KNOWLEDGE MEDIA
KMi
INSTITUTE

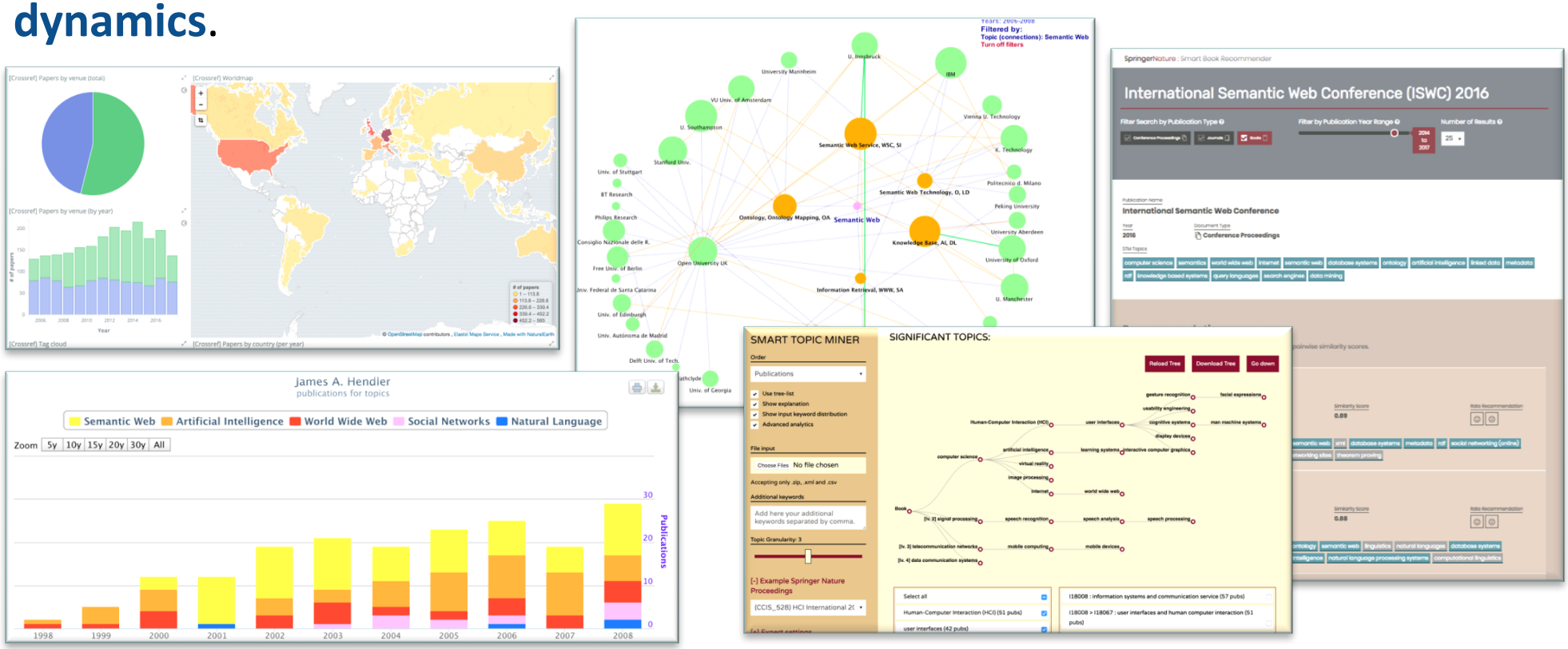
# Understanding Research with Semantic Technologies

Francesco Osborne
SKM$^3$- KMi, The Open University, United Kingdom

http://skm.kmi.open.ac.uk/
http://people.kmi.open.ac.uk/francesco/

# SKM³ - KMi

The **SKM³** team aims at producing innovative approaches leveraging large-scale data mining, semantic technologies, machine learning, and visual analytics for **making sense of scholarly data and forecasting research dynamics**.




We collaborate with **major publishers, universities, and companies** and produce a variety of services supporting researchers, editors, and research polities makers.

http://skm.kmi.open.ac.uk/

# The destination

*The Semantic Web will likely profoundly change the very nature of how scientific knowledge is produced and shared, in ways that we can now barely imagine.*

*T. Berners-Lee*

*Researchers will benefit from better, faster, cheaper access to data related to publications, enhancing the capacity for in silico meta-research.*

*D. Shotton*

*The challenge is to develop an AI system that can make major scientific discoveries in biomedical sciences and that is worthy of a Nobel Prize and far beyond.*

*H. Kitano*

# Fostering knowledge transfer with semantic technologies

1. Understanding the space of research fields
   - What are the main components of a field? How are they linked?
   - How can we classify research products according to them?
   - What are the relevant scientific venues and pieces of literature?

2. Fostering technology transfer
   - How to detect technologies?
   - How to accelerate the knowledge flow and the pace of technology propagation?

3. Describing and assessing research actors and products
   - Who are the researchers, communities, and organizations that will be impacted by a research policy?
   - What are the promising research areas that should be nurtured?

4. Anticipating trends
   - What research topics will emergence in the following years?
   - How to facilitate the creation of new branch of science?

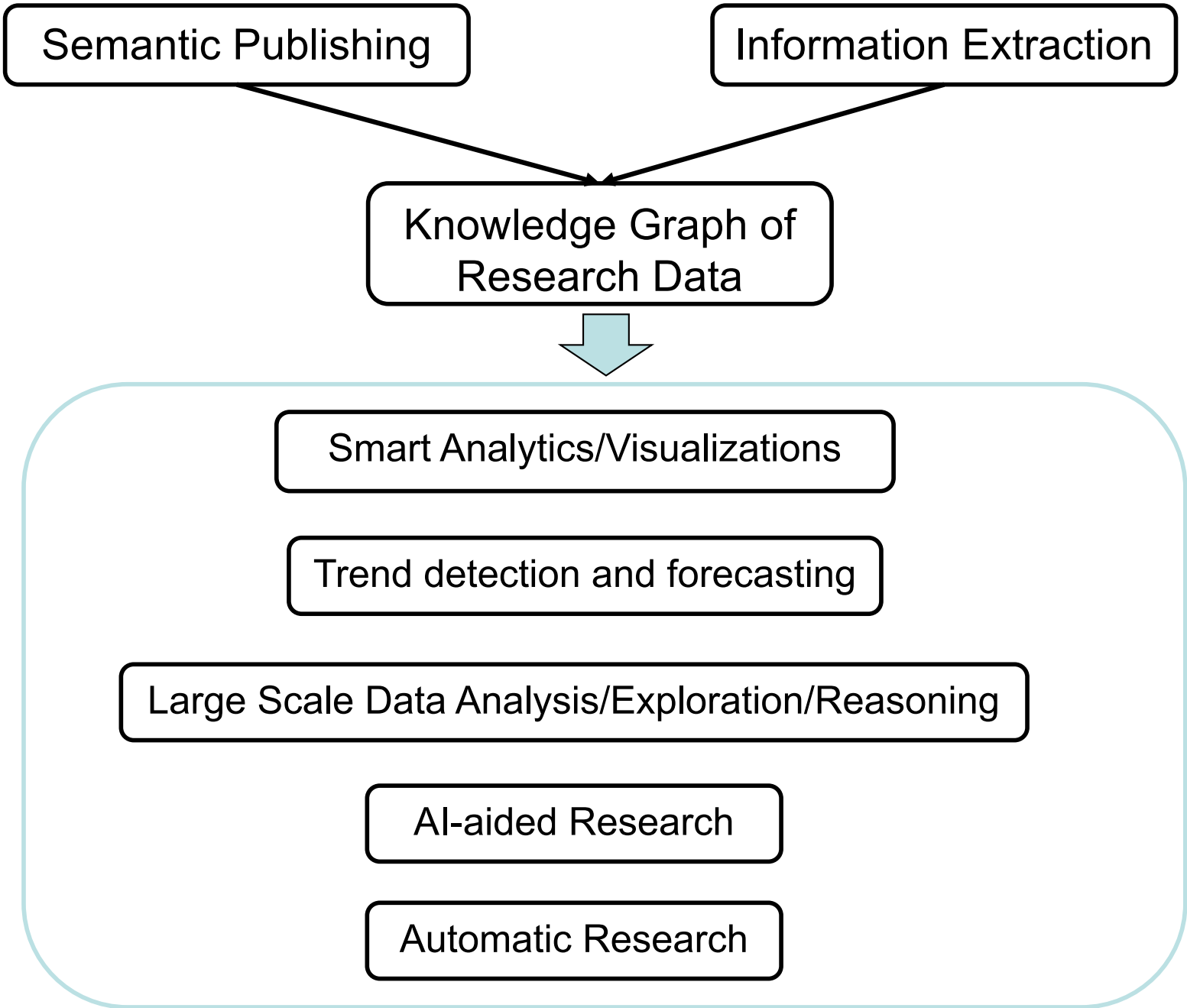
# A variety of semantic solutions

1. Understanding the space of research fields
    - CSO Portal
    - Smart Topic Miner
    - Smart Book Recommender
2. Fostering technology transfer
    - Topic-technology Framework
3. Describing and assessing research actors and products
    - Research Map Builder
    - Rexplore
4. Anticipating trends
    - Augur

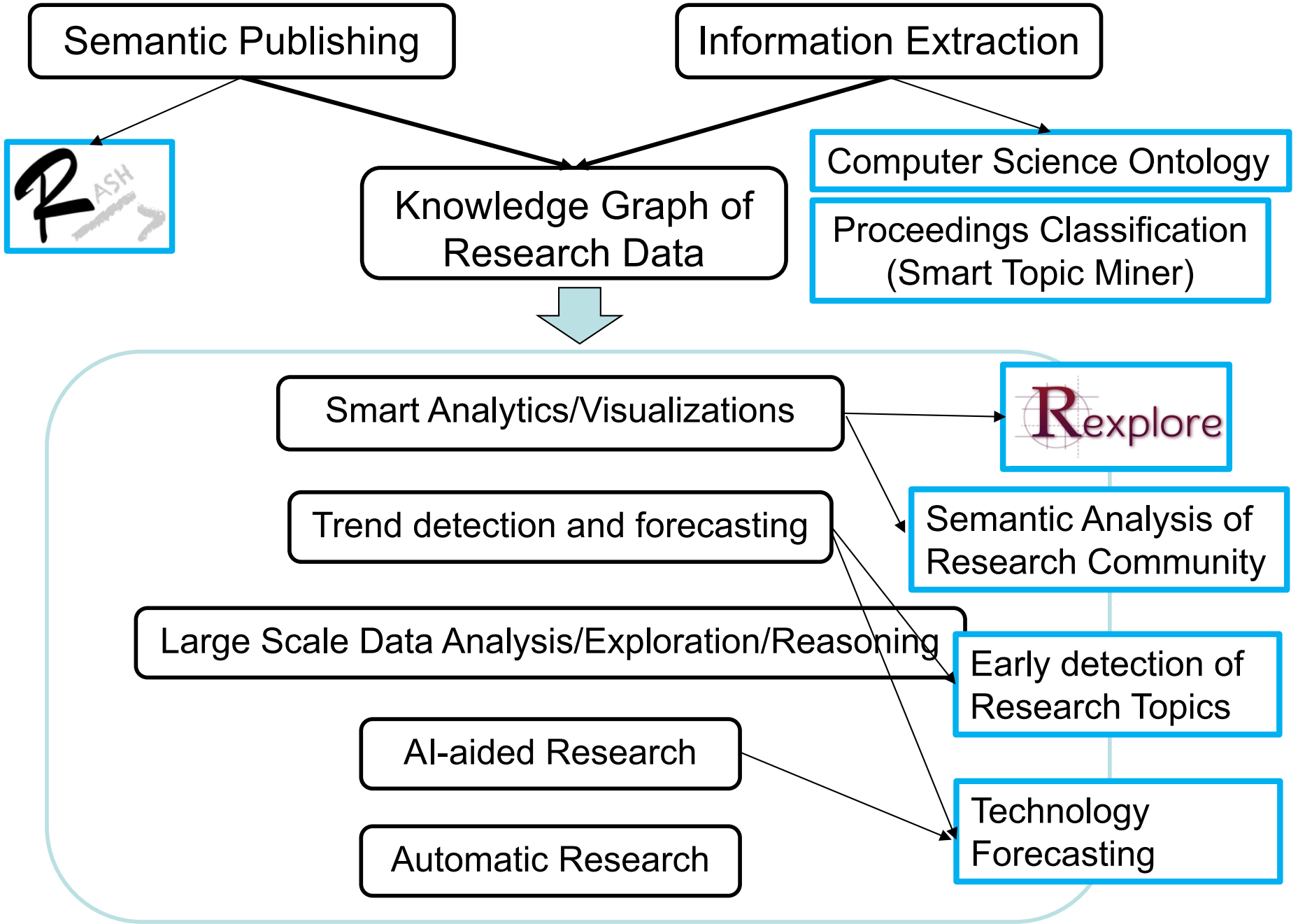
# A variety of semantic solutions (hand-on ones)

1. Understanding the space of research fields
   - **CSO Portal**
   - **Smart Topic Miner**
   - **Smart Book Recommender**
2. Fostering technology transfer
   - Topic-Technology Framework
3. Describing and assessing research actors and products
   - Research Map Builder
   - **Rexplore**
4. Anticipating trends
   - Augur

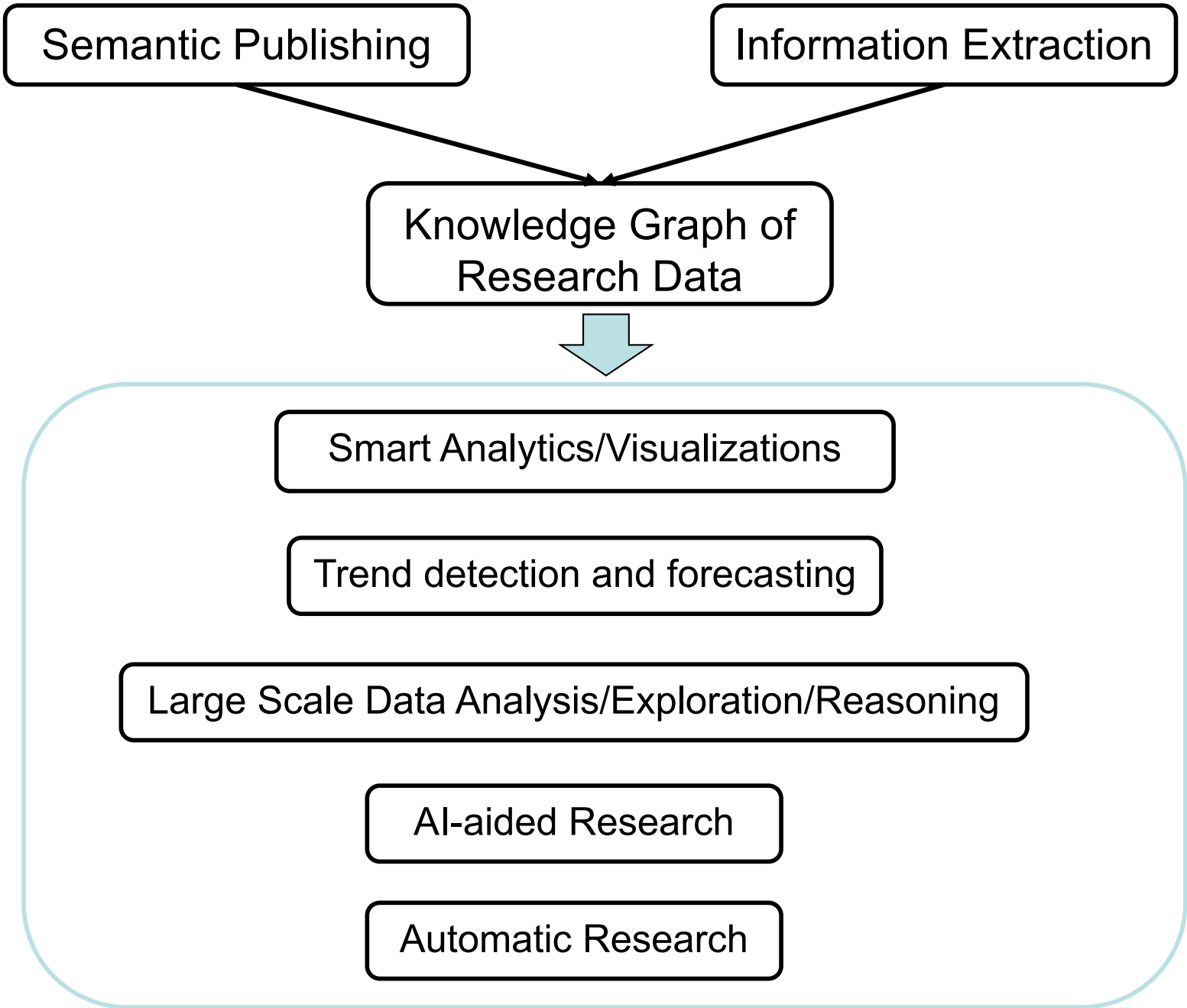
# Research 3.0


Semantic Publishing
Information Extraction
Knowledge Graph of Research Data
Smart Analytics/Visualizations
Trend detection and forecasting
Large Scale Data Analysis/Exploration/Reasoning
AI-aided Research
Automatic Research

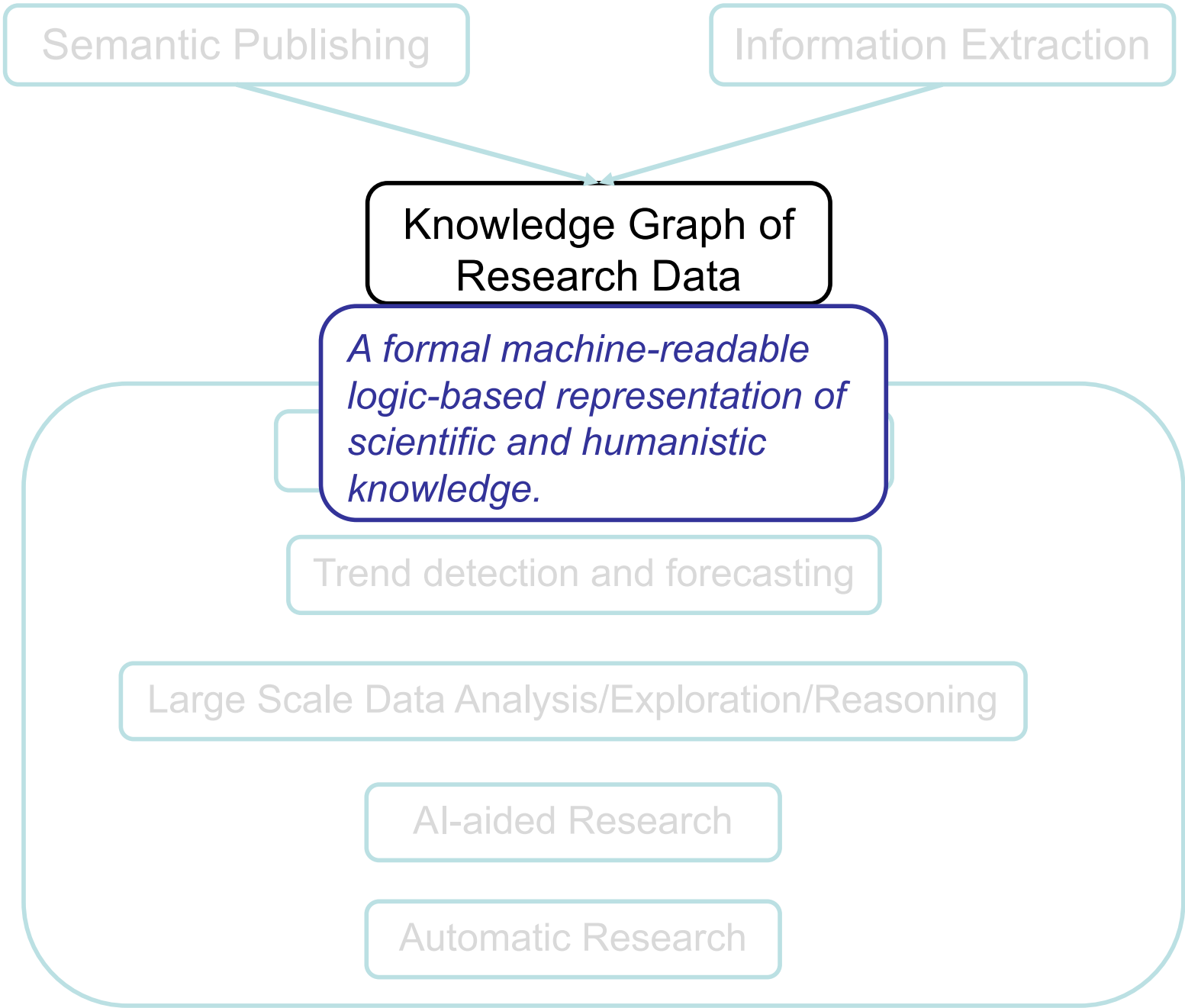
# Research 3.0 – some contributions of SKM3



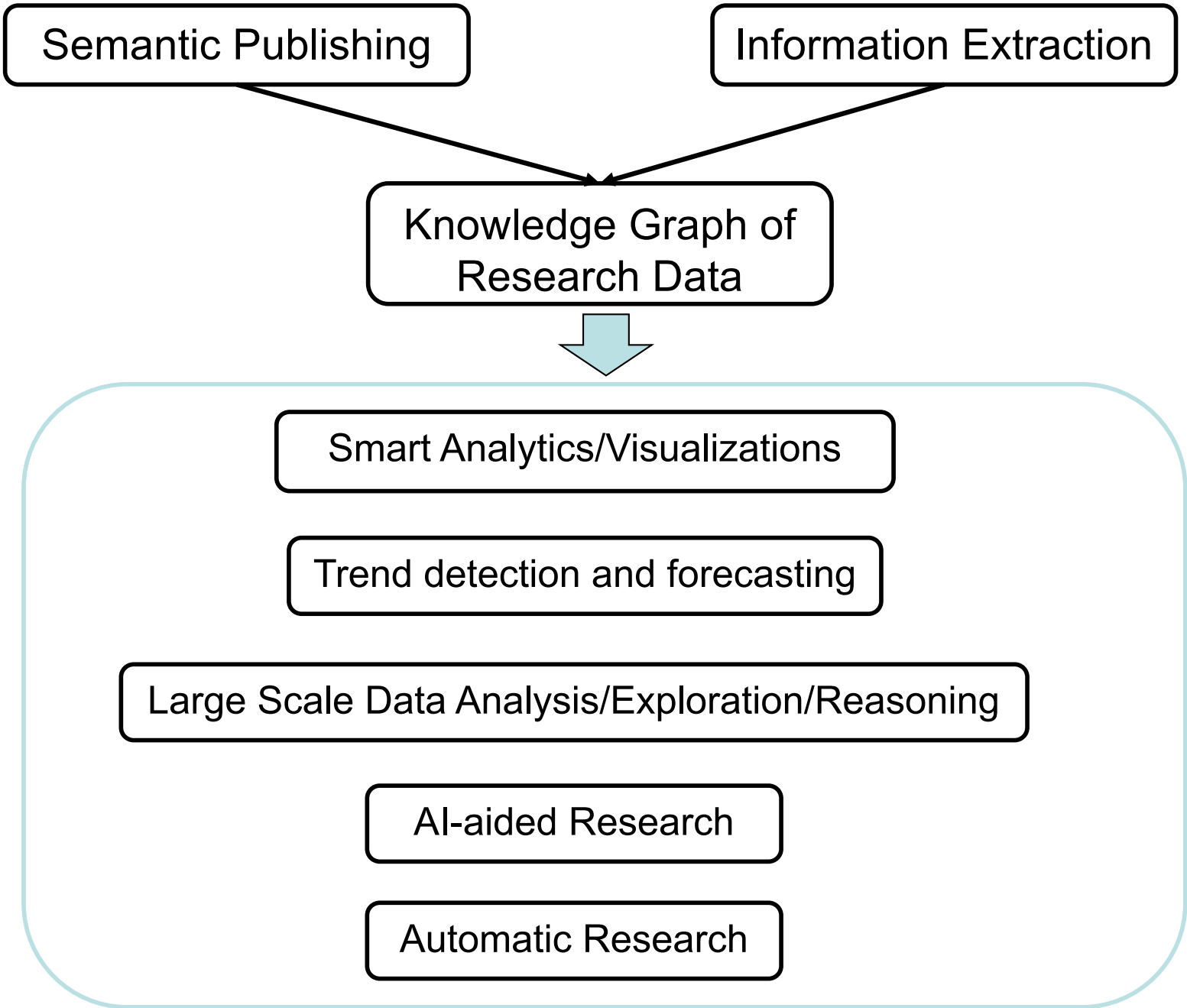
# Research 3.0


Semantic Publishing
Information Extraction
Knowledge Graph of Research Data
Smart Analytics/Visualizations
Trend detection and forecasting
Large Scale Data Analysis/Exploration/Reasoning
AI-aided Research
Automatic Research

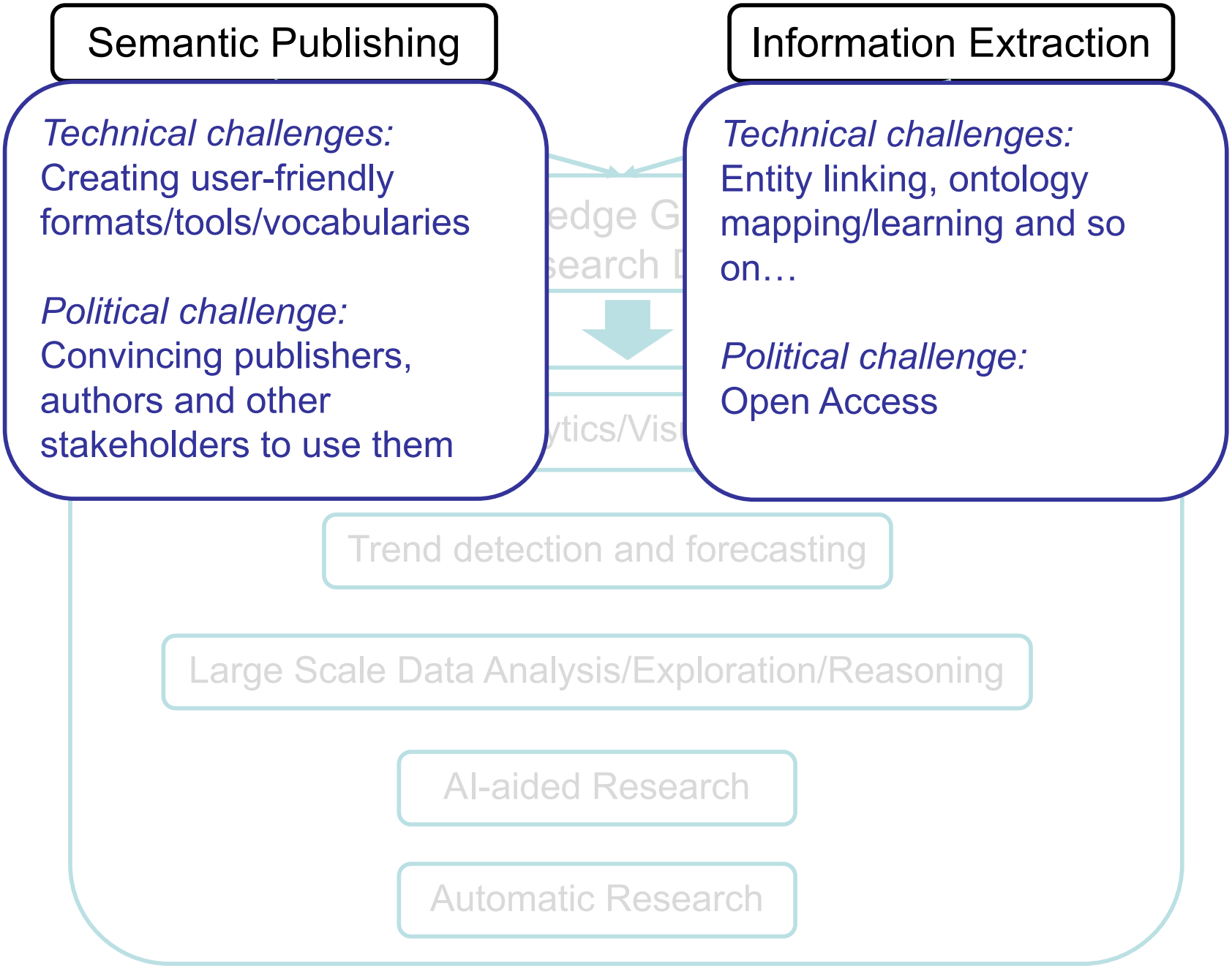
# Research 3.0


Semantic Publishing
Information Extraction
Knowledge Graph of Research Data
A formal machine-readable logic-based representation of scientific and humanistic knowledge.
Trend detection and forecasting
Large Scale Data Analysis/Exploration/Reasoning
AI-aided Research
Automatic Research

# Research 3.0


Semantic Publishing
Information Extraction
Knowledge Graph of Research Data
Smart Analytics/Visualizations
Trend detection and forecasting
Large Scale Data Analysis/Exploration/Reasoning
AI-aided Research
Automatic Research

# Research 3.0 – some contributions of KMi RA Team



# An Exciting Environment…

Scientific Workshops

- SAVE-SD - Semantics, Analytics, Visualisation: Enhancing Scholarly Data (WWW)
- Scientometrics Workshop (ESWC)
- Semantic Web Publishing Challenge (ESWC)
- SemSci: Enabling Open Semantic Science (ISWC)
- WORSP: International Workshop of Mining Scientific Publication (JCDL)
- Emc$^{Sci}$ - Extracting and Modelling Scientific Knowledge from Texts (IC)

Initiatives

- FORCE11, Research Data Alliance (RDA), ORCHID

Companies

- Publishers: Springer Nature, Elsevier ,Thomson Reuters
- Services based on research data: Google, Microsoft, Idex Lab, Altmetric

Online Repositories

- Scholarly.org, Figshare, Zootero, OpenCitation.net, CORE …

# Fostering knowledge transfer with semantic technologies

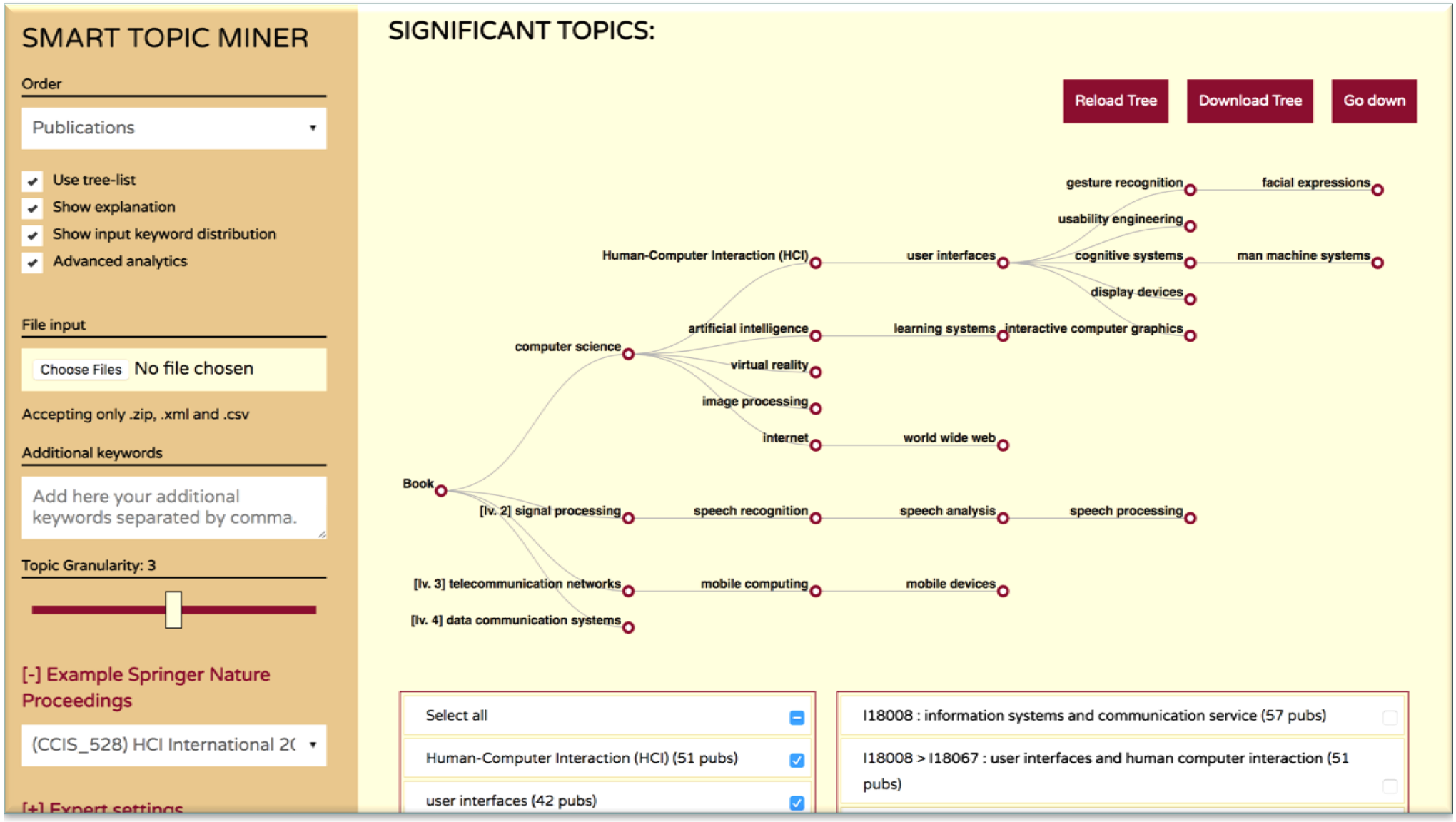
1. Understanding the space of research fields
   - What are the main components of a field? How are they linked?
   - How can we classify research products according to them?
   - What are the relevant scientific venues and pieces of literature?

2. Fostering technology transfer
   - How to detect technologies?
   - How to accelerate the knowledge flow and the pace of technology propagation?

3. Describing and assessing research actors and products
   - Who are the researchers, communities, and organizations that will be impacted by a research policy?
   - What are the promising research areas that should be nurtured?

4. Anticipating trends
   - What research topics will emergence in the following years?
   - How to facilitate the creation of new branch of science?

# 1. Understanding the space of research fields

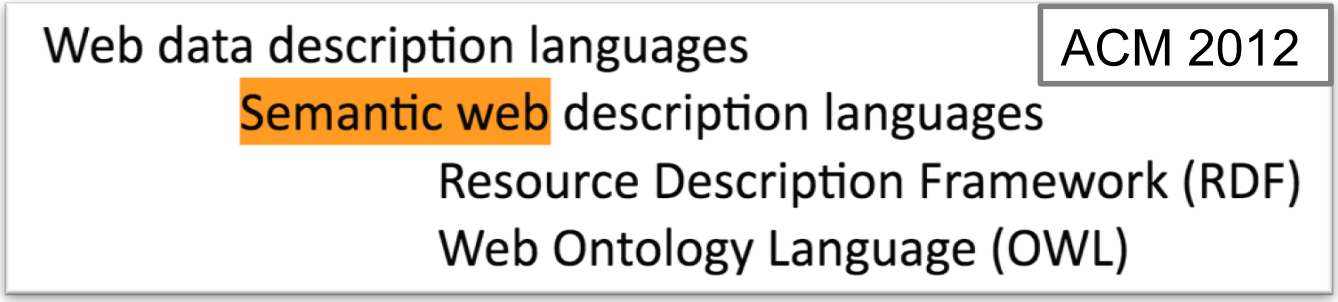



http://skm.kmi.open.ac.uk/cso

http://skm.kmi.open.ac.uk/stm

# Research Topics - a complex issue to solve

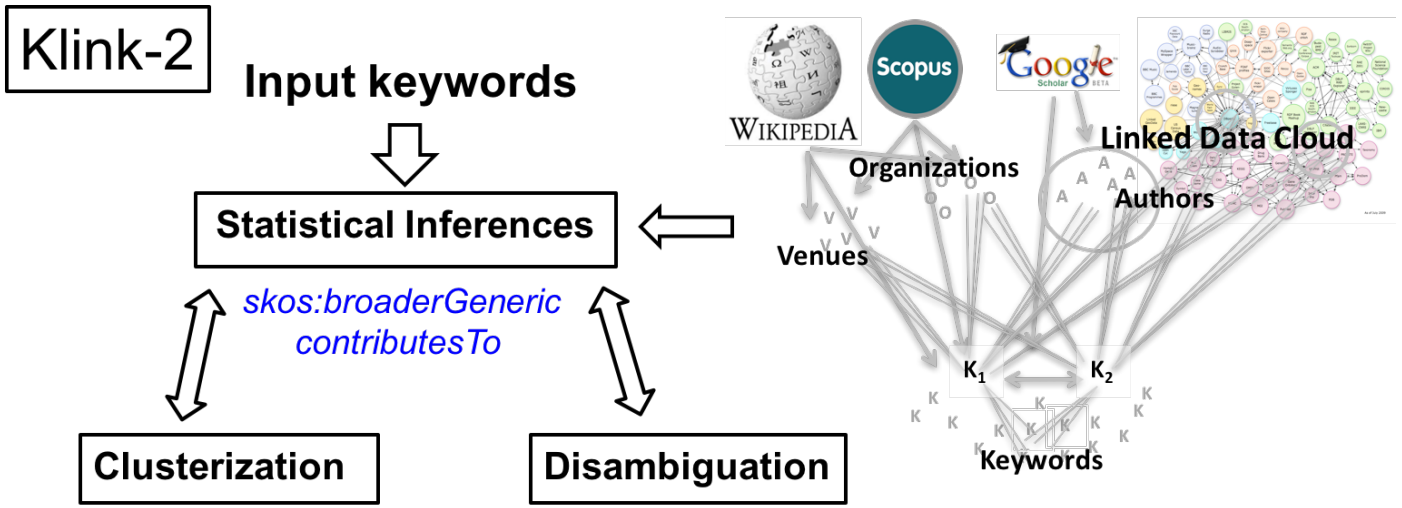
Standard research areas taxonomies/classifications/ontologies such as ACM are not apt to the task.




- Not fine-grained enough.
  - E.g., only 2 topics are classified under Semantic Web
- Static, manually defined, hence prone to get obsolete very quickly.
- The relations between entries are unclear.
  - They are meant to be sub-areas, even if it can be argued that many of them are not really sub-areas.

# The Computer Science Ontology I

The Computer Science Ontology consist of about **15,000 topics** linked by about **70,000 semantic relationships. It was automatically created** and is regularly updated using the **Klink-2 algorithm**.



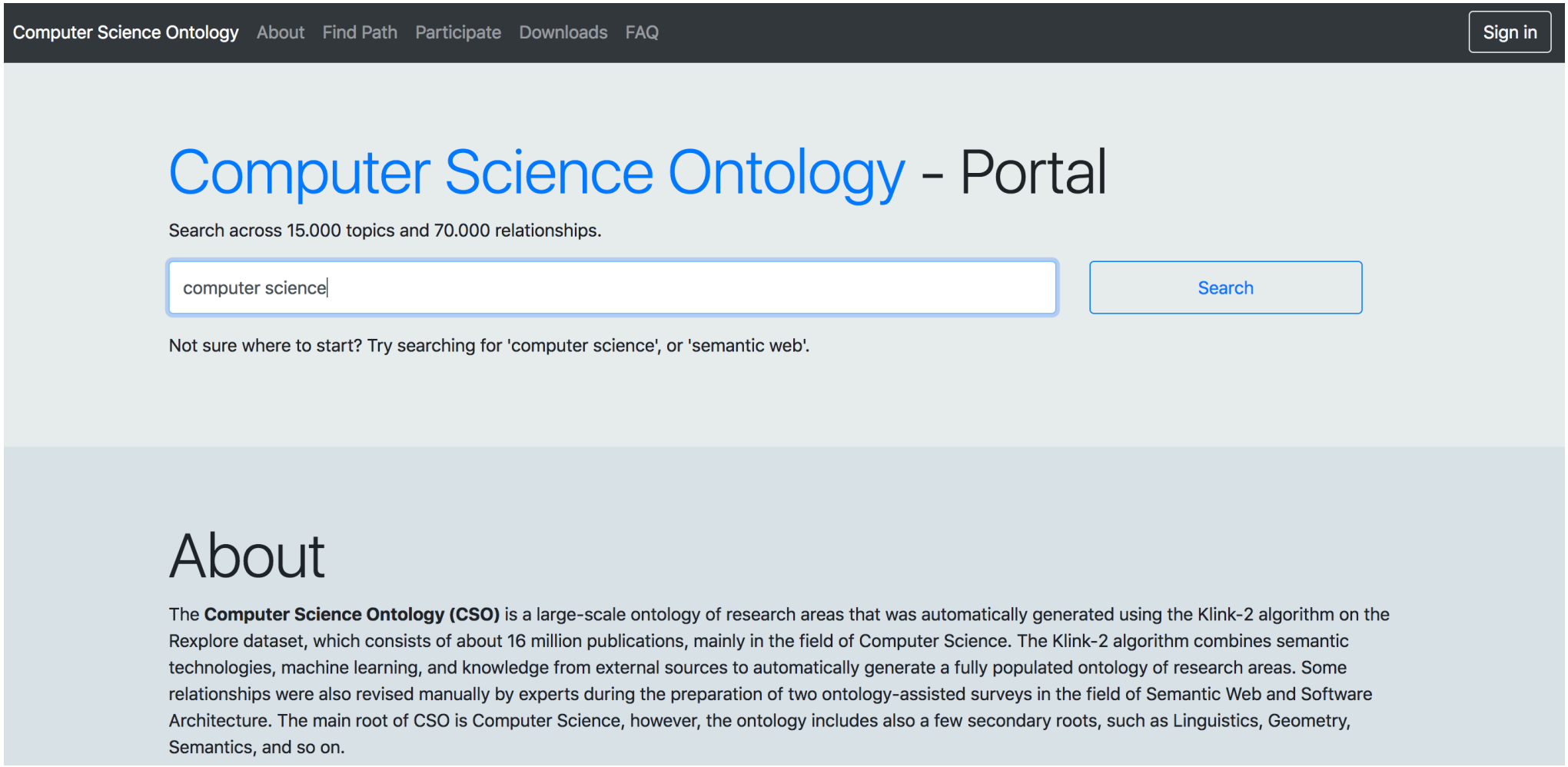

```
<!-- http://kmi.open.ac.uk/technologies/rexplore/ontologies/BiboExtension#semantic_web -->

<owl:NamedIndividual rdf:about="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#semantic_web">
    <rdf:type rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#Topic"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#semantic_web_services"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#semantic_web_technologies"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#web_ontology_language"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#semantic_web_applications"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#semantic_description"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#owl"/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension
    "/>
    <skos:broaderGeneric rdf:resource="http://kmi.open.ac.uk/technologies/rexplore/ontologies/
BiboExtension#resource_description_framework"/>
```

# The Computer Science Ontology II

- It included **very granular** and low level research areas, e.g., Linked open data, Probabilistic packet marking, Synthetic aperture radar imaging
- It can be **regularly updated** by running Klink-2 on a new set of publications.
- It allows for a **research topic to have multiple super-areas** – i.e., the taxonomic structure is a graph rather than a tree, e.g., Inductive Logic Programming is a sub-area of both Machine Learning and Logic Programming.
- This knowledge base is used to **enhance semantically** a variety of information extraction method and data analytics

# CSO Portal I




**http://cso.kmi.open.ac.uk/**

# CSO Portal II

## topics:semantic web

Explore the topic **semantic web** using the following visualisations.

Overview | Compact | Detailed

| Subject | Predicate | Object |
|---|---|---|
| semantic web | parent of | domain ontologies |
| " | parent of | linked datum |
| " | parent of | ontology engineering |
| " | parent of | ontology matching |
| " | parent of | owl |
| " | parent of | rdf |
| " | parent of | semantic annotations |
| " | parent of | semantic descriptions |
| " | parent of | semantic interoperability |
| " | parent of | semantic search |
| " | parent of | semantic technologies |
| " | parent of | |

### semantic web

The Semantic Web is an extension of the Web through standards by the World Wide Web Consortium (W3C). The standards promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Framework (RDF). According to the W3C, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries". The term was coined by Tim Berners-Lee for a web of data that can be processed by machines.While its critics have questioned its feasibility, proponents argue that applications in industry, biology and human sciences research have already proven the validity of the original concept.

**http://cso.kmi.open.ac.uk/topics/semantic%20web**

# Research Topics and Springer Nature

**The Open University** and **Springer Nature** have been collaborating since 2015 in the development of an array of semantically-enhanced solutions for:

- Automatic classification of **proceedings** and other editorial products.

- Automatic selection of the most appropriate **books, journals, and proceedings** to market at a scientific event.

# Proceedings classification in a nutshell

ISWC: International Semantic Web Conference

## The Semantic Web – ISWC 2016

15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part I

Editors (view affiliations)

Paul Groth, Elena Simperl, Alasdair Gray, Marta Sabou, Markus Krötzsch, Freddy Lecue, Fabian Flöck, Yolanda Gil

Conference proceedings
ISWC 2016

| 36 | 23 | 360 | 37k |
|---|---|---|---|
| Citations | Mentions | Readers | Downloads |

Part of the Lecture Notes in Computer Science book series (LNCS, volume 9981)

Table of contents (39 papers) | Other volumes | About these proceedings

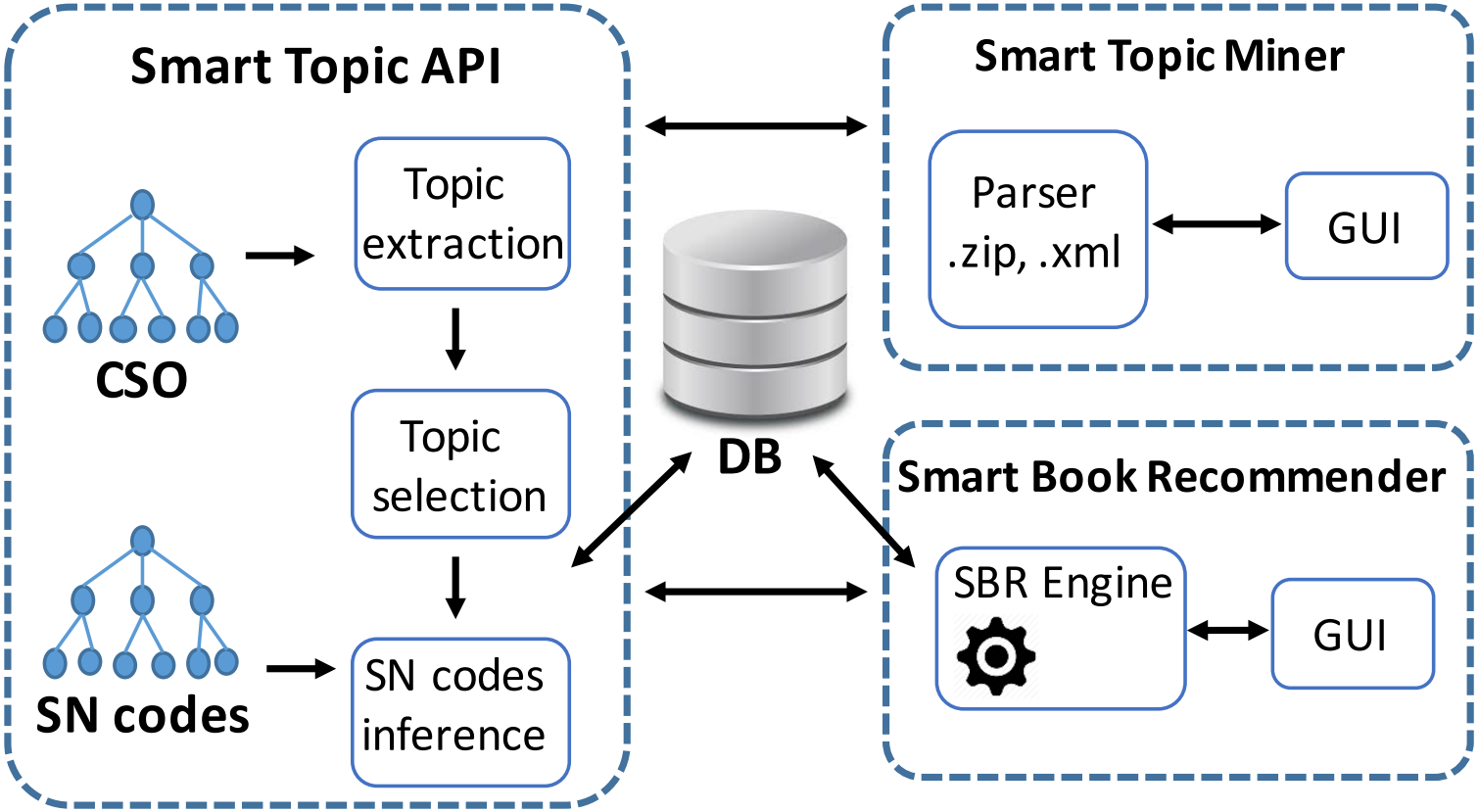
### Introduction

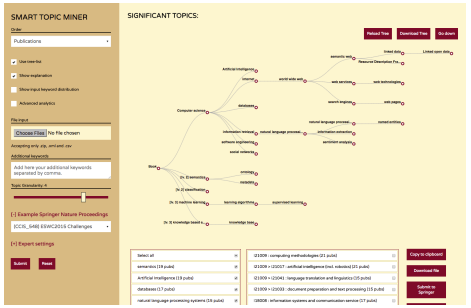
The two-volume set LNCS 9981 and 9982 constitutes the refereed proceedings of the 15th International Semantic Web Conference, ISWC 2016, which was held in Kobe, Japan, in October 2016. The 75 full papers presented in these proceedings were carefully reviewed and selected from 326 submissions.
The International Semantic Web Conference is the premier forum for Semantic Web research...

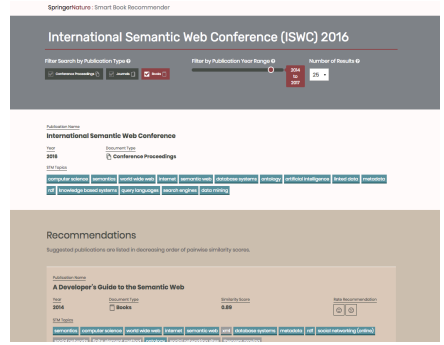
### Keywords

artificial intelligence | entity resolution | information retrieval | instance matching | knowledge based systems | knowledge management | linked data | Linked Open Data (LOD) | Natural Language Processing (NLP) | ontologies | ontology alignment | query processing | reasoning | Resource Description Framework (RDF) | semantic web | semantics | social media | SPARQL | Web Ontology Language (OWL) | World Wide Web

# Smart Semantic Solutions For Springer Nature



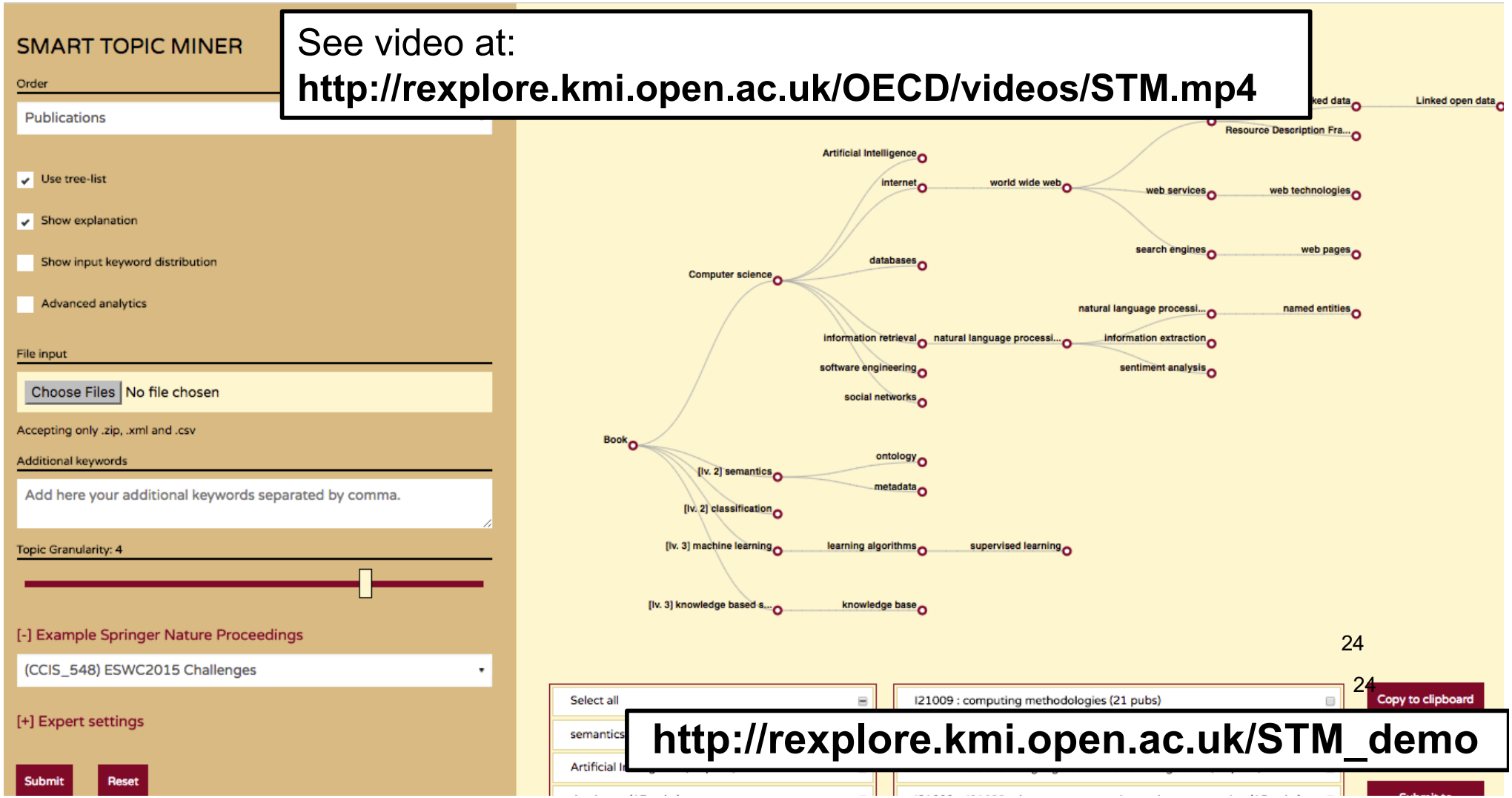




Input: Publication metadata

Output:
- Topic Taxonomy
- SN codes
- Analytics

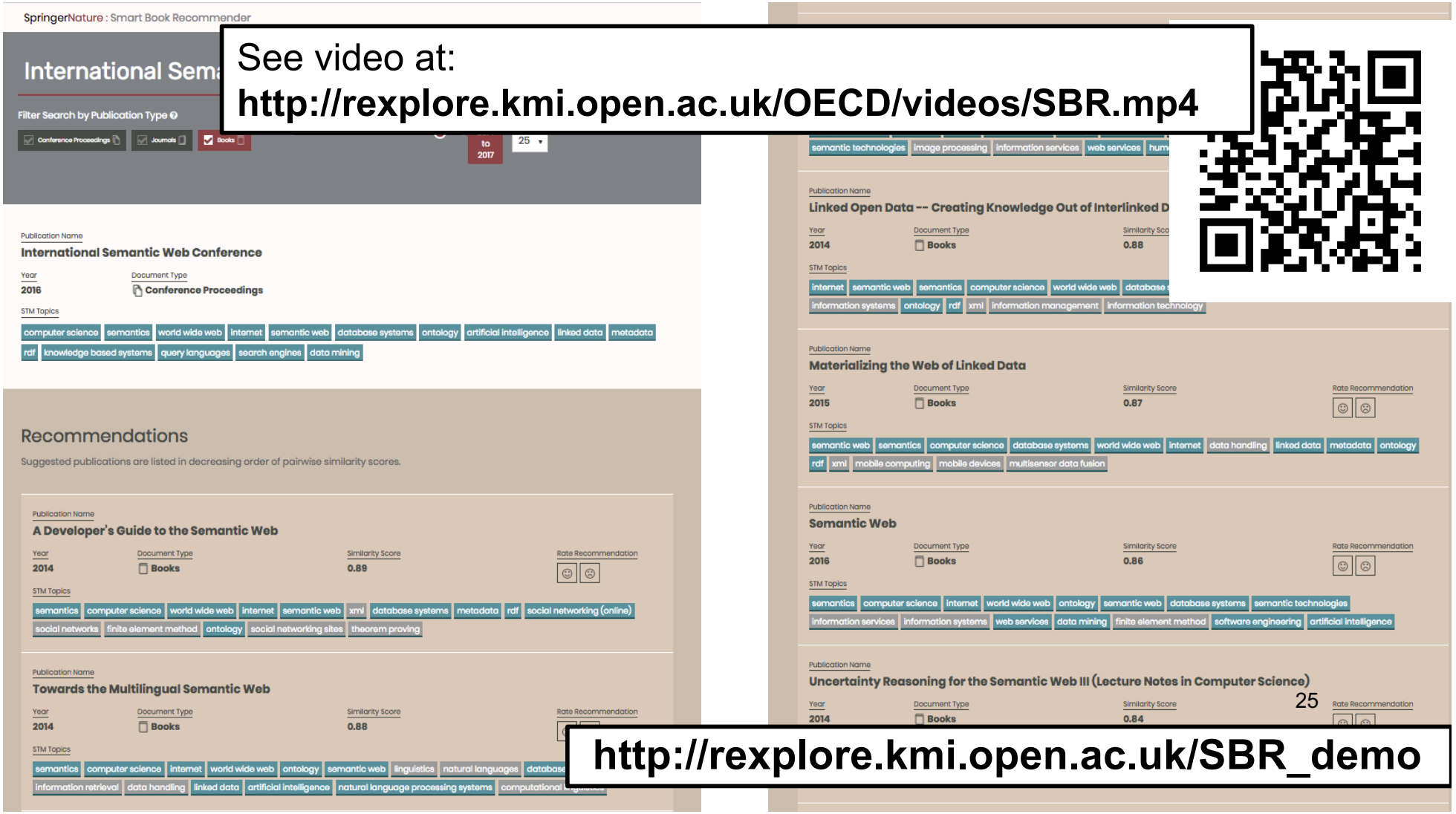
# Smart Topic Miner

**Smart Topic Miner** (STM) is a semantic application designed to support the **Springer Nature Computer Science editorial team** in classifying scholarly publications.



# Smart Book Recommender

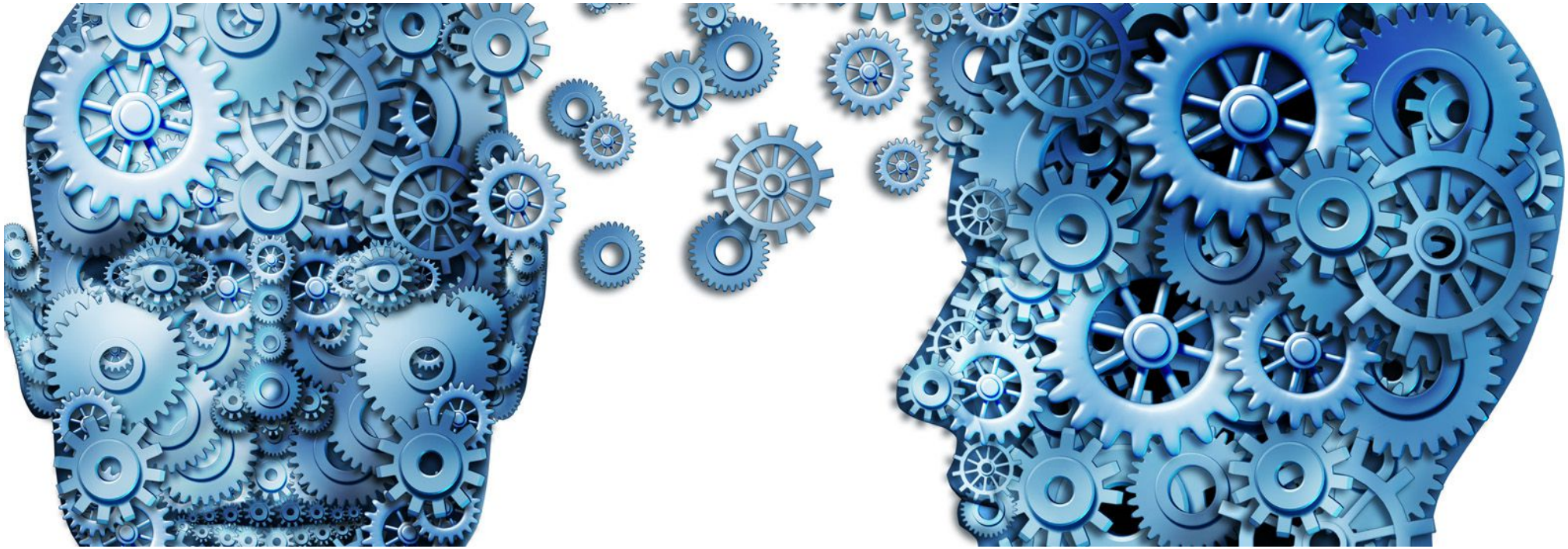
**Smart Book Recommender** (STM) is a web application that takes as input a conference and suggests books, proceedings and journals which address similar topics.




See video at:
**http://rexplore.kmi.open.ac.uk/OECD/videos/SBR.mp4**

**http://rexplore.kmi.open.ac.uk/SBR_demo**

# 2. Fostering technology transfer



# Standing on the Shoulder of Giants

- We constantly reuse ideas, technologies, methods and materials.
- Technologies will usually appear in a research community and then spread to other research areas in the following years.
  - *e.g., SW technologies were created in the field of AI, KBS, WWW and then they spread to Information Retrieval, HCI, Biology and so on.*
- This process is often inefficient and may take several years resulting in a **technology adoption lag**.
- Currently there are no methods to predict technology spreading across research areas.

# Standing on the Shoulder of Giants

- We constantly reuse ideas, **technologies**, methods and materials.
- Technologies will usually appear in a research community and then spread to other research areas in the following years.
  - *e.g., SW technologies were created in the field of AI, KBS, WWW and then they spread to Information Retrieval, HCI, Biology and so on.*
- This process is often inefficient and may take several years resulting in a **technology adoption lag**.
- Currently there are no methods to predict technology spreading across research areas.

**How can we improve the technology transfer?**

**How can we help researchers to track down relevant technologies?**

# TechMiner

TechMiner (TM) is a new approach, which combines NLP, machine learning and semantic technologies, for **mining technologies** applications, systems, languages and formats from research publications.

It generates an **OWL ontology describing technologies** and their relationships with other research entities.

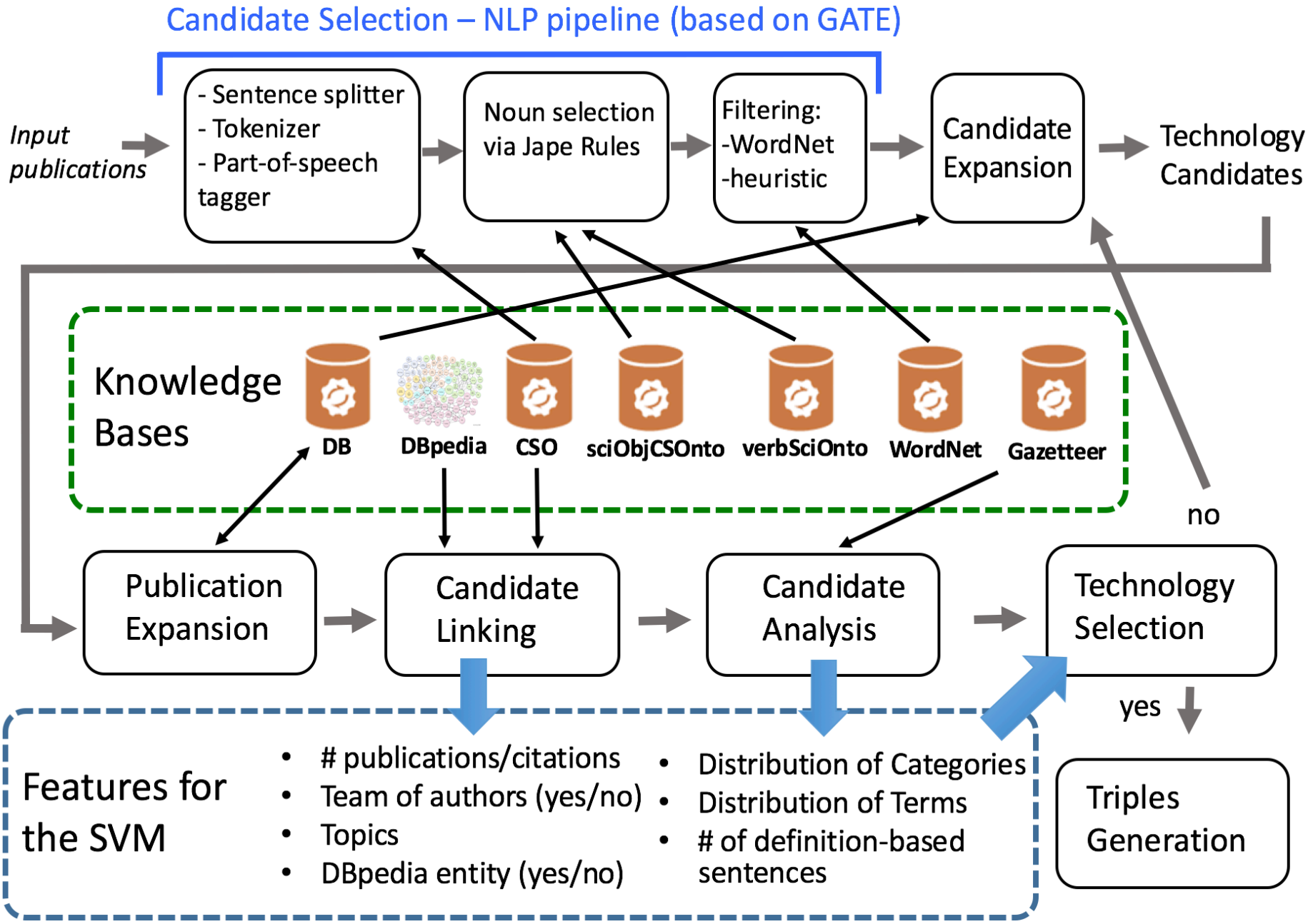
We evaluated TM on a manually annotated gold standard and found that it improves significantly both precision and recall over alternative NLP approaches.

- The **proposed semantic features** significantly **improve both recall and precision**.

# Some example – Tecnologies created by E. Motta

| Label | Debut | pub | Introducers | Main Topics |
|---|---|---|---|---|
| Magpie | 2004 | 3 | Domingue, Motta, Dzbor | Semantic Web, Web Browser, Ontologies, Information Retrieval |
| PowerAqua | 2006 | 4 | Uren, Motta, Lopez, Sabou, D'aquin, Nikolov | Semantic Web, Question Answering, Ontology, Natural Language |
| KnoFuss | 2007 | 3 | Uren, Motta, De Roeck, Nikolov | Ontology, Semantic Web, Information Retrieval |
| Revyu | 2007 | 3 | Motta, Heath | Semantic Web, External Sources, Data Handling, Best Practices |
| Evolva | 2009 | 1 | Zablith, Motta, Sabou, D'aquin | Ontology Evolution, Semantic Web, Background Knowledge, WWW |
| KC-Viz | 2010 | 2 | Motta, Peroni, Manuel G.P., D'aquin […] | Human Computer Interaction, User Interfaces, Ontology Engineering |

# TechMiner - Architecture

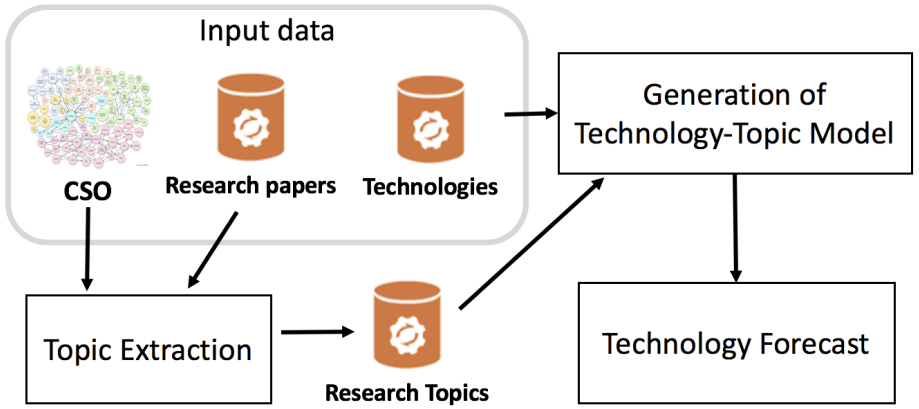


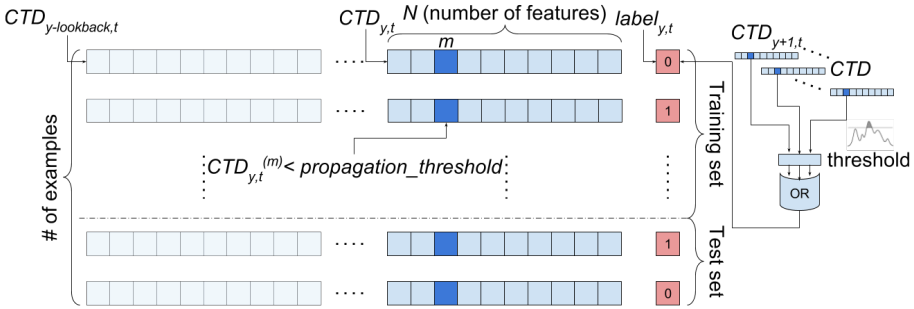
# Technology-Topic Framework

It learn how technologies propagate across research field, predict where a technology should be adopted next, and alert relevant stakeholders.

The aim is helping to **accelerate the knowledge flow** and the pace of technology propagation.

## Architecture




| Technologies | TTF predictions |
|---|---|
| **Latent Dirichlet Allocation (2003)** | object recognition, software engineering, comp. networks, security of data, ontology, semantic web, robotics, imaging systems, data reduction, image compression … |
| **Web Ontology Language (2004)** | multimedia systems, image processing, computer vision, robotics, security systems, computer aided design, genetic, telecomunication services, e-learning … |
| **Extreme Learning Machine (2008)** | infor. retrieval, wireless telecommunication systems, signal processing, data mining, computer vision, robotics, image reconstruction, speech recognition … |

**http://skm.kmi.open.ac.uk/ttf**

# Example of forecasted topics

| |
|---|
| **Extreme learning machine**: infor. retrieval, wireless telecommunication systems, signal processing, security of data, data mining, computer vision, robotics, image reconstruction, speech recognition, object recognition, […] |
| **Markov Logic Network :** information retrieval, software engineering, problem solving, signal processing, image processing, data mining, KB systems, computer vision, theorem proving , image segmentation, […] |
| **Latent Dirichlet allocation:** object recognition, software engineering, computer networks, security of data, ontology, semantic web, robotics, imaging systems, data reduction, image compression, object recognition, […] |
| **Web Ontology language:** multimedia systems, image processing, parallel processing systems, computer vision, robotics, network architecture, security systems, computer aided design, genes (biology), e-learning, […] |
| **SKOS:** software engineering, bioinformatics, computer networks, ML systems, information systems, computer programming languages, linguistics, object oriented programming, software design , e-learning, […] |
| **Semantic Web Rule Language:** quality of service, object oriented programming, software design, mobile telecommunication systems, mobile devices, e-commerce, social networks, computer aided soft. engineering, […] |
| **FOAF:** software engineering, computer networks, machine learning systems, security of data, distributed computer systems, web services, mobile computing, software design, multimedia systems, P2P networks, […] |

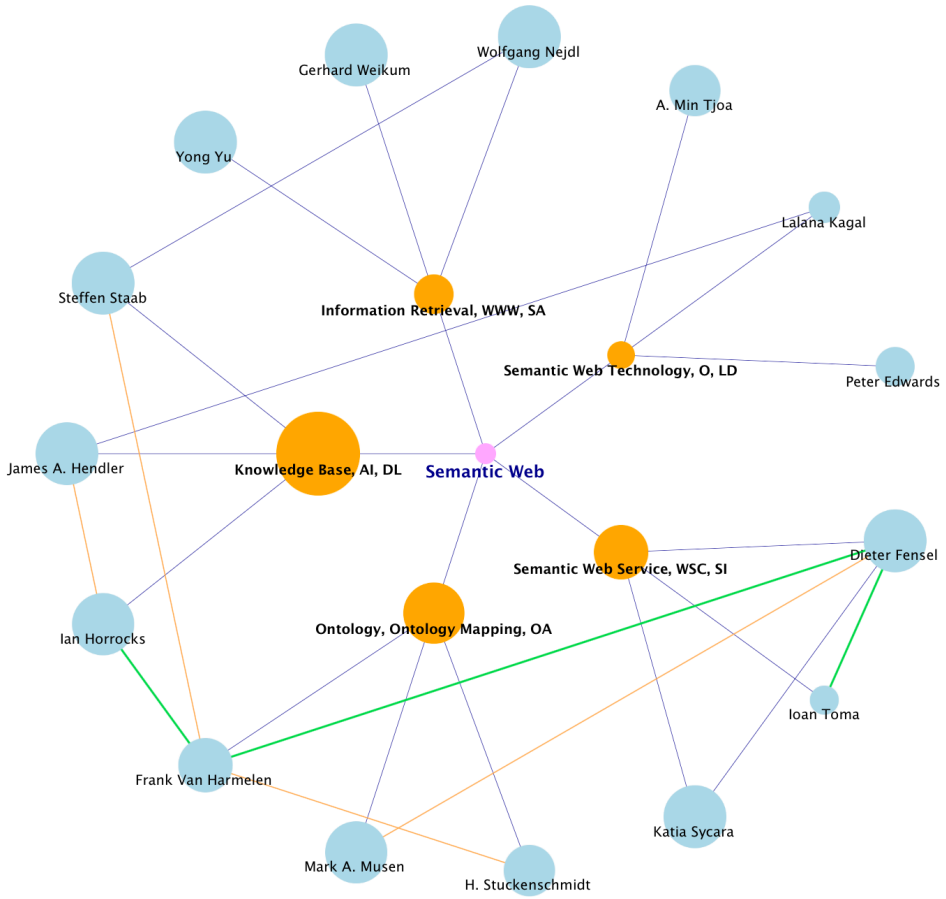
# Conclusions

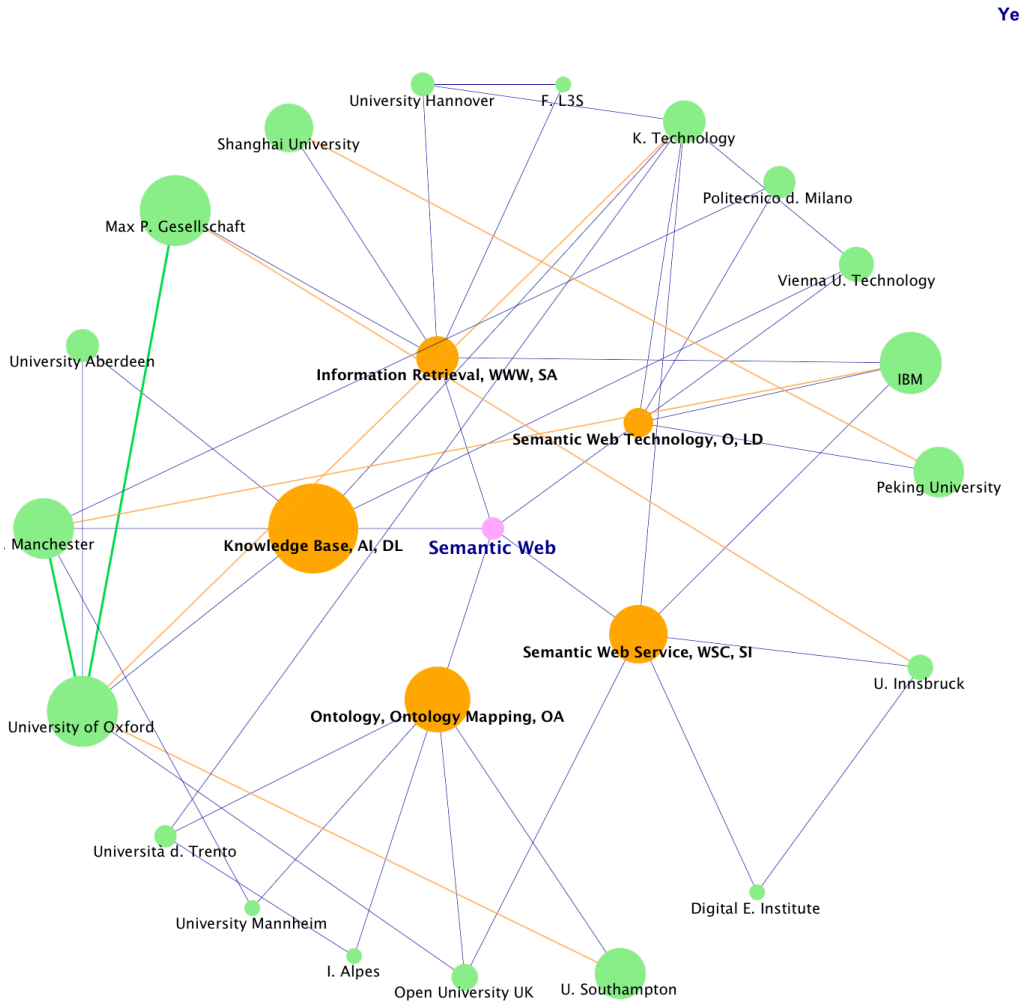
- **It is possible to forecast technology spreading**, at least for some categories of topics.
- We will need to consider other kinds of features if we want to address **other kinds of technology transfers**.
- We can use TTF for **alerting researchers** about **promising new technologies** relevant to their research and shorten the technology adoption lag.

# Next Steps

- Collecting data about more technologies to perform larger scale experiments.
- Creating an ontology of technologies and incorporate it in the analysis.
- Expanding the scope of our work by including other research fields, such as Biology, Social Science, and Engineering.

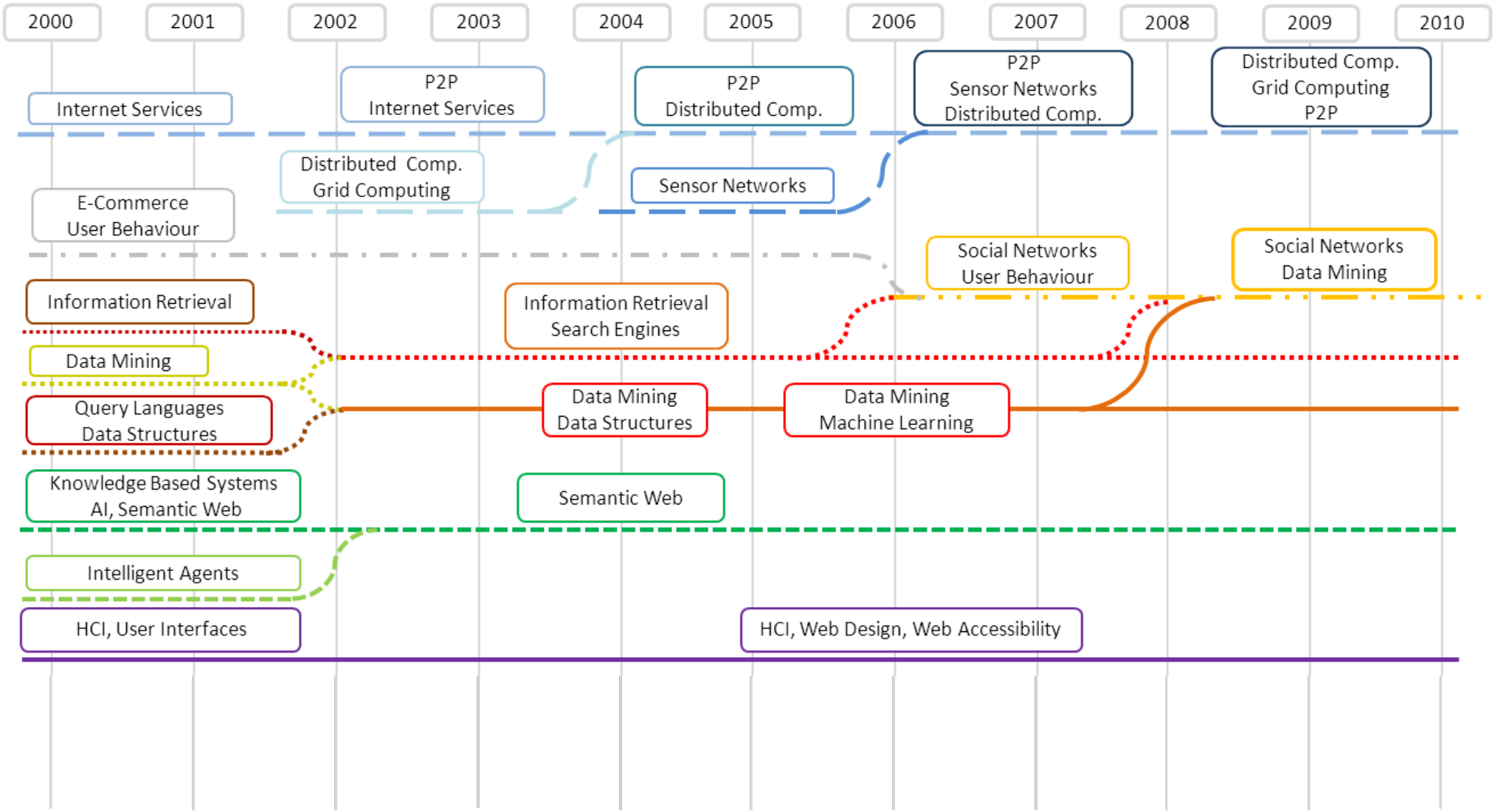
# 3. Describing and assessing research actors and products



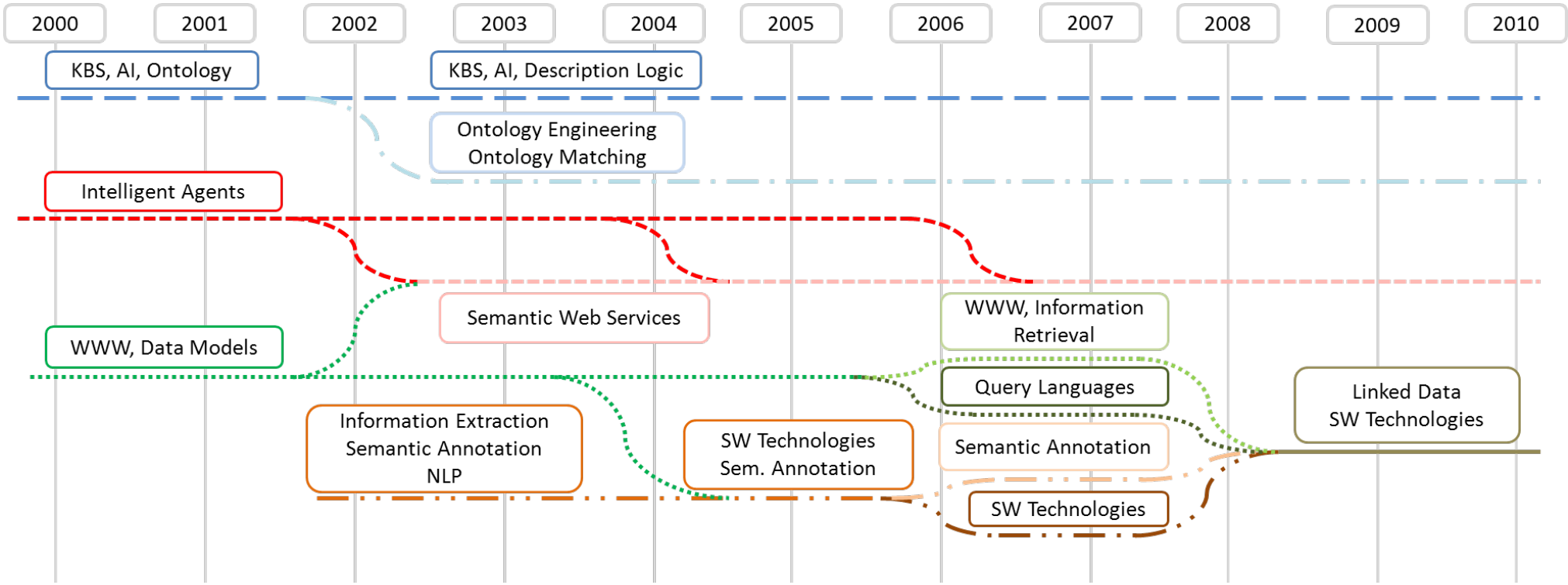




http://skm.kmi.open.ac.uk/rexplore

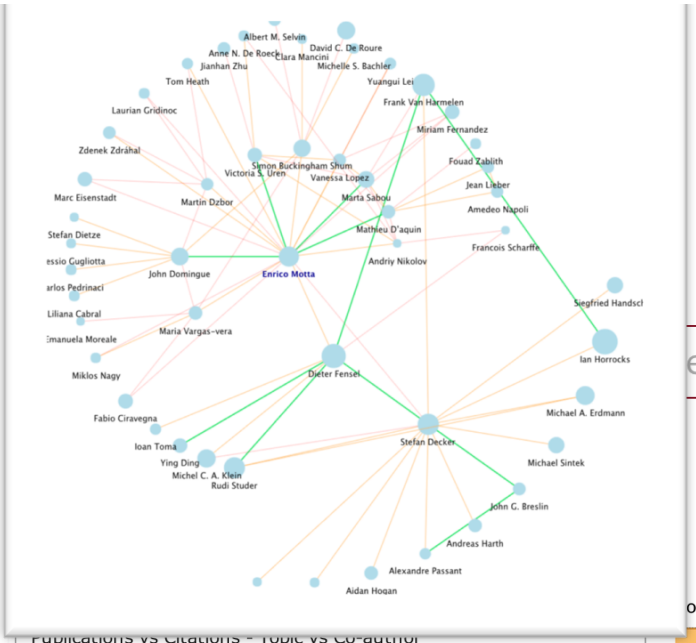
# Research Maps – WWW

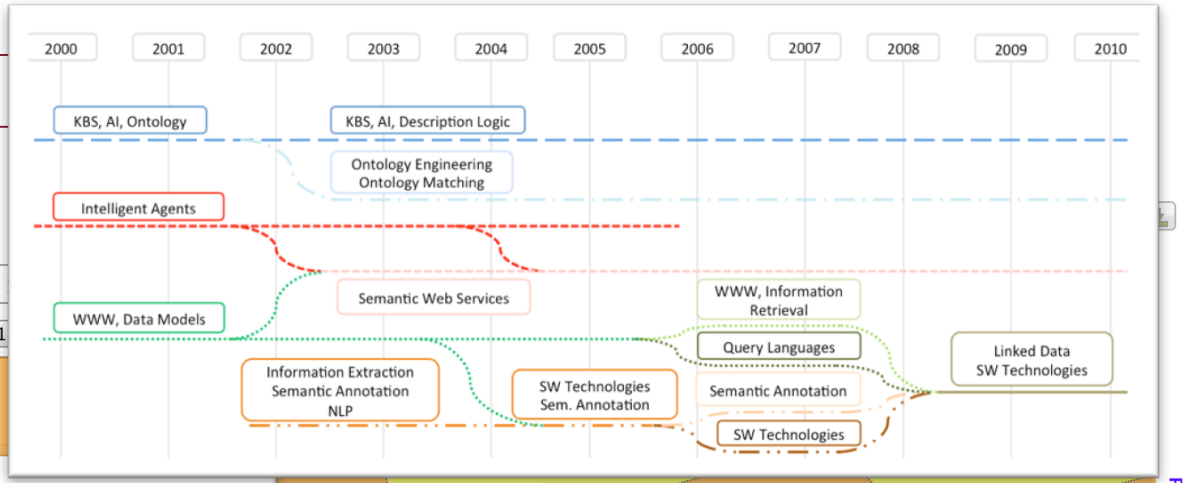


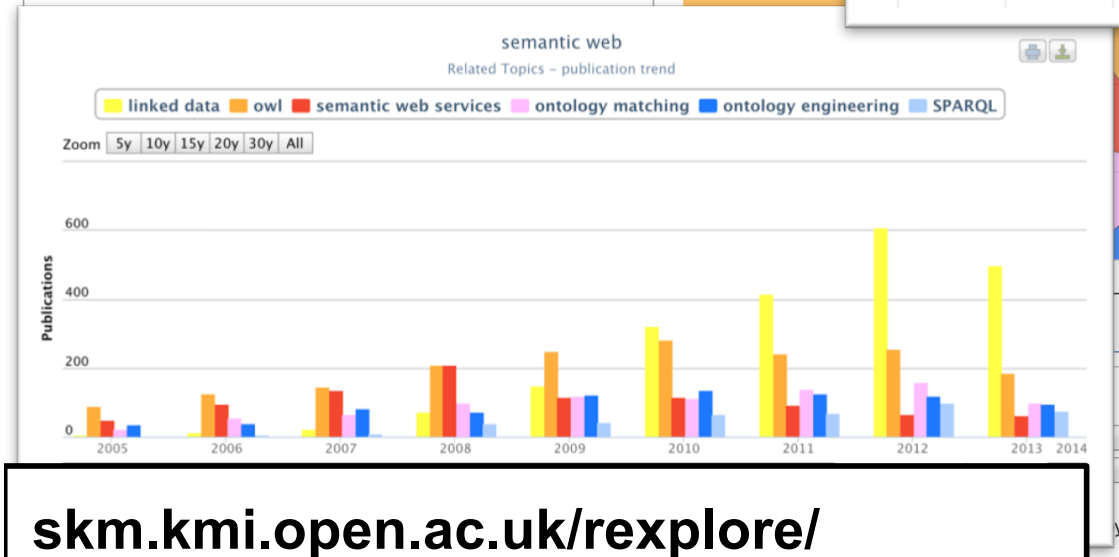
# Research Maps – Semantic Web

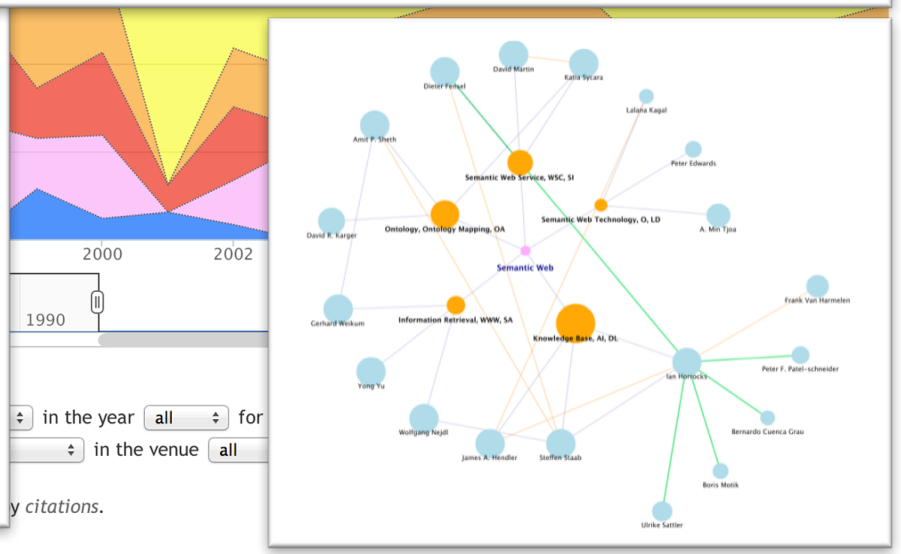


# Rexplore







Topics by Publications:

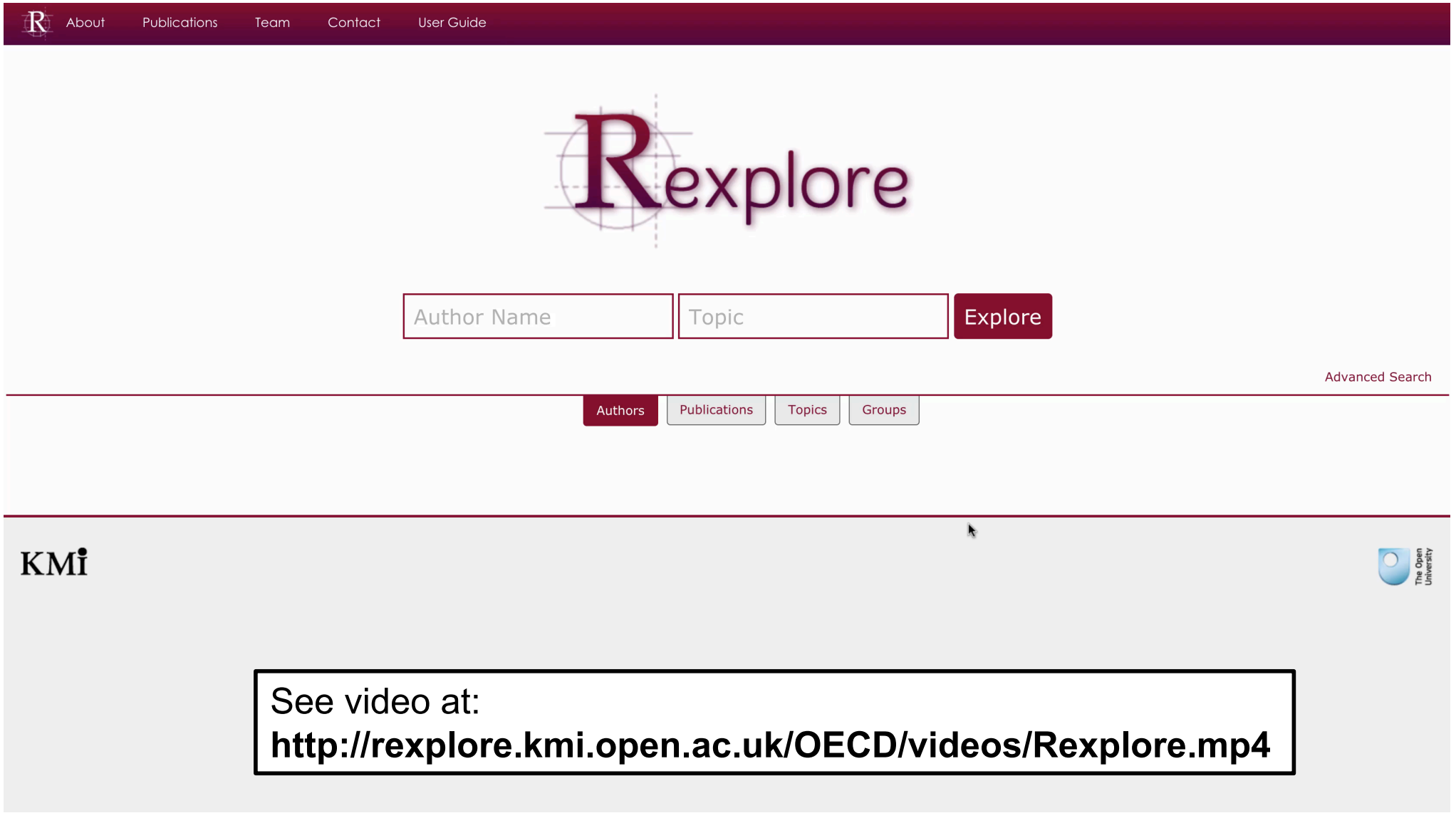
in the year all for

in the venue all

*y citations.*

**skm.kmi.open.ac.uk/rexplore/**

*Osborne, F., Motta, E. and Mulholland, P.*
*Exploring scholarly data with Rexplore.*
*International Semantic Web Conference 2013*

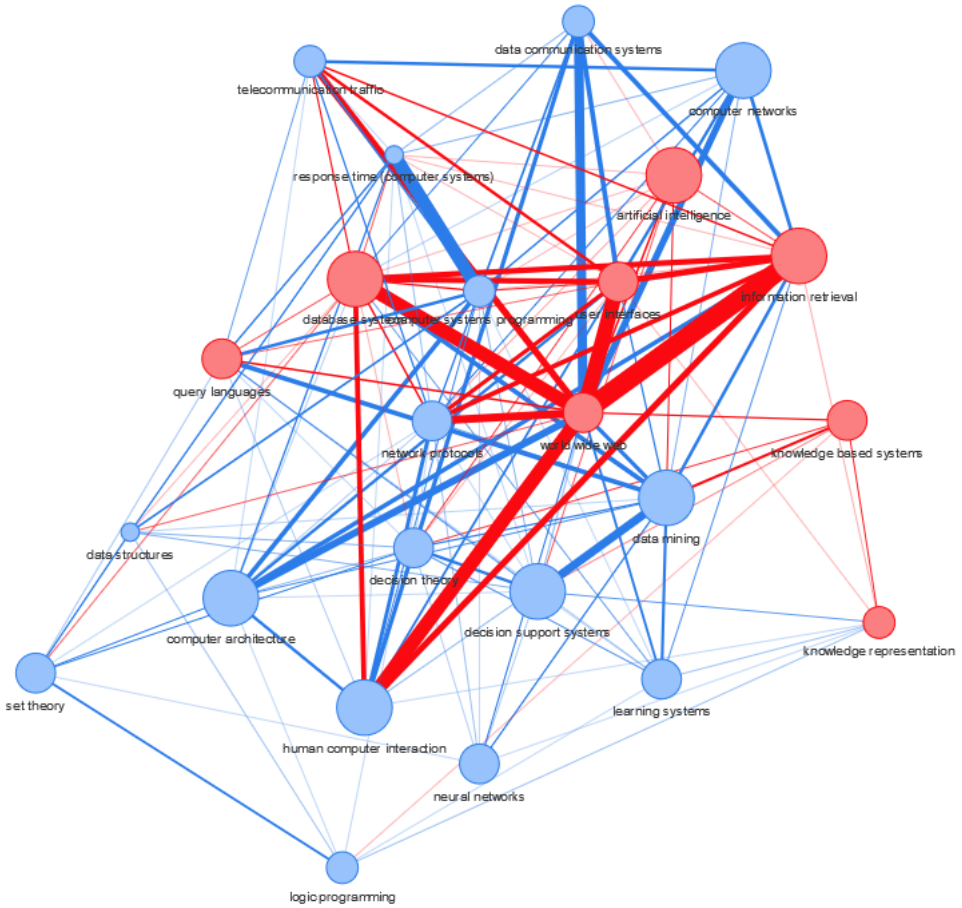


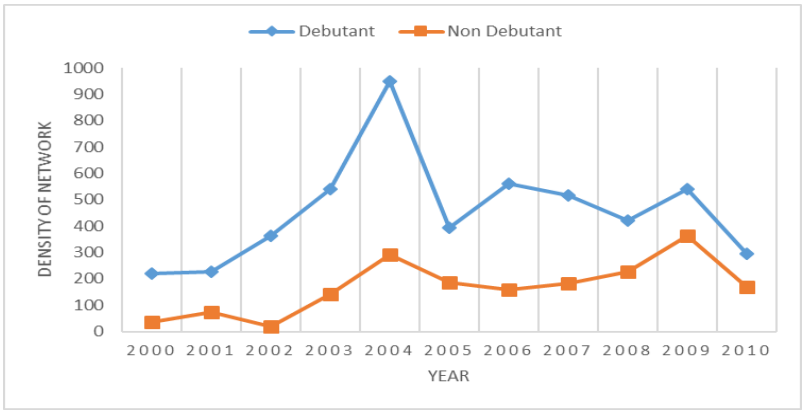

See video at:
**http://rexplore.kmi.open.ac.uk/OECD/videos/Rexplore.mp4**

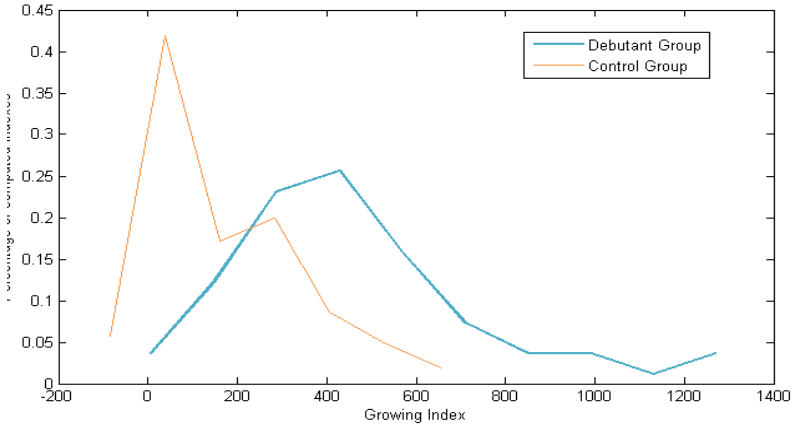
**skm.kmi.open.ac.uk/rexplore/**

# 4. Anticipating trends

***Augur*** is a method for detecting the emergence of research areas at an **embryonic stage**, i.e., before the topic has been consistently labelled by researchers and associated with several publications.



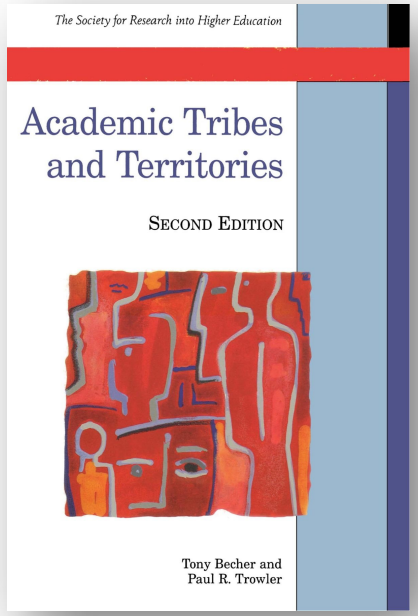




http://skm.kmi.open.ac.uk/augur

# What are research topics?

«As the work and the points of view **grow more specialised**, men in different disciplines have **fewer things in common**, in their background and in their daily problems»

Clark - *The study of Campus Cultures*

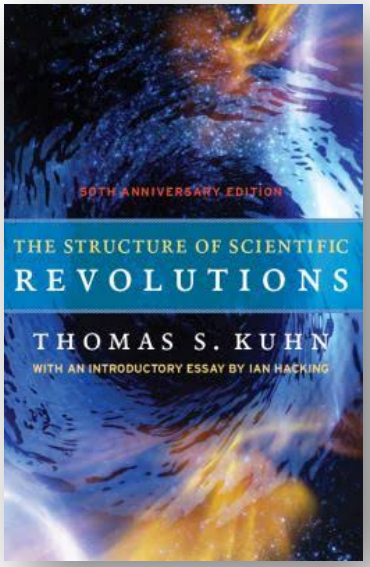



«Sometimes, of course, friendly relations **may be established** to **mutual benefit** ...»

Becher and Throwler - *Academic Tribes and Territories*

# How do they change?

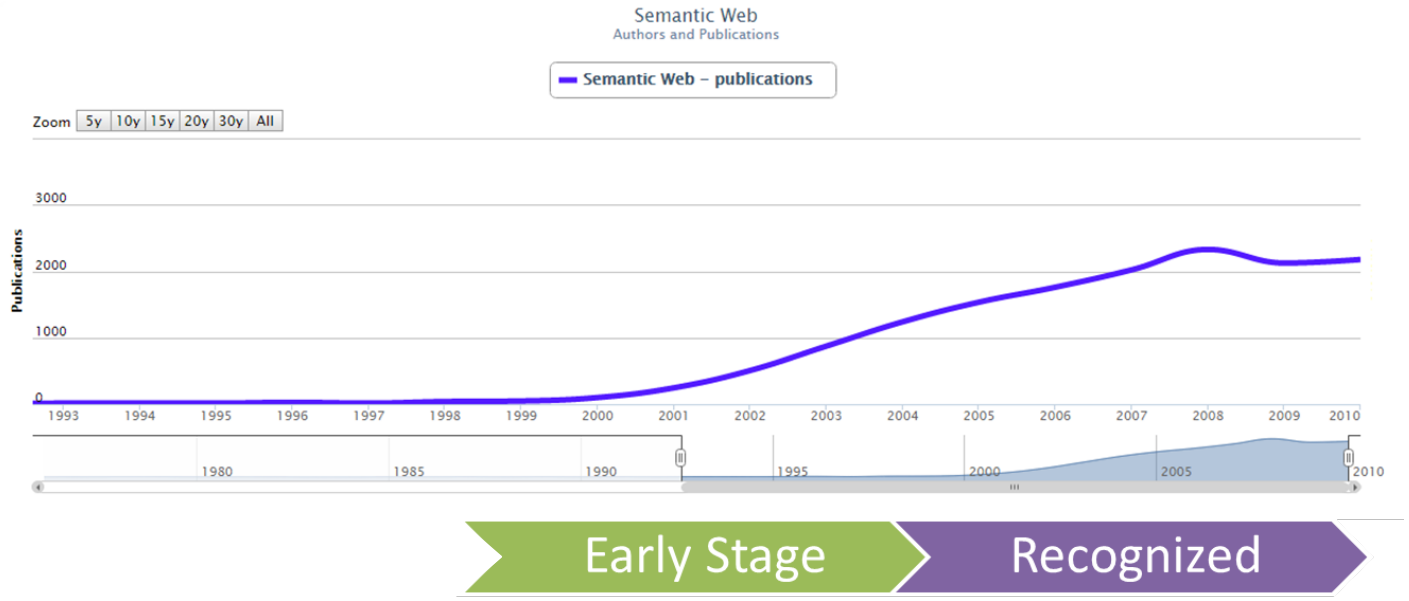





«[…] successive **transition** from one **paradigm** to another via revolution is the usual **developmental pattern** of mature **science**.»

Thomas Kuhn - *The Structure of Scientific Revolutions*

# How are topics born? A study with Springer Nature

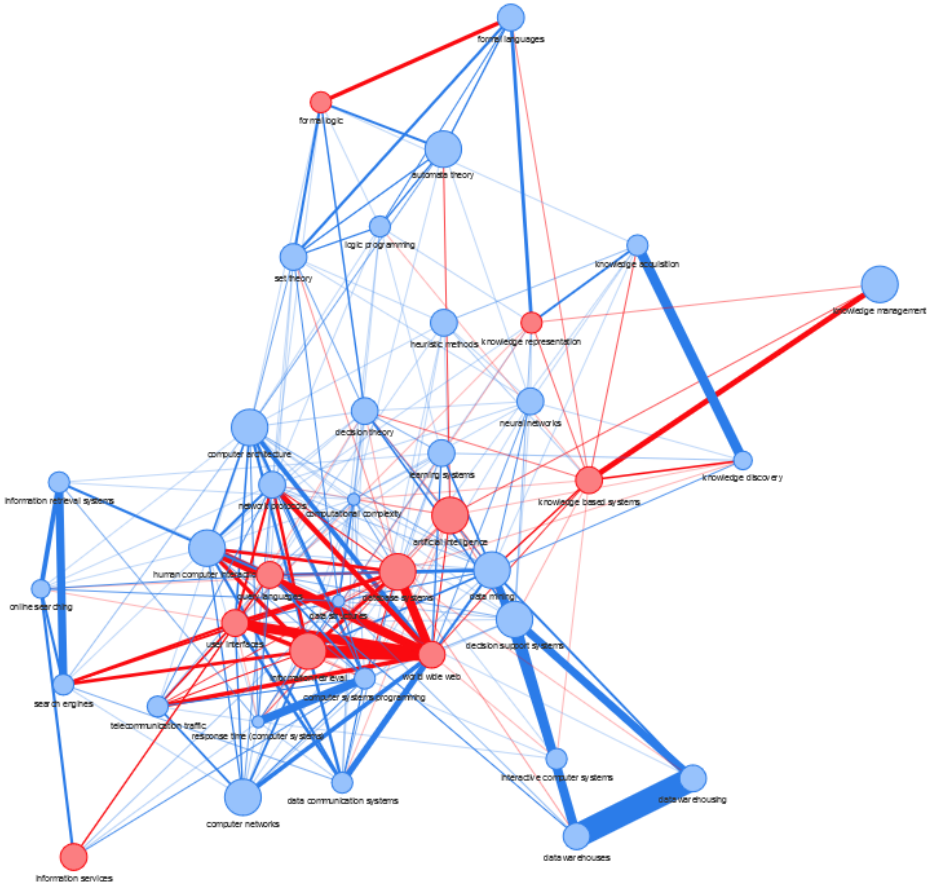


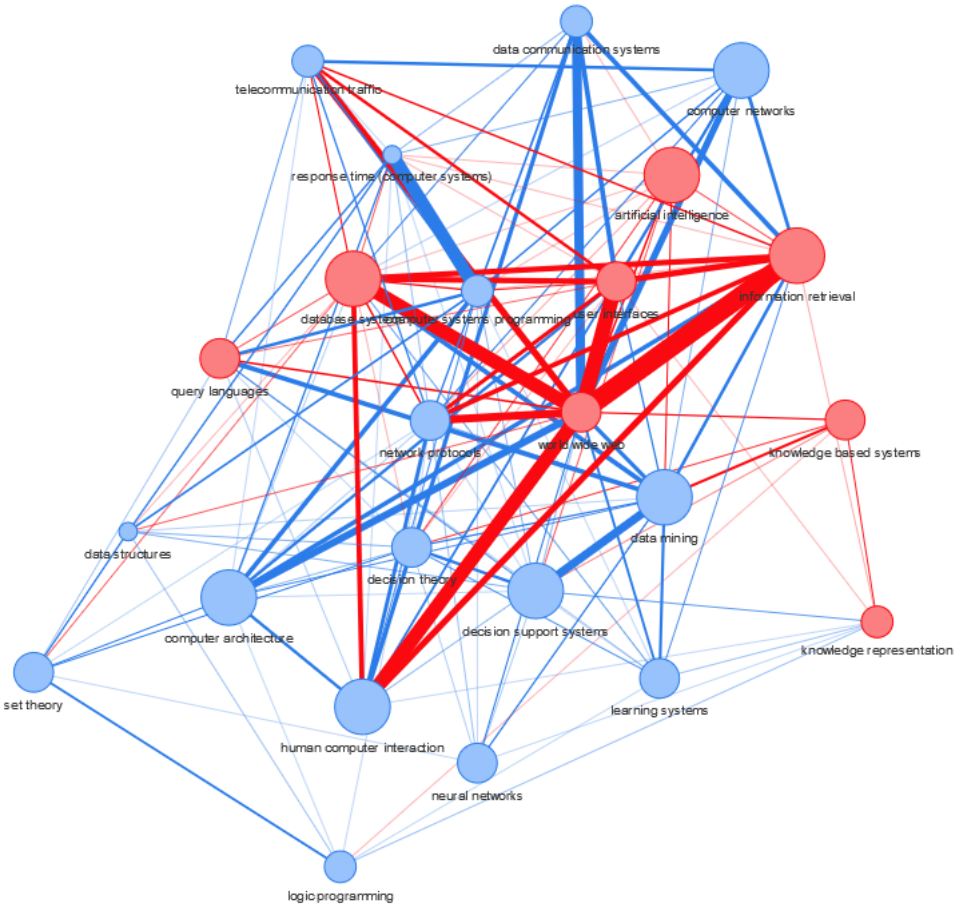

Is it possible to **detect the emergence of new research topics** even before they are consistently labelled by the community?

**What are the dynamics** that anticipate the creation of new research topic?

We analysed the co-occurrence network of research topics preceding a treatment group of debutant topics vs. a control group.

Can the increase of **pace of collaboration** and **density** in a topic network be an indicator of a new research topic?

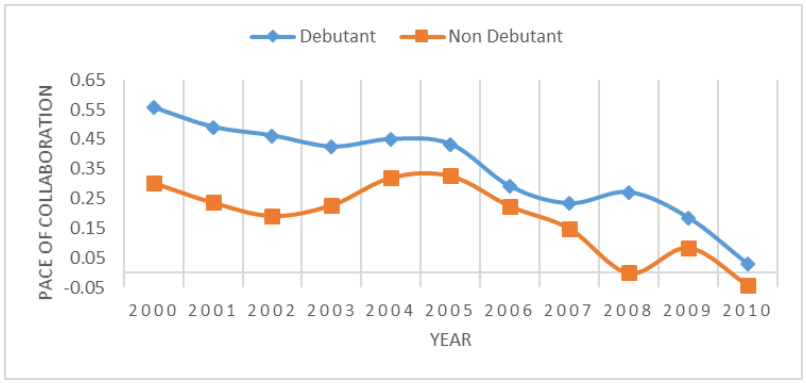


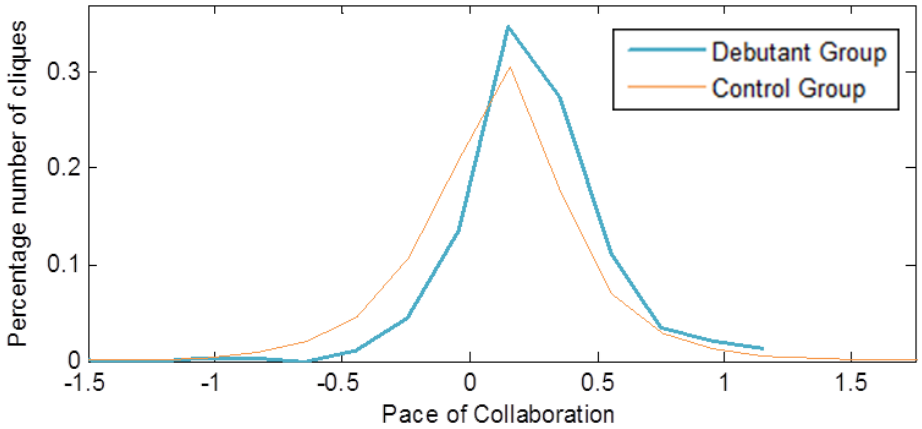



# Results

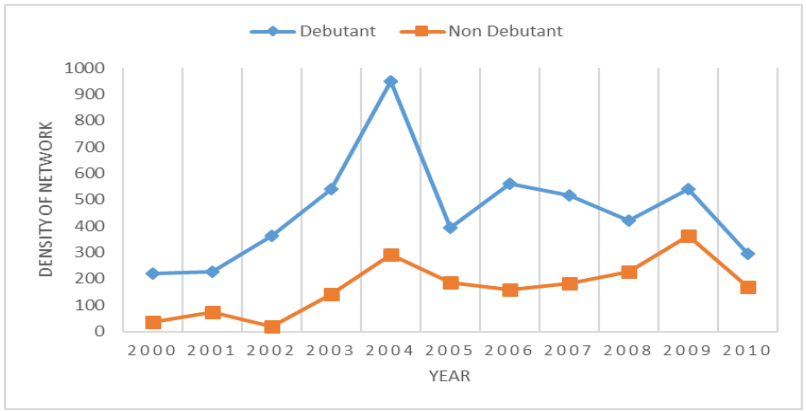
The **dynamics preceding the appearance** of research topics are significantly different from the ones associated to the other portions of the networks (**p<0.0001**).

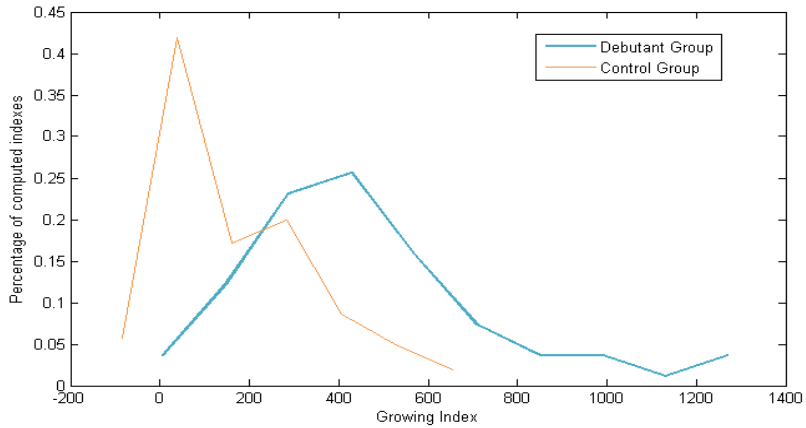
## Pace of Collaboration




## Density Growth

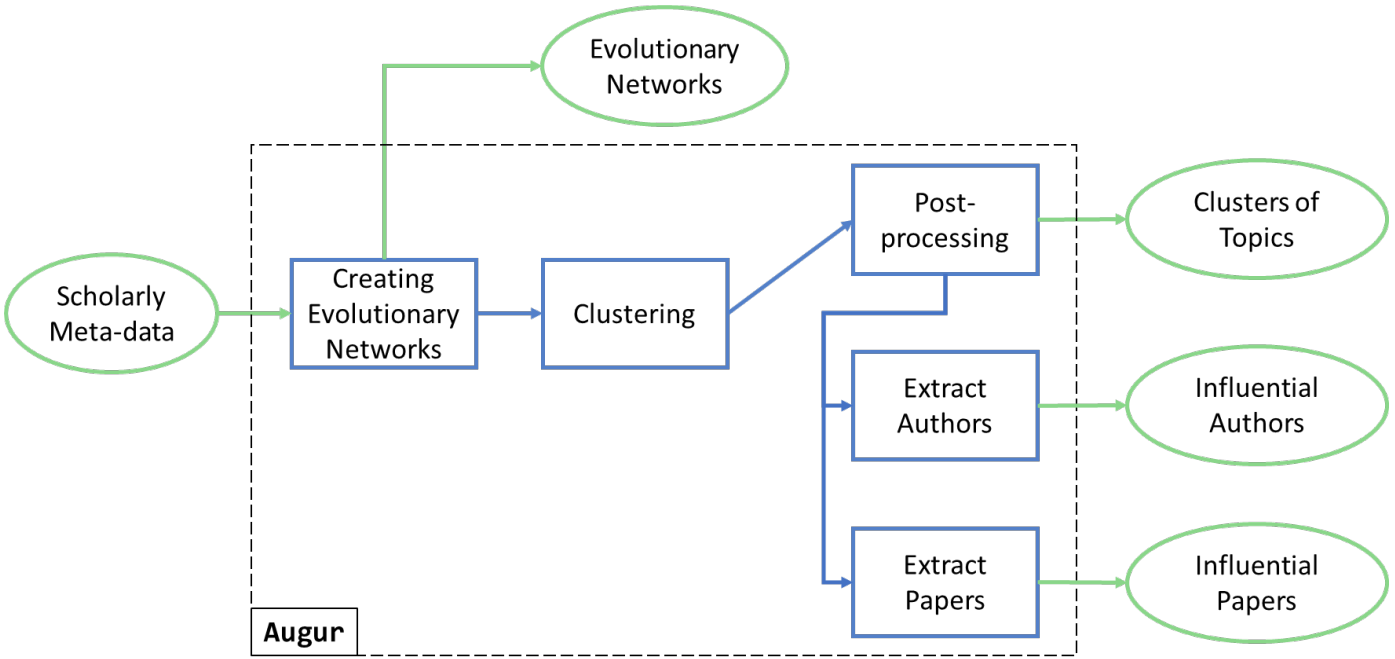


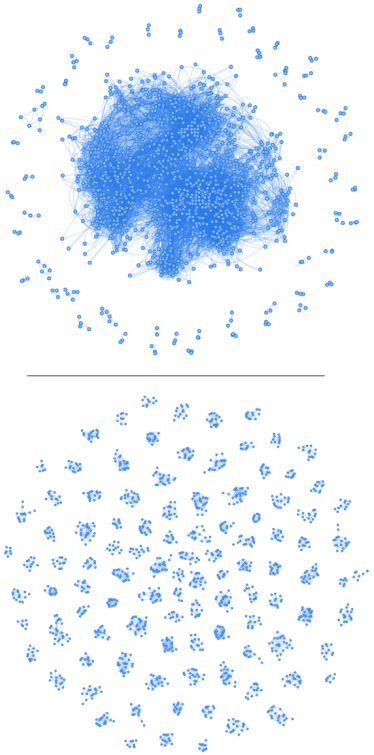



# Augur

***Augur*** is a method for detecting the emergence of research areas at an **embryonic stage**, i.e., before the topic has been consistently labelled by researchers and associated with several publications.

It analyses the relationships between research areas and is able to detect clusters of topics that exhibit dynamics correlated with the emergence of new research topics.



http://skm.kmi.open.ac.uk/augur

# Conclusion

- The study confirmed our hypothesis that it is possible in principle to detect the emergence of a new topic at the embryonic stage.

- It suggest that new topics tend to emerge in an environment in which weakly interconnected research areas begin to cross fertilize.
  - Interestingly, this can be fostered by relevant research policies

# Some Lesson Learnt

- A little semantic goes a long way (Hendler's hypothesis)
- Working together is imperative (scientific community, companies, organizations, political entities)
- Semantic does not just produce analytics, but also a new way to understand how research works.

KNOWLEDGE MEDIA INSTITUTE

SKM 3

Andrea Mannocci

Enrico Motta

Francesco Osborne

Angelo Salatino

Thiviyan Thanapalasingam

# **S**cholarly **K**nowledge **M**odelling, **M**ining and Sense**M**aking

# References


- **Fostering Technology Transfer:** Osborne, F., Mannocci, A. and Motta, E. (2017) Forecasting the Spreading of Technologies in Research Communities, K-CAP 2017, Austin, Texas, USA.
- **Early detection of research trends:** Salatino, A.A., Osborne, F. and Motta, E., (2017) How are topics born? Understanding the research dynamics preceding the emergence of new areas. PeerJ Computer Science, 3, p.e119.
- **Smart Book Recommender:** Osborne, F., Birukou, A., Thanapalasingam, T. , and Motta, E. (2017) Smart Book Recommender: A Semantic Recommendation Engine for Editorial Products. International Semantic Web Conference 2017, Poster Track. Vienna, Austria.
- **Smart Topic Miner:** Osborne, F., Salatino, A., Birukou, A. and Motta, E. (2016) Automatic Classification of Springer Nature Proceedings with Smart Topic Miner. International Semantic Web Conference 2016, Kobe, Japan.
- **CSO ontology creation:** Osborne, F. and Motta, E. (2015) Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks, International Semantic Web Conference 2015, Bethlehem, Pennsylvania, USA
- **Rexplore:** Osborne, F., Motta, E. and Mulholland, P. (2013) Exploring Scholarly Data with Rexplore, International Semantic Web Conference, Sydney, Australia

**Francesco Osborne**
**KMi, The Open University**

francesco.osborne@open.ac.uk


Knowledge Media Institute

skm.kmi.open.ac.uk