Aalto University
School of Business

Data, text and web mining & Bibliometrics

Management IS course

March 19, 2024

Johanna Bragge
Principal University Lecturer of ISS
Dept. of Information & Service Management ISM

Learning objectives

Part 1:

Know how data and text mining differ from each other

Know text-mining business applications

Know the basics of text-mining and visualization

Know how to do “research profiling”

Know what is web mining and sentiment analysis

Part 2:

Hands-on demo of VOSviewer with literature data exported from Scopus

A!

Data and text mining

Data and Text mining defined

Data mining is a process that uses **statistical**, **mathematical**, **AI** and **ML (Machine Learning)** techniques to extract and identify useful information from large databases.

Text from documents, e-communications, and e-commerce activities can also be mined.

> *"**Text mining** or **text analytics** are broad umbrella terms describing a range of technologies for analyzing and processing semi-structured and unstructured text data."* (Delen et al. 2012, p. 30)

A!

Source: Delen et al. 2012 Practical text mining and statistical analysis for non-structured text data applications, Academic Press, Elsevier, p.31

Examples of data mining applications for identifying business opportunities

Retailing and sales. Predicting sales, determining correct inventory levels and distribution schedules among outlets, and loss prevention.

Banking. Forecasting levels of bad loans and fraudulent credit card use and which kinds of customers will best respond to new loan offers.

Manufacturing and production. Predicting machinery failures; finding key factors that control optimization of manufacturing capacity.

Healthcare. Developing better insights on symptoms and their causes and how to provide personalized treatments.

Broadcasting. Predicting which programs are best to air during prime time and how to maximize returns by interjecting advertisements.

Marketing. Classifying customer demographics that can be used to predict which customers will respond to a mailing or Internet banners or buy a particular product as well as to predict other consumer behavior.

A!

Text-Mining

Textual data comprises **up to 80% of all information** collected – important to utilize it too!
Text-mining helps organizations find the “hidden” content of documents, e.g. useful **relationships / patterns, customer sentiments / opinions** etc.

Content that is mined include unstructured data from **documents**, text from **emails**, **social media** and **log data from Internet**, among others.

May be major source of **competitive advantage**!
see e.g. Tech intelligence article at https://www.thevantagepoint.com/resources/articles/CI_May-Jun_05_Brenner.pdf

See also: “Text Analytics. Bridging the gap between quantitative and qualitative information”
http://www.informs.org/ORMS-Today/Public-Articles/June-Volume-39-Number-3/Text-analytics

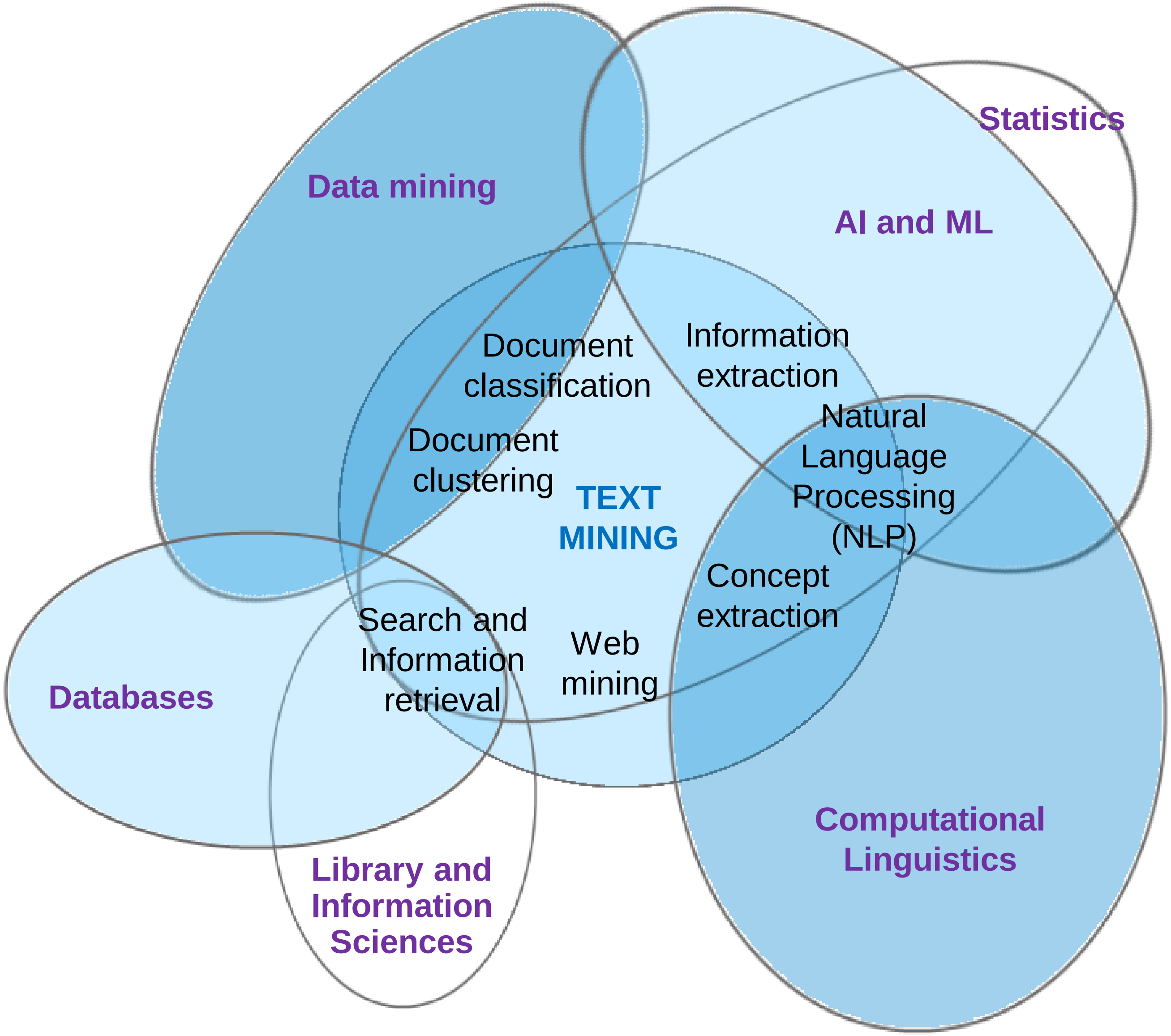
and “Text Analytics: Your Customers are Talking About You”
https://www.cio.com/article/276843/customer-relationship-management-text-analytics-your-customers-are-talking-about-you.html

Intersection of Text Mining and six related fields

Seven technologies or practice areas at the intersections

*"The unifying theme behind each of these 7 technologies is **the need to "turn text into numbers"** so that powerful analytical algorithms can be applied to large document databases."*

(Delen et al. 2012, p. 29)

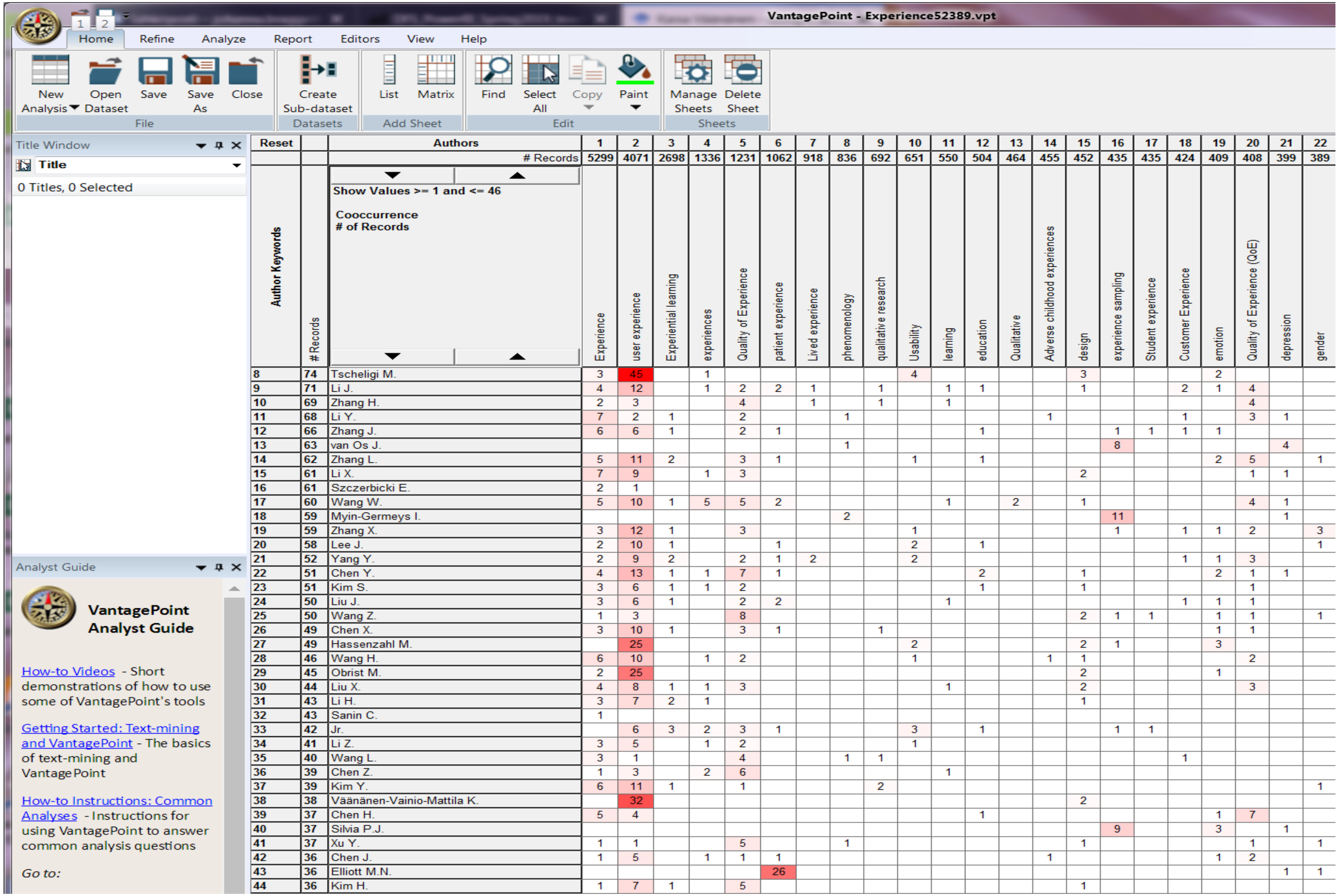


A!

Delen et al. 2012 Practical text mining and statistical analysis for non-structured text data applications, Academic Press, Elsevier, p.31

Example from turning text into numbers: Matrix of Authors & Keywords

data downloaded from Scopus in CSV format (tool used: VantagePoint by SearchTechnology Inc.)



Reset		Authors	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		# Records	5299	4071	2698	1336	1231	1062	918	836	692	651	550	504	464	455	452	435	435	424	409	408	399	389
Author Keywords	# Records		Experience	user experience	Experiential learning	experiences	Quality of Experience	patient experience	Lived experience	phenomenology	qualitative research	Usability	learning	education	Qualitative	Adverse childhood experiences	design	experience sampling	Student experience	Customer Experience	emotion	Quality of Experience (QoE)	depression	gender
8	74	Tscheligi M.	3	45		1						4					3				2			
9	71	Li J.	4	12		1	2	2	1		1		1	1			1			2	1	4		
10	69	Zhang H.	2	3			4		1		1		1									4		
11	68	Li Y.	7	2	1		2			1						1				1		3	1	
12	66	Zhang J.	6	6	1		2	1						1				1	1	1	1			
13	63	van Os J.								1								8					4	
14	62	Zhang L.	5	11	2		3	1				1		1							2	5		1
15	61	Li X.	7	9		1	3										2					1	1	
16	61	Szczerbicki E.	2	1																				
17	60	Wang W.	5	10	1	5	5	2					1		2		1					4	1	
18	59	Myin-Germeys I.								2								11					1	
19	59	Zhang X.	3	12	1		3					1						1		1	1	2		3
20	58	Lee J.	2	10	1			1				2		1										1
21	52	Yang Y.	2	9	2		2	1	2			2								1	1	3		
22	51	Chen Y.	4	13	1	1	7	1						2			1				2	1	1	
23	51	Kim S.	3	6	1	1	2							1			1					1		
24	50	Liu J.	3	6	1		2	2					1							1	1	1		
25	50	Wang Z.	1	3			8										2	1	1		1	1		1
26	49	Chen X.	3	10	1		3	1			1										1	1		
27	49	Hassenzahl M.		25								2					2	1			3			
28	46	Wang H.	6	10		1	2					1				1	1					2		
29	45	Obrist M.	2	25													2				1			
30	44	Liu X.	4	8	1	1	3						1				2					3		
31	43	Li H.	3	7	2	1											1							
32	43	Sanin C.	1																					
33	42	Jr.		6	3	2	3	1				3		1				1	1					
34	41	Li Z.	3	5		1	2					1												
35	40	Wang L.	3	1			4			1	1									1				
36	39	Chen Z.	1	3		2	6						1											
37	39	Kim Y.	6	11	1		1				2													1
38	38	Väänänen-Vainio-Mattila K.		32													2							
39	37	Chen H.	5	4										1							1	7		
40	37	Silvia P.J.																9			3		1	
41	37	Xu Y.	1	1			5			1							1					1		1
42	36	Chen J.	1	5		1	1	1								1					1	2		
43	36	Elliott M.N.						26															1	1
44	36	Kim H.	1	7	1		5										1							

Example of
Natural Language Processing (NLP)

Words in this case from article titles are parsed both as single words and **multi-word phrases*** along with their frequencies (# of instances in the article sample's titles)

*__Multi-word phrases__ (MWEs) refer to phrases that can vary in length and often carry a specific meaning that is not deducible from the individual words.

(Scopus data, tool: VantagePoint)

# Instances	Title (NLP) (Phrases)	Multi Word Phrases
2883	experience	
1068	effects	
1052	role	
937	impact	
725	Learning	
714	user experience	✓
687	effect	
658	Development	
605	study	
600	qualitative study	✓
604	quality	
585	influence	
542	patients	
530	case study	✓
477	children	
464	use	
442	relationship	
411	Case	
372	experiential learning	✓
358	Analysis	
357	research	
325	life	
308	Implications	
297	Evidence	
287	living	
283	perceptions	
277	students	
278	women	
273	practice	
257	adverse childhood experiences	✓
262	time	
237	application	
236	review	
224	challenges	
221	education	
224	care	
215	factors	
214	Knowledge	
210	systematic review	✓
208	attitudes	
206	teaching	
202	China	
192	depression	
188	Assessment	
187	experiential avoidance	✓
188	meaning	

Most popular applications areas of Text Mining

Information extraction

Identification of key phrases and relationships within text.

Topic tracking

Predicting documents of interest to the user, based on user profile and other documents that a user has viewed.

Summarization

Summarizing a document to save time on the part of the reader.

Categorization

Identifying the main themes of a document and then placing the document into a predefined set of categories based on those themes.

Clustering

Grouping similar documents without having a predefined set of categories.

Concept linking

Connects related documents by identifying their shared concepts.

Question answering

Finding the best answer to a given question via knowledge-driven pattern matching.

Source: Sharda, R., Delen, D., Turban, E., Aronson, J., & Liang, T. (2014). Business intelligence and analytics. *System for Decision Support*.

Text-Mining business cases

Benefits of text mining are obvious in **areas**, **where very large amounts of textual data** are being generated, such as

- **law** (court orders),
- **finance** (quarterly reports),
- **management** (annual reports),
- **technology** (patent files),
- **marketing** (customer comments) and
- **academic research** (research articles).

E.g. **customer reviews or complaints** can be used to identify product and service characteristics that need to be developed. Also, **market outreach programs** and **focus groups** generate large amounts of data.
Text mining can be used to **classify and filter junk e-mail**, automatically **prioritize e-mail** based on importance level and also **generate automatic responses**.

Earning calls data from: https://datahub.aalto.fi/en/node/281

Source: Sharda, R., Delen, D., Turban, E., Aronson, J., & Liang, T. (2014). Business intelligence and analytics. *System for Decision Support*.

Text mining is nowadays integrated to many web survey products for analyzing open-ended questions – now with enhanced AI capabilities


WEBROPOL

THE OFFICIAL WEBROPOL BLOG

AT YOUR SURVEYS

Monday, 7 December 2009

Webropol brings Text Mining to its online survey software

Welcome to At Your Surveys!

We have some exciting news for you! As first in the world, Webropol brings Text Mining capabilities to online survey software. Working as part of the Webropol toolset, the Text Minin analysis, classification and group based on occurring themes, keyw background data. Thanks to this can now process and analyse unstructured survey thoroughly, and objectively than ever before.

"It is unfortunate how often one runs into situations whe ask customers for free feedback, but will not do so in th sheet after sheet of answers. The Text Mining solution ways to use unstructured data for business intelligence worth of manual work is done in minutes.", says UK Co Bassi from Webropol.

By comparing textual answers based on other response data managers get an insight into how feedback, sugge depending on e.g. employee satisfaction or customer lif help organisations set up and fine-tune their numeric cu metrics to measure the things that are the most relevan

"The Text Mining solution suits excellently both short su research, especially when you are asking respondents or advice. It is virtually impossible to collect all ideas un in free form", Mukesh points out.

The Text Mining solution works as part of the Webropol can be brought for analysis from external sources as we

Sources: http://atyoursurveys.blogspot.com/2009/12/webropol-brings-text-mining-to-its.html

http://w3.webropol.com/our-product/analyse-and-visualise/

https://webropol.co.uk/webropol-modules/ai-text-analysis/

The manager's agenda

"It goes without saying that the most immediate agenda with respect to Big Data is operational.*

*People responsible for market research, process engineering, pricing, risk, logistics, and other complex functions need to master an entirely new set of **statistical techniques.***

Non-specialist managers need to understand enough about the possibilities and pitfalls of Big Data to translate its output into practical business benefits.

***Data visualization** is emerging as critical **interface** between the specialist and the non-specialist."*

*Text data is one form of Big data

Source: Evans, Philip (2015), "Reinventing the Company in the Digital Age. From Deconstruction to Big Data: How Technology is Reshaping the Corporation", OpenMind / BBVA

A!

Increased importance of visualization literacy

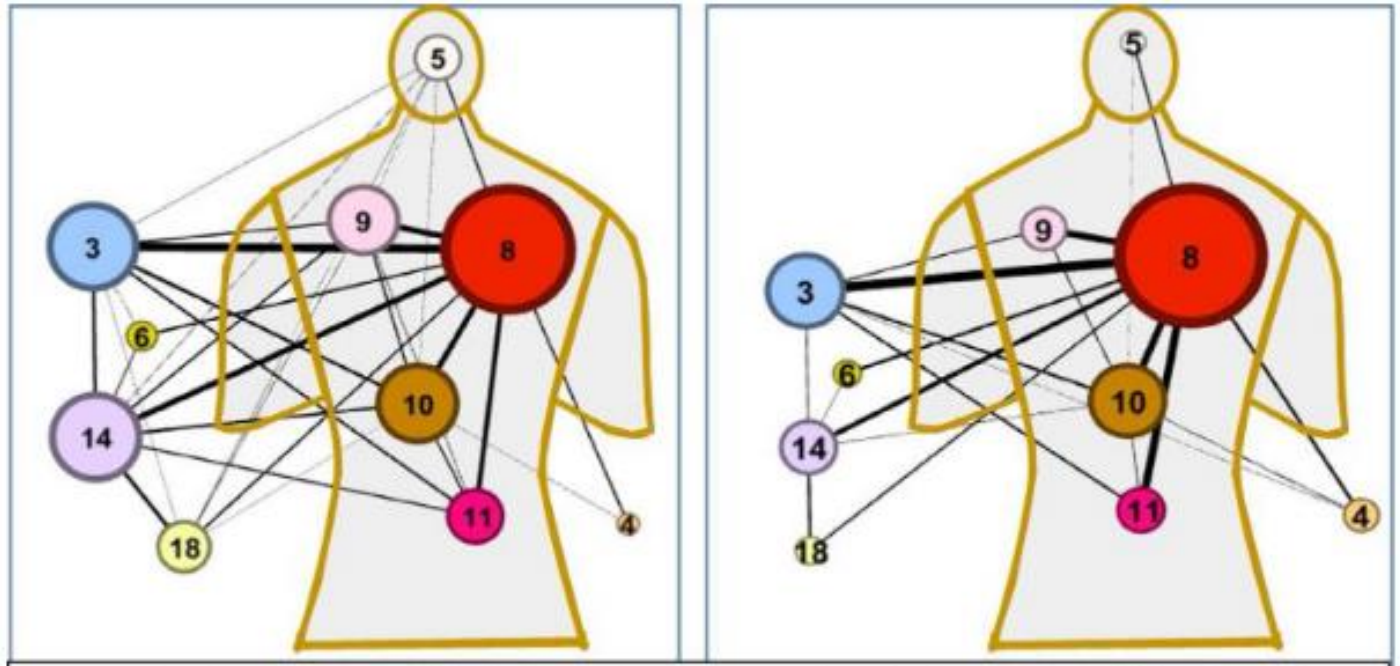
"In the information age, a person's ability to read and make data visualizations is nearly as important as being able to read and write text".

"The amount of data in our world is increasing radically, and the capability to analyze datasets is becoming a key basis for all citizens to be data-literate decision makers."

"We define data visualization literacy as the ability to make meaning from and interpret patterns, trends and correlations in visual representations of data".

Börner, K. et al. (2016), Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors, Information Visualization, 15(3), 198-213. https://journals.sagepub.com/doi/10.1177/1473871615594652

Data mining & viz example: A multimorbidity* network analysis of electronic medical records (EMR) by gender



2a. Female Organ Comorbidity Network   2b. Male Organ Comorbidity Network

1	Infectious and parasitic diseases
2	Neoplasms
3	Endocrine, nutritional and metabolic diseases, and immunity disorders
4	Diseases of the blood and blood-forming organs
5	Mental disorders
6	Diseases of the nervous system
7	Diseases of the sense organs
8	Diseases of the circulatory system
9	Diseases of the respiratory system
10	Diseases of the digestive system
11	Diseases of the genitourinary system
12	Complications of pregnancy, childbirth, and the puerperium
13	Diseases of the skin and subcutaneous tissue
14	Diseases of the musculoskeletal system and connective tissue
15	Congenital anomalies
16	Certain conditions originating in the perinatal period
17[a]	Symptoms, signs, and ill-defined conditions
18	Injury and poisoning

[a] Not considered in the analysis.

The study develops multimorbidity networks for males and females based on ICD-9 codes of diagnoses. **The network comprises diseases connected based on the co-occurrences of diseases** in **22.1 million patient records in the US** (spanning 17 years).
Knowing the relationships between diseases at the network level enhances our understanding about disease associations at the patient population level.

*** multimorbidity = the simultaneous presence of two or more diseases or medical conditions in a patient**

Kalgotra, P., Sharda, R., & Croff, J. M. (2017). Examining health disparities by gender: a multimorbidity network analysis of electronic medical record. *International journal of medical informatics*, *108*, 22-28. Available at https://www.sciencedirect.com/science/article/pii/S138650561730237X

Deloitte data- & text-mining study example on 4651 US and global firms listed in NYSE

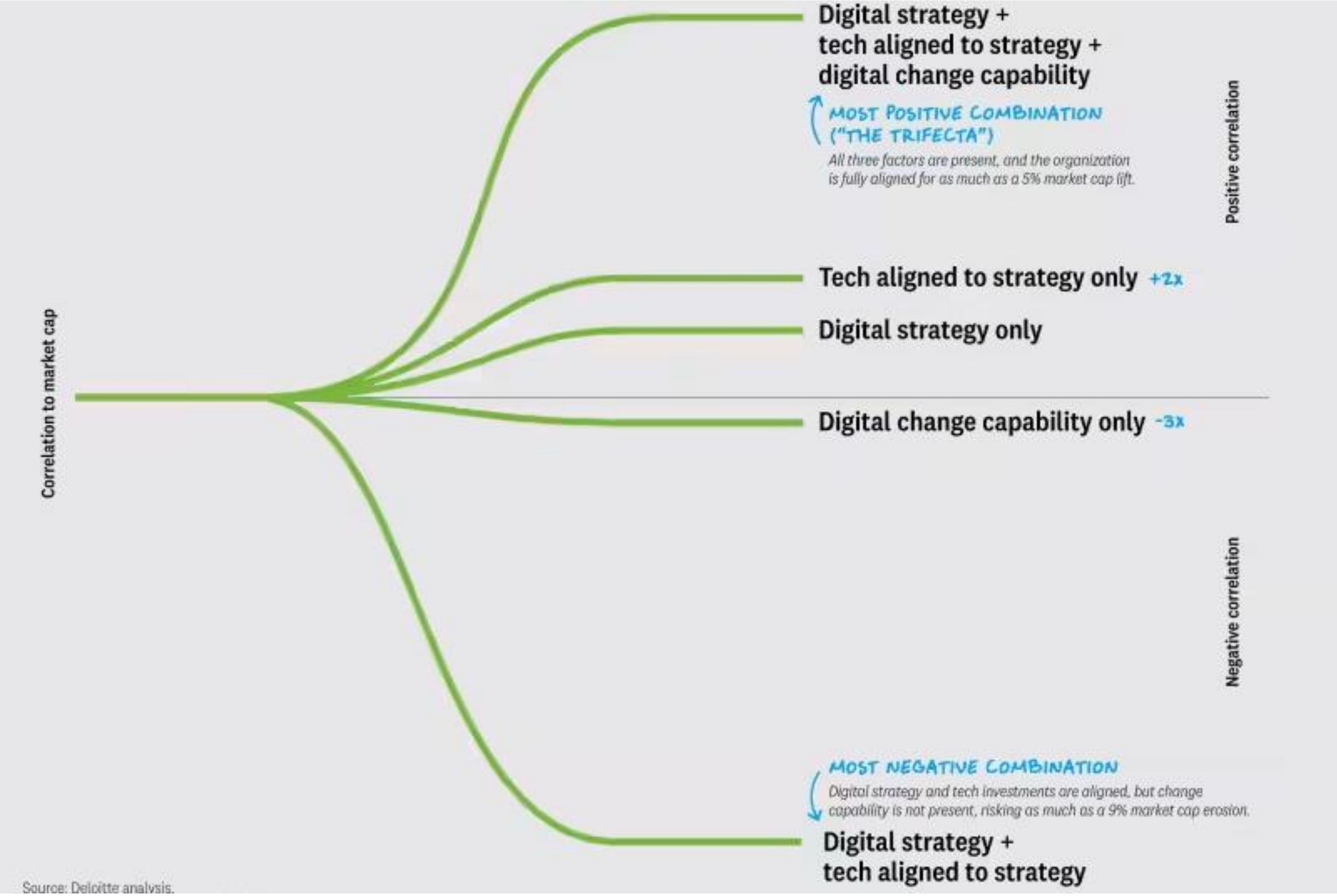
Analyzed financial disclosures to ascertain how companies talked about their digital transformation actions—i.e., how they spoke to
(1) implementing a digital strategy;
(2) their discrete, strategically aligned technology investments;
(3) their efforts to prepare their people and processes for digital transformation.

The link between strategy and action was found the determining factor in a company's ability to derive the most value from its digital transformation. Research showed these actions can increase enterprise value if executed with intent, yet not all actions are created equal.

A!

Source: Smith et al. (2023) "Unleashing value from digital transformation: Paths and pitfalls", available at https://www2.deloitte.com/us/en/insights/topics/digital-transformation/digital-transformation-value-roi.html

Deloitte: Visualizing how digital transformation factors correlate to market capitalization (2)

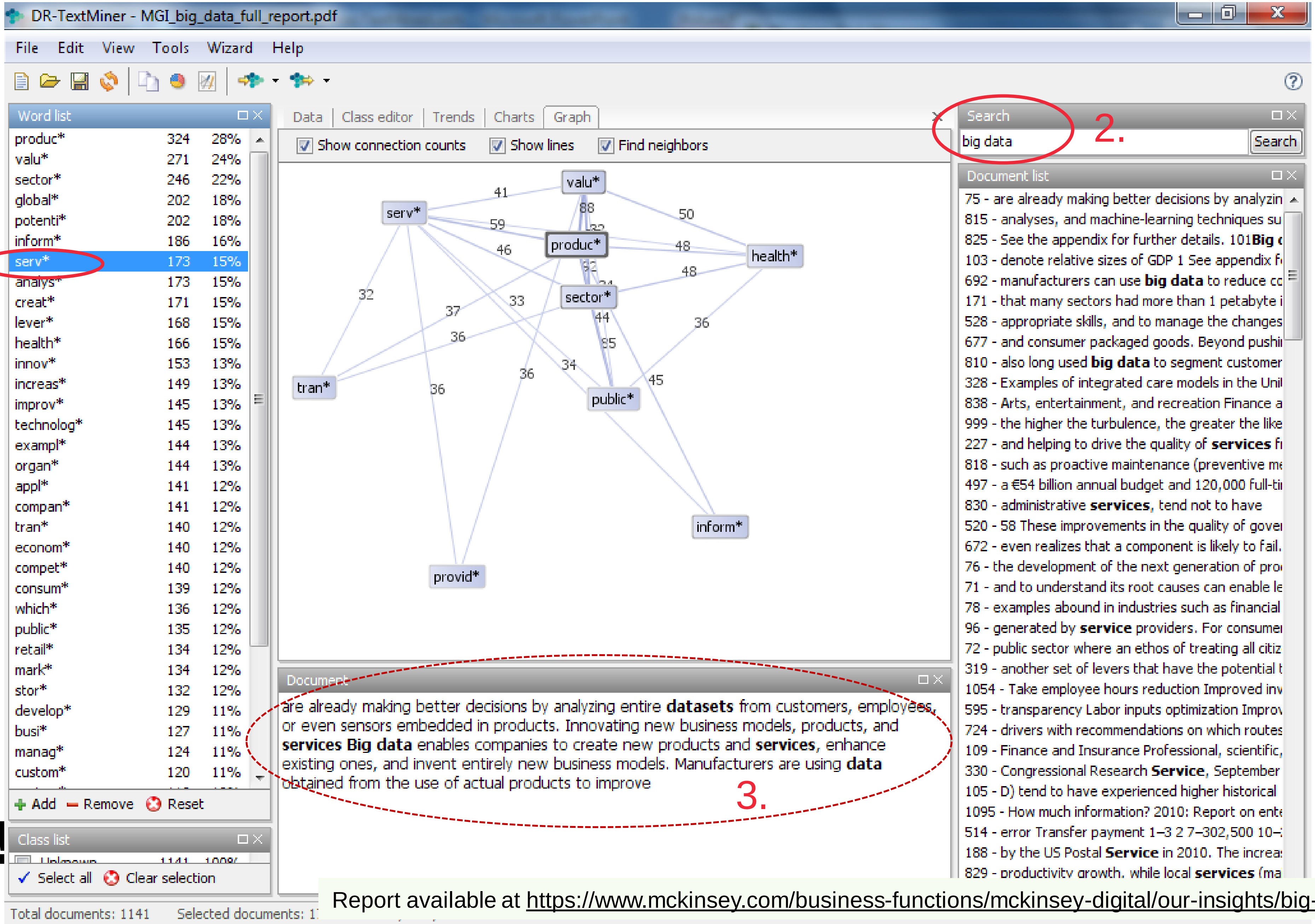


A!

Source: Smith et al. (2023) "Unleashing value from digital transformation: Paths and pitfalls", available at https://www2.deloitte.com/us/en/insights/topics/digital-transformation/digital-transformation-value-roi.html

Text mining & viz example 1 **(mining McKinsey Global Institute’s Big Data report PDF). Tool used: TextMiner by https://www.datarangers.fi)**

1.

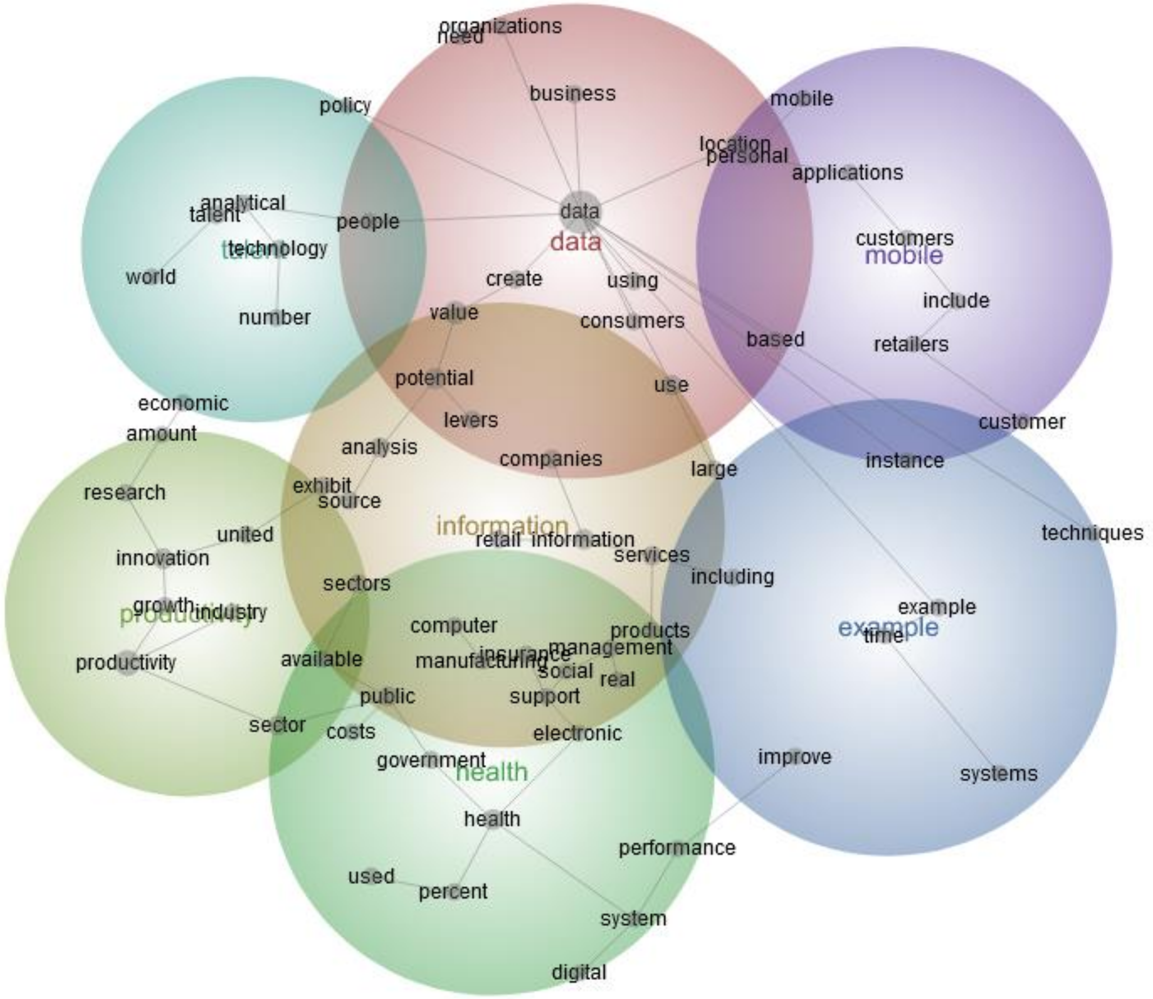


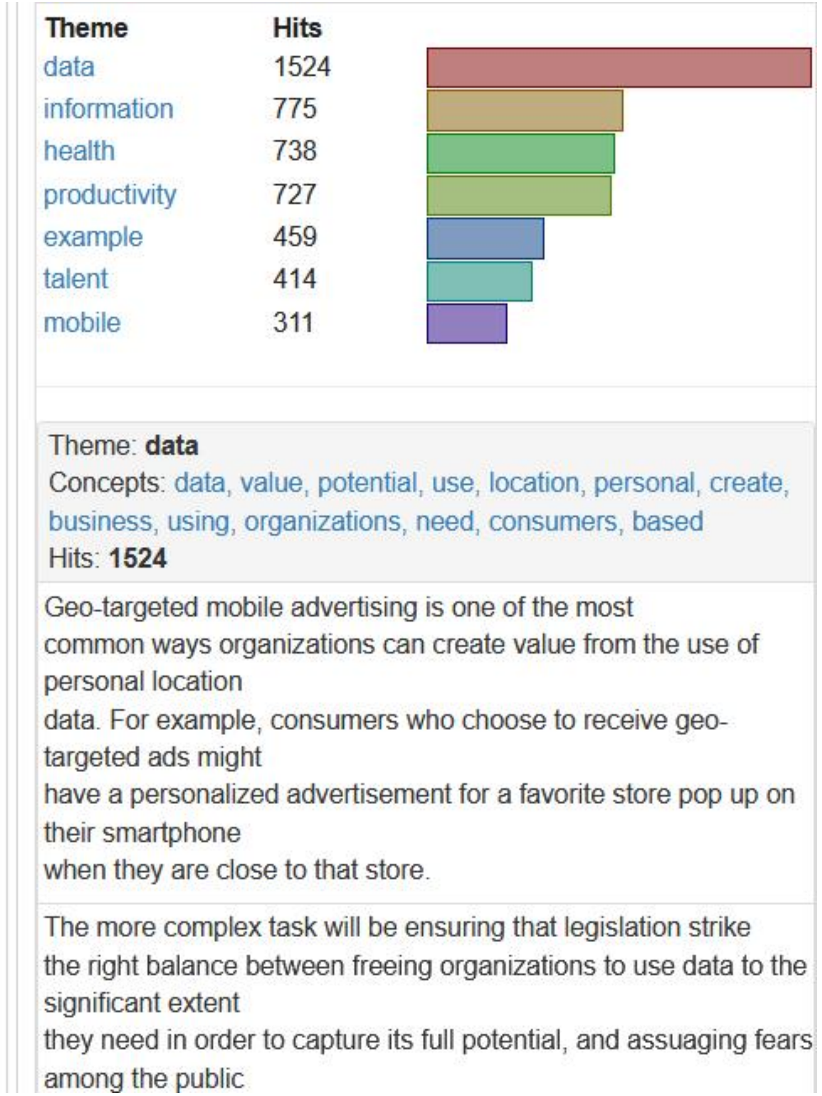
Example of querying content, where **big data** (2.) is mentioned in connection with **services** (1.).

Seeing all results as a sentence list on the right-hand column, and one-by-one at the bottom (3.) pane.

Report available at https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Text mining & viz example 2: Mining the same MGI's Big Data report PDF (tool used: Leximancer's LexiDesktop 5.0)





Theme	Hits
data	1524
information	775
health	738
productivity	727
example	459
talent	414
mobile	311

Theme: **data**
Concepts: data, value, potential, use, location, personal, create, business, using, organizations, need, consumers, based
Hits: **1524**

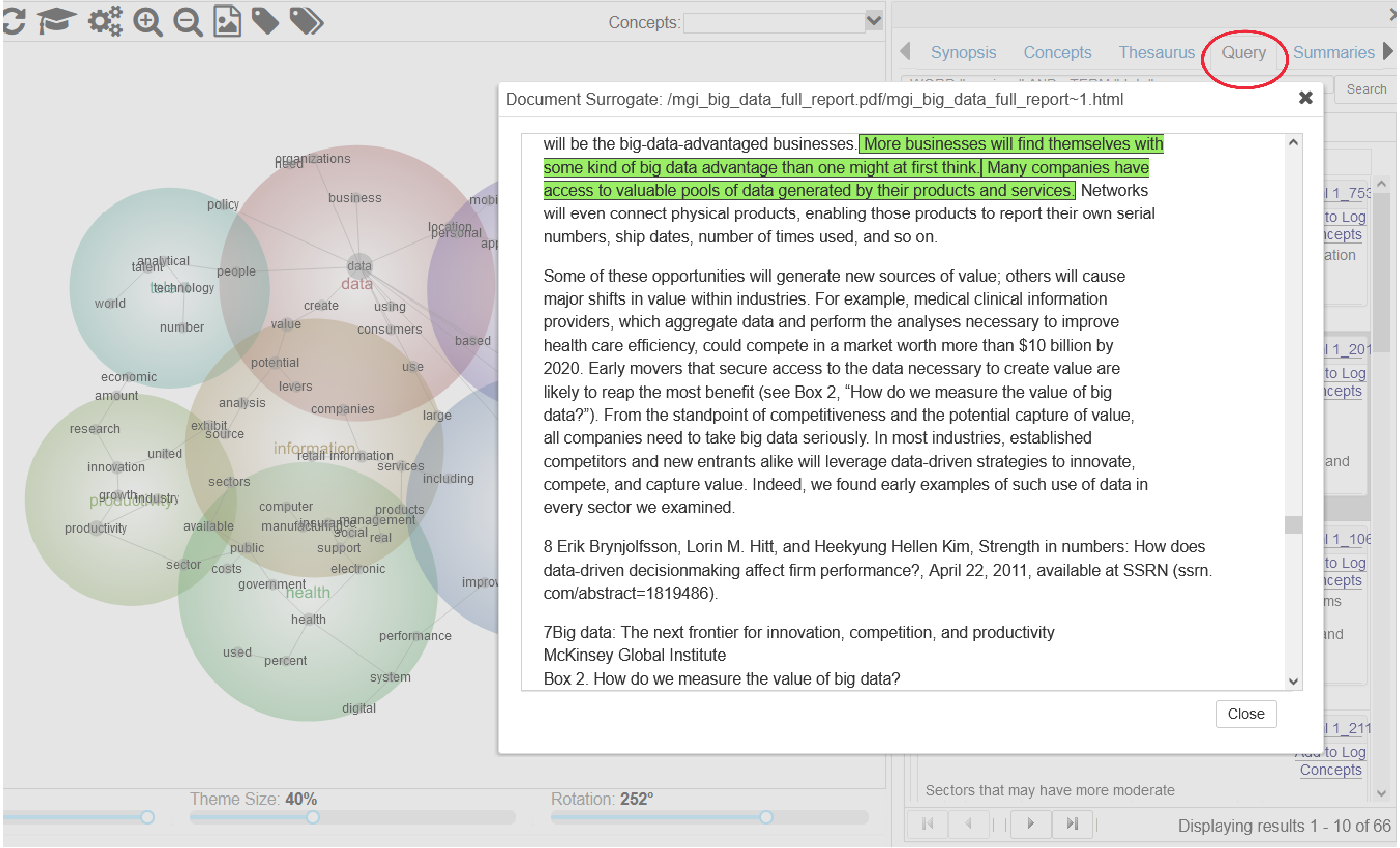
Geo-targeted mobile advertising is one of the most common ways organizations can create value from the use of personal location data. For example, consumers who choose to receive geo-targeted ads might have a personalized advertisement for a favorite store pop up on their smartphone when they are close to that store.

The more complex task will be ensuring that legislation strike the right balance between freeing organizations to use data to the significant extent they need in order to capture its full potential, and assuaging fears among the public

A!

Report PDF available at https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Text mining example 2b: Querying the results for *services* and *big data*



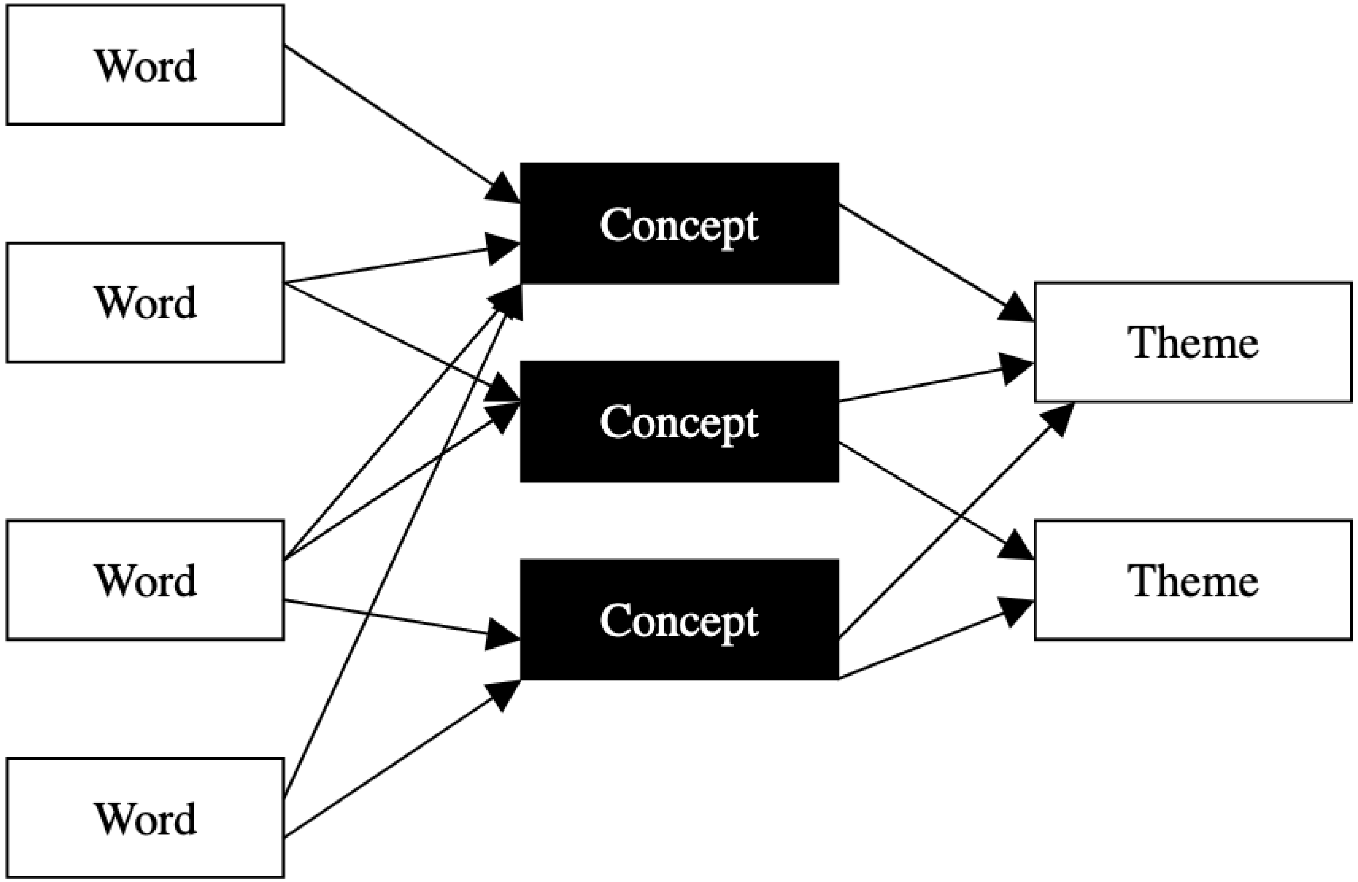
Using the Query function to find out where **services** appear together with **(big) data**

*Hint for the last assignment: See User guide’s **Query section 3.5.4** for details how to build the query sentences!*

https://www.leximancer.com/s/Leximancer-User-Guide-5.pdf

Report PDF available at https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Simplified Model of Leximancer's Semantic Pattern Extraction



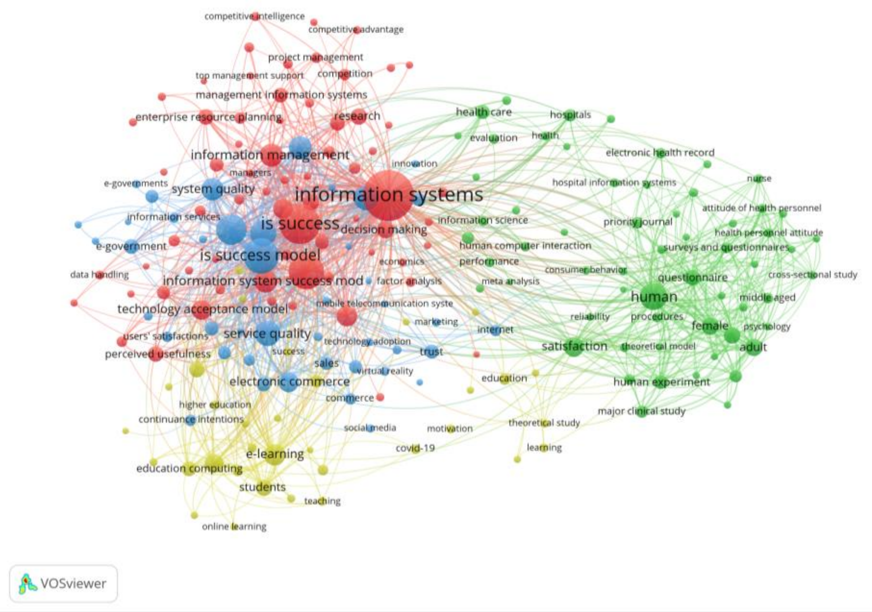
Source: Crofts, K. and Bisman, J. (2010) Interrogating accountability: An illustration of the use of Leximancer software for qualitative data analysis. *Qualitative Research in Accounting & Management* 7, 180–207

Text Visualization Browser https://textvis.lnu.se/

Web-based UI of the **Text Visualization Browser** survey tool. The interaction panel on left allows users to look for specific viz techniques and filter entries with respect to a set of categories (i.e. the taxonomy built by Kucher and Kerren, 2015).
Details are shown by clicking a thumbnail image

Currently contains 440 viz techniques.

Kucher, K., & Kerren, A. (2015, April). Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific visualization symposium (pacificVis)* (pp. 117-121). IEEE.



Research profiling aka bibliometric literature reviews

“Bibliometrics (statistical bibliography) or scientometrics is the application of mathematical and statistical methods to books and other media of communication”

Porter, A.L., Kongthon, A. and J.-C. Lu (2002), "Research Profiling: Improving the Literature Review", *Scientometrics*, 53(3), 351-370.

Hood, W.W. and Wilson, C.S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics, *Scientometrics,* 52(2), 291-314.

Pritchard, A. (1969). Statistical Bibliography or Bibliometrics?, *Journal of Documentation*, 25, 348-349

How I got interested in "research profiling"?

PhD in MS/OR in 1997 (HSE/Aalto)

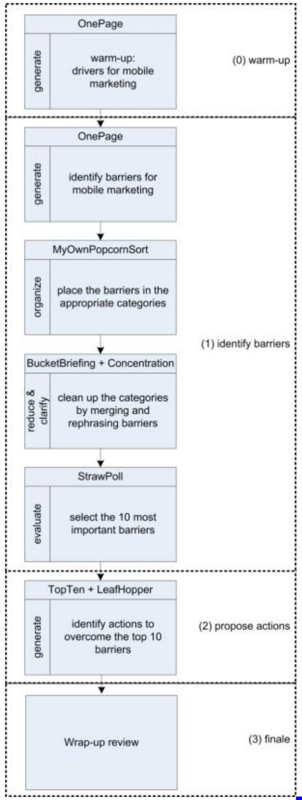
Analyzing and mediating the CO_2/energy taxation dispute in Finland

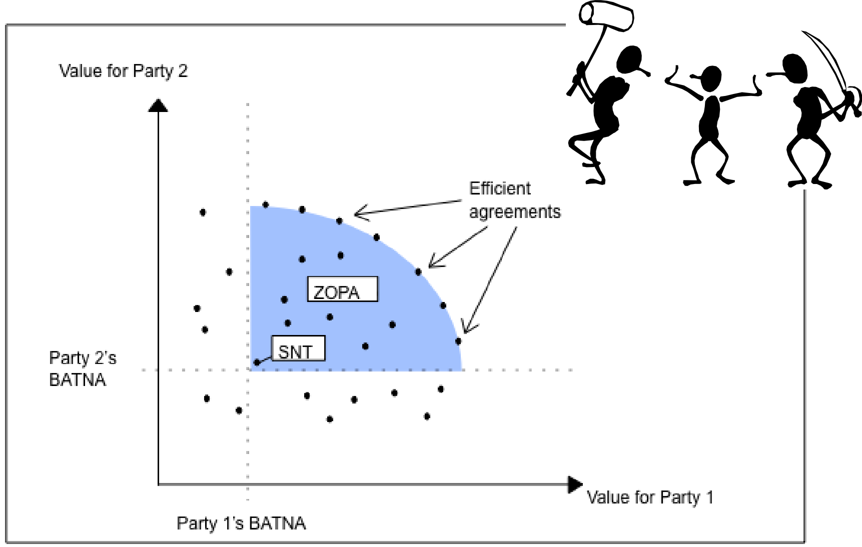
Based on my mediator background, my first task in ISS was to choose a *Group Support System* for teaching

GSS produce a lot of text!!

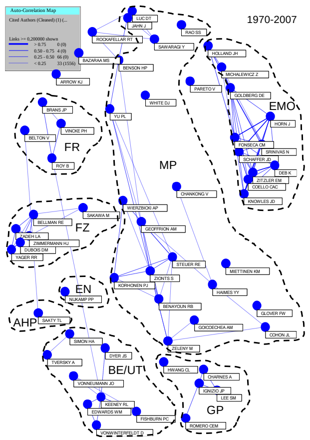
Intensive course in Aalto on text-mining

> Porter's research profiling method utilizing *bibliographic* or patent data

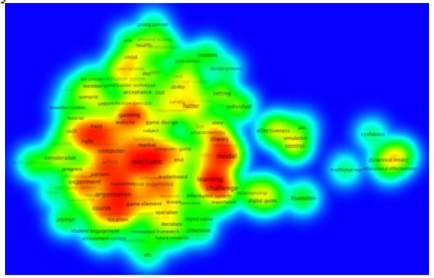


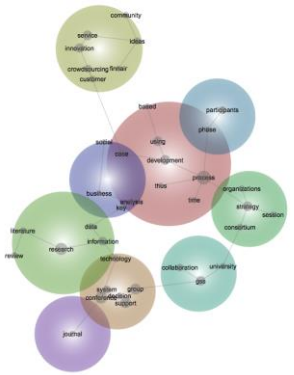


https://www.sciencedirect.com/science/article/pii/S0377221700001740



	1986-1990	1991-1995	1996-2000	2001-2005
distributed-*	15	65	62	63
group-decision-making	10	30	57	58
decision-making	9	28	35	38
face-to-face*	9	23	28	27
consensus-*	6	31	31	16
virtual-*	0	7	32	35
information-technology	5	25	18	15
idea-generation*	7	10	25	13
web-based-*	0	0	14	33
Internet	0	4	25	18
anonymity	5	10	21	8





A!

What is *bibliographic* data?

The data used for research profiling covers typically all *except the full-text* of the publication, i.e. **authors, affiliations, outlets, keywords, article titles, abstracts, times cited, references**, etc.

Supply Chain Management • Open Access • Volume 27, Issue 7, Pages 49 - 63 • 19 December 2022

Analysing competing logics towards sustainable supplier management

León Bravo, Verónica[a] ; Jaramillo Villacrés, Mariuxy[b]; Silva, Minelle E.[c]

Save all to author list

[a] Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy
[b] School of Hospitality and Tourism, Universidad de Las Américas, Quito, Ecuador
[c] Supply Chain, Purchasing and Project Management Department, Excelia Business School, La Rochelle, France

Abstract

Purpose: To understand the context surrounding the sustainable supplier management (SSM) process (i.e. selection, development and evaluation), this paper aims to explore institutional logics existing in the Ecuadorian cocoa supply chain (SC). By considering local characteristics and sustainability practices, this study illustrates how competing logic influences SSM. Design/methodology/approach: This paper uses a multiple-case study method for which the authors interviewed different cocoa SC members in Ecuador and used a ground-up approach to analyse the data and reveal singularities influencing sustainability management. Findings: The analysis uncovered two main logics operating within the Ecuadorian cocoa SC SSM process: a commercial logic (e.g. potential for market access, product traceability) and a sustainability logic (e.g. local development and traditions/cultural issues). These logics address market demand requirements; however, some local producers' needs that impact SSM remains unexplored such as the existence of a regional ancestral culture that poses sustainability as a dominant logic with meaning beyond the triple bottom line. While the two logics have influenced supplier sustainability performance, this paper finds that, of the three SSM sub-processes (selection, development and evaluation), supplier development was the most relevant sub-process receiving attention from SC managers in the studied context. Practical implications: By understanding the differences in logic and needs, SC managers can better develop strategies for SSM. Originality/value: The study highlighted in this paper investigated the underexplored topic of the effects that competing logic may have on SSM. This paper focusses on the supplier's point of view regarding sustainability requirements, addressing a consistent research gap in the literature. © 2021, Verónica León Bravo, Mariuxy Jaramillo Villacrés and Minelle E. Silva.

Author keywords

Food industry; Global supply chain; Supply chain management; Sustainability

References (80)

View in search results format >

All Export Print E-mail Save to PDF Create bibliography

1 Adesanya, A., Yang, B., Bin Iqdara, F.W., Yang, Y.
Improving sustainability performance through supplier relationship management in the tobacco industry (Open Access)
(2020) *Supply Chain Management*, 25 (4), pp. 413-426. Cited 17 times.
http://www.emeraldinsight.com/info/journals/scm/scm.jsp
doi: 10.1108/SCM-01-2018-0034
ViewIt@Aalto View at Publisher

2 (2020) *Organic certification*. Cited 2 times.
(accessed, Agrocalidad: 20 January 2020
https://organicos.agrocalidad.gob.ec/

3 (2015) *Cacao nacional*. Cited 3 times.
(accessed, Asociación Nacional de Exportadores de Cacao-Ecuador: 18 January 2019
www.anecacao.com/es/quienes-somos/cacao-nacional.html

4 Annala, L., Polsa, P.E., Kovács, G.
Changing institutional logics and implications for supply chains: Ethiopian rural water supply
(2019) *Supply Chain Management*, 24 (3), pp. 355-376. Cited 9 times.
http://www.emeraldinsight.com/info/journals/scm/scm.jsp
doi: 10.1108/SCM-02-2018-0049
ViewIt@Aalto View at Publisher

5 Awaysheh, A., Klassen, R.D.
The impact of supply chain structure on the use of supplier socially responsible practices (Open Access)
(2010) *International Journal of Operations and Production Management*, 30 (12), pp. 1246-1268. Cited 332 times.
doi: 10.1108/01443571011094253
ViewIt@Aalto View at Publisher

6 Baquero Méndez, D., Mieles López, J.D.
(2014) *Los 'booms' en perspectiva cacao y banano*
Foro Economía Ecuador
http://foroeconomiaecuador.com/fee/los-booms-en-perspectiva-cacao-banano/?pdf=1557

7 Besharov, M.L., Smith, W.K.
Multiple institutional logics in organizations: Explaining their varied nature and implications (Open Access)
(2014) *Academy of Management Review*, 39 (3), pp. 364-381. Cited 646 times.
http://amr.aom.org/content/39/3/364.full.pdf+html
doi: 10.5465/amr.2011.0431

Bibliographic data is *semi-structured* (fielded) text data from literature databases. Example record below from the Web of Science database

```
PT J
AU Park, EM
   Seo, JH
   Ko, MH
AF Park, Eun-Mi
   Seo, Joung-Hae
   Ko, Mi-Hyun
TI The effects of leadership by types of soccer instruction on big data
   analysis
SO CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND
   APPLICATIONS
LA English
DT Article
DE Big data; Crawling; Textmining; Leadership; Korea nation football team
ID ATHLETICS; BEHAVIOR; SPORTS
AB The purpose of the present study is to figure out football coaches' leadership styles. So far, numerous of coaches have coached South Korea's
national team. Compared to other countries, the Korea Republic national team has changed coaches relatively often. In particular, owing to the
result-centric Korean culture, if the national team had deplorable results in a specific match, the head coach would be fired right away. Of
course, there were some successful and popular coaches. However, many other coaches ended up in a failure in the Korean national team. Therefore,
there must be a difference in leadership styles between the successful and unsuccessful coaches. In this context, it would be critical to find
out the traits of the successful coaches' leadership. Using text-mining techniques, the present study aims to establish different leadership type
of football coaches. To this end, we analyzed the South Korean national football team coaches' leadership styles using text-mining techniques
applied to the analysis of NAVER news. Our results suggest that successful leaders have important leadership elements, such as communication,
trust, and belief.
C1 [Park, Eun-Mi; Seo, Joung-Hae] Kyungpook Natl Univ, Dept Business Adm, 80 Daehakro, Daegu, South Korea.
   [Ko, Mi-Hyun] Korea Inst Sci & Technol, Dept Policy Res, 245 Daehak Ro, Daejeon, South Korea.
RP Ko, MH (reprint author), Korea Inst Sci & Technol, Dept Policy Res, 245 Daehak Ro, Daejeon, South Korea.
EM issack38317@naver.com; johseo@knu.ac.kr; mihyungo@kisti.re.kr
CR Bass B.M., 1985, BASS STOGDILLSHDB LE
   Bass M., 1995, MULTIFACTOR LEADERSH
   Burns J. M., 1978, LEADERSHIP
   Byung E.Y., 2014, J SPORT LEIS STUD, V56, P133
   CHELLADURAI P, 1984, J SPORT PSYCHOL, V6, P27
   Chelladurai P., 1982, TASK CHARACTERISTICS
   CHELLADURAI P, 1983, J SPORT PSYCHOL, V5, P371
   Cho B.N., 2006, BUS EC REV, V37, P1229
   Cho B.S., 2010, J COACH DEV, V12, P83
   Cho W.J., 2006, KOREA J SPORTS SCI, V15, P317
   Choi B.A., 2007, J COACH DEV, V9, P381
   CONGER JA, 1987, ACAD MANAGE REV, V12, P637, DOI 10.2307/258069
   DANIELSON RR, 1975, RES QUART, V46, P323
   Doherty AJ, 1996, J SPORT MANAGE, V10, P292
   Erle F.J., 1981, THESIS
   Fiedler Fred Edward, 1967, THEORY LEADERSHIP EF
   House P.J., 1971, ADM SCI Q, V16, P321
   Jin SC, 2015, CLUSTER COMPUT, V18, P999, DOI 10.1007/s10586-015-0452-x
```

PT=Publication type
AU=Authors
AF=Authors with full names
TI=Title of the article
SO=Source (Journal or other source)
AB=Abstract
CR=Cited references,
Etc.

Another example from Scopus database
(selected fields exported only as a CSV file)

	A	B	C	D	E
1	Year	Cited by	DOI	Link	Abstract
2	2022	2	10.1108/O	https://w	Purpose: Numerous educational
3	2022	1	10.1108/JS	https://w	Purpose: Consumers today active
4	2022		10.3389/fp	https://w	Under the background of global
5	2022	2	10.1108/K	https://w	Purpose: This study aims to propo

Comparison of bibliometric analysis and visualization tools

	Thematic Network	Author Network	Reference Network	Other Networks	Evolution	Performance	Burst Detection	Spectrogram	Geospatial	Visualization
Science Mapping Analysis Tools										
Bibexcel	•	•	•	•	•	•			•	External software
Biblioshiny	•	•	•	•	•	•	•	•	•	Network, three-fields plot, word cloud, tree map, historiograph, strategic diagram, evolution map, and world map
BiblioMaps	•	•	•	•		•			•	Network
CiteSpace	•	•	•	•		•	•		•	Tree ring, geospatial map
CitNetExplorer			•							Network
SciMAT	•	•	•	•	•	•				Strategic diagram, cluster network, overlapping map, evolution map
Sci² Tool	•	•	•	•			•		•	Temporal, geospatial map, topical, network
VOSviewer	•	•	•	•		•				Network, overlay, density
Libraries										
Bibliometrix	•	•	•	•	•	•	•	•	•	Network, three-fields plot, word cloud, tree map, historiograph, strategic diagram, evolution map, and world map
BiblioTools	•	•	•	•		•			•	Network
Citan						•				Bars, bow plots, and pie chart
Metaknowledge	•	•	•	•			•	•		Timeline graph, spectrogram, and network
scientoText		•				•				
SxientoPy					•	•				Timeline graph, bar graph, evolution graph, and word cloud

Aras and B (2023): Digital Transformation Journey Guidance: A Holistic Digital Maturity Model Based on a Systematic Literature Review, based on Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la Información*, *29*(1).

What is research profiling (RP)*?

It augments traditional literature reviews, and should be done at the beginning of every new research project!

It gives you the big picture or helicopter view of a research area.

It can be done quickly using the analysis tools in databases (*Scopus* or *Web of Science*) OR

with more time using external text-mining and visualization tools (such as Leximancer[1] or VOSviewer[2] or VantagePoint[3]).

*Porter, A.L., Kongthon, A. and J.-C. Lu (2002), "Research Profiling: Improving the Literature Review", *Scientometrics*, 53(3), 351-370. https://link.springer.com/article/10.1023/A:1014873029258

A!

[1]BIZ Campus license, [2] Free software [3] ISM Department licence for 5 Windows users
see also Google's free https://openrefine.org for cleaning bibliometric data

Comparison of traditional and research profiling reviews

Old (Traditional Literature Review)	New (Research Profiling)
Micro focus (paper-by-paper)	Macro focus (**patterns** in the literature as a body)
Narrow range (~20 references)	**Wide range** (~100 – 20,000 references)
Tightly restricted to the topic	Encompassing the **topic + related** areas
Text discussion	**Text, numerical &** **visual** depiction

A!

Porter, A.L., Kongthon, A. and J.-C. Lu (2002), "Research Profiling: Improving the Literature Review", *Scientometrics*, 53(3), 351-370. https://link.springer.com/article/10.1023/A:1014873029258

Simple research profiling process e.g. with Scopus or Web of Science

Select database

- *Scopus*
- *Web of Science*

Open databases:

- *Lens.org*
- *Dimensions.ai*

Formulate search phrase

Conduct search

Analyze results within the tool, or using external visualization / science mapping tools such as VOSviewer or Leximancer.

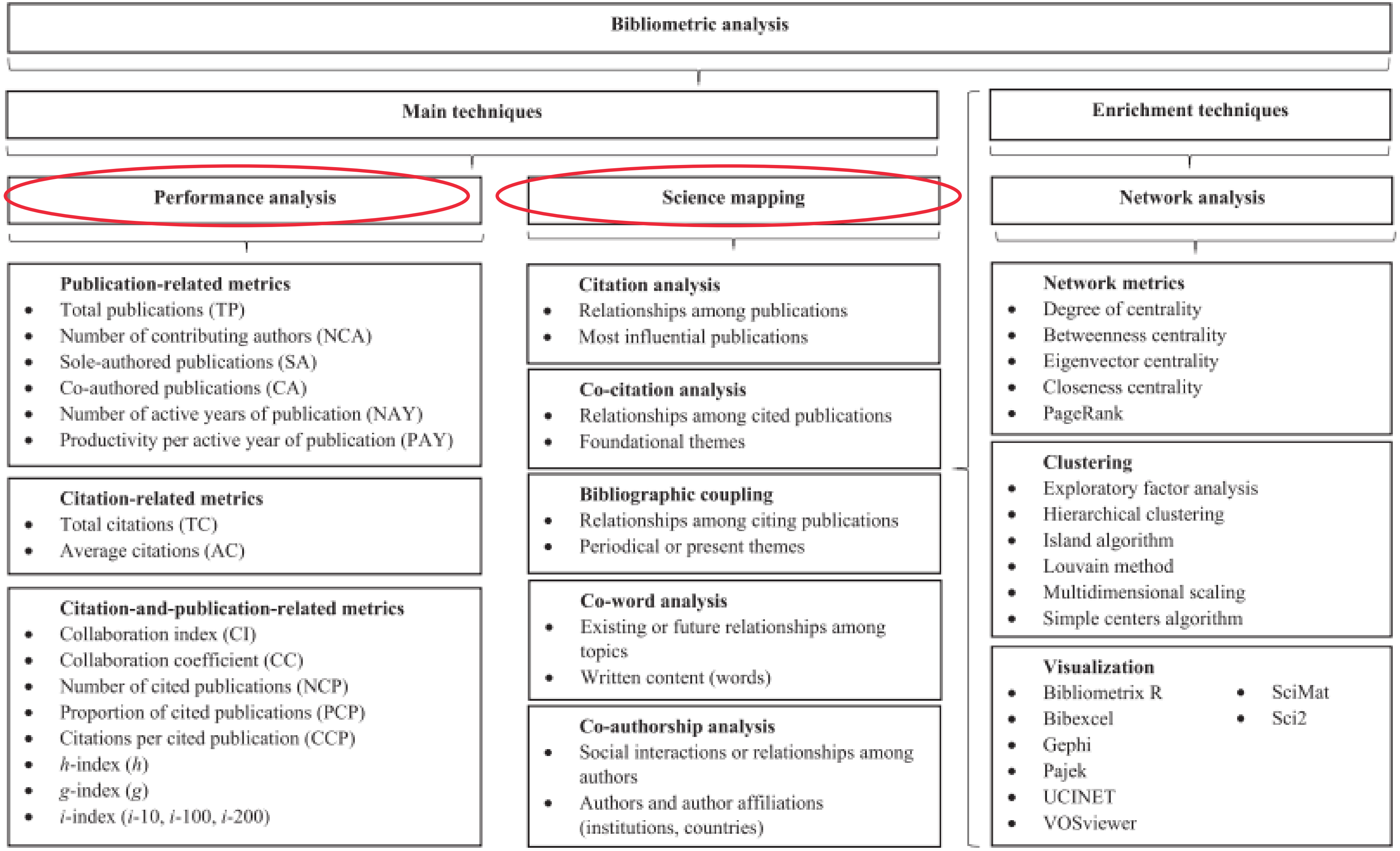
Report the results

Aalto University School of Business

Scopus and **Web of Science** are so-called ***reference databases***, where you can find links (ViewIt@Aalto) to full-text articles in actual article databases

See Aalto's Learning Centre pages (Business Guide/Articles).

Techniques for bibliometric analyses



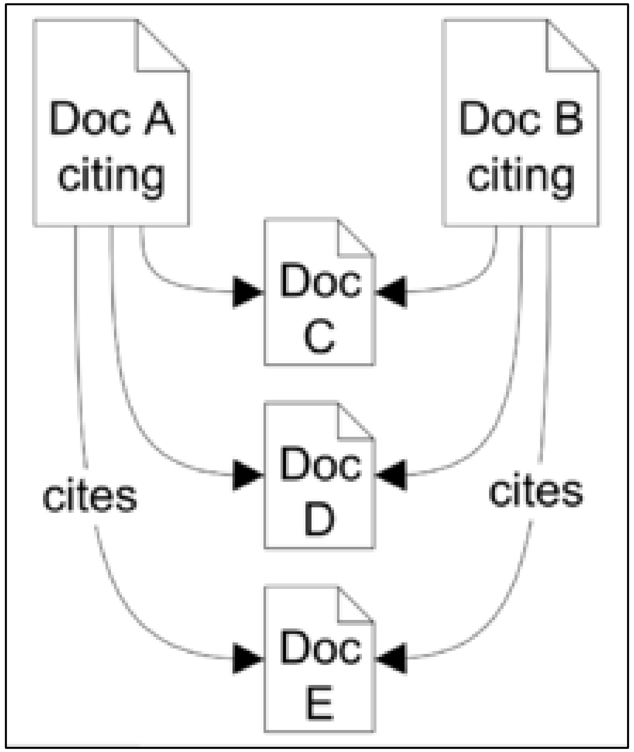
Source: Donthu et al. 2021 (Figure 2), How to conduct a bibliometric analysis: An overview and guidelines, *J. of Business Research*, 133, p. 289.
Available at: https://www.sciencedirect.com/science/article/pii/S0148296321003155

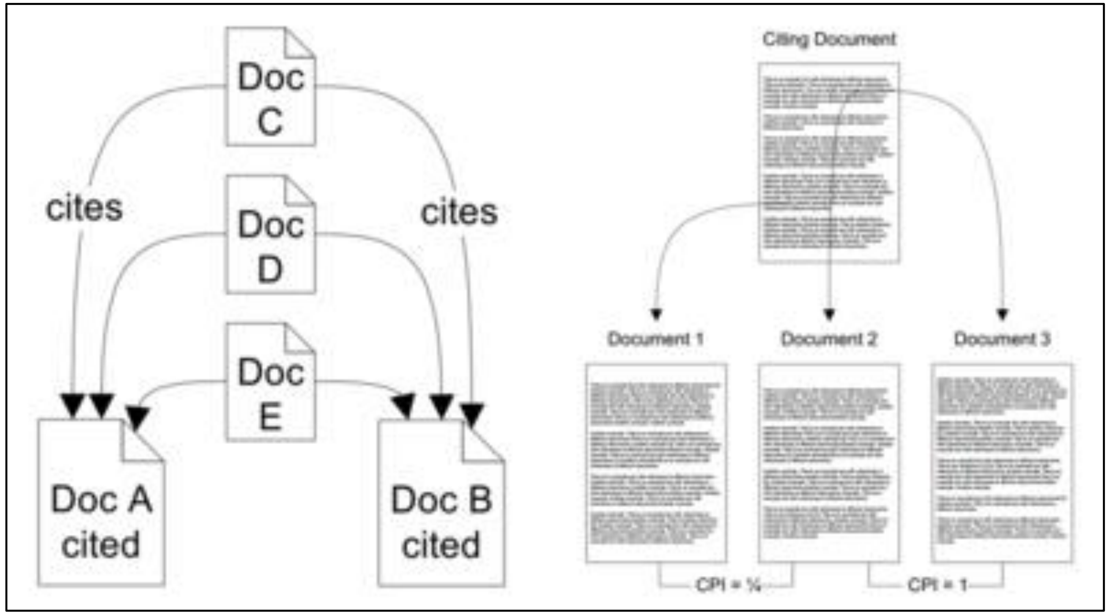
Techniques for science mapping

TECHNIQUE	USAGE	UNIT OF ANALYSIS	DATA REQUIREMENTS	EXAMPLE
Citation analysis	To analyze the relationships among publications by identifying the *most influential publications* in a research field.	Documents	Author name, Citations, Title, Journals, DOI, References	Podsakoff et al. (2005)
Co-citation analysis	To analyze the relationships among **cited** publications to understand the *development of the foundational themes in a research field*.	Documents	References	Fahimnia et al. (2015)
Bibliographic coupling	To analyze the relationships among citing publications to understand the *periodical or present development* of themes in a research field.	Documents	Author name, Title, Journals, DOI, References	Donthu et al. (2020b)
Co-word analysis	To explore the *existing or future relationships among topics* in a research field by focusing on the written content of the publication itself.	Words	Title, Abstract, Author keywords, Index keywords, Full text	Emich et al. (2020)
Co-authorship analysis	To examine the *social interactions or relationships among authors and their affiliations* and equivalent impacts on the development of the research field.	Authors Affiliations	Author, Affiliation (institution and country)	Acedo et al. (2006)

Source: Donthu et al. 2021 (Table 3), How to conduct a bibliometric analysis: An overview and guidelines, *J. of Business Research*, 133, p. 289. Available at: https://www.sciencedirect.com/science/article/pii/S0148296321003155

Bibliographic coupling | Co-citation





Bibliographic coupling occurs when two works reference a common third work in their bibliographies. It is an indication that a probability exists that the two works treat a related subject matter.

Co-citation is a similarity measure for documents that makes use of citation relationships. Co-citation is defined as the frequency with which **two documents are cited together** by other documents.

http://en.wikipedia.org/wiki/Co-citation & http://en.wikipedia.org/wiki/Bibliographic_coupling

Focus in bibliometric reviews

“Bibliometric reviews do not deal with theories, methods, and constructs as much as they usually do with authors, affiliations, countries, citations and co-citations, etc.”

(Paul and Criado, 2020)

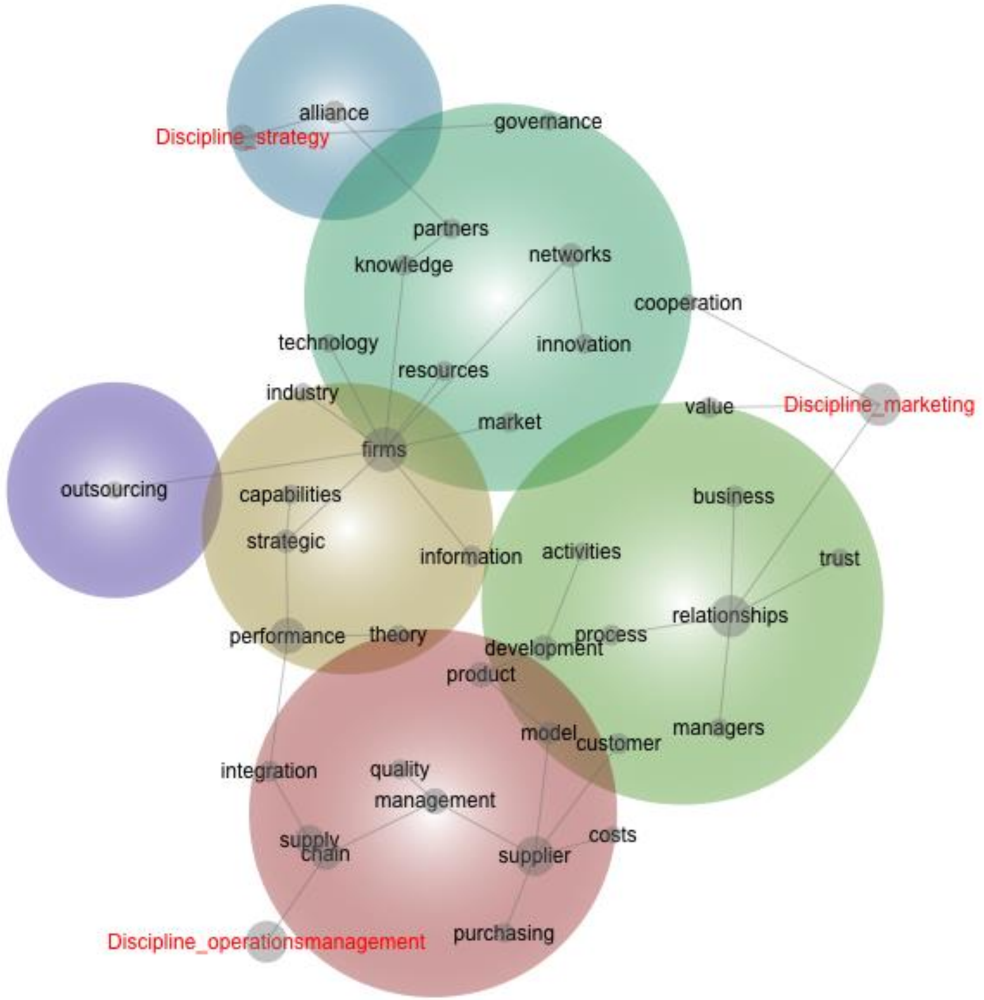
- This follows directly from the type of data that is indexed in databases – they do NOT index theories, methods and constructs (unfortunately).
- Manual categorization & hybrid method is needed for those!

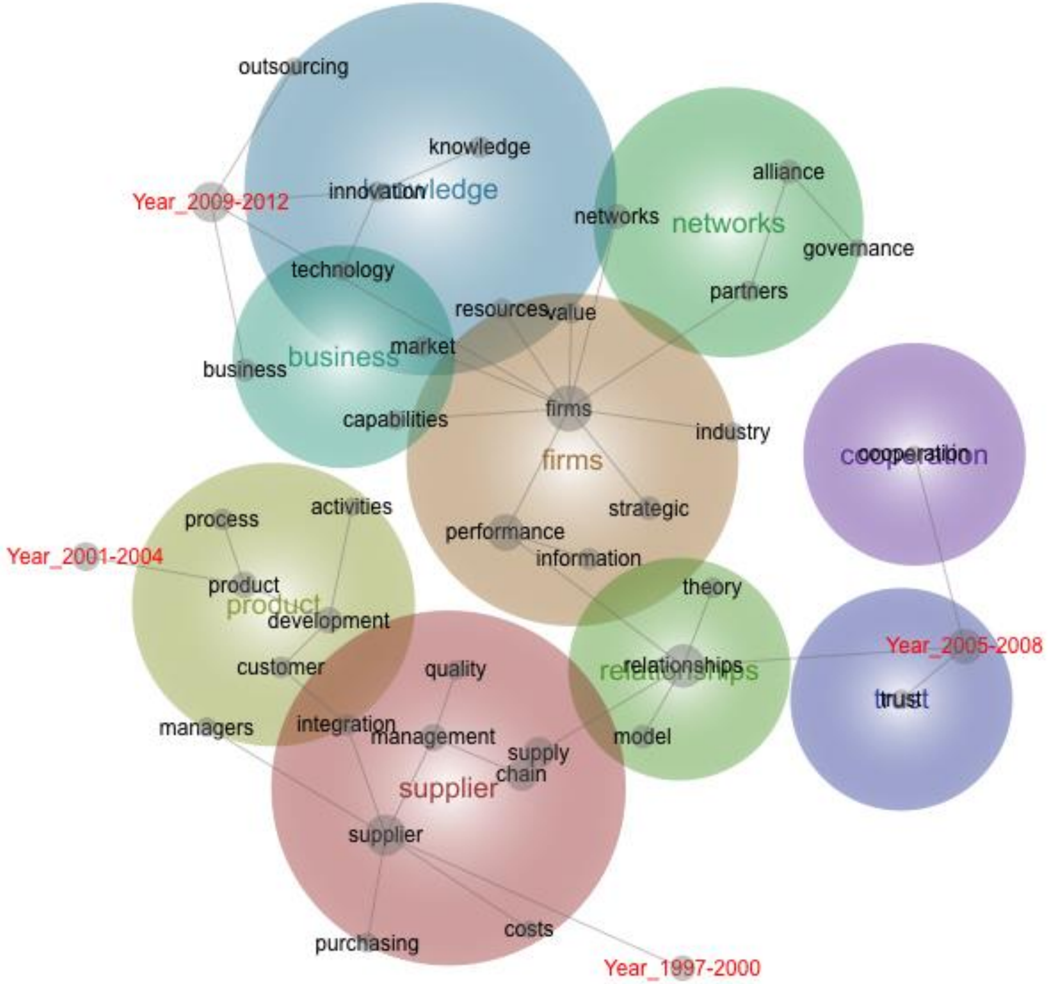
A!

Paul, J. and Criado, A. R. (2020), The art of writing literature review: What do we know and what do we need to know? *International Business Review*, 29(4), 1-7, https://www.sciencedirect.com/science/article/pii/S0969593120300585

Text mining & Research Profiling example 1 (on ERM): Secondary fields (here discipline, or time period) in the data (= abstracts in excel format) can be used as TAGS / additional descriptors (tool used: Leximancer)

In questionnaire data you might use the respondent category, in interviews the interviewee names as tags, etc.

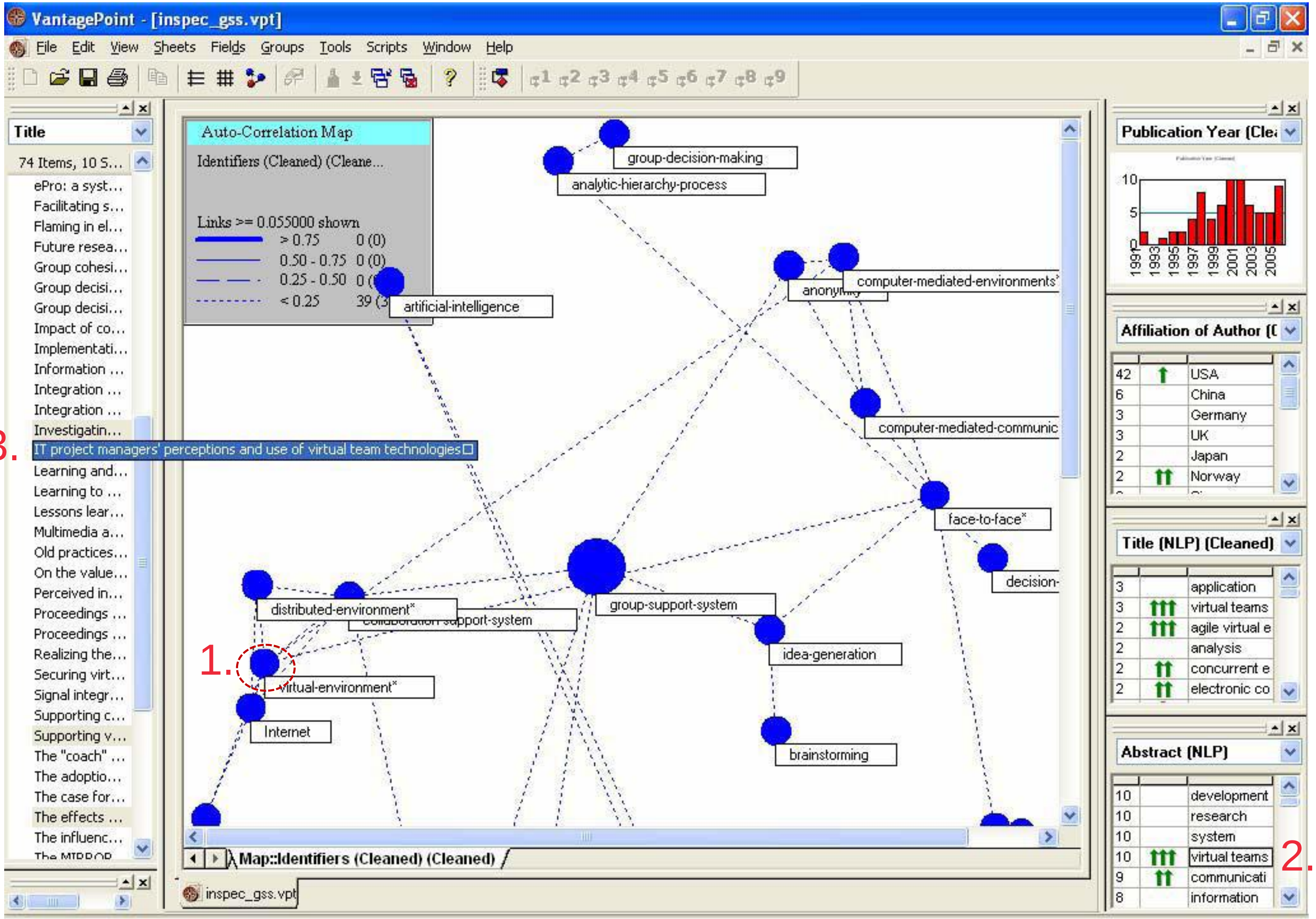




The data is from 601 abstracts of **external resource management (ERM)** research in three scientific disciplines from 1997-2012

© Tanskanen, K., Ahola, T., Aminoff, A., Bragge, J., Kaipia, R. and Kauppi, K. (2017), "Towards evidence-based management of external resources: Developing design propositions and future research avenues through research synthesis", *Research Policy*, 46(6), 1087-1105, available at http://dx.doi.org/10.1016/j.respol.2017.04.002

Text-mining & Research Prof. example 2: Mining keywords on Group Support Systems research (2.000 articles from IEEE Inspec database, tool used: VantagePoint by SearchTechnology Inc.)



3.

1. The user clicks the virtual-environment node and all four detail fields update on the right-hand side columns.

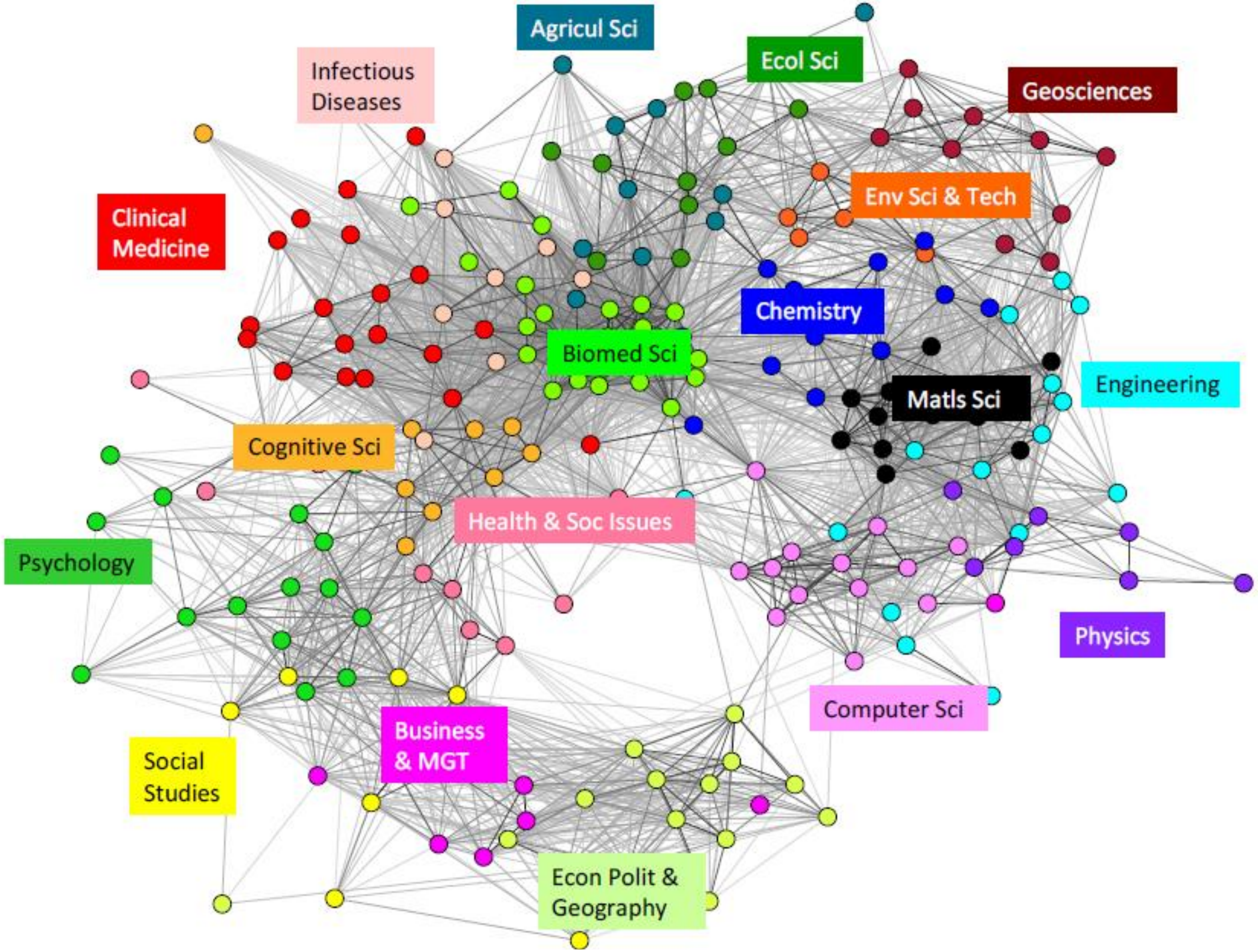
2. The user clicks "virtual teams" from the Abstract (NLP) detail field table.

3. The user can see four highlighted articles on the left that consider virtual teams in a virtual environment, and she can double-click to open the full bibliographic information of those articles.

A!

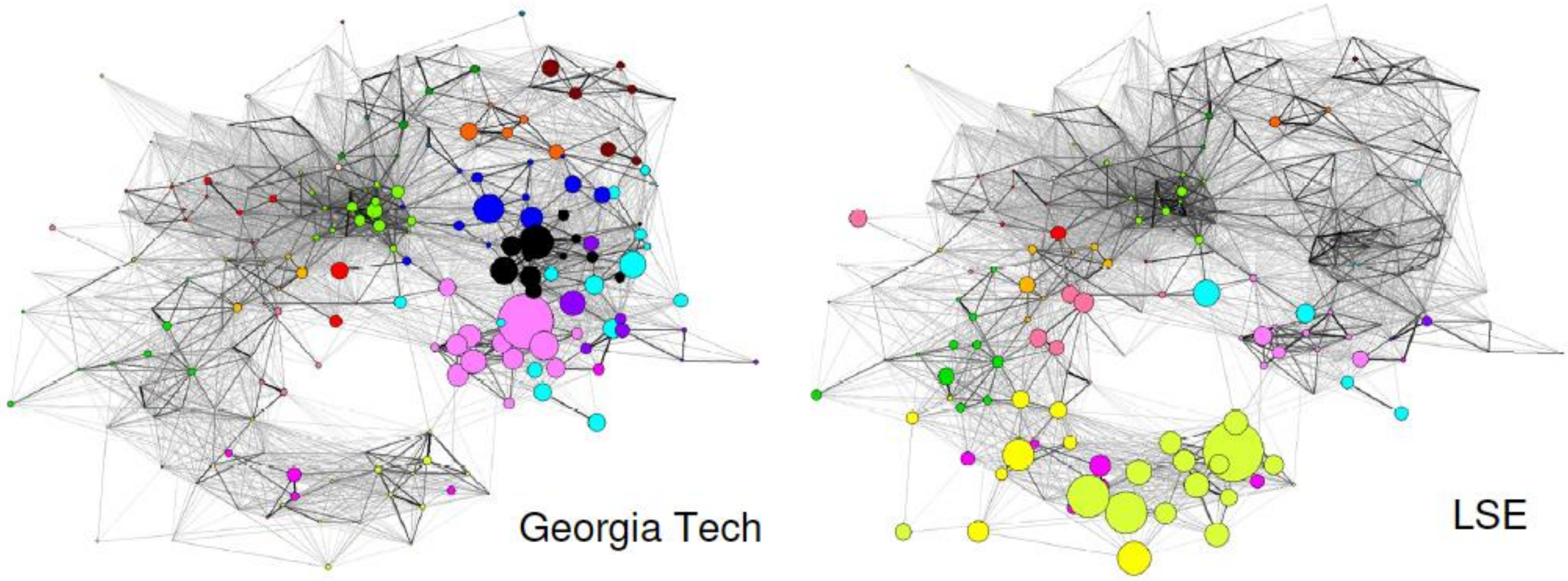
Source: Bragge, J., Relander, S., Sunikka, A. and Mannonen, P. (2007), "Enriching Literature Reviews with Computer-Assisted Research Mining. Case: Profiling Group Support Systems Research", PDF available at https://www.computer.org/csdl/proceedings/hicss/2007/2755/00/27550243a.pdf

Text-mining & research profiling example 3: Global science map from 2007 - based on citing similarities among Web of Science (WoS) subject categories



Source: Rafols, Porter and Leydesdorff 2010, "Science overlay maps: a new tool for research policy and library management", *Journal of the American Society for Information Science and Technology,* Available at http://www.leydesdorff.net/overlaytoolkit/overlaytoolkit.pdf

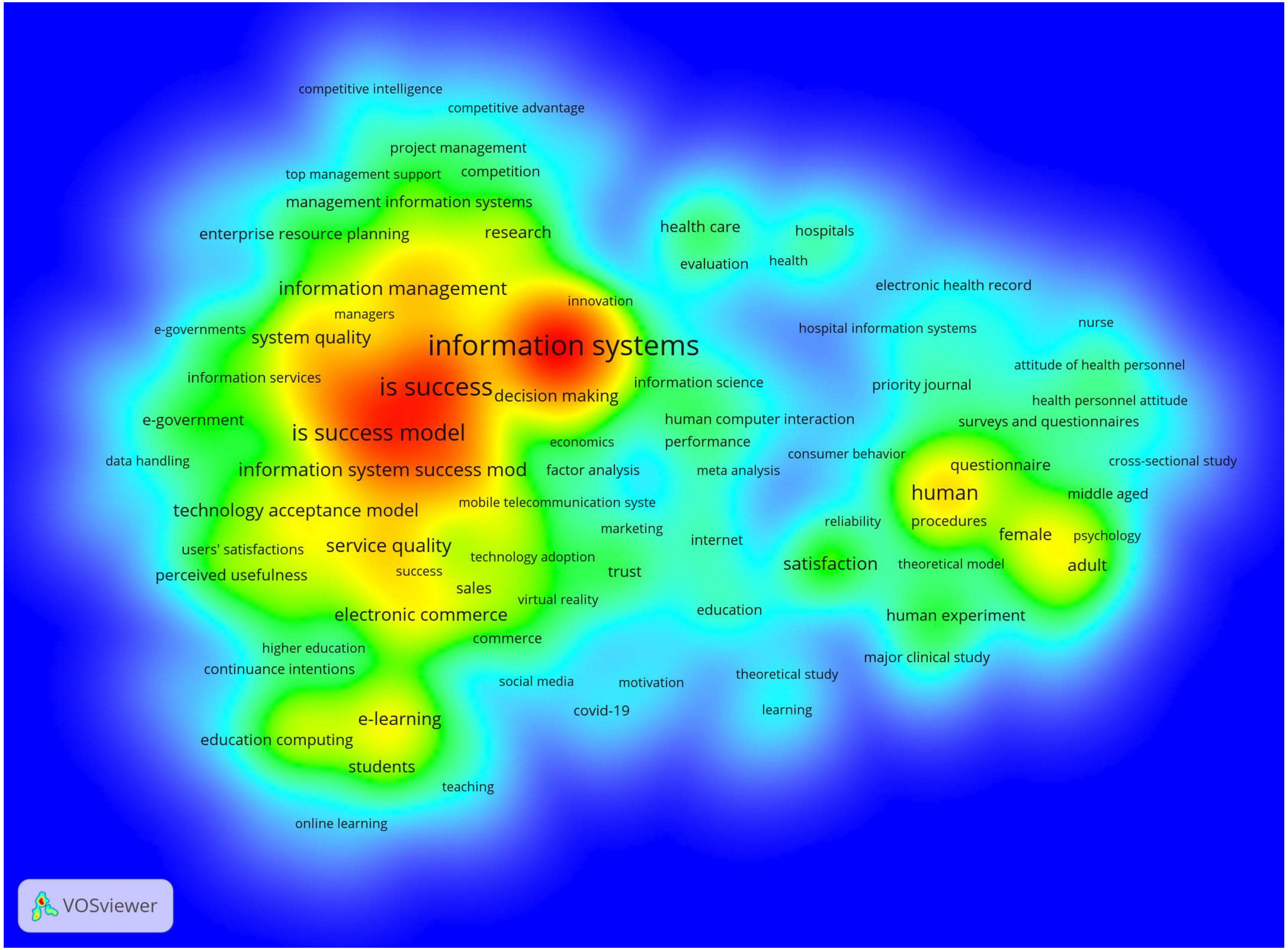
Text mining & RP example 3b: Research published by *Georgia Tech* and *LSE* scholars <u>overlayed</u> on the previous global science map – easy to compare institution profiles!



A!

Source: Rafols et al. 2010, ”Science overlay maps: a new tool for research policy and library management”, *JASIST*, http://www.leydesdorff.net/overlaytoolkit/overlaytoolkit.pdf

Text-mining & Research Profiling example 4: Research heatmap on "*IS Success*" literature (ca. 1900 articles' index keywords from Scopus, using the free VOSviewer.com tool)



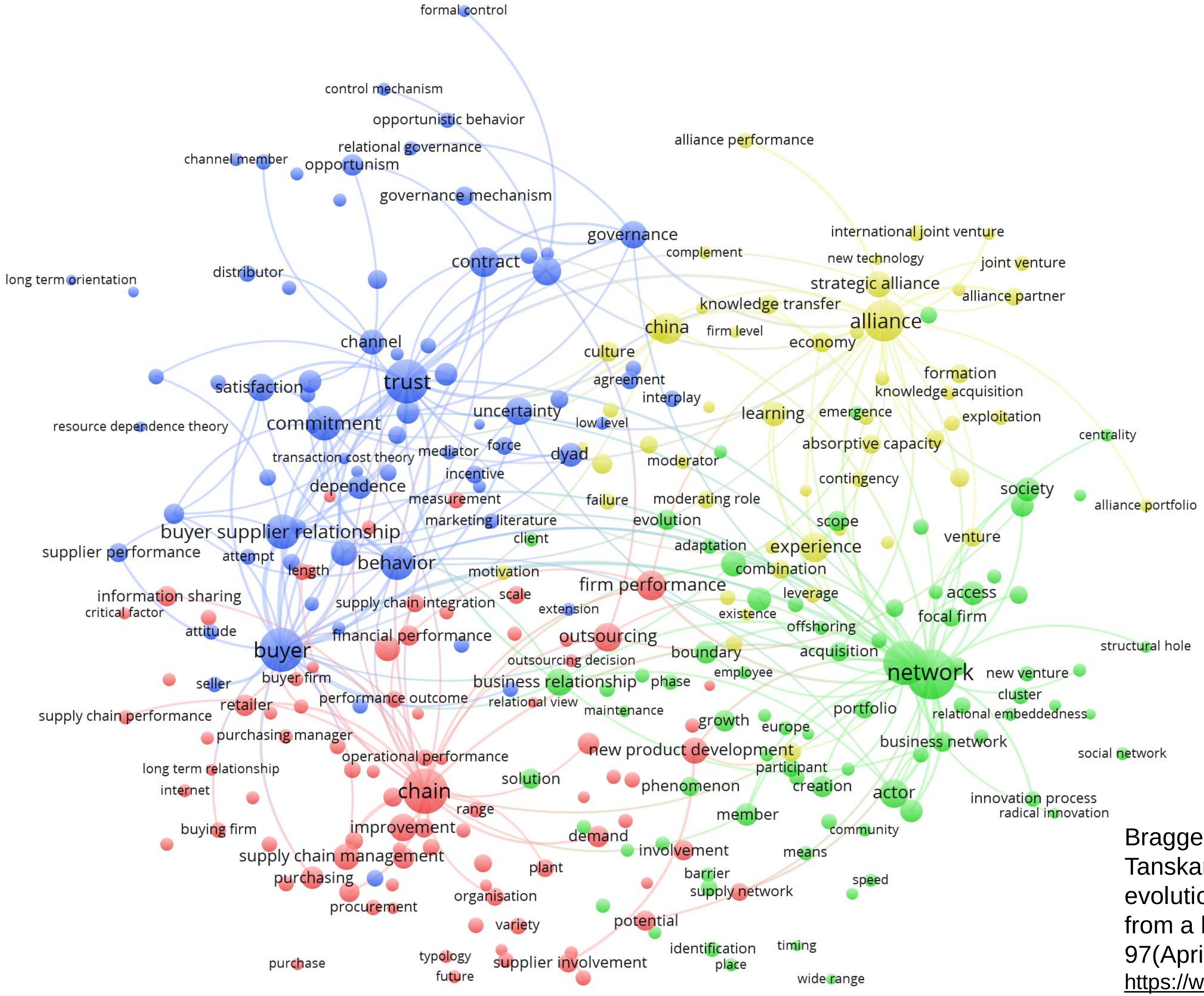
VOSviewer = **V**isualization **O**f **S**imilarities viewer

Free, university-based software tool for visualizing bibliometric networks.
The tool also offers a text-mining functionality that can be used to visualize any co-occurrence networks.
- Can be used with other text data as long as the file structure (CSV) is similar than exported from literature databases.

https://www.vosviewer.com/

Advice booklet for researchers what kind of questions can be analyzed with VOSviewer:
https://pure.tudelft.nl/ws/files/40154954/AIDA_Booklet_V2.2.pdf

Text-mining & RP example 5: Cluster / network map on “*External Resource Management*” research (1290 articles’ abstracts & titles from Scopus, using VOSviewer tool)



Bragge, J., Kauppi, K., Ahola, T., Aminoff, A., Kaipia, R. and Tanskanen, K. (2019), “Unveiling the intellectual structure and evolution of external resource management research: Insights from a bibliometric study”, *Journal of Business Research*, 97(April), 141-159. available at: https://www.sciencedirect.com/science/article/pii/S0148296318306696

Text-mining & RP example 5b: Interpreting the previous Cluster / network map on ERM research

Trust, commitment and power - The interorganizational relationship view
Trust, Buyer, Commitment, Behaviour

Relationship performance and relational governance
Fawcett et al., 2012; Partanen et al., 2008; Singh and Mitchell, 2005
- Factors that influence relationship performance: trust, relationship dynamics
- Socialization and social capital

The outcomes of inter-organizational collaboration
Fawcett et al., 2012; Singh and Mitchell, 2005; Whipple et al., 2010
- Collaboration as an antecedent of relationship success and supply chain performance
- Relationship development and its impact on performance

Formal governance mechanisms
Carson and John, 2013; Parker and Brey, 2015
- Contracts, incomplete contracting
- Contrasting formal and relational governance mechanisms

Antecedents of interorganizational collaboration
Johnston et al., 2004; Laaksonen et al., 2008; Moore, 1998; Narayandas and Rangan, 2004; Nyaga et al., 2010; Skarmeas and Katsikeas, 2001; Wong et al., 2005
- Factors that benefit collaboration
- Interdependence, trust, commitment, and balance between them

Factors that may harm relationships
Belaya et al., 2009; Kang and Jindal, 2015; Samaha et al., 2011
- Opportunism or its threat
- Conflicts and solving them

Knowledge transfer and learning - The alliance view
Alliance, Experience, Strategic Alliance, Learning

The antecedents of interorganizational learning and knowledge development
Ahuja, and Katila, 2001; Lane et al., 2001; Sampson, 2007
- Absorptive capacity: the ability to achieve, assimilate, and utilize new external knowledge
- Balancing between acquiring or exchanging knowledge
- Learning critical skills or capabilities from alliance partners
- The need to protect oneself from losing core proprietary assets or capabilities

Knowledge search and transfer
Daniel et al., 2002; Fang, 2008; Gallego et al., 2013; McGinnis and Vallopra, 1999; Oke et al., 2008; Primo and Amundson, 2002; Vasudeva and Anand, 2011; Wagner, 2012
- The impact of ties on knowledge transfer in various competitive environments
- Knowledge searching and search strategies and transfer
- External actors as knowledge sources

Alliance formation
Oke et al., 2008; Phene and Tallman, 2014; Sampson, 2007
- Alliance partners' selection
- Strength of ties between partners

Integration and operational effectiveness - The chain view
Chain, Firm performance, Outsourcing, Improvement

Integration in supply chains
Cai et al., 2009; Cagliano et al., 2006; Cousins et al., 2006; Enz and Lambert, 2012; Funda and Robinson, 2005; Li et al., 2005; Rajaguru and Matanda, 2013; Sanders 2007; Yan and Wang, 2012; Yao et al., 2009
- IT integration
- Socialization of managers between firms, end-customer orientation
- Governance mechanisms of quasi-integration, cross-functional and cross-firm teams

Supply chain management practices
Fullerton and Wempe, 2009; Gonzales-Benito et al., 2003; Lösch and Lambert, 2007; Narayanan et al., 2011; Wisner and Tan, 2000
- Effects of practices to firm performance, operational performance dimensions and social & environmental performance
- The antecedents to the adoption of these practices
- Lean and Total Quality Management

Other
Carter and Kaufmann, 2007; Jap, 2007; Lösch and Lambert, 2007; Paulraj and Chen, 2007
- The use of market mechanisms for reducing costs, e.g. electronic reverse auctions
- Outsourcing decisions

Risk of disruptions in supply chains
Cannon and Homburg, 2001; Ellis et al., 2010
- The antecedents of supply disruption
- Supply risks

Innovation and value creation - The network view
Network, Innovation, Actor, Business relationship

Structure and design of networks
Baum et al., 2000; Capaldo, 2007; Chen and Chiang, 2011; Koka and Prescott, 2008; Lavie, 2007; Phelps, 2010; Shipilov, 2006; Terpend and Ashenbaum, 2012; Zaheer and Bell, 2005
- Different network structures: single sourcing, alliance networks, alliance portfolios
- Structure and design to enhance innovation, supply chain performance and firm performance

Dynamism in inter-organizational networks
Lee, 2007; Madhavan et al., 1998; Soda et al., 2004
- Changes to network design
- Structural holes and their impact on firm performance
- Centrality and structural holes

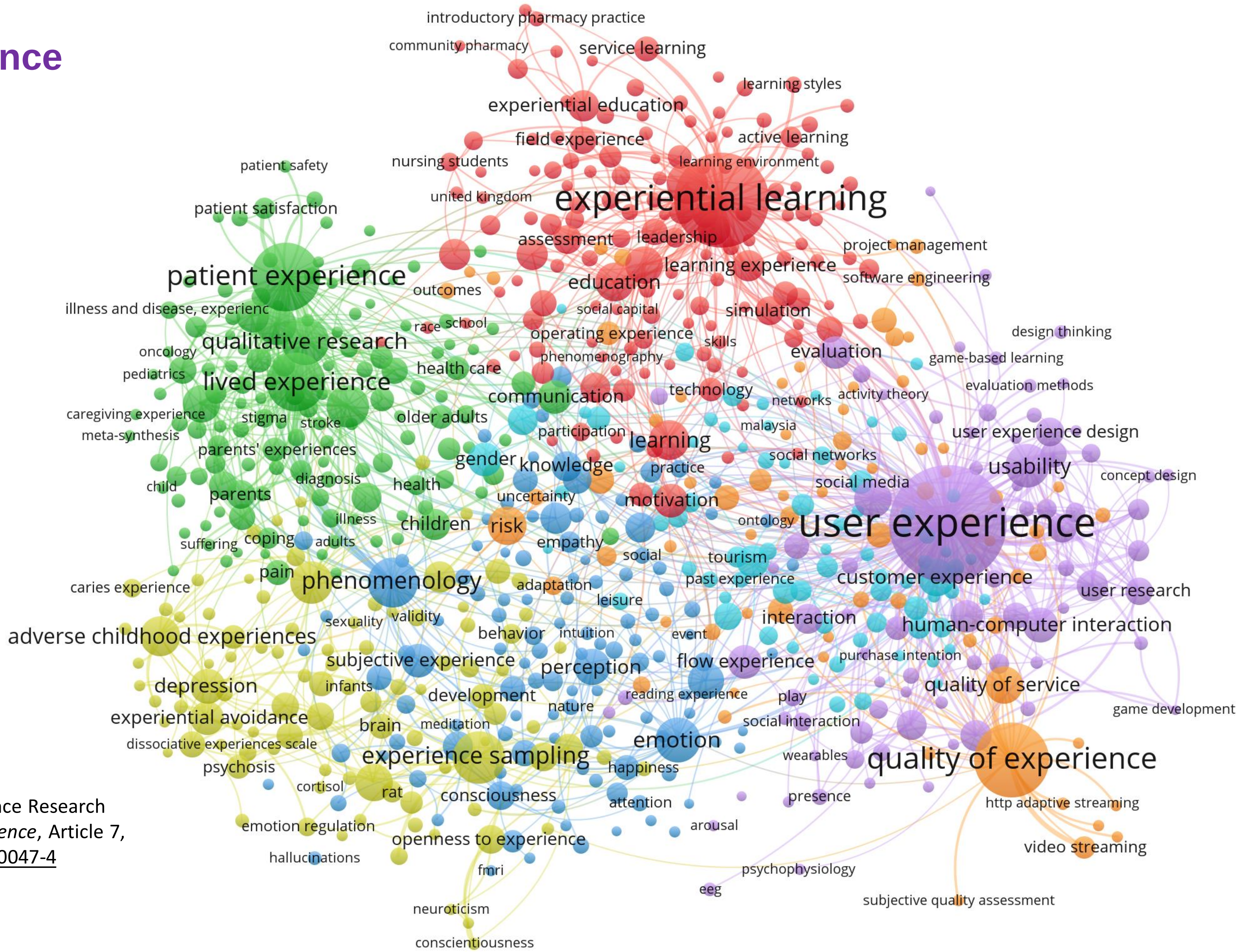
Value creation in networks
Cova and Salle, 2008; Guenzi and Troilo, 2007; Möller and Rajala 2007
- Customer value, supplier value
- Value co-creation

Other
Ho and Pollack, 2014; Koka and Prescott, 2008; Smith and Lohrke, 2008
- Entrepreneurial benefits resulting from interorganizational linkages
- Enduring linkages, linkage strength

Bragge, J., Kauppi, K., Ahola, T., Aminoff, A., Kaipia, R. and Tanskanen, K. (2019), "Unveiling the intellectual structure and evolution of external resource management research: Insights from a bibliometric study", *Journal of Business Research*, 97(April), 141-159. Available at: https://www.sciencedirect.com/science/article/pii/S0148296318306696

Text-mining & RP example 6: Cluster / network map on Experience research's author keywords (52.000 articles from Scopus, analyzed using VOSviewer)

No such research area as 'Experience' exists.
Still, thousands of scientific publications come out every year with **author keywords** experience or experiential



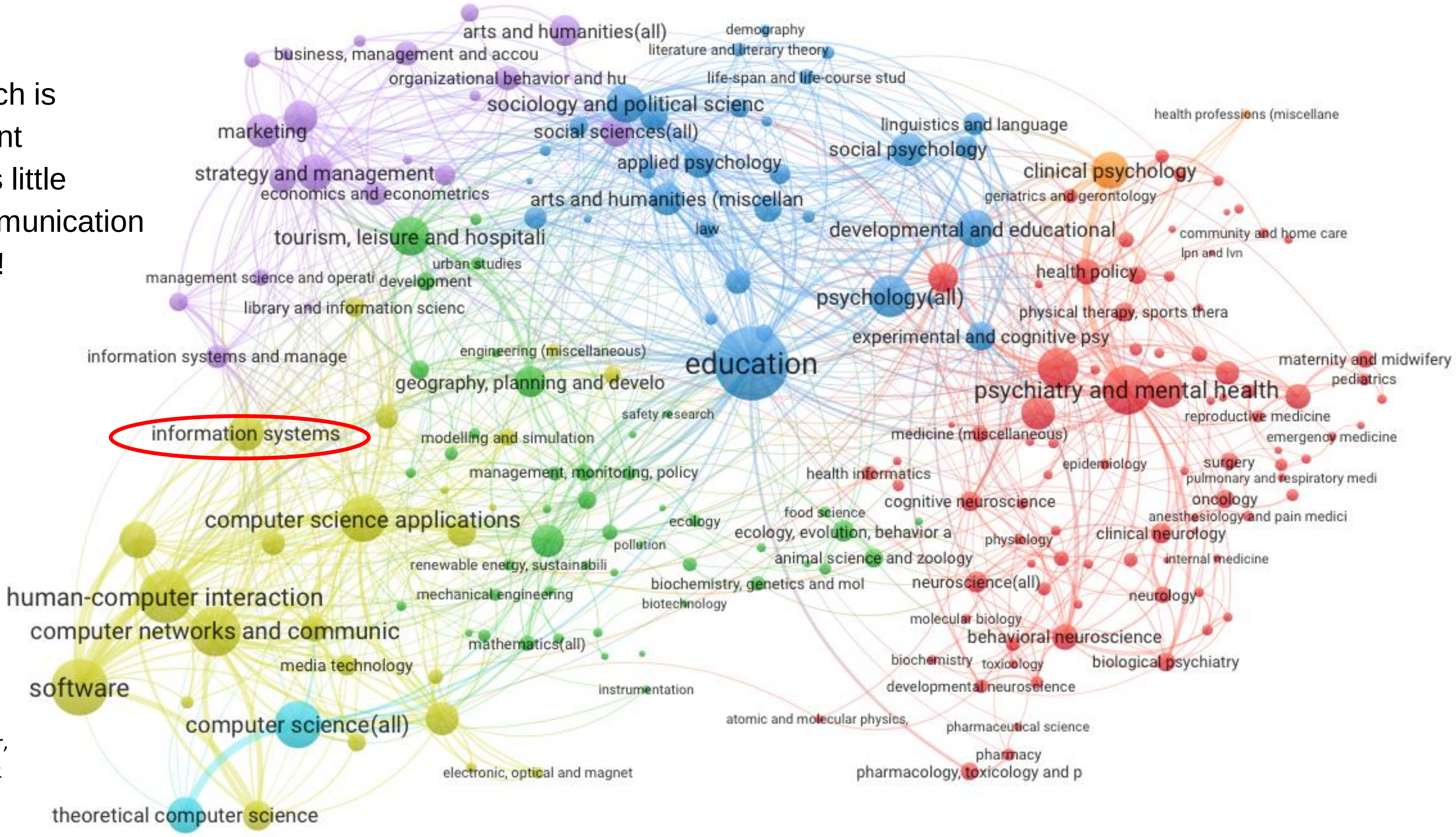
Roto V., Bragge, J., Lu, Y., & Pacauskas, D. (2021): "Mapping Experience Research Across Disciplines: Who, Where and When", *Quality and User Experience*, Article 7, September, https://link.springer.com/article/10.1007/s41233-021-00047-4

A!

Online map: https://app.vosviewer.com/?map=https://users.aalto.fi/~bragge/experiencemaps/Keywords_min40wordsMap726items.txt&network=https://users.aalto.fi/~bragge/experiencemaps/Keywords_min40wordsNetwork726items.txt

TM & RP example 7: Overview of experience research fields (based on ASJC classes*)

Experience research is scattered in different disciplines: there is little awareness or communication between the areas!

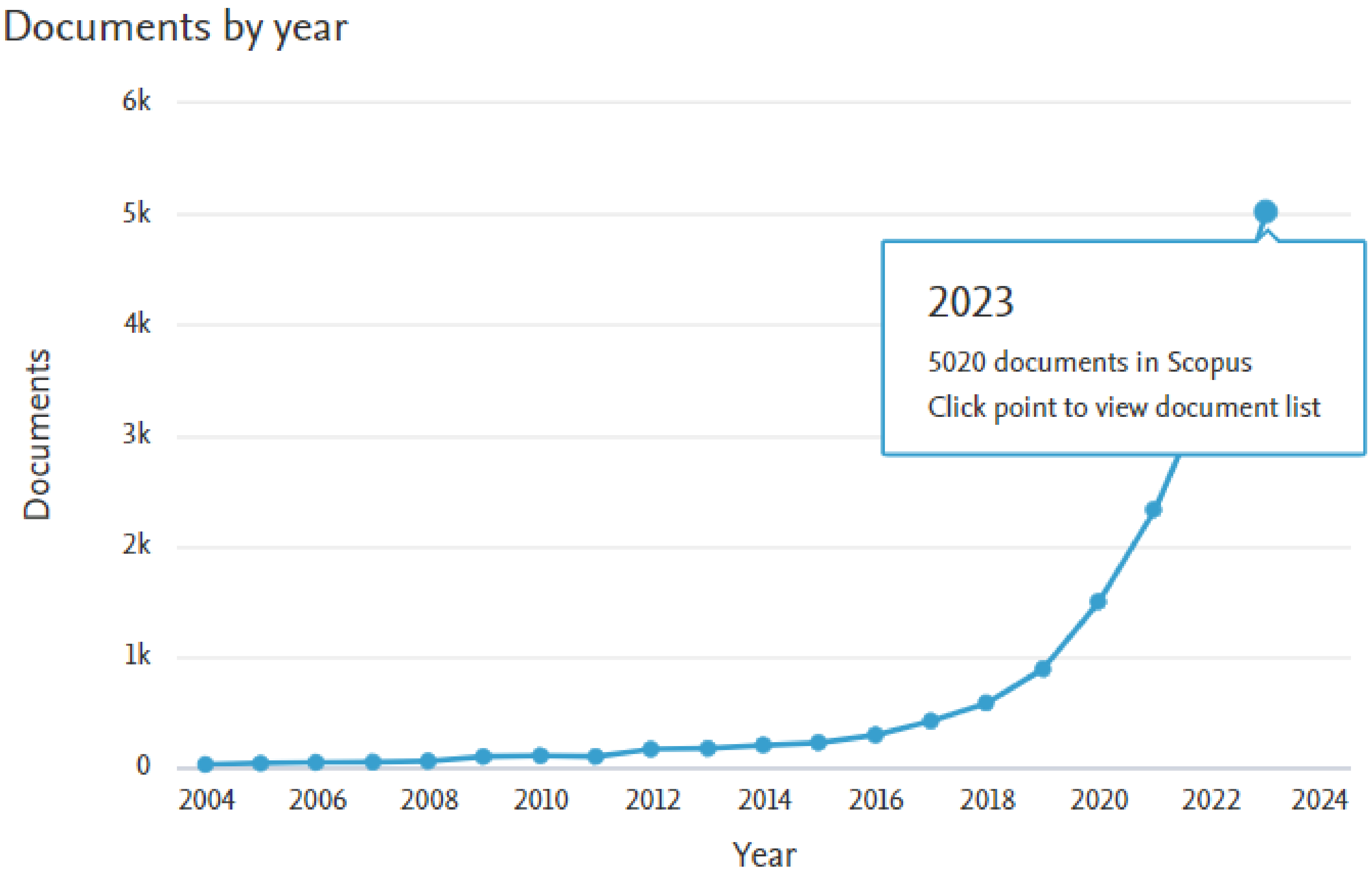


Roto V., Bragge, J., Lu, Y., & Pacauskas, D. (2021): "Mapping Experience Research Across Disciplines: Who, Where and When", *Quality and User Experience*, Article 7, September, https://link.springer.com/article/10.1007/s41233-021-00047-4

A!

*ASJC = All Science Journal Classification by Scopus, see https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/ Online map at: https://app.vosviewer.com/?map=https://users.aalto.fi/~bragge/experiencemaps/ASJCLow212mapfile.txt&network=https://users.aalto.fi/~bragge/experiencemaps/ASJCLow212networkfile.txt

Article search in Scopus:
(bibliometric OR scientometric OR “research profiling”) AND review



Publications trends 2004-2023:Source: Scopus, March 18, 2024

16 profiling articles since 2007, selected list:

- ✓ Roto, R., Bragge, J. Lu, Y. and Pacauskas, D. (2021), ”Mapping **Experience research** across disciplines: who, where, when”, *Quality and User Experience*,6(7), 1-26 https://link.springer.com/article/10.1007/s41233-021-00047-4ic47-4
- ✓ Bragge, J., Kauppi, K., Ahola, T., Aminoff, A., Kaipia, R. and Tanskanen, K. (2019), “Unveiling the intellectual structure and evolution of **External Resource Management** research: Insights from a bibliometric study”, *Journal of Business Research*, 97(4), 141-159. https://www.sciencedirect.com/science/article/pii/S0148296318306696
- ✓ Naukkarinen, O. and Bragge, J. (2016), “**Aesthetics** in the age of digital humanities”, *Journal of Aesthetics & Culture*, 8(1), https://www.tandfonline.com/doi/abs/10.3402/jac.v8.30072
- ✓ Bragge, J., Korhonen, P., Wallenius, H. and Wallenius, J. (2012) “Scholarly Communities of Research in **Multiple Criteria Decision Making**: A Bibliometric Research Profiling Study”, *International Journal of Information Technology and Decision Making*, 11(2), 401-426.
- ✓ Sunikka, A. and Bragge, J. (2012) “Applying Text-Mining to Profile **Personalization and Customization** Research – Who, What and Where?”*, Expert Systems with Applications*, 39(11)
- ✓ Leone, R., Robinson, L., Bragge, J. and Somervuori, O. (2012) “A Citation and Profiling Analysis of **Pricing Research** in 19 Marketing Journals from 1980-2010”. *Journal of Business Research*, 65(7), 1010–1024
- ✓ Bragge, J., Thavikulwat, P. and Töyli, J. (2010), Profiling 40 Years of Research in **Simulation & Gaming**, *Simulation & Gaming,* 41(6), 869-897.
- ✓ Bragge, J., and Storgårds, J, (2007) “Profiling Academic Research on **Digital Games** Using Text Mining Tools”, Proceedings of the *Digital Games Research Association’s Conference*, DiGRA.
- ✓ Bragge, J., Relander, S., Sunikka, A. and Mannonen, P. (2007) “Enriching Literature Reviews with Computer-Assisted Research Mining. Case: Profiling **Group Support Systems** Research”, *Proceedings of the 40th HICSS conference*.

A!

My first profiling article in 2007 was from my research focus at that time (GSS), next ones with doctoral students on their dissertation topics, and latest ones with various professors on their fields of expertise. One was also commissioned by the Editor of *Simulation & Gaming* for its 40-year anniversary issue.

Web mining

Sentiment Analysis

See an early 2010 text mining / big data example related to mobile phone brand discussion in Twitter at
http://www.youtube.com/watch?v=PSq7hZ0shLs

Web mining

Web mining is the application of data mining techniques to discover actionable and meaningful patterns, profiles, and trends **from Web resources**.

Web mining is used to understand customer behavior, evaluate a Web site's effectiveness, and quantify the success of a marketing campaign.

Research example: Bragge, J., Kallio, H. and Sunikka, A. (2012), "An Exploratory Study on Customer Responses to Personalized Banner Messages in the Online Banking Context", *Journal of Information Technology Theory and Application,* 13(3).

- Experiment in 2006 on a Finnish online bank, **3 different bank services were promoted with personalized marketing messages on the online bank** (after authentication to the service), ***see next slide***
- **Predictive analytics** was used for selecting experiment customers for group 2.
- **Click-stream analysis** was used to study the results (comparisons were made to default banners, and to direct mail marketing campaigns).

A!

Research example of web mining, continued

Table 1: Three Online Study Groups

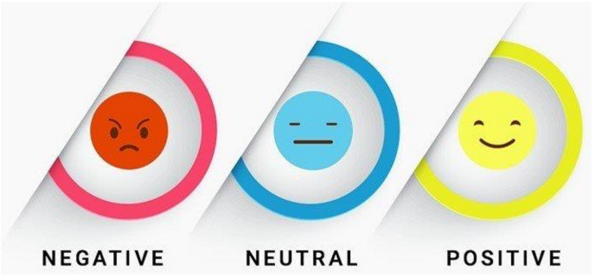
	Group 1. Net bank statement (NBS)	**Group 2: Loan**	**Group 3: X-card**
Promotional message	Message 1 (M1): Problems with archiving? Switch your bank account statement to the net.	M2: Have you considered that credit loans from banks are less expensive?	M3: Your X-card is about to expire. You can switch easily to Y-card on the net.
Number of customers	281	300	293
Criterion for customer selection	Customers had no electronic bank account statement service.	Customers only had mortgage loans from the bank.	The bank card that the customers were using was about to be withdrawn from the market.
Picture used in the personalized banner			

Click-stream analysis was used to study the results: comparisons were made to default banners, and to direct mail marketing campaigns.

Personalized banners were more effective than other options, except in Group 3.

Source: Bragge, Sunikka and Kallio (2012), *JITTA*

Sentiment analysis



"**Sentiment analysis** is a type of text research aka mining. It applies a mix of **statistics**, **natural language processing** (NLP), and **machine learning** to identify and extract subjective information from text files, e.g., a reviewer's feelings, thoughts, judgments, or assessments about a particular topic, event, or a company and its activities.

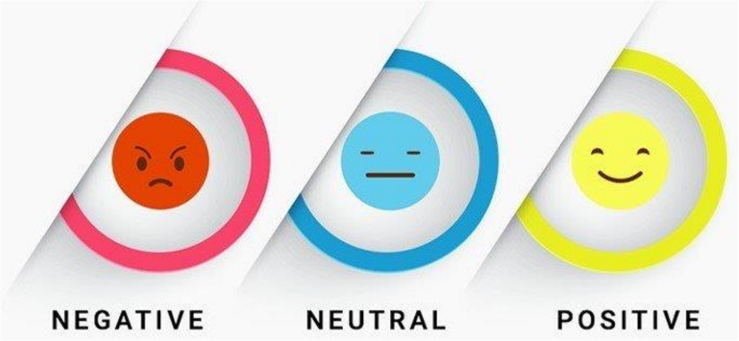
This analysis type is also known as ***opinion mining*** (with a focus on extraction) or ***affective rating***. Some use the term *sentiment classification* and *extraction* as well.

Regardless of the name, the goal of sentiment analysis is the same:
to know a user or audience opinion on a target object by analyzing a vast amount of text from various sources."

A!

Source: "Sentiment analysis: types, tools and use cases", https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases, 21.8.2018, 11-minute read

Use cases of sentiment analysis



- Brand monitoring
- Competitive research
- Flame detection and customer service prioritization
- Product analysis
- Market research and insights into industry trends
- Workforce analytics / employee engagement monitoring

A!

Source: ”Sentiment analysis: types, tools and use cases”,
https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases, 21.8.2018

Research applications in Sentiment analysis

Research applications	References
Spam detection	Jacob et al. [149], Tida and Hsu [150], Magdy et al. [151], Rodrigues et al. [148], Oswald et al. [152], Kanmani and Balasubramanian [153]
Health reviews	Ramirez et al. [154], Akhtyamova et al. [155], Babu and Kanaga [156], Basiri et al. [157], Egger et al. [146], Edara et al. [158]
Business analytics	Desai et al. [159], Habbat et al. [160], Ahmed et al. [161], Luo et al. [162], Kanan et al. [163]
Recommendation system	Serrano et al. [164], Prabakaran et al. [165], Karn et al. [166], Choudhary et al. [167], An and Moon [168]
Market research analysis	Rambocas and Pacheco [169], Micu et al. [170], Puavualoaia et al. [171], Kyaw et al. [172]
Stock market prediction	Xu and Keselj [173], Jin et al. [174], Wu et al. [175], Jing et al. [176], Zhao and Yang [177], Bl and Br [178]

Sahoo, C., Wankhade, M., & Singh, B. K. (2023). Sentiment analysis using deep learning techniques: a comprehensive review. *International Journal of Multimedia Information Retrieval*, *12*(2), 41. https://link.springer.com/article/10.1007/s13735-023-00308-2

Research applications in Sentiment analysis

Reviews [146], Business Data Analysis [147], and Spam Detection [148]. The popular research applications in sentiment analysis are outlined in Table 5.

6.1 Applications

Spam detection In light of technological advancements and the fourth industrial revolution, the majority of businesses and organizations have acquired electronic commerce platforms, which have expanded the use of online marketing through user reviews [149]. Sometimes the reviews are fake and can mislead the customers about a particular product or service [150]. There are several techniques that use deep learning methods to detect spam emails with better accuracy results [151].

Health reviews Sentiment analysis is becoming increasingly popular in the medical domain as it helps us to access information about mental disorders, epidemics, and patient emotions in order to provide them with better healthcare facilities [154]. The researchers in [155] proposed an extensive set of CNN as a method of predicting drug safety which is based on user feedback from healthcare discussion forums. In another study, Nirmal et al. [156] used deep learning methods, and their analysis measured the individual's depression scale by analyzing and retrieving emotions as text features from various social media platforms.

Business analytics In business intelligence, sentiment analysis offers several benefits as businesses may use sentiment analysis information to enhance services, evaluate customer suggestions, and generate new marketing objectives. Most often, sentiment analysis is utilized in business intelligence to assess customer perceptions of a product or service, which helps the customers to improve their decision-making skills [159]. Moreover, business intelligence technologies are useful for identifying and comparing the relevant topics and patterns between various social media posts or products [160].

Recommendation system Recommender systems use deep learning-based algorithms to predict an item's rating or preference for a certain user [164]. It is employed to handle any type of online overloading problem that arises between customers and enterprises. There are many examples of recommender systems that have gained popularity in recent years, such as the ones employed by Amazon and Netflix [165]. Alatrash et al. [179] emphasized the formation of a recommendation system by using deep learning-based methods. They demonstrated that their approach produces high-quality results in a number of recommendation contexts and is also capable of incorporating diverse recommendations.

Market research analysis Analysis of market research is one of the most prevalent applications of sentiment classification. Companies apply deep learning techniques to examine customer feedback and opinions regarding their goods or services [169]. By doing so, they can gain a deeper understanding of their target audience and their needs, and recognize current trends and sentiments, enabling them to upgrade their products and services. The target of the market research is to find the leading competitor and compare marketing campaigns.

Stock market prediction According to recent studies, various deep learning algorithms have exhibited high accuracy in predicting and classifying stock prices. Xu and Keselj [173] used LSTM as the deep learning method, which was developed to estimate the next day's stock price return. Their experimental results indicate that these approaches can be used to ensure a sustained profit in the short run, even in developed markets. A hybrid approach that merges deep learning-based methods was developed by [176] to predict stock prices, and results showed that it outperformed a single deep learning method in terms of prediction accuracy.

Sahoo, C., Wankhade, M., & Singh, B. K. (2023). Sentiment analysis using deep learning techniques: a comprehensive review. *International Journal of Multimedia Information Retrieval*, *12*(2), 41. https://link.springer.com/article/10.1007/s13735-023-00308-2

Opinion mining and sentiment analysis, book and other resources
by Distinguished Computer Science Prof. Bing Liu

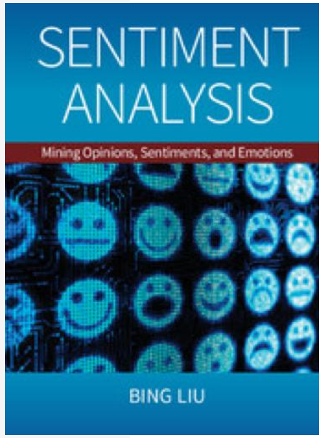
https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

Search

Opinion Mining, Sentiment Analysis, and Opinion Spam Detection

Feature-Based Opinion Mining and Summarization
(or Aspect-Based Sentiment Analysis and Summarization)
Detecting *Fake Reviews*
(Media coverage: The New York Times, The Economist, BusinessWeek and more ...)

Opinion Lexicon --------- Datasets --------- Talks --------- Publications



***New Book*: Sentiment Analysis: mining opinions, sentiments, and emotions. Cambridge University Press, 2015.**

Book: Sentiment Analysis and Opinion Mining (Introduction and Survey), Morgan & Claypool, May 2012.

See "Feature-Based Opinion Mining and Summmarization" in Microsoft Live/Bing Search and Google Product Search (paper).

- **Note**: I don't know the techniques used by **Microsoft Live/Bing** (9/28/2007), but Google has a paper. To see the model, please check out (Hu and Liu, KDD-2004) and (Liu et al, WWW-2005) below, or the books above (better). Try search for a camera and click on reviews. You will see summarized user opinions on product features/aspects in a bar chart.

NLP Handbook Chapter: Sentiment Analysis and Subjectivity, 2nd Edition, Eds: N. Indurkhya and F.J. Damerau, 2010.

***Opinion Parser*: my sentiment analysis system has been licensed to two companies.**

- The system analyzes **sentiments, opinions** and **emotions**, extracts **sentiment targets**: **entities, topics** and their **aspects/features**, and handles **comparative sentences**.
- I cannot make the system open-source due to its commerical use. If you want to know how it works, please read my new sentiment analysis book, which gives a lot of details.

Tutorial: Sentiment Analysis Tutorial - (references), given at AAAI-2011, August 8, 2011 - (Check out the new book)

Interesting Piece from New Republic: If you want to be a successful novelist, should you be sentimental in your writing or not?

Recent Keynote and Invited Talks (not updated) (Older Talks)

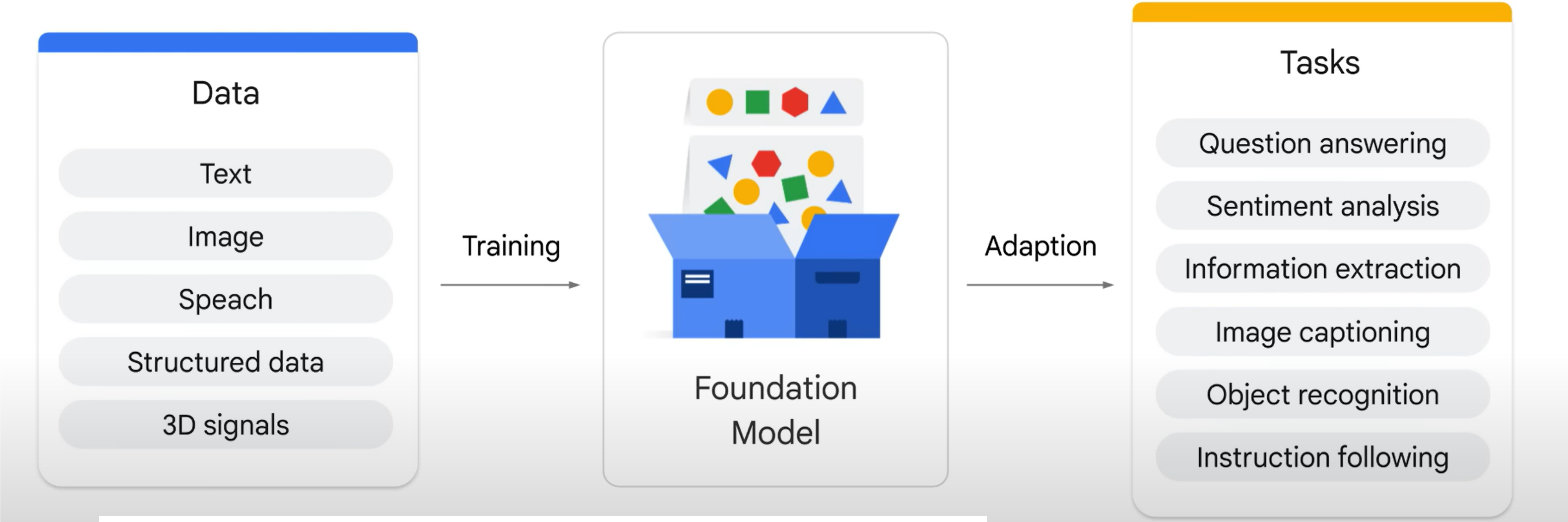
1. Invited Talk. “Sentiment Analysis with Lifelong Learning.” ETS, December 7, 2015.
2. Invited Talk. “Sentiment Analysis with Lifelong Learning.” Brigham Young University, December. 3, 2015.
3. Keynote speech. “Sentiment Analysis, Lifelong Learning and Intelligent Personal Assistants.” The 2015 Conf. on Technologies and Applications of Artificial Intelligence (TAAI-2015). Taiwan, Nov. 20-22, 2015.
4. Invited talk. “Sentiment analysis and lifelong machine learning.” Frontiers in Computational Mathematics: AMS Central Fall Sectional Meeting, October 2-4, 2015.

A!

https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

Sentiment analysis using generative AI

While generative AI models offer significant advantages, they may not completely replace traditional SA methods but rather **complement them**. Traditional methods still have their place, especially in scenarios where interpretability, transparency, and domain-specific knowledge are crucial. The integration of generative AI into sentiment analysis represents an evolution of the field, enhancing the tools available to researchers and practitioners.



Screenshot from Introduction to Generative AI (youtube.com), at 17:25 mins

MS Co-pilot (18.3.2024): Prompt “Can generative AI models replace traditional methods in sentiment analysis?”

See also:; Krugmann, J. O., & Hartmann, J. (2024). Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, *11*(1), 1-19.

The generative AI application Landscape

But while generative models can achieve incredible results, they aren’t the best choice for all types of data. For tasks that involve making predictions on structured data, like the tabular data in a spreadsheet, generative AI models tend to be outperformed by traditional machine-learning methods, says Devavrat Shah (Professor in Electrical Engineering and Computer Science at MIT).

“**The highest value they have, in my mind, is to become this terrific interface to machines that are human friendly.**

Previously, humans had to talk to machines in the language of machines to make things happen. Now, this interface has figured out how to talk to both humans and machines,” says Shah.

Explained: Generative AI | MIT News | Massachusetts Institute of Technology

A!

Text-Mining today

“Industrial needs involving text mining and natural language processing (NLP) have grown in demand, triggering the development and growth of algorithms which could run on unstructured data.”

“Chatbots traditionally used NLP to respond to queries raised by the user, while mapping it to the best possible response sets available in the system. In order to provide real time feedback to customers, **chatbots have adopted language models along with deep learning** while addressing NLP problems. The recent launch of OpenAI’s ChatGPT significantly extends the capabilities of chatbots via the integration of deep learning and language models based on the **Generative Pre-training Transformer** (**GPT**) architecture. **Language models attempt to predict the likelihood of a sequence of words** a typical human interaction is likely to create through generative and discriminative algorithms, typically through the application of deep learning and transformer architectures of neural networks. **ChatGPT uses a combination of unsupervised pre-training and supervised fine-tuning to generate human-like responses to queries and provide responses to topics that resemble that of a human expert**.”

Source: Dwivedi, Y. et al. (2023),“So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy”, *International Journal of Information Management*, 71, 102642 https://www.sciencedirect.com/science/article/pii/S0268401223000233/

Information
Technology
Program
Aalto University

Minor program for the Digital Future

WHAT IS ITP?

The Information Technology Program (ITP) is a masters-level, summer minor program, organized by the Aalto School of Business.

It gathers a multidisciplinary and international cohort with the goal of solving a real-life business challenge in 3 months.

How many credits? 24 - 30 ECTS

Where? Aalto Campus

When? Summer 2024. From 3.6. to 30.8.
Lectures 4 days/week, from 9 - 12

A!

Aalto University School of Business

Information Technology Program

SERVICE & EXPERIENCE DESIGN

User-centered design and UX specialization track.

Design Strategy

Designed to provide students with advanced design-thinking practices and methods and how to transmit a design-driven approach in teams and organizations.

Service Creation

Hands-on experimentation with the process of Service Design – with a twist towards experimental, concrete, collaborative, and visual ways of working.

UX Design

The course covers the basics of UX in the digital product development process from evaluation and user discovery to user interface and visual design principles. Using Figma software, you take learning into practice by building and presenting prototypes and other design deliverables.

A!

Aalto University School of Business

Information Technology Program

INFORMATION & SERVICE BUSINESS

IT consulting and business models, new digital business opportunities and the information ecosystem

Digital Service Innovations

The course provides students with skills to manage the business side of a software business, focusing on the key issues faced by SME’s and start-ups over the lifecycle of a software venture.

Strategic IT management

This course explores the implementation of business models, strategies and tactics made possible by modern information technology.

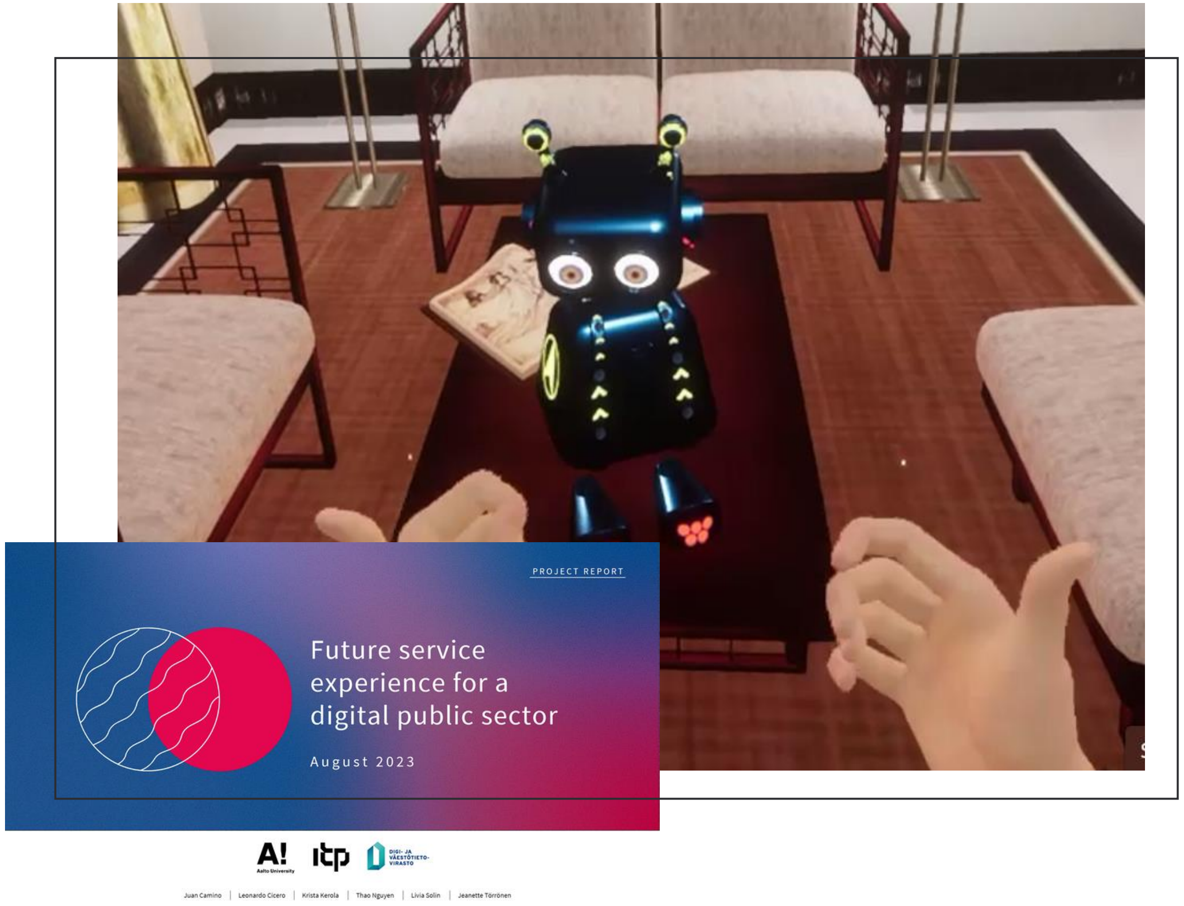
Data Driven Decision making

On this course you will learn how to identify different data resources, differentiate between numbers and numbers that matter, basics of data visualization and how to persuade with data.

ITP BUSINESS PROJECT

Provides a key learning platform for interdisciplinary teamwork in real-life industry projects, lasting approximately 3 months

Business partners assign students to an open-ended project to develop innovative concepts using design thinking, data analytics, and agile methodologies.



Previous ITP projects topics include:

Competitor analysis & positioning
Robotics process automation in communication
Systems architecture for cross-collaboration in the public sector

Assesment of AI maturity
Mapping digital service user needs
ESG data visualization
UX design of public services
& more

INTERESTED?

MORE INFO & APPLICATIONS

NEWSLETTER

Information Technology Program
Aalto University

Minor program for the Digital Future

Victor de Bruin
Program manager

victor.debruin@aalto.fi
+358 50 475 7314

Nikolay Efimov
ISB Track coordinator

nikolay.efimov@aalto.fi
+358 40 353 8085

Juan Camino
SED Track coordinator

juan.camino@aalto.fi
+358 46 558 1446

BREAK 10 minutes

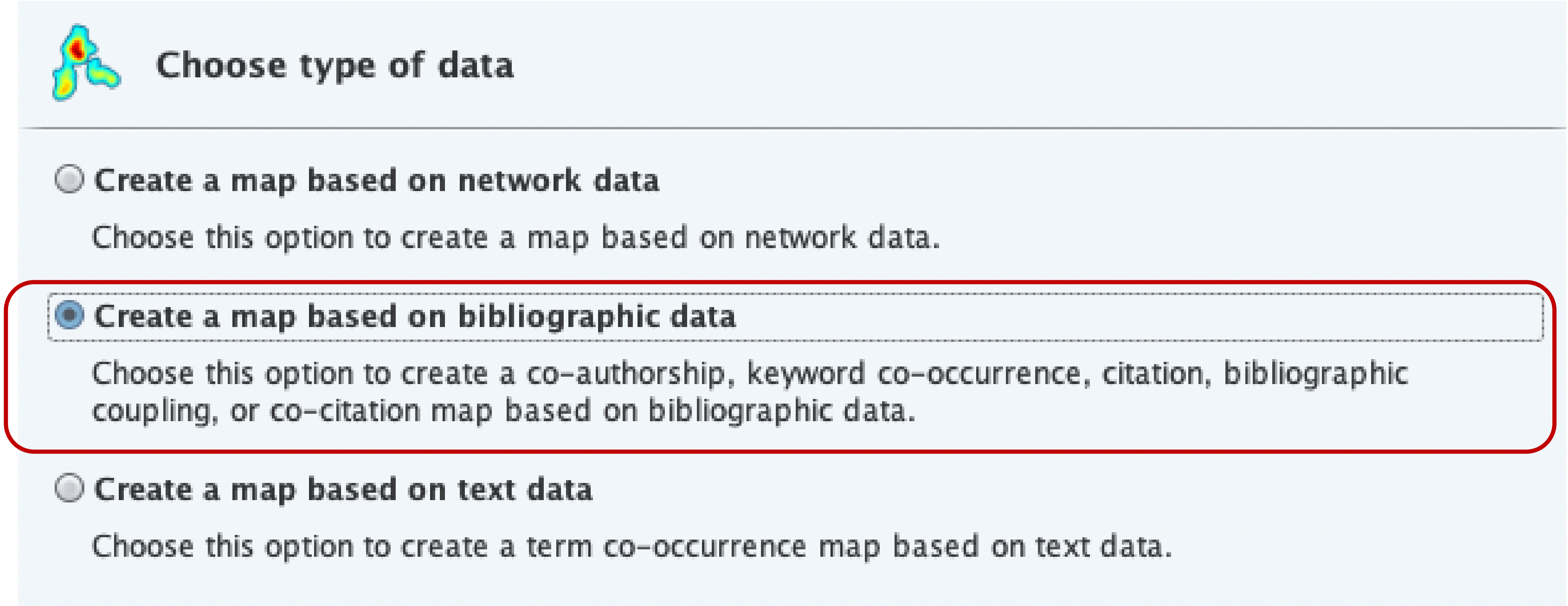
Demo and further options

https://www.aalto.fi/en/learning-centre > search Scopus
https://libguides.aalto.fi/business > Articles tab > Scopus
Search phrase from Assignment 7 (from 2010->): **"digital sustainability" OR "sustainable digitalization" OR "sustainable digitalisation" OR "twin transformation" OR "twin transition" OR "dual transformation" OR "dual transition" OR digitainability OR (“digital transformation” AND “sustainable transformation”)**

Note: the Scopus CSV demo file can be downloaded from the web directly!
https://users.aalto.fi/bragge/scopus (digital sustainability 2010 onwards).csv

For later reference:
My 17 min tutorial video on the basics of research profiling with Scopus:
https://aalto.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=d53e46d3-85f5-4d33-ac5a-aa2b00a5e9b5

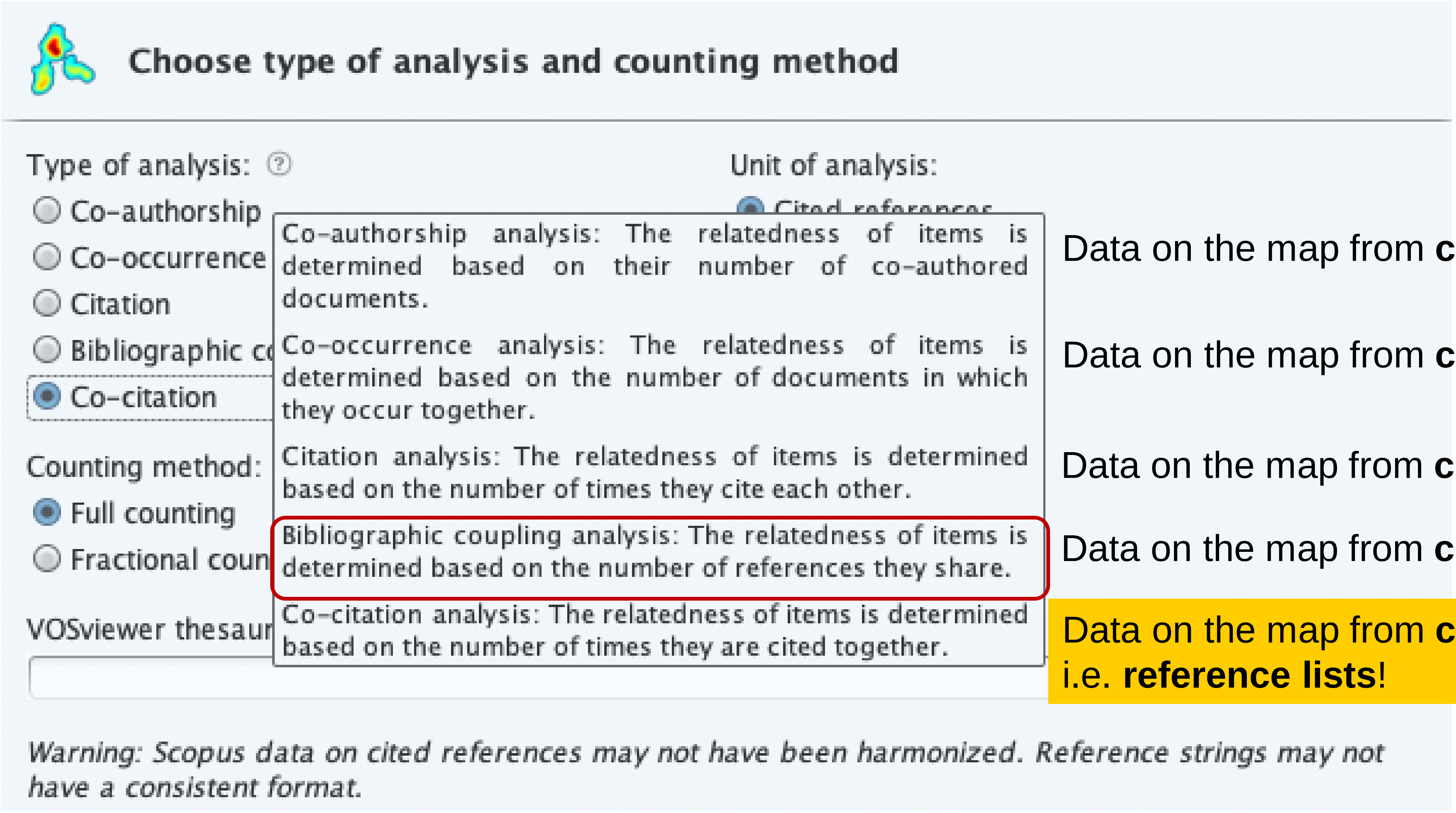
Bibliographic analysis options in VOSviewer 1/2



"VOSviewer is a software tool for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature". https://www.vosviewer.com/

A!

Bibliographic analysis options in VOSviewer 2/2



Data on the map from **citing** research

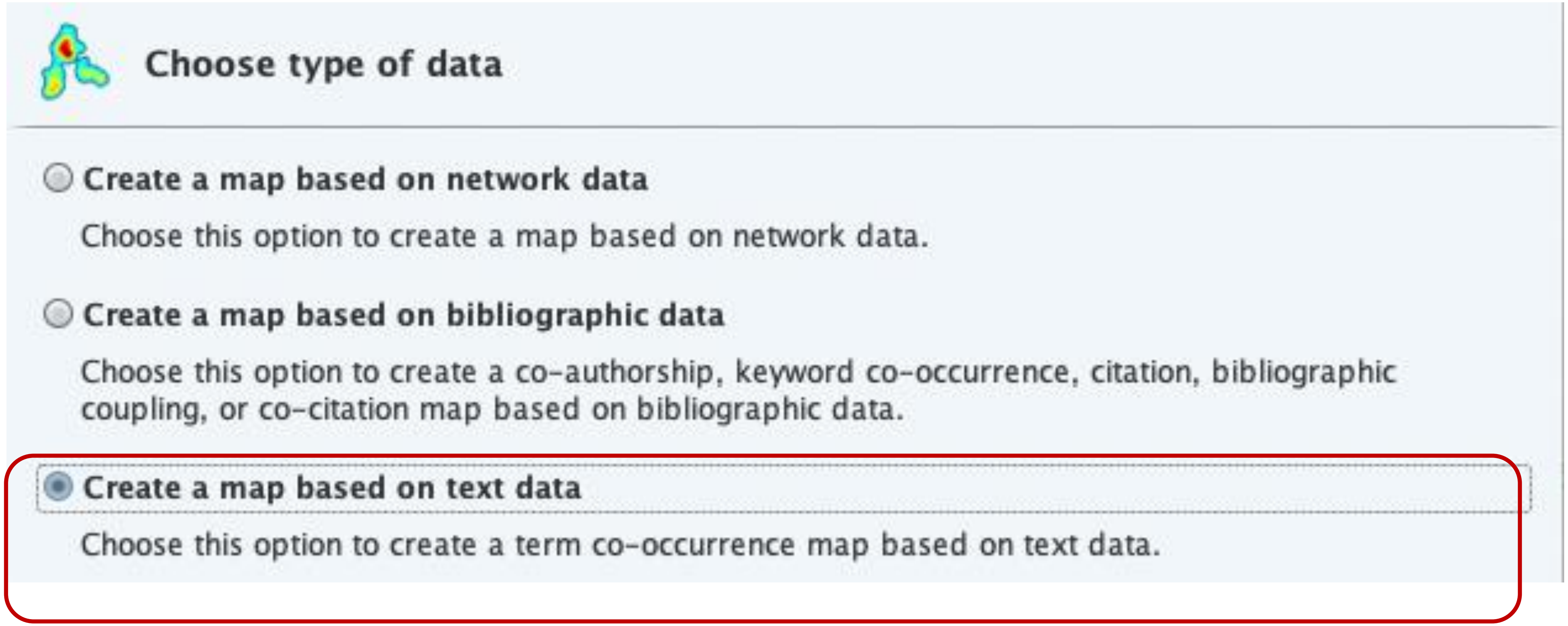
Data on the map from **citing** research

Data on the map from **citing** research

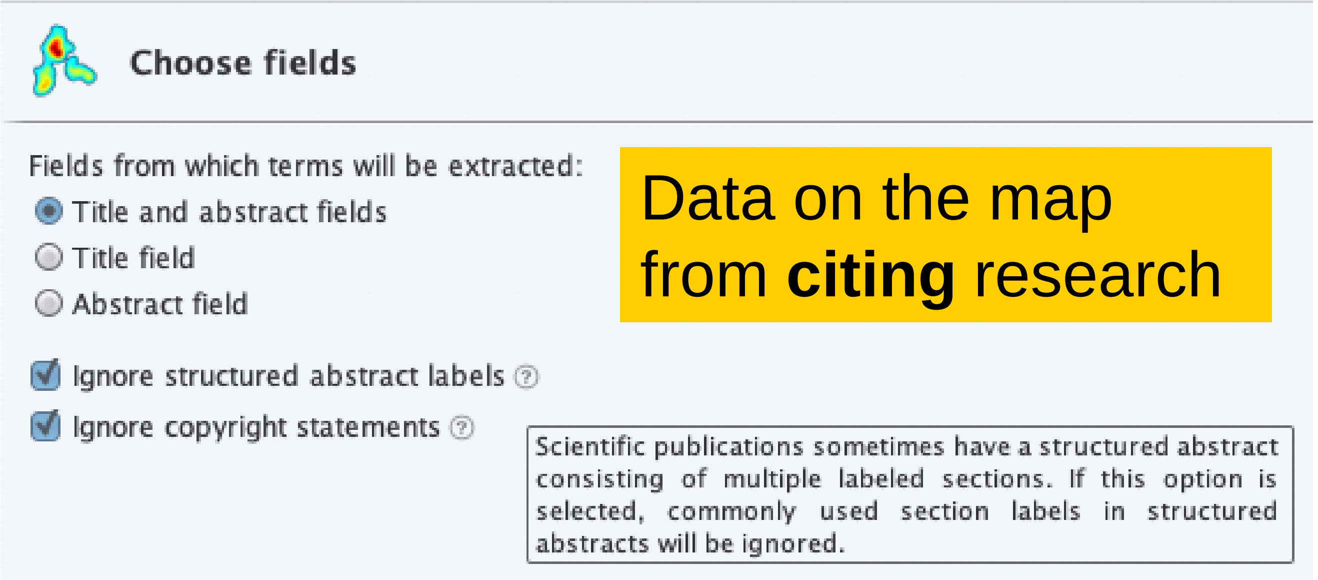
Data on the map from **citing** research

Data on the map from **cited** research, i.e. **reference lists**!

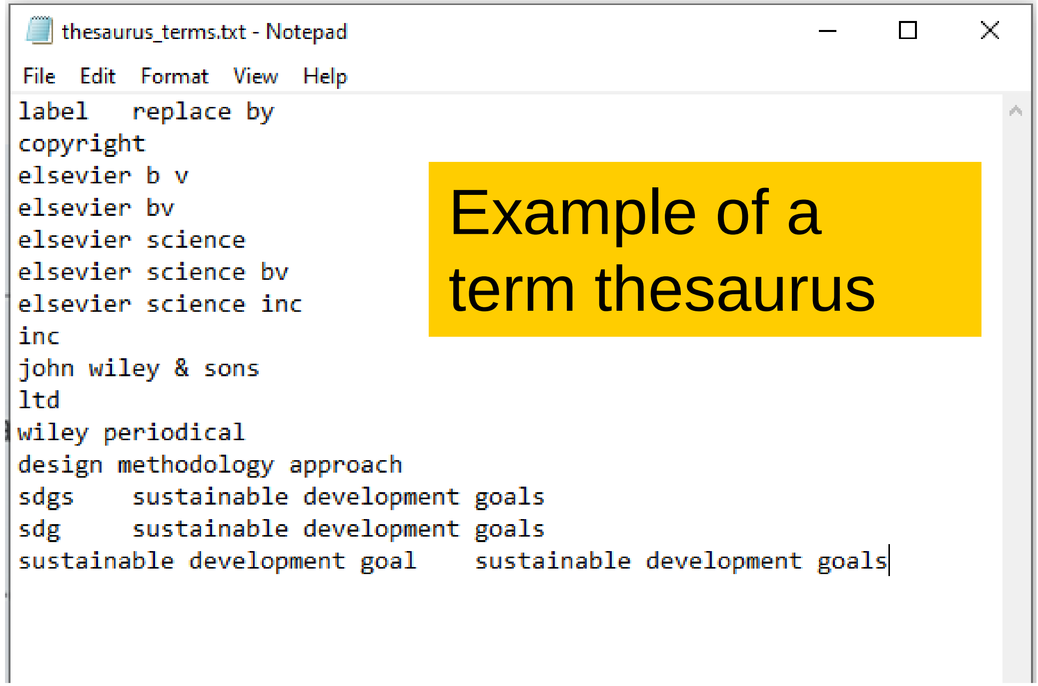
Term co-occurrence analysis in VOSviewer 1/2



Term co-occurrence analysis in VOSviewer 2/2



Data on the map from **citing** research



Example of a term thesaurus

Kiitos

aalto.fi