Information Services & Use 39 (2019) 319–333
DOI 10.3233/ISU-190063
IOS Press

Artificial intelligence in scholarly communications: An elsevier case study

Ann Gabriel[*]
Senior Vice President, Global Strategic Networks, Elsevier, 230 Park Ave., Suite 800, New York, NY 10169, USA

Abstract. This paper is an adaptation of presentation given by the author at the NFAIS Conference on Artificial Intelligence that was held in Alexandria, VA from May 15–16, 2019. It provides an overview on the need for increased Artificial Intelligence (AI) usage in scholarly communications for both information providers and the research community. It also includes an introduction to how Elsevier transitioned from print to electronic to information solutions (P – E – S) and how some of its tools employ AI. In addition, it covers two case studies showcasing how Elsevier incorporated Machine Learning (ML) and Natural Language Processing (NLP) to create two technological and data-based solutions for researchers, as well as a summary of the solutions' positive outcomes.

Keywords: Scholarly communications, Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), information solutions, Elsevier, ScienceDirect

1. Introduction

Andrew Ng, Stanford computer science professor and co-founder of Coursera, once said, “I have a hard time thinking of an industry that I don't think AI will transform in the next several years [1]”.

The scholarly publishing and information industries are no exception. Smart publishers are beginning to embrace AI, weaving it into the core of their business - to help them enter new markets or source new content, to inform and improve existing content, and for new product development - which otherwise might have required significant investment and business risk. Publishers are also using AI to reduce costs in their editorial processes. Many use Natural Language Processing (NLP) to provide sematic enrichment and content recommendations for users.

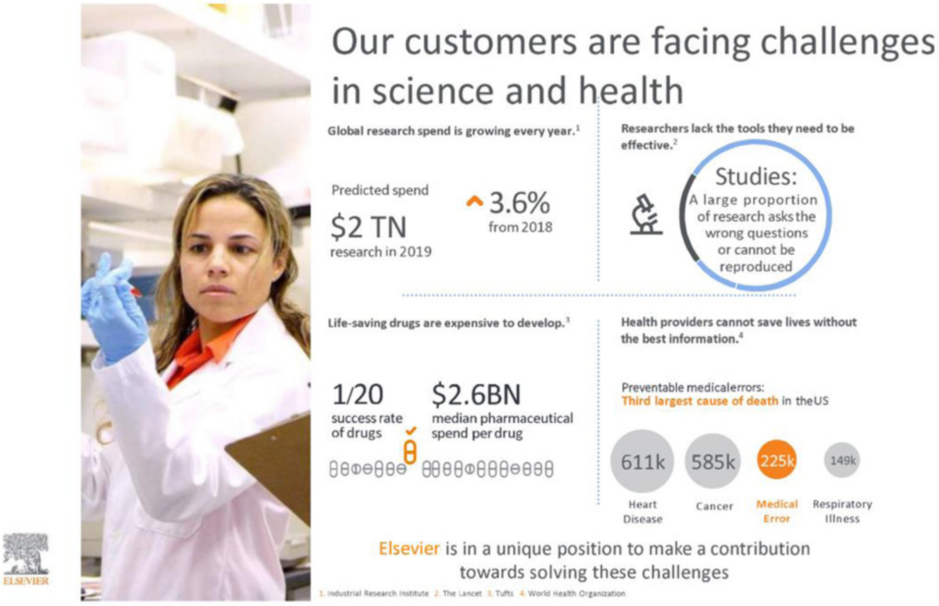
Today, the need for academic publishers to adopt AI has become undeniable. There are several reasons for this, the most obvious of which is that scholarly output across the world has tripled in the last two decades. As the number of articles, journals, and publishers grows, it's reasonable to assume that the value of this content will decline.

Additionally, while global research spend is growing quickly every year - $2 trillion today, which is nearly triple the spending in 2000 [2] - we're hearing from researchers that they still lack effective tools. A number of recent studies show that a large proportion of scientific research asks the wrong questions, or when the research is completed, it cannot be reproduced.

[*] E-mail: a.gabriel@elsevier.com.

On the health side, the median cost of developing a drug is now $2.6 billion [3]; yet the success rate is only one in twenty. Furthermore, alongside heart disease and cancer, the third largest cause of death is medical error.

There is a huge opportunity in these spaces for publishers to leverage AI, to curate and enhance content, and ro help researchers and medical professionals find what they are looking for more quickly and accurately.



Our customers are facing challenges in science and health

Global research spend is growing every year.[1]

Predicted spend

$2 TN

research in 2019

3.6%

from 2018

Researchers lack the tools they need to be effective.[2]

Studies:

A large proportion of research asks the wrong questions or cannot be reproduced

Life-saving drugs are expensive to develop.[3]

1/20

success rate of drugs

$2.6BN

median pharmaceutical spend per drug

Health providers cannot save lives without the best information.[4]

Preventable medicalerrors:

Third largest cause of death in theUS

611k Heart Disease

585k Cancer

225k Medical Error

149k Respiratory Illness

Elsevier is in a unique position to make a contribution towards solving these challenges

1. Industrial Research Institute 2. The Lancet 3. Tufts 4. World Health Organization

2. Elsevier’s P – E – S transition

That’s why Elsevier employs more than one thousand technologists today, who are rapidly developing new capabilities in ML and NLP. But before I go into more detail on what we are doing today, it might be useful to go back in time and review how we got here in the first place.

Elsevier’s history as a publisher of physical books and journals goes back hundreds of years, but when the Internet started to become mainstream in the 1990s, our core products suddenly seemed outdated. In fact, in 1995, *Forbes* even released an article predicting that we would be the “Internet’s first victim” [4]. We did survive the digital revolution, however, because of our willingness to embrace technological innovation. As industry leaders, we realized that we had to transform ourselves from being a traditional print content publisher to a provider of e-content and information solutions.

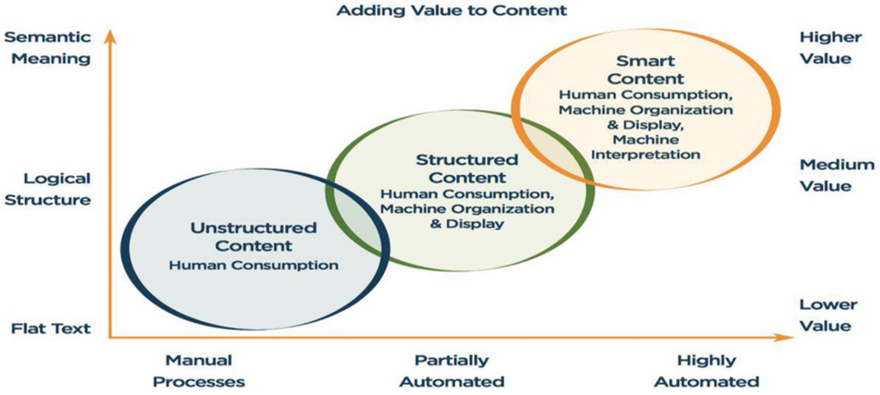
In the print environment, Elsevier published and sold raw content in fixed quantities, without really knowing how it was being used. Then, in the internet era of the 1990s, we pivoted to electronic distribution. This enabled us to consolidate all of our content into a digital repository (ScienceDirect), where library and research customers could easily search and pick out what they needed.

This shift from print to electronic was momentous, but it was only the beginning. Once we had all our content digitized, we started to understand that we had even more value on our hands than we’d originally thought - and this is where the topic of data analytics and AI comes in. With all of this digital content, we realized that Elsevier had the world’s largest, highest-quality source of machine-readable STM data and metadata. With the right analytics, this data could be leveraged to create a suite of AI and ML-powered solutions. For instance, we provide tools to help doctors in making the right diagnosis, exploratory drillers in finding oil efficiently, pharmaceutical drug companies in bringing new drugs to market, scholars in tracking the most important research in their field, universities in benchmarking their performance, and much more.

In other words, Elsevier's trajectory was one of "P – E – S"; we started off as a print publisher (P), became a digital publisher of electronic content (E), and ultimately evolved into a data-based solutions (S) provider. Today, print sales account for less than 10% of Elsevier's revenue - the vast majority now coming from data-based solutions that help turn static information into actionable knowledge.

3. Case study 1: ScienceDirect topic pages

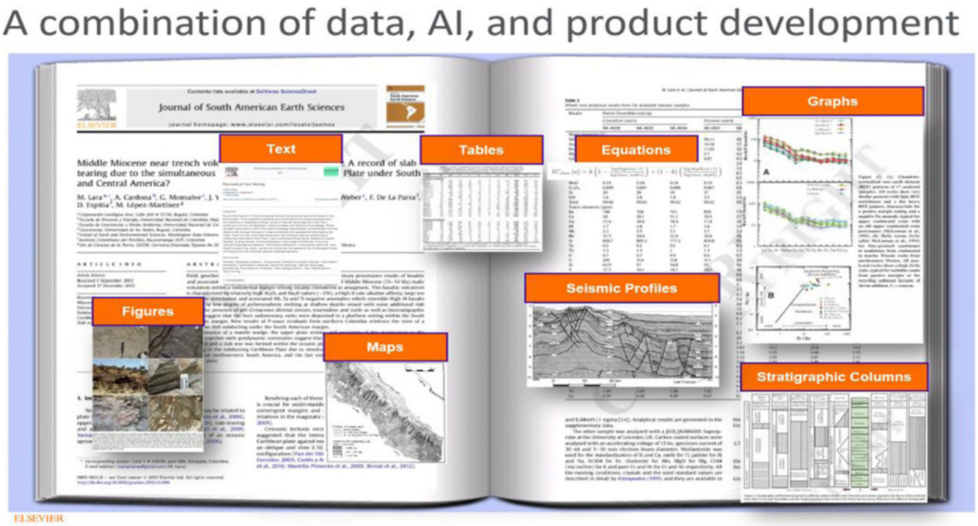
NLP (the "machine" used to turn natural language into structured semantic text that is machine-readable and parse-able) and ML enable truly Smart Content. We are currently seeing a gradual, industry-wide migration from manual processes and flat text to using highly-automated means to derive greater semantic meaning from content.



Elsevier uses NLP and semantic technologies extensively in all of its solutions. For example, we can transform the source text to noun phrases, apply eleven domain-specific thesauri and structured vocabularies, extract concepts, and create fingerprints from any text – such as an abstract or a grant description. We've recently added a Chemistry thesaurus with hundreds of thousands of compound names.

So, let's take a look at how these NLP and semantic technologies are being applied in Elsevier tools. I'd like to focus on one example from our online database of scientific and medical research, ScienceDirect, which will also serve as a concrete demonstration of our shift from electronic content to solutions.

Below is an illustrative example in the field of Earth Sciences of what types of compositional content we might seek to extract from previously static text.



A combination of data, AI, and product development

While the text in this article may be the most obvious form of extractable data, there are also many other data points in different formats: namely, tables, equations, graphs, figures, maps seismic profiles, and stratigraphic columns.

Here I think it would be appropriate to quote David Smith (Head of Product Solutions at the Institution of Engineering and Technology), in his *Scholarly Kitchen* article on AI and scholarly communications:

> "The research article, frankly, isn't a very good "raw material" as things currently stand. It's not written to be consumed by a machine. Its components can't be easily decoupled and utilized; they lack enough context and description and organization to collect at scale and, oh yeah, often times the data is wrong… Because research is a journey involving scholars trying to become slightly less uncertain about the worlds they are trying to understand" [5].

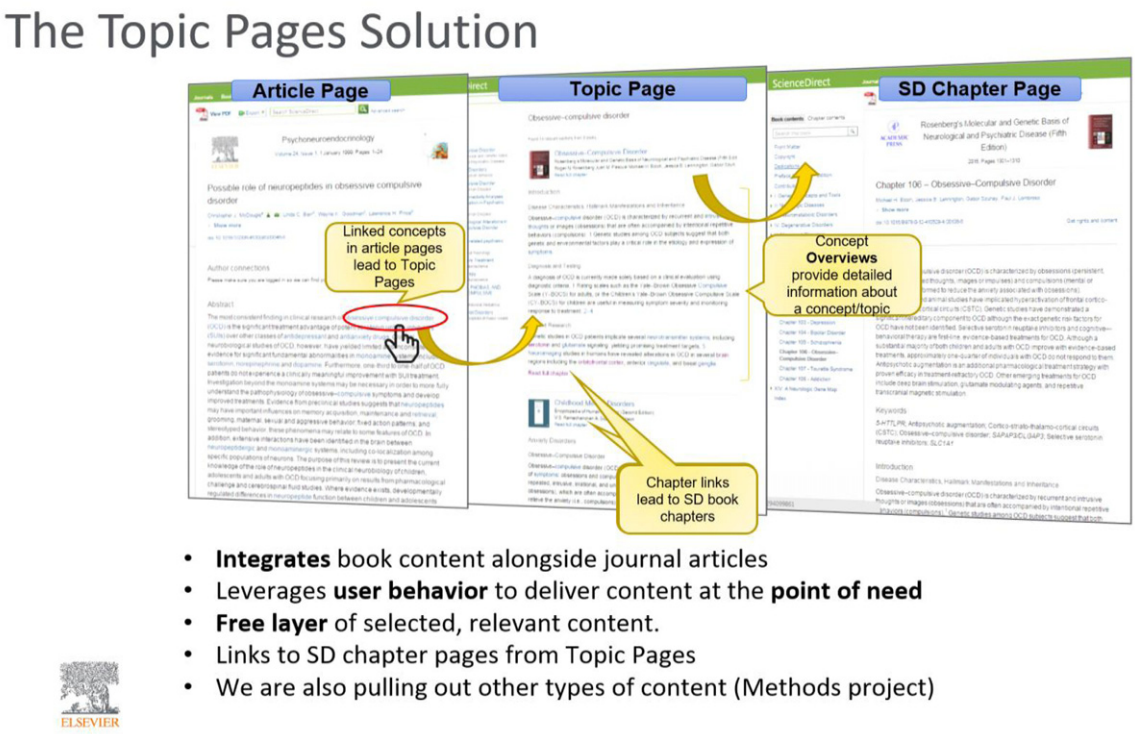
However, adding further structure to the content in an article - which would otherwise have remained trapped in a PDF - enables it to be leveraged for use cases that enhance research productivity.

For instance, if researchers reading an article were to stumble across a term with which they were not familiar, our research showed that they would stop reading and search the term on other internet sources (e.g. Wikipedia) to get definitions and background information about the concept. The problem is that information from sources like Wikipedia is not always authoritative - which is why we decided to deploy AI to help our end-users understand articles better and more efficiently.

Our objective was to create an automated, scalable model for extracting and highlighting definitions and concepts at the point of use, so that we could give researchers the trusted and citable information they needed when they needed it. So how did we accomplish this?

Creating a viable solution was not easy, and there were several serious roadblocks we had to overcome before we could proceed. For example, we have a massive corpus of data; and within this corpus, most sentences are not definitions and many scientific concepts are ambiguous. Therefore, we sought to develop strong predictive models, including incorporating human feedback to ensure that all concept definitions were sound.

What finally emerged from the end-user need to understand the article was a feature on ScienceDirect (SD) called "Topic Pages" as shown in the following figure



The Topic Pages Solution

Article Page
Topic Page
SD Chapter Page
Psychoneuroendocrinology
Possible role of neuropeptides in obsessive compulsive disorder
Obsessive–compulsive disorder
Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease (Fifth Edition)
Chapter 106 – Obsessive–Compulsive Disorder

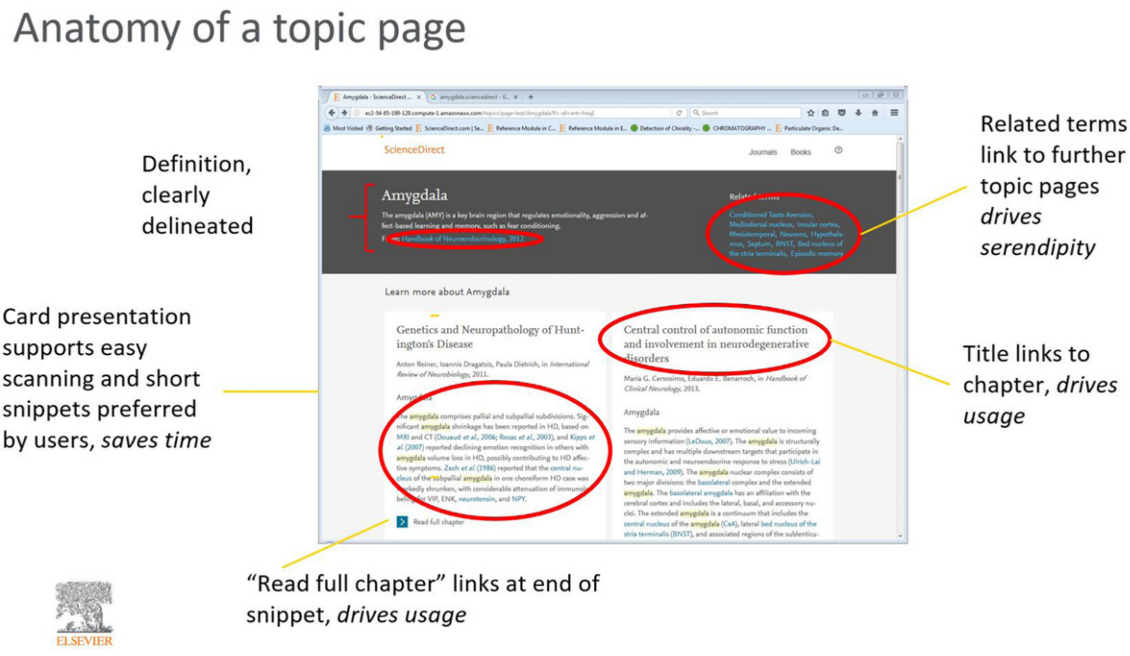
Linked concepts in article pages lead to Topic Pages

Concept Overviews provide detailed information about a concept/topic

Chapter links lead to SD book chapters

- **Integrates** book content alongside journal articles
- Leverages **user behavior** to deliver content at the **point of need**
- **Free layer** of selected, relevant content.
- Links to SD chapter pages from Topic Pages
- We are also pulling out other types of content (Methods project)

Concise definitions for terms help to orient users to a subject quickly. Relevant excerpts are highlighted across all literature, and we meet researchers where they are - we list Topic Pages in journal articles as well as in search engine results, and they are all freely-accessible.



Anatomy of a topic page

Amygdala

The amygdala (AMY) is a key brain region that regulates emotionality, aggression and affect-based learning and memory, such as fear conditioning.

From: Handbook of Neuroendocrinology, 2012

Related terms

Conditioned Taste Aversion, Mediodorsal nucleus, Insular cortex, Mesiotemporal, Neurons, Hypothalamus, Septum, BNST, Bed nucleus of the stria terminalis, Episodic memory

Learn more about Amygdala

Genetics and Neuropathology of Huntington's Disease

Anton Reiner, Ioannis Dragatsis, Paula Dietrich, in *International Review of Neurobiology*, 2011.

Amygdala

The amygdala comprises pallial and subpallial subdivisions. Significant amygdala shrinkage has been reported in HD, based on MRI and CT (Douaud *et al.*, 2006; Rosas *et al.*, 2003), and Kipps *et al.* (2007) reported declining emotion recognition in others with amygdala volume loss in HD, possibly contributing to HD affective symptoms. Zech *et al.* (1986) reported that the central nucleus of the subpallial amygdala in one choreiform HD case was markedly shrunken, with considerable attenuation of immunolabeling for VIP, ENK, neurotensin, and NPY.

Read full chapter

Central control of autonomic function and involvement in neurodegenerative disorders

Maria G. Cersosimo, Eduardo E. Benarroch, in *Handbook of Clinical Neurology*, 2013.

Amygdala

The amygdala provides affective or emotional value to incoming sensory information (LeDoux, 2007). The amygdala is structurally complex and has multiple downstream targets that participate in the autonomic and neuroendocrine response to stress (Ulrich-Lai and Herman, 2009). The amygdala nuclear complex consists of two major divisions: the basolateral complex and the extended amygdala. The basolateral amygdala has an affiliation with the cerebral cortex and includes the lateral, basal, and accessory nuclei. The extended amygdala is a continuum that includes the central nucleus of the amygdala (CeA), lateral bed nucleus of the stria terminalis (BNST), and associated regions of the sublenticu-

Definition, clearly delineated

Related terms link to further topic pages *drives serendipity*

Card presentation supports easy scanning and short snippets preferred by users, *saves time*

Title links to chapter, *drives usage*

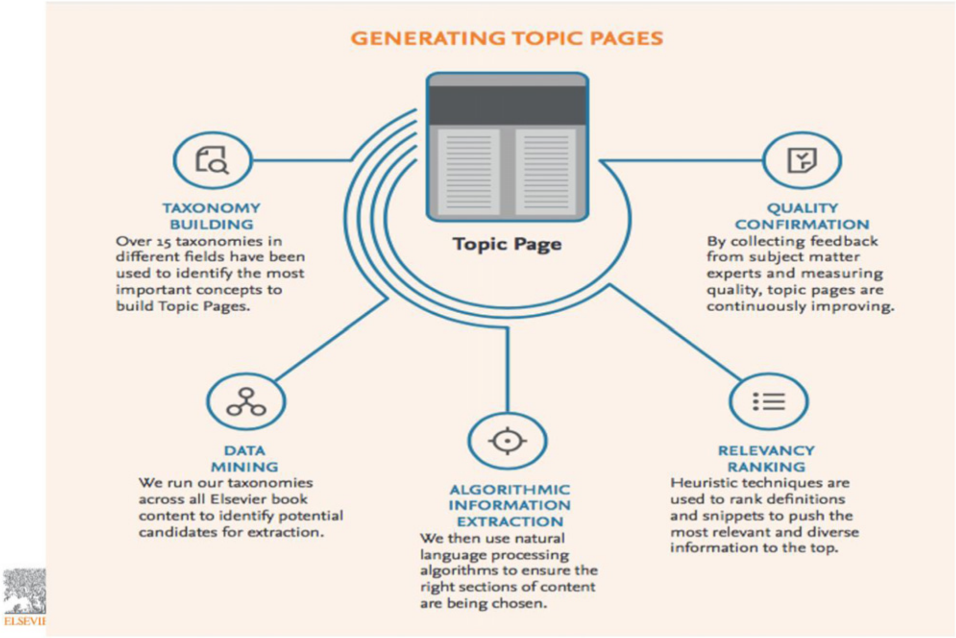
"Read full chapter" links at end of snippet, *drives usage*

Using ML, we have generated more than one hundred thousand Topic Pages from reference content on ScienceDirect. Over one million articles are being enriched with links to Topic Pages, and we are planning to expand to more domains.

This achievement would not have been possible without the right technology, including:

(1) **Taxonomy building:** Over fifteen taxonomies in different fields have been used to identify the most important concepts to build Topic Pages.
(2) **Data mining:** We run our taxonomies across all Elsevier book content to identify potential candidates for extraction.
(3) **Algorithmic information extraction:** We then use NLP to ensure that the right sections of content are being chosen.
(4) **Relevancy ranking:** We use heuristic techniques to rank definitions and snippets to push the most relevant and diverse information to the top.
(5) **Quality confirmation:** By collecting feedback from subject matter experts, measuring quality, and testing the User Experience (UX), Topic Pages are continuously improving.

The response within the research community and across the broader public has been astounding. Since its launch in 2017, Topic Pages have received more than thirty-three million visits, which in less than one year grew from two hundred and seventy thousand per page to an average of six million per month. We have a 60% return rate among users, and 55% of users return to view more pages. 92% of surveyed researchers agree that Topic Pages have helped them to achieve their goals.



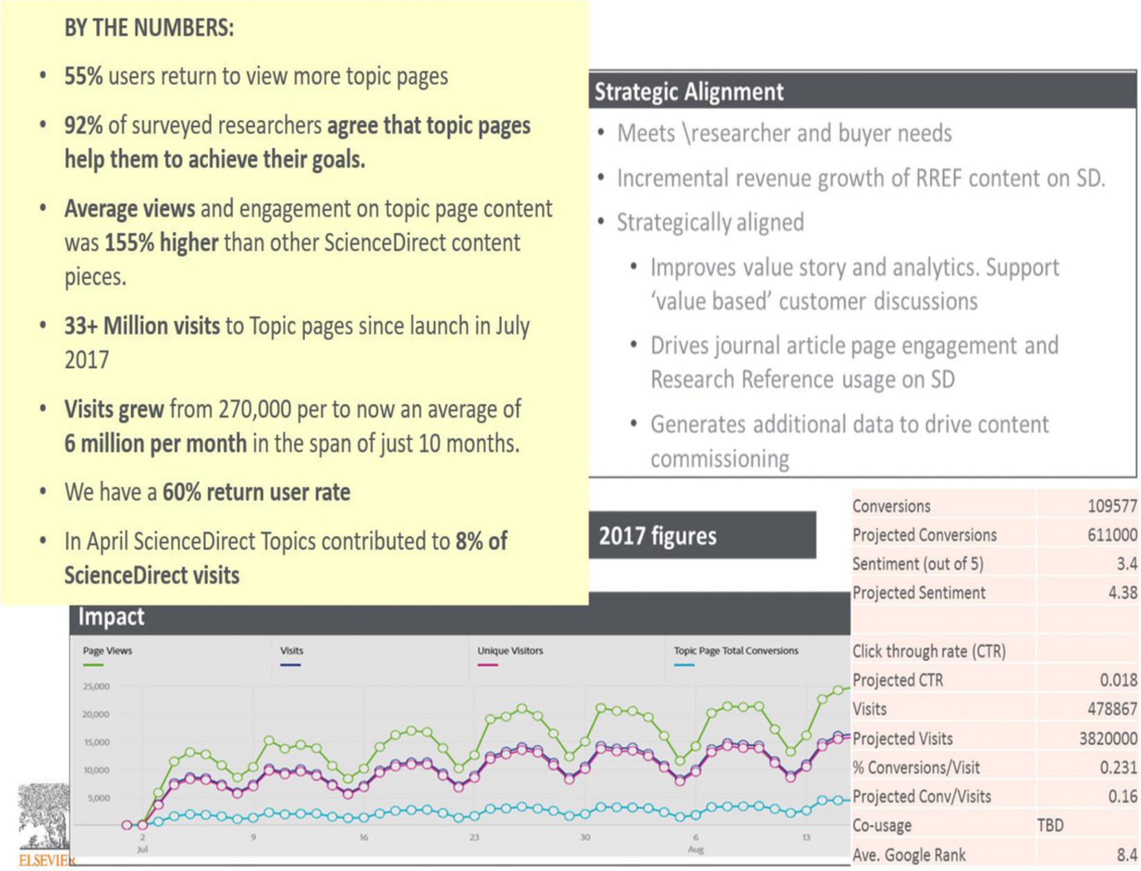
Over 15 taxonomies in different fields have been used to identify the most important concepts to build Topic Pages.

By collecting feedback from subject matter experts and measuring quality, topic pages are continuously improving.

We run our taxonomies across all Elsevier book content to identify potential candidates for extraction.

We then use natural language processing algorithms to ensure the right sections of content are being chosen.

Heuristic techniques are used to rank definitions and snippets to push the most relevant and diverse information to the top.



BY THE NUMBERS:

- **55%** users return to view more topic pages
- **92%** of surveyed researchers **agree that topic pages help them to achieve their goals.**
- **Average views** and engagement on topic page content was **155% higher** than other ScienceDirect content pieces.
- **33+ Million visits** to Topic pages since launch in July 2017
- **Visits grew** from 270,000 per to now an average of **6 million per month** in the span of just 10 months.
- We have a **60% return user rate**
- In April ScienceDirect Topics contributed to **8% of ScienceDirect visits**

Strategic Alignment

- Meets \researcher and buyer needs
- Incremental revenue growth of RREF content on SD.
- Strategically aligned
  - Improves value story and analytics. Support 'value based' customer discussions
  - Drives journal article page engagement and Research Reference usage on SD
  - Generates additional data to drive content commissioning

2017 figures

Conversions		109577
Projected Conversions		611000
Sentiment (out of 5)		3.4
Projected Sentiment		4.38
Click through rate (CTR)		
Projected CTR		0.018
Visits		478867
Projected Visits		3820000
% Conversions/Visit		0.231
Projected Conv/Visits		0.16
Co-usage	TBD	
Ave. Google Rank		8.4

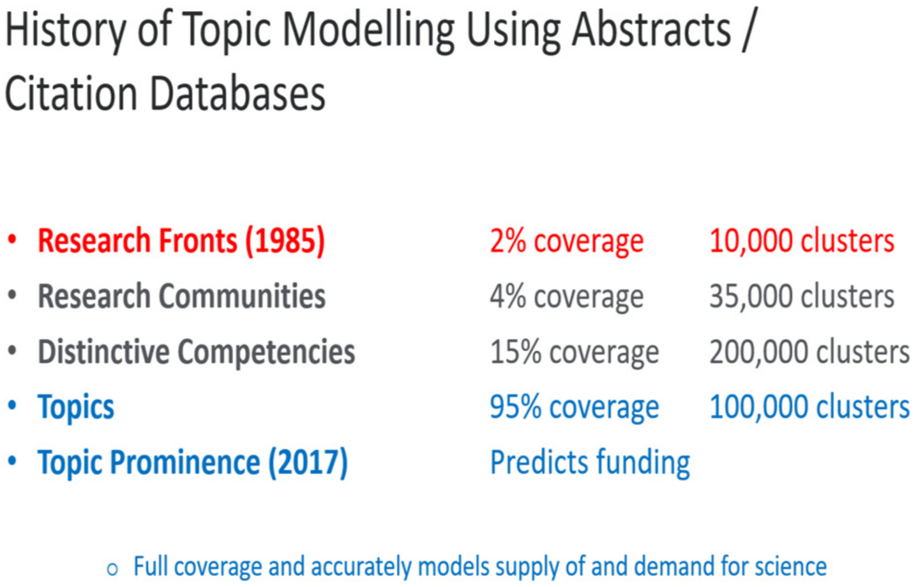
Impact

Last year, ScienceDirect Topic Pages was even a CODiE Award finalist for the category of "Best Artificial Intelligence/Machine Learning Solution".

4. Case study 2: Topics of prominence

The idea of modelling science using computers and citation linkages is not a new idea. Eugene Garfield and Henry Small took this approach back in the 1980s, and their pioneer work played a role in the development of what is now known as "Topic Prominence".

Back in 1985, it took eight weeks to analyze 2% of the citation base in order to come up with a list of hot research topics - namely, those in the Top 1%. Now, we use the entire corpus of literature to create a comprehensive view of the roughly one hundred thousand global research topics currently extant.



History of Topic Modelling Using Abstracts / Citation Databases

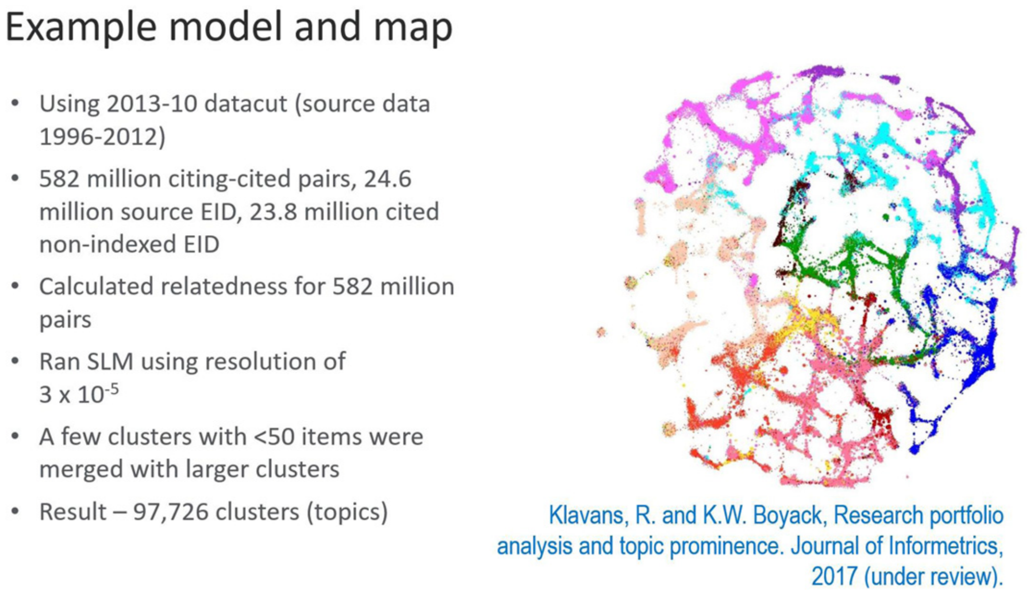
• Research Fronts (1985)	2% coverage	10,000 clusters
• Research Communities	4% coverage	35,000 clusters
• Distinctive Competencies	15% coverage	200,000 clusters
• Topics	95% coverage	100,000 clusters
• Topic Prominence (2017)	Predicts funding	

- Full coverage and accurately models supply of and demand for science

When we think about modeling research, there are a few key design principles to consider: full coverage, the right level of topic granularity, accuracy of topics that contain the right papers, and the stability of topics over time.

Based on researcher needs and aimed at portfolio analysis, we chose to identify roughly one hundred thousand topics in science using direct citation on citation linkages (including those to cited non-indexed items) in the full Scopus database.

As I mentioned previously, it is important to use the full corpus of literature when modeling. Elsevier's Topics of Prominence (TOP) tool does this. I won't go into too much detail here, but the following figure should give you an idea of the level of calculation required for such an analysis - calculation of over half a billion cited-citing pairs - to yield around ninety-seven thousand specific research topics (topic clusters) across every field of research, from basic science to highly applied fields related to manufacturing and commercial technologies.

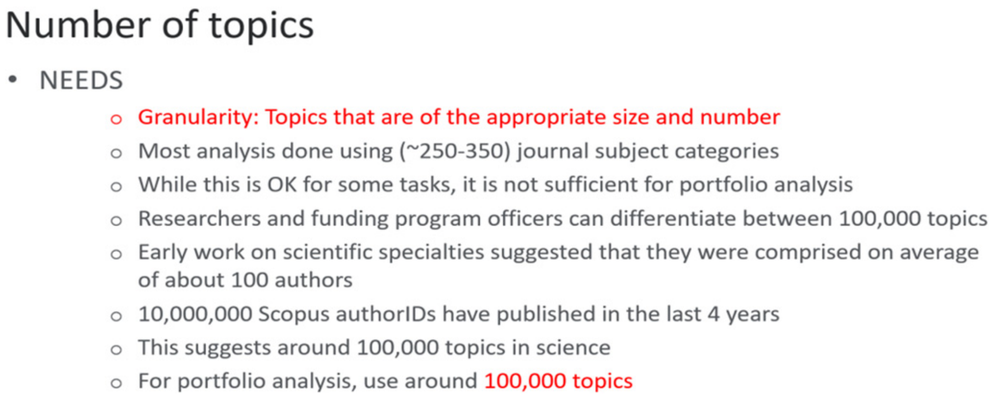


Example model and map

- Using 2013-10 datacut (source data 1996-2012)
- 582 million citing-cited pairs, 24.6 million source EID, 23.8 million cited non-indexed EID
- Calculated relatedness for 582 million pairs
- Ran SLM using resolution of 3×10^{-5}
- A few clusters with <50 items were merged with larger clusters
- Result – 97,726 clusters (topics)

Klavans, R. and K.W. Boyack, Research portfolio analysis and topic prominence. Journal of Informetrics, 2017 (under review).

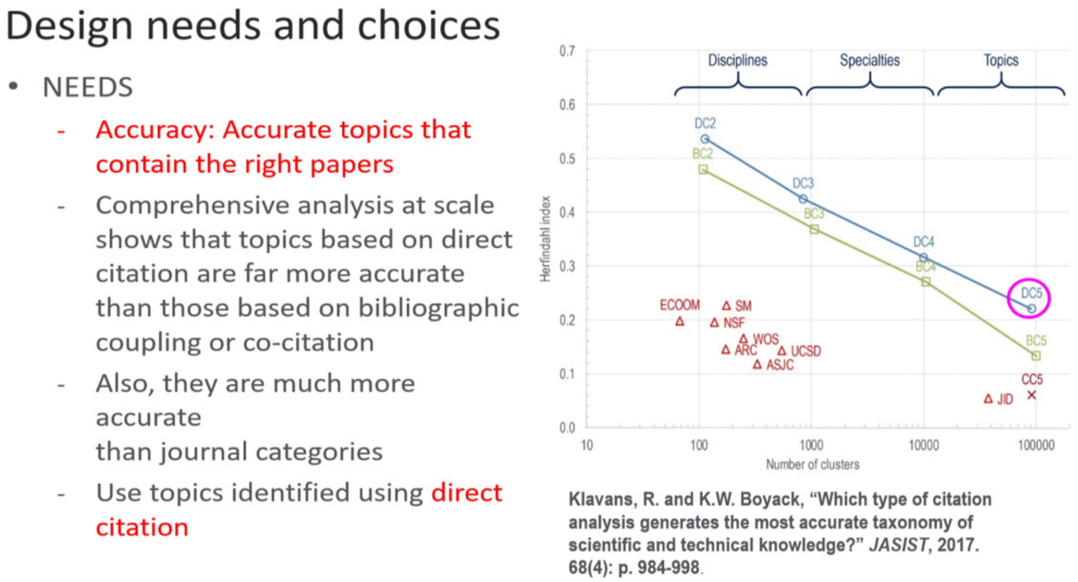
To determine the right level of granularity, prior research and expert interviews were used to look at the approximate size of scientific research questions.



Number of topics

- NEEDS
  - Granularity: Topics that are of the appropriate size and number
  - Most analysis done using (~250-350) journal subject categories
  - While this is OK for some tasks, it is not sufficient for portfolio analysis
  - Researchers and funding program officers can differentiate between 100,000 topics
  - Early work on scientific specialties suggested that they were comprised on average of about 100 authors
  - 10,000,000 Scopus authorIDs have published in the last 4 years
  - This suggests around 100,000 topics in science
  - For portfolio analysis, use around 100,000 topics

Perhaps one of the most important elements in any model is how accurately that model captures scientific activity. Using direct citation analysis, we have a higher accuracy at 105 scale than most of the current classification schema have using around one hundred to three hundred categories. No pre-existing categories are assumed, unlike journal-based classification schema, and the clusters are calculated from the bottom up. These clusters are small enough so that we can see manufacturing-oriented research in specific geographies, and multiple topics around a single larger specialty or discipline. This also captures interdisciplinary research in a fundamentally different way.

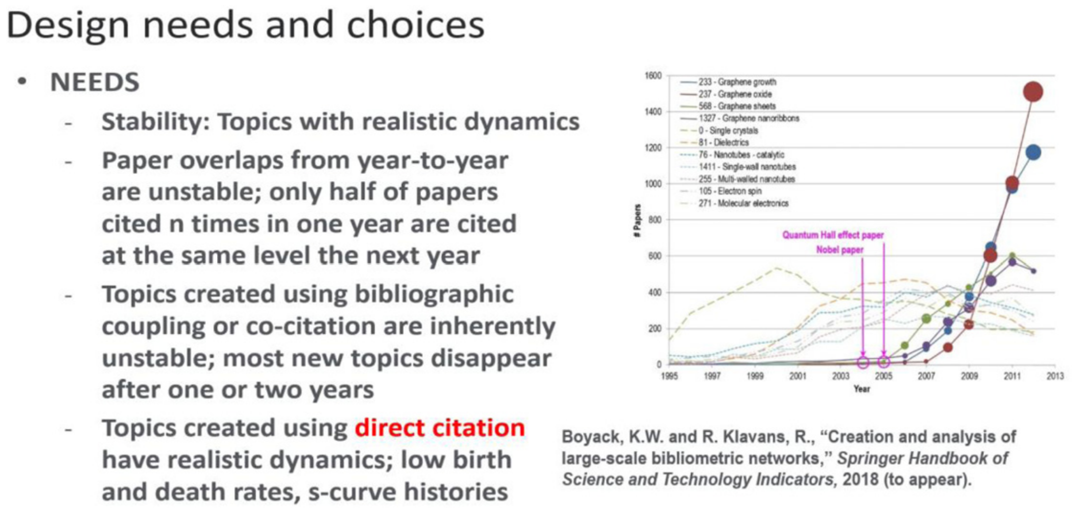


Design needs and choices

- NEEDS
  - Accuracy: Accurate topics that contain the right papers
  - Comprehensive analysis at scale shows that topics based on direct citation are far more accurate than those based on bibliographic coupling or co-citation
  - Also, they are much more accurate than journal categories
  - Use topics identified using direct citation

Klavans, R. and K.W. Boyack, "Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?" *JASIST*, 2017. 68(4): p. 984-998.

Finally, the topics need to be stable year-to-year for long-term trend analysis. The model also meets this criterion. Only about five hundred new clusters per year are born, and older ones persist. Indeed, we can see the growth dynamics of specific topics as they develop. Below are four topics related to graphene, and one can trace the inflection points to papers by Andre Geim and his colleagues at the University of Manchester.

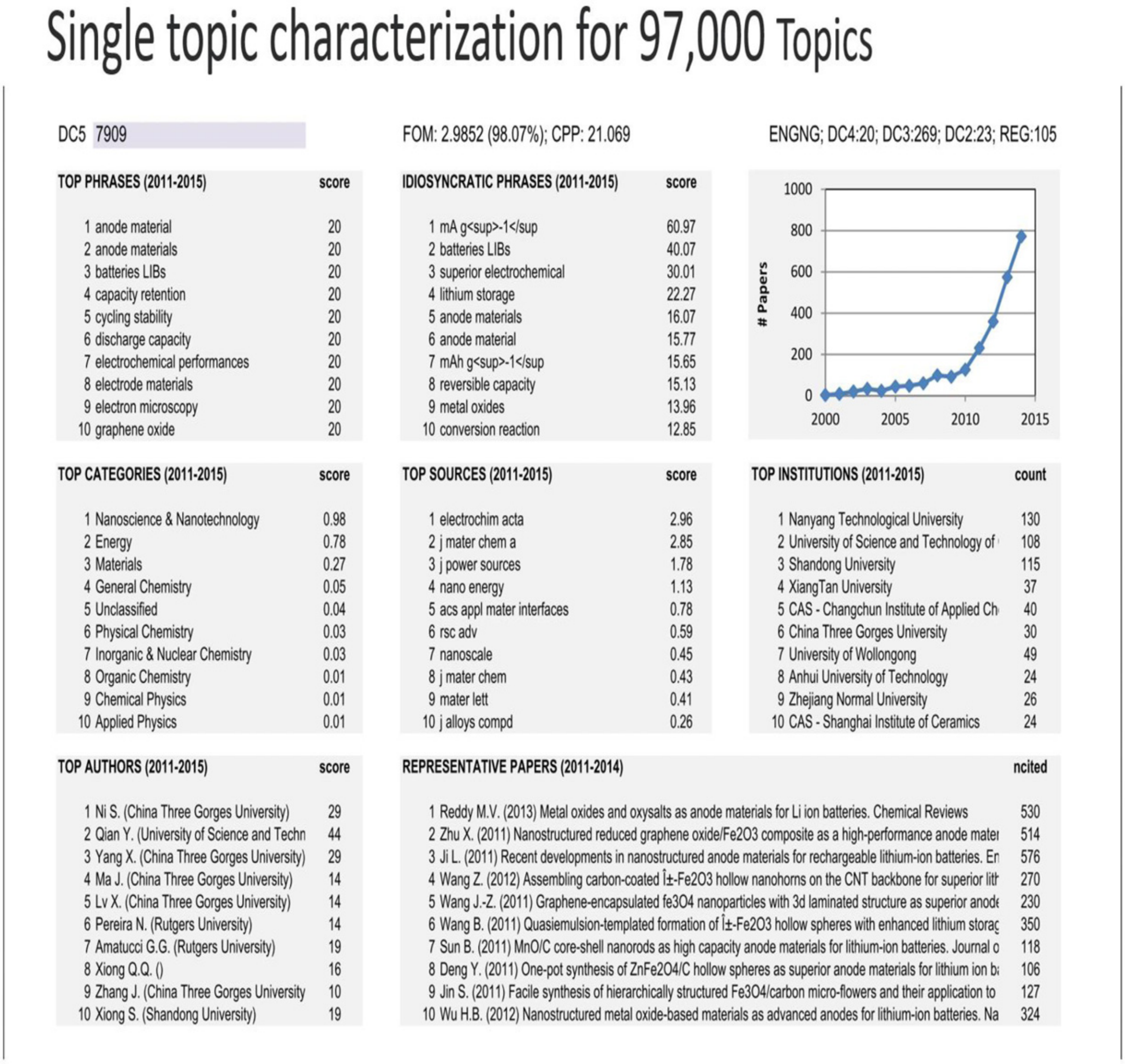


Design needs and choices

- NEEDS
  - Stability: Topics with realistic dynamics
  - Paper overlaps from year-to-year are unstable; only half of papers cited n times in one year are cited at the same level the next year
  - Topics created using bibliographic coupling or co-citation are inherently unstable; most new topics disappear after one or two years
  - Topics created using direct citation have realistic dynamics; low birth and death rates, s-curve histories

Boyack, K.W. and R. Klavans, R., "Creation and analysis of large-scale bibliometric networks," *Springer Handbook of Science and Technology Indicators*, 2018 (to appear).

Another primary benefit to this approach is the richness of detail per topic. Leading authors, institutions, semantic terms, journals, and representative papers are generated for one hundred and five topics. We see the top idiosyncratic phrases, top authors, leading institutions, and the different fields from which this particular interdisciplinary topic is drawn. Using semantic analysis, we can even identify the most representative papers for this topic.



Single topic characterization for 97,000 Topics

DC5 7909

FOM: 2.9852 (98.07%); CPP: 21.069

ENGNG; DC4:20; DC3:269; DC2:23; REG:105

TOP PHRASES (2011-2015)	score
1 anode material	20
2 anode materials	20
3 batteries LIBs	20
4 capacity retention	20
5 cycling stability	20
6 discharge capacity	20
7 electrochemical performances	20
8 electrode materials	20
9 electron microscopy	20
10 graphene oxide	20

IDIOSYNCRATIC PHRASES (2011-2015)	score
1 mA g<sup>-1</sup	60.97
2 batteries LIBs	40.07
3 superior electrochemical	30.01
4 lithium storage	22.27
5 anode materials	16.07
6 anode material	15.77
7 mAh g<sup>-1</sup	15.65
8 reversible capacity	15.13
9 metal oxides	13.96
10 conversion reaction	12.85

TOP CATEGORIES (2011-2015)	score
1 Nanoscience & Nanotechnology	0.98
2 Energy	0.78
3 Materials	0.27
4 General Chemistry	0.05
5 Unclassified	0.04
6 Physical Chemistry	0.03
7 Inorganic & Nuclear Chemistry	0.03
8 Organic Chemistry	0.01
9 Chemical Physics	0.01
10 Applied Physics	0.01

TOP SOURCES (2011-2015)	score
1 electrochim acta	2.96
2 j mater chem a	2.85
3 j power sources	1.78
4 nano energy	1.13
5 acs appl mater interfaces	0.78
6 rsc adv	0.59
7 nanoscale	0.45
8 j mater chem	0.43
9 mater lett	0.41
10 j alloys compd	0.26

TOP INSTITUTIONS (2011-2015)	count
1 Nanyang Technological University	130
2 University of Science and Technology of	108
3 Shandong University	115
4 XiangTan University	37
5 CAS - Changchun Institute of Applied Ch	40
6 China Three Gorges University	30
7 University of Wollongong	49
8 Anhui University of Technology	24
9 Zhejiang Normal University	26
10 CAS - Shanghai Institute of Ceramics	24

TOP AUTHORS (2011-2015)	score
1 Ni S. (China Three Gorges University)	29
2 Qian Y. (University of Science and Techn	44
3 Yang X. (China Three Gorges University)	29
4 Ma J. (China Three Gorges University)	14
5 Lv X. (China Three Gorges University)	14
6 Pereira N. (Rutgers University)	14
7 Amatucci G.G. (Rutgers University)	19
8 Xiong Q.Q. ()	16
9 Zhang J. (China Three Gorges University	10
10 Xiong S. (Shandong University)	19

REPRESENTATIVE PAPERS (2011-2014)	ncited
1 Reddy M.V. (2013) Metal oxides and oxysalts as anode materials for Li ion batteries. Chemical Reviews	530
2 Zhu X. (2011) Nanostructured reduced graphene oxide/Fe2O3 composite as a high-performance anode mater	514
3 Ji L. (2011) Recent developments in nanostructured anode materials for rechargeable lithium-ion batteries. En	576
4 Wang Z. (2012) Assembling carbon-coated Î±-Fe2O3 hollow nanohorns on the CNT backbone for superior lith	270
5 Wang J.-Z. (2011) Graphene-encapsulated fe3O4 nanoparticles with 3d laminated structure as superior anode	230
6 Wang B. (2011) Quasiemulsion-templated formation of Î±-Fe2O3 hollow spheres with enhanced lithium storag	350
7 Sun B. (2011) MnO/C core-shell nanorods as high capacity anode materials for lithium-ion batteries. Journal o	118
8 Deng Y. (2011) One-pot synthesis of ZnFe2O4/C hollow spheres as superior anode materials for lithium ion ba	106
9 Jin S. (2011) Facile synthesis of hierarchically structured Fe3O4/carbon micro-flowers and their application to	127
10 Wu H.B. (2012) Nanostructured metal oxide-based materials as advanced anodes for lithium-ion batteries. Na	324

To give a quick flavor of how this works, let's look at the U.S. Elsevier's SciVal tracks over eight thousand institutions and countries, including major corporates that publish R&D, government labs such as Los Alamos or CNRS, and over seven thousand universities.

Below is 2012–2016 data from the U.S. Looking at the U.S.' output as a country, we can see that it conducts research in every field and has strengths across the board, although it has a higher proportion of biochemistry, genetics, and biomedical research as a percentage of all output than do most other countries.

5. Overall research performance: United States country – output (2012–2016)

Publications:	3,210,189
Citations:	24,767,636
Authors:	2,560,062
Citation/Publication:	7.7
Field-weighted Citation Impact:	1.47

Research Disciplines:

Medicine: 21.9%
Biochem, Genetics, Mol. Biology: 8.7%
Engineering: 8.3%
Social Science: 6.7%
Physics & Astronomy: 5.8%
Computer Science: 5.6%
Agriculture & Biological Science: 4.4%
Material Science: 4.1%
Chemistry: 3.4%
Arts & Humanities: 3.3%
Mathematics: 3.2%
Environmental Science: 2.8%
Earth & Planetary Sciences: 2.8%
Psychology: 2.5%
Neuroscience: 2.2%
Other: 14.5%

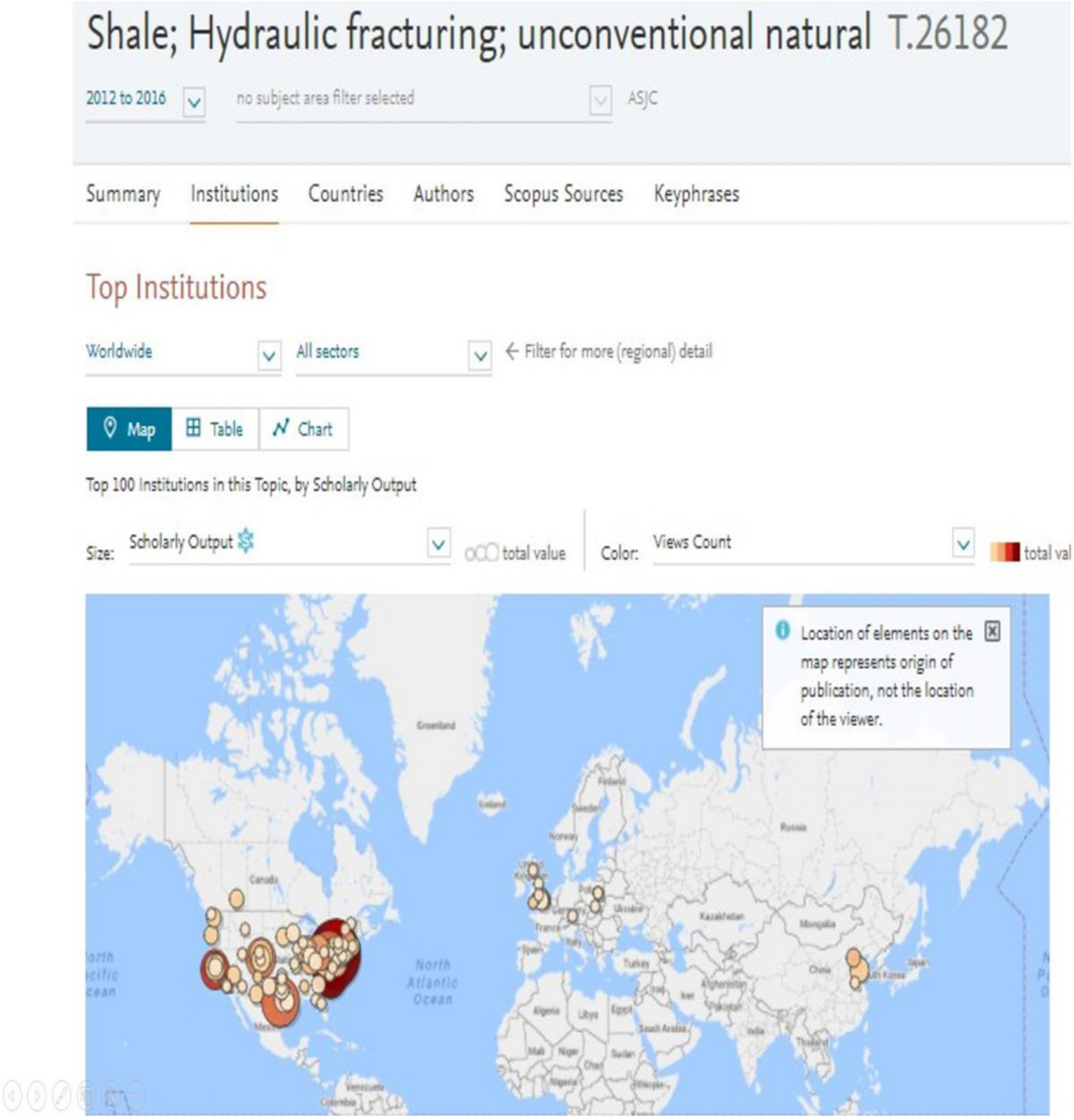
Looking at a few of the U.S.' most prominent topics, we can see that these correlate with areas of research strength. Many topics can easily be linked with specific technologies and even manufacturing capabilities, such as graphene growth and solar cell heterojunctions. Between 2012 and 2017, the U.S. contributed to more than ninety thousand of the approximately ninety-seven thousand topics pursued globally.

6. United States: Researchers in the USA have contributed to 90,988 topics between 2012 to 2017

Topic	Scholarly output	Publication share (in the US)	Field-weighted citation impact	Prominence percentile worldwide
Analgesics, Opioids, Prescriptions (T.248)	1,989	69.59%	2.53	99.762
Brain, Magnetic Resonance Imaging (T.219)	1,689	47.19%	2.97	99.940
Genome, RNA (T.456)	1,829	45.49%	5.94	99.998
Planet (T.131)	1,545	62.15%	2.8	99.740

Again, we see the striking correlation between research activity in a topic and its successful commercial application. The U.S. is almost entirely dominant in this area of applied geology.

7. U.S. topics of prominence - shale; hydraulic fracturing



Shale; Hydraulic fracturing; unconventional natural T.26182

2012 to 2016

no subject area filter selected ASJC

Top Institutions

Worldwide

All sectors

← Filter for more (regional) detail

Top 100 Institutions in this Topic, by Scholarly Output

Size: Scholarly Output

total value

Color: Views Count

total val

Location of elements on the map represents origin of publication, not the location of the viewer.

It should be noted that not all topics dominated by the U.S. are basic science topics or removed from commercial applications.

To summarize, we have accurately modeled approximately one hundred thousand topics in all areas of science and social science, calculated an indicator that measures the current growth momentum of a topic and correlates it with funding, and can help institutions make more informed decisions about their investments in research.

8. Conclusion

Despite all the progress academic publishers and information providers have made, it is important to note two things: first, that there is a long way to go before we reach a point where AI has certain abilities that are still uniquely human, such as the ability to generate a research hypothesis; and there is no conclusive evidence on whether or not AI will either wholly benefit or wholly harm humans. According to Ray Kurzweil, thanks to AI, by 2045 “we will have multiplied the human intelligence of our civilization a billion-fold” [6]. But according to Stephen Hawking, “The development of full AI could spell the end of the human race” [7].

The second is that as in any industry, there has to be more discussion about how to create the necessary checks and balances to ensure that new AI-based technologies are built with representative datasets and with as little bias as possible. By allowing these technologies to develop without agreeing on a guiding philosophy now, we risk a world in which humans cannot understand how AI draws its conclusions, or set limits on the technology’s power.

There is still a lot of work to be done, and it is important for researchers and research solutions providers alike to monitor and analyze emerging AI scholarship - which will ultimately help us predict trends and preempt potential problems. For more information on the state of global AI research, I encourage you to explore Elsevier’s recent AI report [8] - which explains the terminology around, and brings clarity to, the global AI research landscape. For more information on Elsevier’s research and health solutions, please visit us at https://www.elsevier.com/.

About the Author

Ann Gabriel, Senior Vice President, Global Strategic Networks at Elsevier, along with Elsevier’s global team, engages with key stakeholders across the research enterprise to establish strategic collaborations and to use analytics and data to address societal challenges in the area of sustainability, diversity and inclusion, and open science. Prior to her current role, she held a variety of positions at the forefront of scholarly communication - most recently as Elsevier’s Publishing Director for journals in Computer Science and Engineering, as well as electronic product development for Elsevier’s ScienceDirect platform. She was previously with Cambridge University Press. Ann represents Elsevier on several STM (Scientific, Technical & Medical) industry committees, including the CHORUS Board and RA21 (Resource Access for the 21st Century), each of which has a mission to enhance access to scientific data and publication. Ann holds a master’s degree in communications from the University of Pennsylvania, Philadelphia, PA, USA.

References

[1] S. Lynch, Andrew Ng: Why AI is the new electricity, *Insights by Stanford Business*, March 11, 2017, available at: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity, accessed September 28, 2019.

[2] Congressional Research Service, Global Research and Development Expenditures: Fact Sheet, September 19, 2019, available at: https://fas.org/sgp/crs/misc/R44283.pdf, accessed September 28, 2019.

[3] T. Sullivan, A tough road: Cost to develop one new drug is $2.6 billion; approval rate for drugs entering clinical development is less than 12%, *Policy & Medicine* (2019), available at: https://www.policymed.com/2014/12/a-tough-road-cost-to-develop-one-new-drug-is-26-billion-approval-rate-for-drugs-entering-clinical-de.html, accessed September 28, 2019.

[4] J. Doebele, Another Look, *Forbes*, December 14, 1998, available at: https://www.forbes.com/global/1998/1214/0119009a.html#16d29a8bb560, accessed September 28, 2019.

[5] A. Michael, AI and scholarly communications, *Scholarly Kitchen*, available at: https://scholarlykitchen.sspnet.org/2019/04/25/ask-chefs-ai-scholarly-communications/, accessed September 28, 2019.

[6] C. Reedy, Kurzweil claims that the singularity will happen by 2045: Get ready for humanity 2.0, *Futurism* (2017), available at: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045, accessed September 28, 2019.

[7] Stephen Hawking warns artificial intelligence may supersede humans, disrupt economy, *AIReligion*, available at: https://aireligion.org/?p=368, accessed September 28, 2019.

[8] ArtificiaI Intelligence: How knowledge is created, transferred, and used Trends in China, Europe, and the United States, available at: https://www.elsevier.com/research-intelligence/resource-library/ai-report, accessed September 28, 2019.