...

- EPFL cycled through 6 *variants* while testing for best dimension combinations & weights
- A/B testing was used on 6 variants, each measured during a fixed usage window
  - Compare to existing related articles

Results

- The 6<sup>th</sup> variant provided ~65% increase in CTR (click-through-rate) over Related Articles
- Conclusion: Recommendation based on co-usage outperforms (the current) recommendation based on topical similarity... when properly tuned.

variant	desc	REL-CTR	REC-CTR	relDiff	relDiff@90%
V06	W300:C + F[10]:C:R[10]:P[10]	1.17%	1.60%	36.52%	27.54%
V04	W300:C + C:F[10]:R[10]:P[10]	1.55%	2.31%	48.84%	43.01%
V07	W900K1:C + F[10]:C:R[10]:P[10]	0.99%	1.35%	35.92%	26.47%
V05	W900K1:C(raw) + C(raw):F	1.66%	2.87%	72.70%	64.43%
V08	W900K1:C(raw)S + S:C(raw):F	1.53%	1.11%	-27.23%	34.53%
V09	W900K1:C(raw)S + C(raw):S:F	1.22%	2.12%	73.39%	65.06%



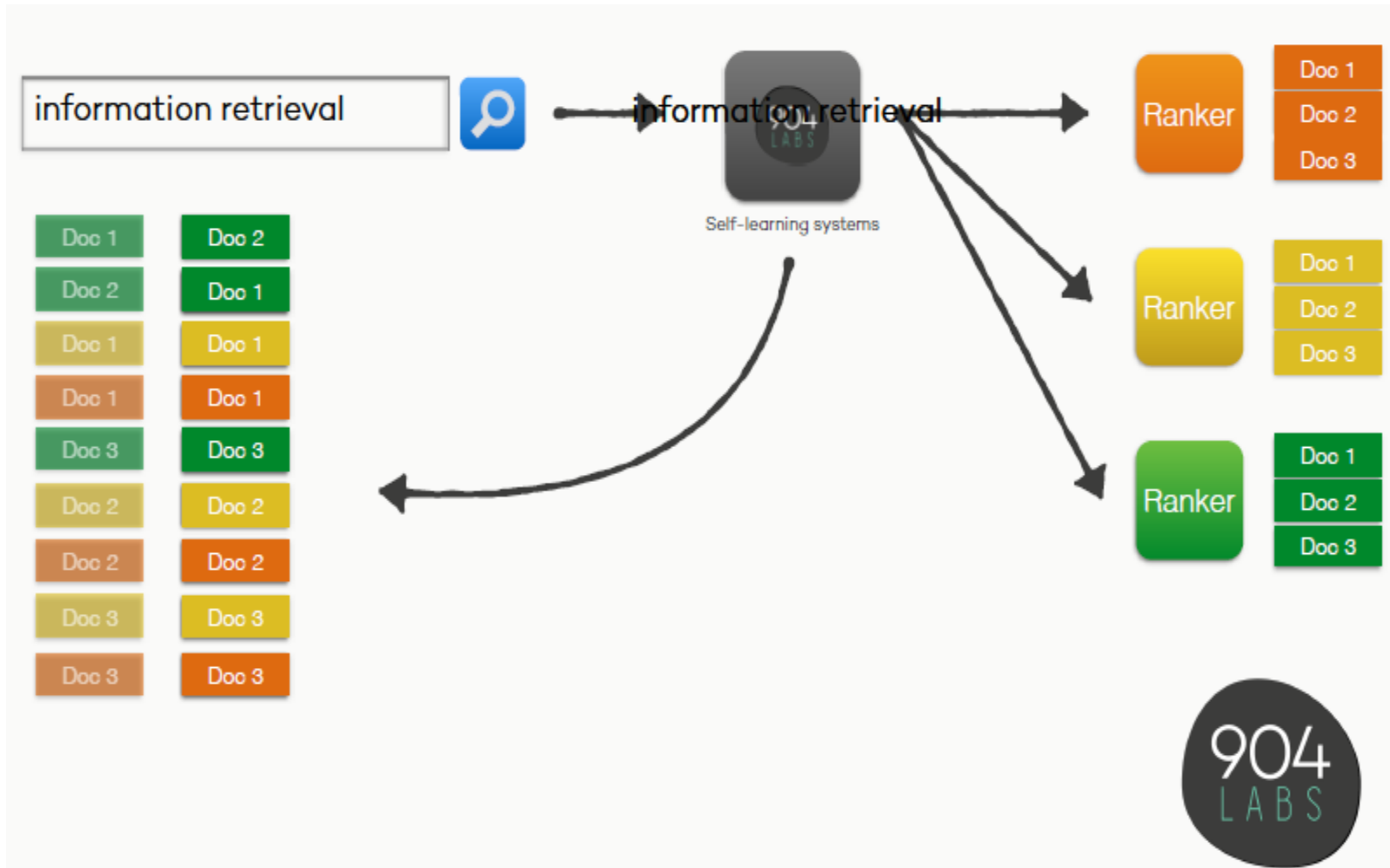
ELSEVIER



# Application::Search

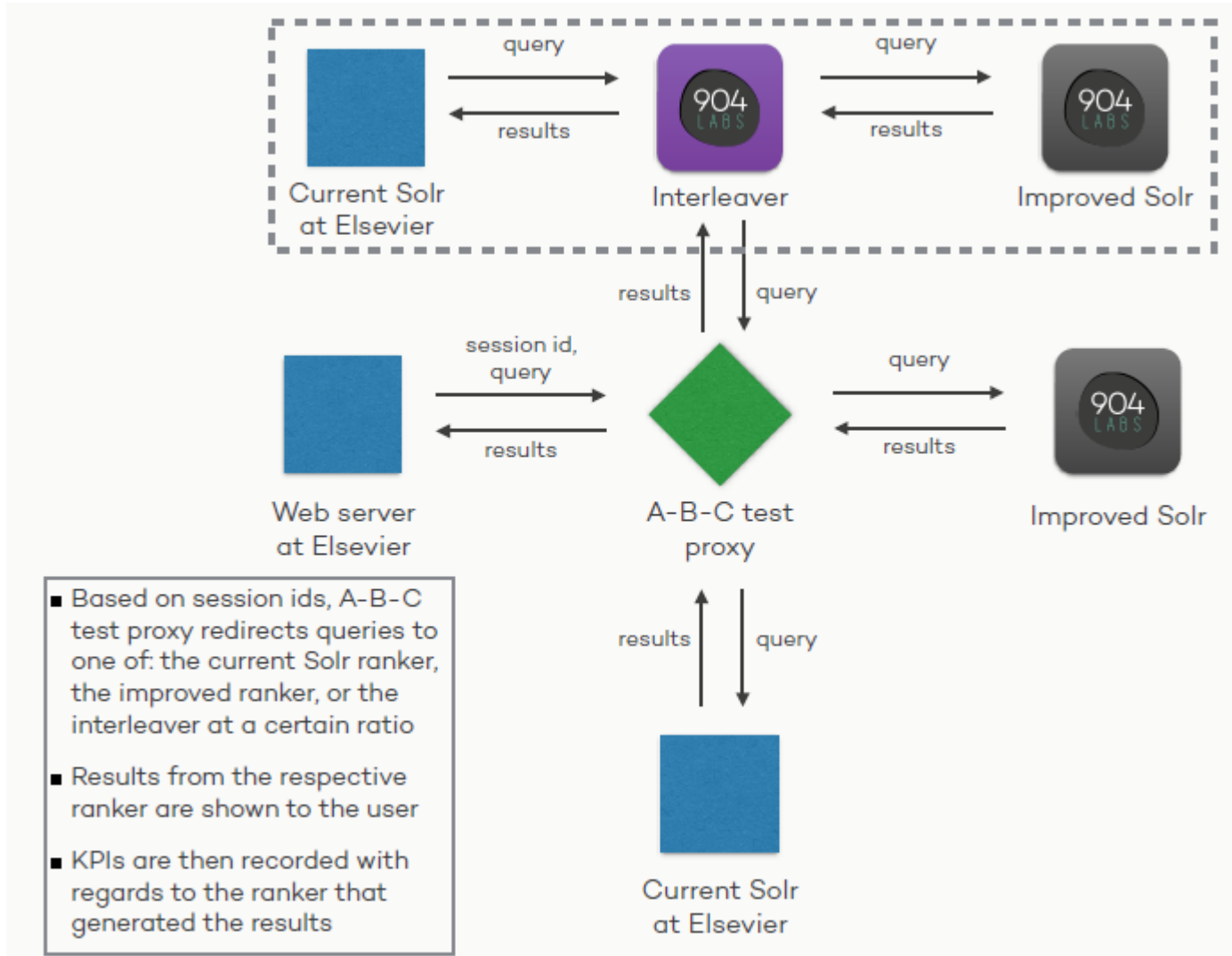


# 904Labs: Learning to Rank, interleaving & A/B/C





# Evaluation method



- Based on session ids, A-B-C test proxy redirects queries to one of: the current Solr ranker, the improved ranker, or the interleaver at a certain ratio
- Results from the respective ranker are shown to the user
- KPIs are then recorded with regards to the ranker that generated the results

# LanguageTools: Autosuggest

## LanguageTools Demo (ScienceDirect, integrated 2)

### Keywords

data structure design|

#### **data structure**

- data structure > design
- data structure > design considerations of the new interface
- data structure > design interfaces
- data structure > designof
- data structure > design and description of applied dependencies
- data structure > design and implementation of web service
- data structure > design and inherent limitations
- data structure > design and interface design
- data structure > design and planning assumptions
- data structure > design and traversal
- data structure > design guidelines
- data structure > design included normalization and entity-relationship
- data structure > design of algorithms
- data structure > design of free curve and surface reconstruction
- data structure > design of middleware interface
- data structure > design of specific layer interfaces
- data structure > design of unified interface
- data structure > design of web database
- data structure > designing algorithms
- data structure > designing new algorithms
- data structure > designing numerical algorithms
- data structure > designing program

#### **structure of the database**

#### **structure of the data**

#### **database structure**

#### hierarchical **data structure**

#### appropriate **data structure**

#### efficient **data structure**

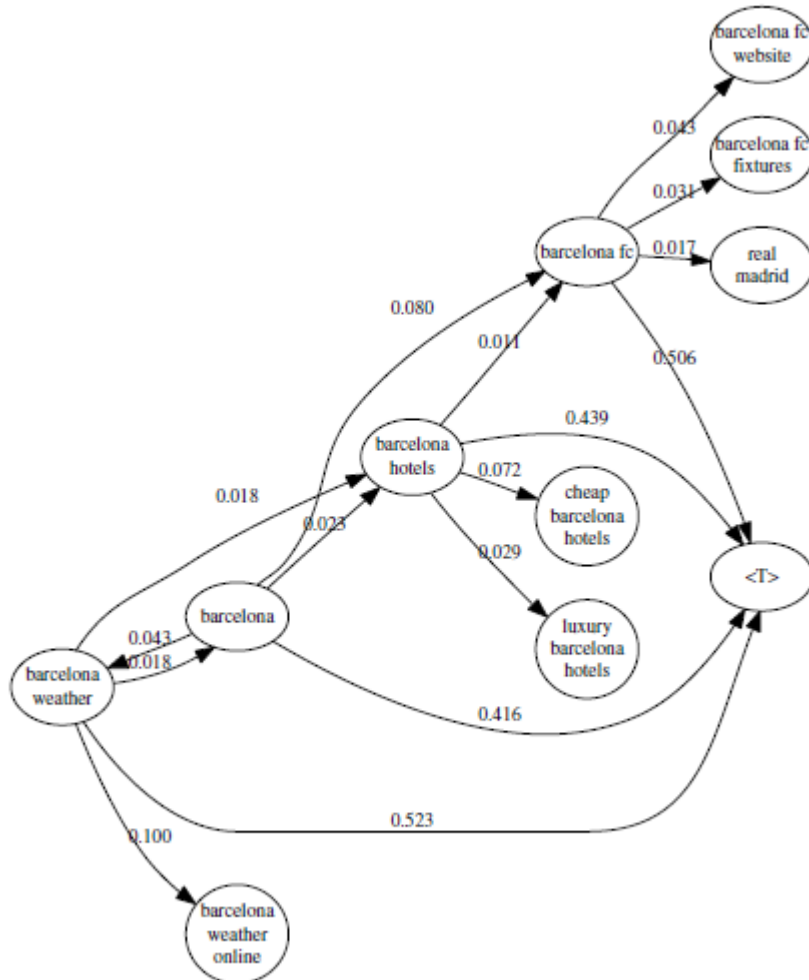
#### complex **data structure**

#### nuclear **structure data file**

venue-f-mysql.txt

Text Filters copy.zip

# UvA: Zero results/query suggestion



## Xinyi Li (University of Amsterdam) - Academic Failures and Struggles: Query Suggestions in Academic Search

Failed queries, in particular zero result queries, present a great challenge for search engines. In academic search, the failure rate is even higher than web search, where approximately 1 out of 10 queries leads to an empty search result page. In this study, we characterize the failures in academic search to understand what failures are there and how they happen. We then propose to help address this issue by providing query suggestions tailored for this setting. We apply different query suggestion models to sessions when failure happens and compare the performance. This study represents an ongoing work in collaboration with Elsevier.



ELSEVIER

# Application:: Entity Fingerprinting



# What, again, is a Fingerprint?



## *Entities: structured representation of unstructured data*

### Influenza A/H1N1v in pregnancy: An investigation of the characteristics and management of affected women and the relationship to pregnancy outcomes for mother and infant

► Show Grant Application Details

Edit Grant Application

#### Abstract

Influenza infection during pregnancy is associated with adverse maternal and fetal outcomes, including probable increases in the risk of maternal pneumonia and possible increases in risks of certain congenital malformations. Recent US H1N1 pandemic experience as well as data from previous influenza pandemics indicates higher morbidity and mortality among pregnant women. DESIGN: Systematic review and a subsequent prospective observational cohort study using several different sources to identify pregnant women needing treatment or prophylaxis. Information about pregnancy management and outcomes will be collected directly from health professionals, as well as women themselves where infection is managed in primary care. COHORT IDENTIFICATION: All women in the UK identified with confirmed or suspected influenza H1N1v or requiring antiviral prophylaxis, identified through the UK Teratology Information Service (UKTIS), the Reproductive Medicine and Childbirth Research Network, the Primary Care Research Network, the HPA Regional Microbiology Laboratory Network, self reporting by patients and active negative surveillance through the UK Obstetric Surveillance System (UKOSS). Information about comparison women will be obtained from the existing UKOSS database of over 1200 women giving birth in the UK. VIROLOGICAL CONFIRMATION OF H1N1: Women will undergo H1N1 testing arranged through the research network staff or by a self testing kit if not done already. DATA GATHERING: 1. Consenting women identified through UKTIS will be contacted by telephone (a) two weeks after the reported exposure/ first contact. Details of them and their illness will be collected (b) Patients who remain unwell will be followed up at two weekly intervals until recovery (c) Final follow up will be of maternal and pregnancy outcome two weeks after birth. 2. Nominated UKOSS reporting clinicians will report anonymised details of all pregnant women admitted with confirmed or suspected H1N1v infection via a specific web-based rapid reporting and data collection system. STUDY SIZE: We anticipate identifying 500-1000 affected pregnancies during the 6 month initial study period. Information on 1200 comparison women is available from existing UKOSS data. ANALYSIS: We will investigate the relationship between demographic, pregnancy characteristics, management and clinical outcomes in order to generate immediate recommendations for changes in practice to improve outcomes for this vulnerable group.

#### Applicants

Name

Date Added

Thomas, Simon H L

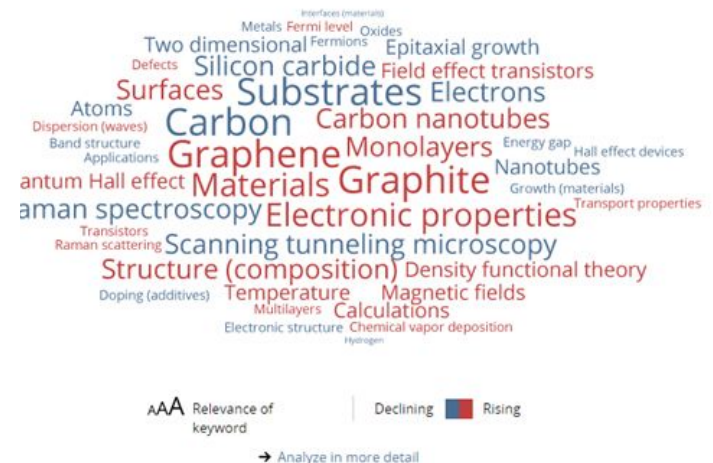
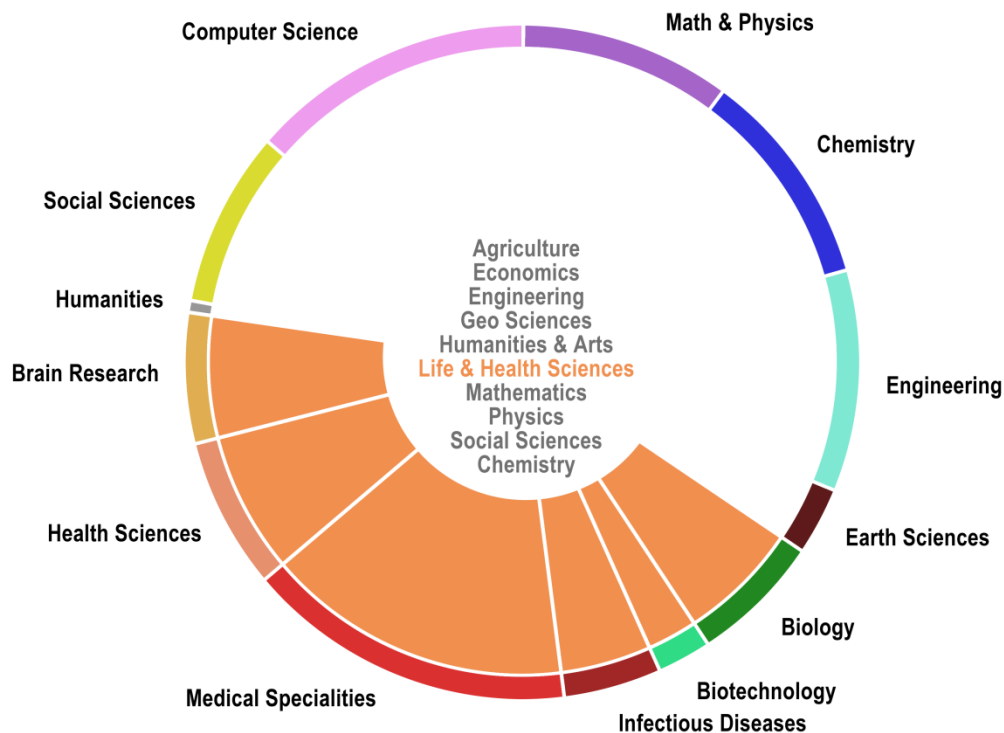


2/3/2010

Fingerprint	weight	Req
Statistical Data Interpreta...		<input checked="" type="checkbox"/> 100
Epidemiology		<input checked="" type="checkbox"/> 100
Women		<input type="checkbox"/> 100
Human Influenza		<input type="checkbox"/> 86
Obstetrics		<input type="checkbox"/> 85
Teratology		<input type="checkbox"/> 67
Pregnant Women		<input type="checkbox"/> 62
Pregnancy		<input type="checkbox"/> 59
Parturition		<input type="checkbox"/> 52
Information Services		<input type="checkbox"/> 48
Pregnancy Outcome		<input type="checkbox"/> 44
Primary Health Care		<input type="checkbox"/> 39
Infection		<input type="checkbox"/> 36
Disease Outbreaks		<input type="checkbox"/> 36
Research		<input type="checkbox"/> 36
Reproductive Medicine		<input type="checkbox"/> 27
Microbiology		<input type="checkbox"/> 24
Telephone		<input type="checkbox"/> 23

# Indexing Scientific Content with the Fingerprint Engine

- Compute semantic content representation of scientific text
- Support of multiple vocabularies / ontologies is important  
Engine's value – agile adaptation to new thesauri



## Find the perfect journal for your article

Elsevier® Journal Finder helps you find journals that could be best suited for publishing your scientific article.

Powered by the [Elsevier Fingerprint Engine™](#), Elsevier Journal Finder uses smart search technology and field-of-research specific vocabularies to match your article to Elsevier journals.

Simply insert your title and abstract and select the appropriate field-of-research for the best results.

### Paper title

astronomy

### Paper abstract

Astronomy is a natural science which is the study of celestial objects (such as stars, galaxies, planets, moons, and nebulae), the physics, chemistry, and evolution of such objects, and phenomena that originate outside the atmosphere of Earth, including supernovae explosions, gamma ray bursts, and cosmic microwave background radiation. A related but distinct subject, cosmology, is concerned with studying the universe as a whole.[1]

Astronomy is one of the oldest sciences. Prehistoric cultures have left astronomical artifacts such as the Egyptian monuments and Nubian monuments, and early civilizations such as the Babylonians, Greeks, Chinese, Indians, Iranians and Maya performed methodical observations of the night sky. However, the invention of the telescope was required before astronomy was able to develop into a modern science. Historically, astronomy has included disciplines as diverse as [astrometry](#), celestial navigation, observational astronomy and the making of calendars, but professional astronomy is nowadays often considered to be synonymous with astrophysics.[2]

## Search results (10)

Journal title **Sort by Match** ▼ Impact Factor Open Access Editorial Times Acceptance Production Times

### Physics Reports

**22.910** - **9 weeks** **100 %** **17 weeks**  
Match Impact Factor Open Access Editorial Times Acceptance Production Times

[View Scope and more information](#) ▼ [Full aims and scope on journal homepage](#) ↗ [Submit Your Paper](#) ↗



ELSEVIER

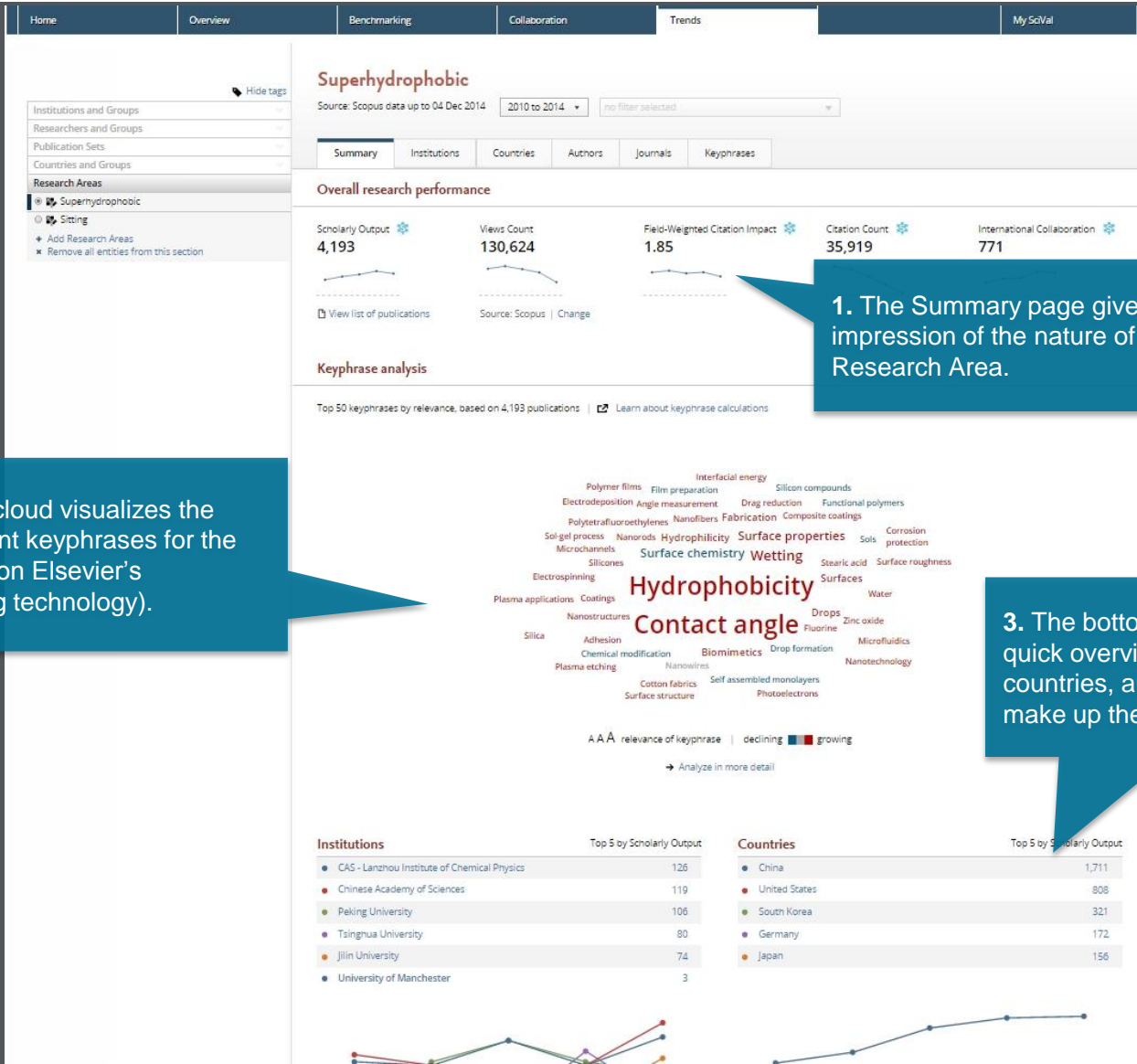


# Application: Research Trends





# Scival: Summary tab



1. The Summary page gives you a quick impression of the nature of and trends within the Research Area.

2. The word cloud visualizes the most important keyphrases for the field. (based on Elsevier's Fingerprinting technology).

3. The bottom of page provides a quick overview of the institutions, countries, authors and journals that make up the field.

# Scival : Institutions tab – map

Home Overview Benchmarking Collaboration Trends My SciVal

## Superhydrophobic

Source: Scopus data up to 04 Dec 2014 2010 to 2014 no filter selected

Summary **Institutions** Countries Authors Journals Keyphrases

### Top Institutions

Map Table Chart

Top 100 Institutions in this Research Area, by Scholarly Output

Worldwide All sectors Filter for more (regional) detail

Size: Scholarly Output total value Color: Views Count total value

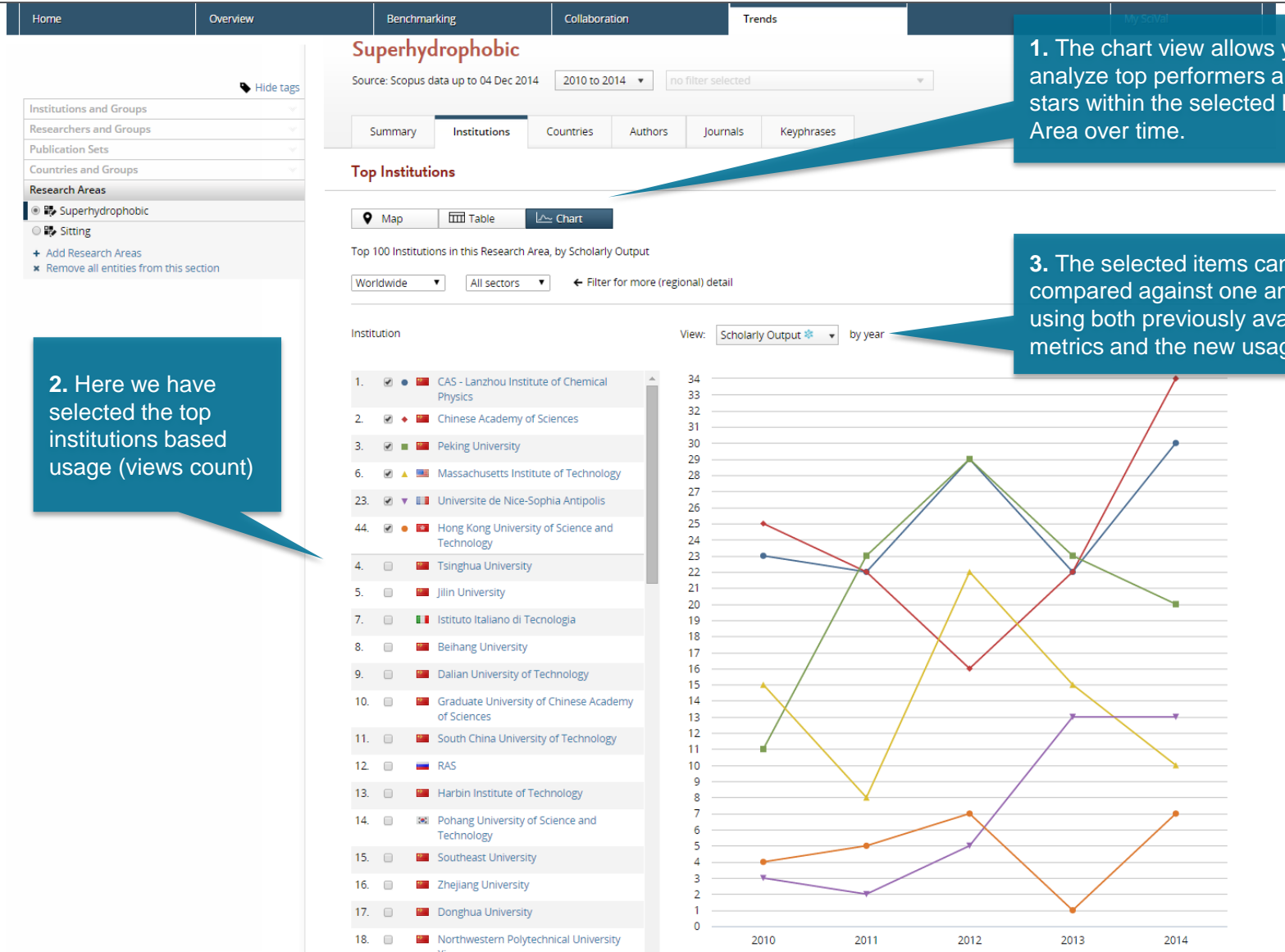
Location of elements on the map represents origin of publication, not who is viewing what.

1. The map view gives you a view on the world where you can see all citation and usage activity within your Research Area

2. View the geographical location of the top performing institutions and rising stars of this Research Area

3. 'Top performers' and 'rising stars' can be visualized using citation, output and usage metrics.

# Scival: Institution tab - chart



2. Here we have selected the top institutions based usage (views count)

1. The chart view allows you to analyze top performers and rising stars within the selected Research Area over time.

3. The selected items can be compared against one another using both previously available metrics and the new usage metrics.

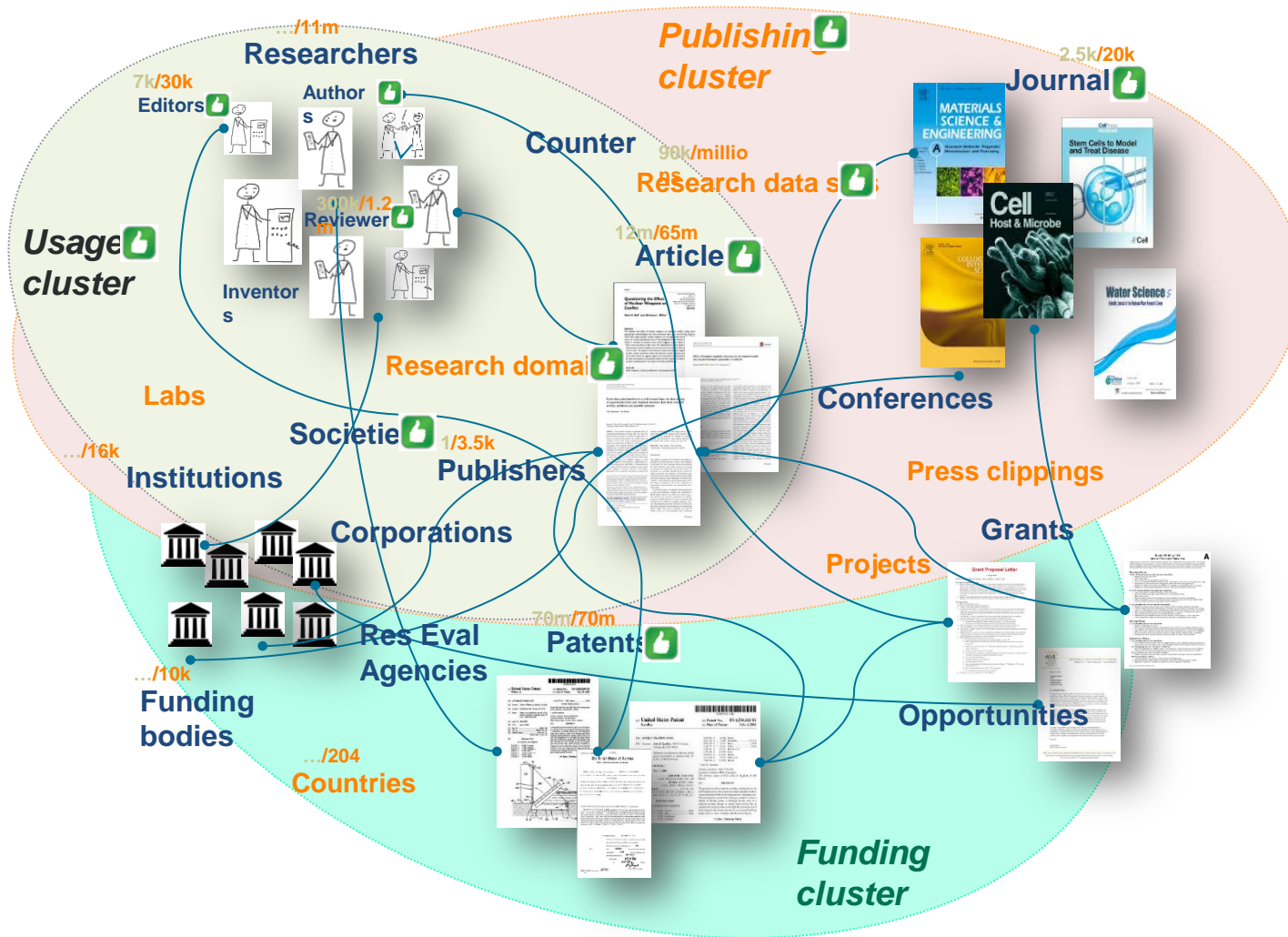


ELSEVIER

# Application:: Entity Disambiguation



# Entities in the Social Network of Science



- Typical needs of the entities:
- Develop research strategy
  - Obtain funding
  - Acquire talent
  - Prioritize activities
  - Working efficiently
  - Measure/Demonstrate societal impact
  - Develop career
  - Publish
  - Find reviewers
  - Build reputation
  - Ranking
  - ...

The Researcher Disambiguation Program is the place where those entities are given a true identity, with the aim to reduce the errors in attribution

# Searching in Scopus relies on disambiguating authors and their affiliations

- Scopus Author Identifier algorithm based on:
- Affiliation
  - Address
  - Subject area
  - Source title
  - Dates of publication
  - Citations
  - Co-authors

Author Last Name:  E.g., smith

Initials or First Name:  E.g., j.l.

Show exact matches only

Affiliation:  E.g., university of toronto

To determine which author names should be grouped together under a single identifier number, the Scopus Author Identifier uses an algorithm that matches author names based on their affiliation, address, subject area, source title, dates of publication, citations, and co-authors. Documents with insufficient data may not be matched, this can lead to more than one entry in the results list for the same author. By default, only details pages matched to more than one document in Scopus are shown in search results. [About Scopus Author Identifier](#)

Author results: 38 of 50 [Show Profile Matches with One Document](#) Go to page: 1 of 2  | [Next >](#)

Refine results:

Source Title:  Analytical Chemistry (3) >  Nature (3) >  Science (3) >  Biochemical and Biophysical Research Communications (2) >  American Journal of Roentgenology Radium Therapy and Nuclear Medicine (2) > [View more](#)

Affiliation:  University of Leeds (3) >  Albany Medical College (2) >  North Carolina State University (2) >  University of Pennsylvania (2) >  University of Vermont (2) >

With selected:  All  Page  Show documents  View citation overview  Request to merge authors

Sort by: Document Count (Descending)

Authors	Documents	Subject Area	Affiliation	City	Country
1 <a href="#">Holt, Peter John</a> Holt, P. J.	221 <a href="#">Show Last Title</a>	Physics and Astronomy ; Engineering ; Biochemistry, Genetics and Molecular Biology; ...	European Organization for Nuclear Research	Geneve	Switzerland
2 <a href="#">Holt, John A G</a> Holt, J. Holt, John Holt, J. A G	126 <a href="#">Show Last Title</a>	Medicine ; Biochemistry, Genetics and Molecular Biology ; Pharmacology, Toxicology and Pharmaceutics; ...	Inst. Radiother./Oncol. W. Australia	Perth	Australia
3 <a href="#">Holt, John M.</a> Holt, J. Holt, J. M. Holt, John	58 <a href="#">Show Last Title</a>	Earth and Planetary Sciences ; Physics and Astronomy ; Decision Sciences; ...	MIT Haystack Observatory	Westford	United States
4 <a href="#">Holt, John C.</a> Holt, J. Holt, John Holt, J. C.	49 <a href="#">Show Last Title</a>	Biochemistry, Genetics and Molecular Biology ; Medicine ; Immunology and Microbiology; ...	University of Kentucky	Lexington	United States
5 <a href="#">Holt, John W.</a> Holt, J. Holt, John Holt, J. W.	41 <a href="#">Show Last Title</a>	Earth and Planetary Sciences ; Multidisciplinary ; Computer Science; ...	University of Texas at Austin	Austin	United States

# Authors like “Michael Smith” can be problematic



All of these records appear to be for the same “Michael Smith,” but they have separate Author IDs in Scopus

68	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Medicine ; Pharmacology, Toxicology and Pharmaceutics ; Decision Sciences; ...	The University of British Columbia
69	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Medicine ; Pharmacology, Toxicology and Pharmaceutics ; Decision Sciences; ...	The University of British Columbia
70	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Medicine ; Pharmacology, Toxicology and Pharmaceutics ; Decision Sciences; ...	The University of British Columbia
71	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Immunology and Microbiology ; Medicine	The University of British Columbia
72	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Biochemistry, Genetics and Molecular Biology	The University of British Columbia
73	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Biochemistry, Genetics and Molecular Biology	The University of British Columbia
74	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Biochemistry, Genetics and Molecular Biology ; Medicine	The University of British Columbia
75	Smith, Michael Smith, M.	1 <a href="#">Show Last Title</a>	Biochemistry, Genetics and Molecular Biology ; Pharmacology, Toxicology and Pharmaceutics ; Medicine; ...	The University of British Columbia

# Authors like “Wei Wu” are also a challenge

This “Wei Wu” appears to have 1642 documents, but they span all of these subject areas:

- Engineering
- Materials Science
- Computer Science
- Physics and Astronomy
- Chemistry
- Medicine
- Biochemistry, Genetics and Molecular Biology
- Chemical Engineering
- Mathematics
- Pharmacology, Toxicology and Pharmaceutics
- Energy
- Immunology and Microbiology
- Neuroscience
- Environmental Science
- Social Sciences
- Agricultural and Biological Sciences
- Earth and Planetary Sciences
- Business, Management and Accounting
- Multidisciplinary
- Decision Sciences
- Dentistry
- Health Professions
- Veterinary
- Arts and Humanities
- Nursing

1,642 document results | [Analyze results](#) | [Show all abstracts](#) | Sort by [Date \(Newest\)](#)

[All](#) | [Download](#) | [Export](#) | [View citation overview](#) | [View Cited by](#) | [Request to remove documents from author](#) | [More...](#)

Page	Document title	Author(s)	Date	Source title	Cited by
1	<a href="#">How to achieve non-repudiation of origin with privacy protection in cloud computing</a>	Wu, W., Zhou, J., <a href="#">Kiang, Y.</a> , Xu, L.	2013	<i>Journal of Computer and System Sciences</i> 79 (8), pp. 1200-1213	0
2	<a href="#">Solidification of nanostructured lipid carriers (NLCs) onto pellets by fluid-bed coating: Preparation, in vitro characterization and bioavailability in dogs</a>	Tian, Z., Yi, Y., Yuan, H., Han, J., Zhang, X., Xie, Y., Lu, Y., (...), Wu, W.	2013	<i>Powder Technology</i> 247, pp. 120-127	0
3	<a href="#">Coupling influence of ship dynamic flexure on high accuracy transfer alignment</a>	Wu, W., Qin, S., Chen, S.	2013	<i>International Journal of Modelling, Identification and Control</i> 19 (3), pp. 224-234	0
4	<a href="#">Liposomes interiorly thickened with thermosensitive nanogels as novel drug delivery systems</a>	Zhang, B., Chen, J., Lu, Y., Qi, J., Wu, W.	2013	<i>International Journal of Pharmaceutics</i> Article in Press	0
5	<a href="#">One-pot reaction to synthesis and characterization of two-stage zinc oxide hexagonal microprism on nickel thin films</a>	Wu, W., Zhang, S., Xiao, X., Jiang, C.	2013	<i>Advanced Materials Letters</i> 4 (8), pp. 610-614	0
6	<a href="#">Inhibitory Effects of Green Tea Polyphenols on the Retrogradation of Starches from Different Botanical Sources</a>	Xiao, H., Lin, Q., Liu, G.-Q., Wu, Y., Wu, W., Fu, X.	2013	<i>Food and Bioprocess Technology</i> 6 (8), pp. 2177-2181	0
7	<a href="#">DVFS energy-saving scheduling of navigation receiver based on equilibrium optimization</a>	Wu, W., Ge, R., Ni, S.-J., Wang, F.-X.	2013	<i>Lecture Notes in Electrical Engineering</i> 243 LNEE, pp. 617-626	0
8	<a href="#">A study on influential factors of high-phosphorus wastewater treated by electrocoagulation-ultrasound</a>	Li, J., Song, C., Su, Y., Long, H., Huang, T., Yeabah, T.O., Wu, W.	2013	<i>Environmental Science and Pollution Research</i> 20 (8), pp. 5397-5404	0
9	<a href="#">Study on the removal of algae from lake water and its attendant water quality changes using ultrasound</a>	Li, J., Long, H., Song, C., Wu, W., Yeabah, T.O., Qiu, Y.	2013	<i>Desalination and Water Treatment</i> Article in Press	0
10	<a href="#">Synthesis of paclitaxel-conjugated <math>\beta</math>-cyclodextrin polyrotaxane and its antitumor activity</a>	Yu, S., Zhang, Y., Wang, X., Zhen, X., Zhang, Z., Wu, W., Jiang, X.	2013	<i>Angewandte Chemie - International Edition</i> 52 (28), pp. 7272-7277	0





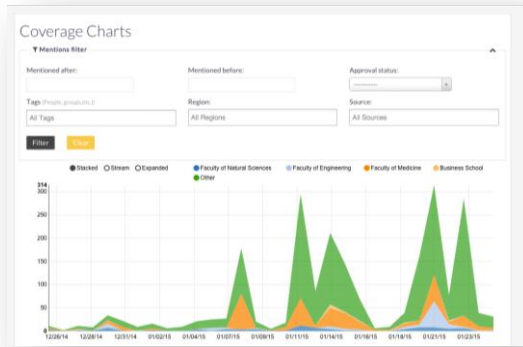
ELSEVIER



# Application: Stream of News



# Platform capabilities



Benchmarking & analytics



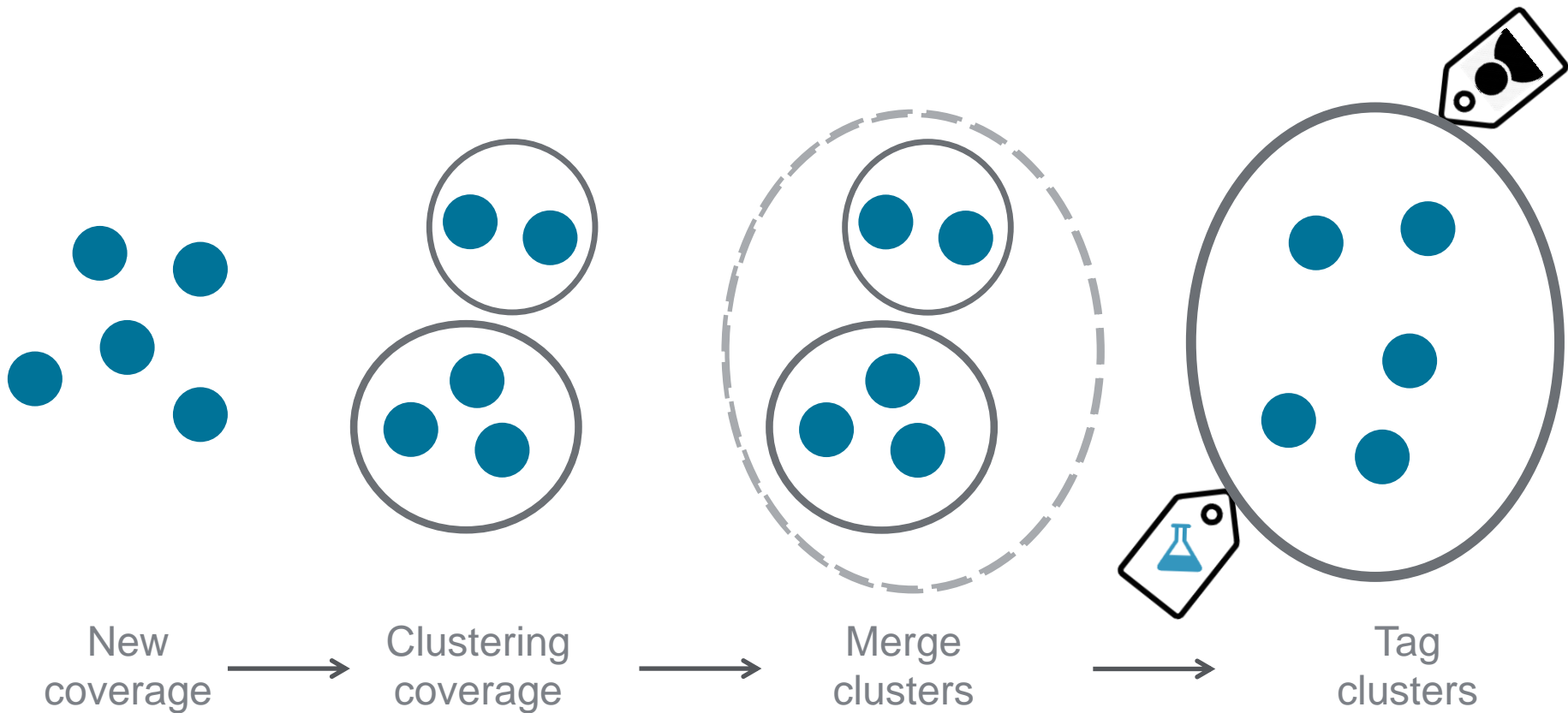
Impact metrics

Reporting

Researcher profiles

# Clustering and tagging

- Process takes **5** minutes for typical institution

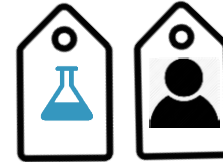


# Entity identification



Articles

500,000 per day

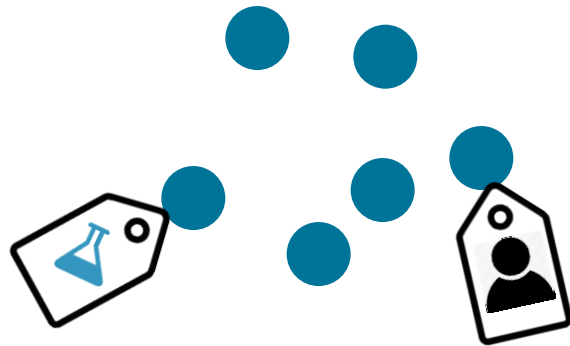


Tags(Researchers, institutions, papers)

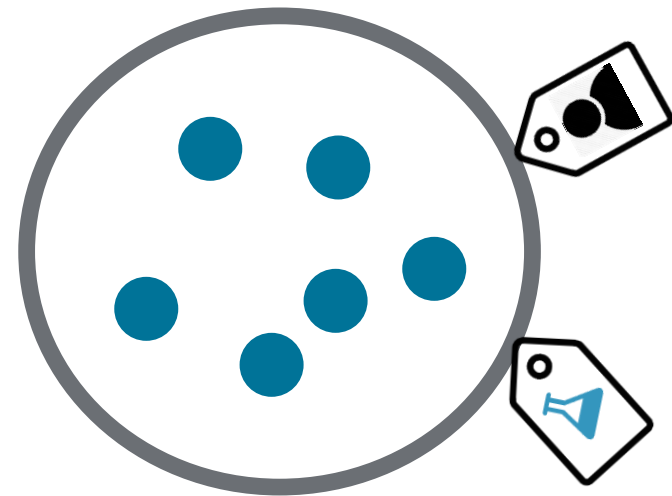
~5,000+ per institution  
×  
~4,000 institutions

# Tagging accuracy

	Imperial	Leicester
<i>Stories per day</i>	13.2	7.6
<i>Mentions per story</i>	5.4	4.5
<i>Stories with DOIs</i>	<b>17%</b>	12%
<i>Mentions with DOIs</i>	<b>2%</b>	3%



Standard approach



Newsflo approach



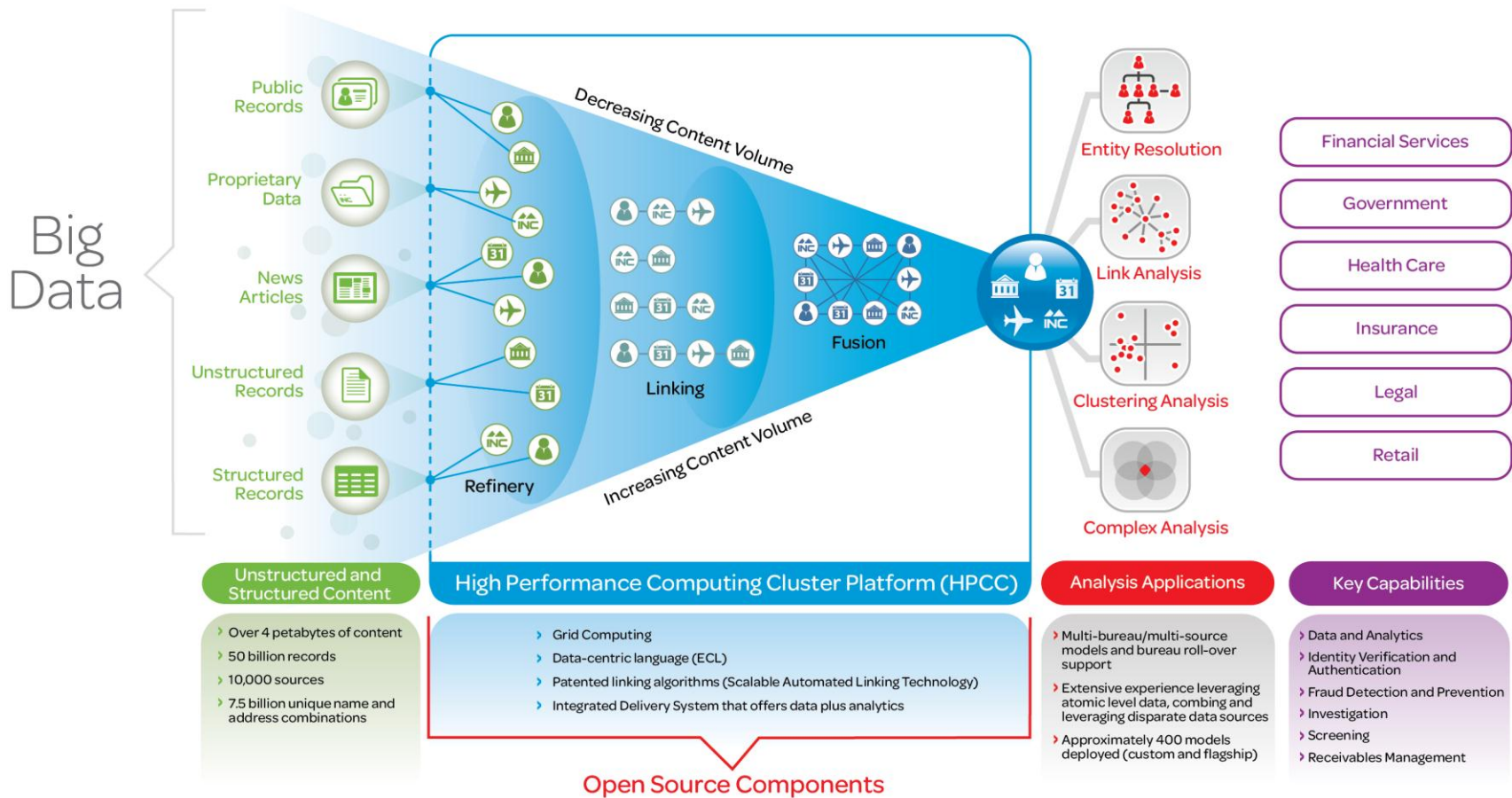
ELSEVIER



# Platform::HPCC

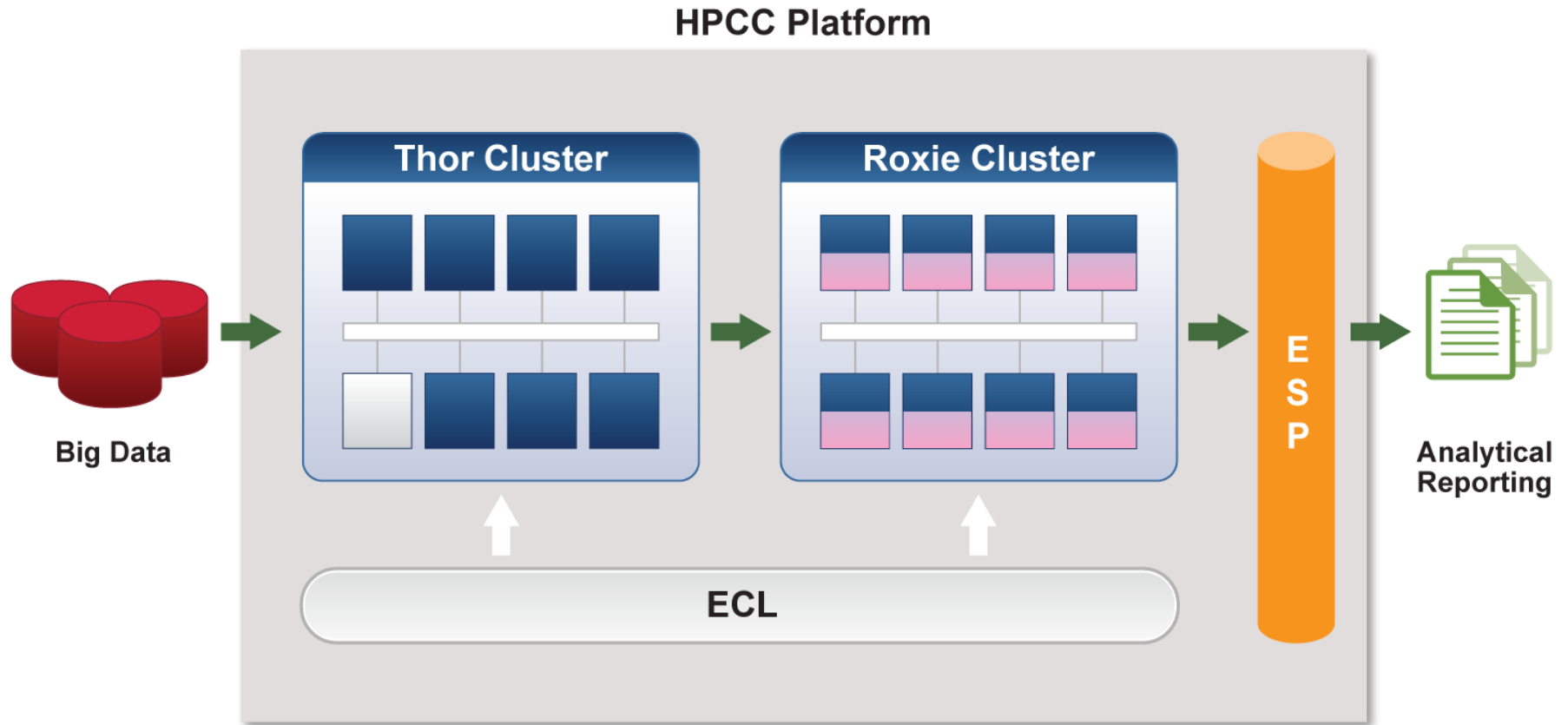


# HPCC Systems: Proven with Enterprise Customers



- **High Performance Computing Cluster Platform (HPCC)** enables data integration on a scale not previously available and real-time answers to millions of users. Built for big data and proven for 10 years with enterprise customers.
- **Offers a single architecture**, two data platforms (query and refinery) and a consistent data-intensive programming language (ECL)
- **ECL Parallel Programming Language** optimized for business differentiating data intensive applications

# HPCC Architecture – Open source





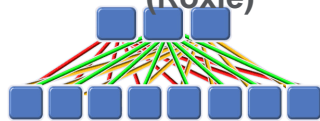
# HPCC Systems: Proven with Enterprise Customers

## 1 HPCC Systems Data Refinery (Thor)



- Massively Parallel Extract Transform and Load (ETL) engine
  - Built from the ground up as a parallel data environment. Leverages inexpensive locally attached storage. Doesn't require a SAN infrastructure.
- Enables data integration on a scale not previously available:
  - Current LexisNexis person data build process generates 350 Billion intermediate results at peak
- Suitable for:
  - Massive joins/merges
  - Massive sorts & transformations
  - Any  $N^2$  problem
  - *"identify and catalog all the DNA in the oceans"*

## 2 HPCC Systems Data Delivery Engine (Roxie)



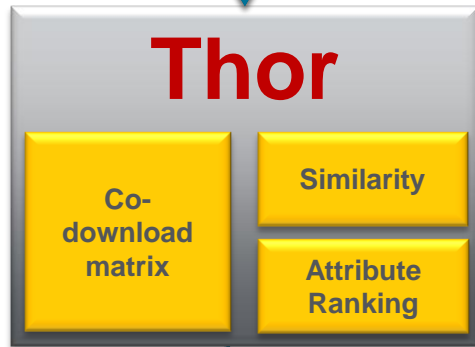
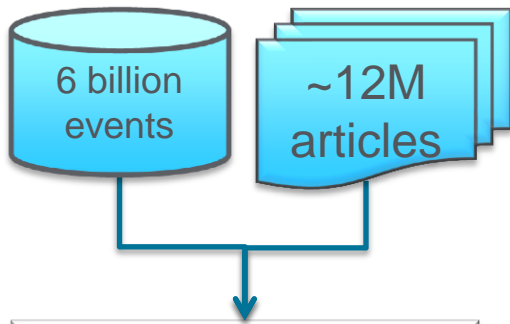
- A massively parallel, high throughput, structured query response engine
- Ultra fast due to its read-only nature.
- Allows indices to be built onto data for efficient multi-user retrieval of data
- Suitable for
  - Volumes of structured queries
  - Full text ranked Boolean search
  - "I want that fish there"*

## 3 Enterprise Control Language (ECL)

- An easy to use, data-centric programming language optimized for large-scale data management and query processing
- Highly efficient; Automatically distributes workload across all nodes.
  - Industry analysts estimate **80% more efficient than C++, Java and SQL** and 1/3 reduction in programmer time to maintain/enhance existing applications
  - Benchmark against SQL (**5 times more efficient**) for code generation
- Automatic parallelization and synchronization of sequential algorithms for parallel and distributed processing
- Large library of efficient modules to handle common data manipulation tasks

# Recommendation Engine process at a glance

5 years of SD usage data/events  
 All SD XML Articles  
 Journal Rankings



Twice weekly, move to daily



ScienceDirect

Home + Recent Actions | Publications | Search | My settings | My alerts

Download PDF | Export citation | Jump to references | More options...

Information Processing Letters  
 Volume 83, Issue 4, 31 August 2002, Pages 211–220

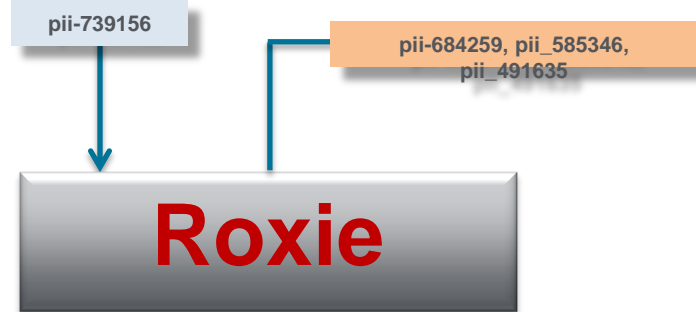
**Mining association rules using inverted hashing and pruning**\*

John D. Holt, Soon M. Chung

Department of Computer Science and Engineering, Wright State University, Dayton, OH 45435, USA

Abstract  
 In this paper, we propose a new algorithm named Inverted Hashing and Pruning (IHP) for mining association rules between items in transaction databases. The performance of the IHP algorithm was evaluated for various cases and compared with those of two well-known mining algorithms, Apriori algorithm [Proc. 20th VLDB Conf., 1994, pp. 487–499] and Direct Hashing and Pruning algorithm [IEEE Trans. on Knowledge Data Engng. 9 (6) (1997) 813–825]. It has been shown that the IHP algorithm has better performance for databases with long transactions.

Recommended articles:  
 TBAR: An efficient method for assoc...  
 2001, Data & Knowledge Engineering  
 Show more information  
 High-performance data mining with s...  
 2002, Parallel Computing  
 Show more information  
 CFP-tree: A compact disk-based stru...  
 2007, Information Systems  
 Show more information



DocID	Recommendations		
pii_484710	pii_789506	pii_622431	pii_588765
pii_564568	pii_788931	pii_212389	pii_100865
pii_739156	pii_684259	pii_585346	pii_491635
pii_758426	pii_116895	pii_986532	pii_456218
...	...	...	...

# ELSEVIER LABS - AREAS OF INTEREST

- Information extraction for scholarly text
- **Knowledge Graphs**
  - **Entity linking / disambiguation**
  - **Machine Reading**
- Information retrieval / Search
- Question answering
- Large Scale Data Management
- Deep Multimodal Networks for Image and Text Analysis
- Future of research communication

# RECENT PUBLICATIONS

- Sara Magliacane, Philip Stutz, Paul Groth, Abraham Bernstein, foxPSL: A Fast, Optimized and eXtended PSL implementation, International Journal of Approximate Reasoning (2015).
- Ron Daniel, "Large Scale Text Analytics with Apache Spark"; Presented at Text Analytics World 2015, San Francisco.
- L. Moreau, P. Groth, J. Cheney, T. Lebo, S. Miles, The rationale of PROV, Web Semantics: Science, Services and Agents on the World Wide Web (2015).
- Marcin Wylot, Philippe Cudré-Mauroux, and Paul Groth. "Executing Provenance-Enabled Queries over Web Data", Proceedings of the 24th International World Wide Web Conference (WWW 2015), 2015.
- Elshaimaa Ali, Michael Lauruhn. Wikipedia-based Extraction of Lightweight Ontologies for Concept Level Annotation, International Conference on Dublin Core and Metadata Applications (DC 2014), Austin, Tx
- Ron Daniel, "Predicting Citation Counts"; Research Trends; June 2014.



ELSEVIER

Join Elsevier if  
you are  
serious about  
Big Data and  
Machine  
Learning