UNLOCKING THE VALUE OF PATENT DATA

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO)
FIRST PATENT ANALYTICS COMMUNITY OF PRACTICE (COP) ANNUAL SYMPOSIUM

17-18 SEPTEMBER 2024
GENEVA, SWITZERLAND

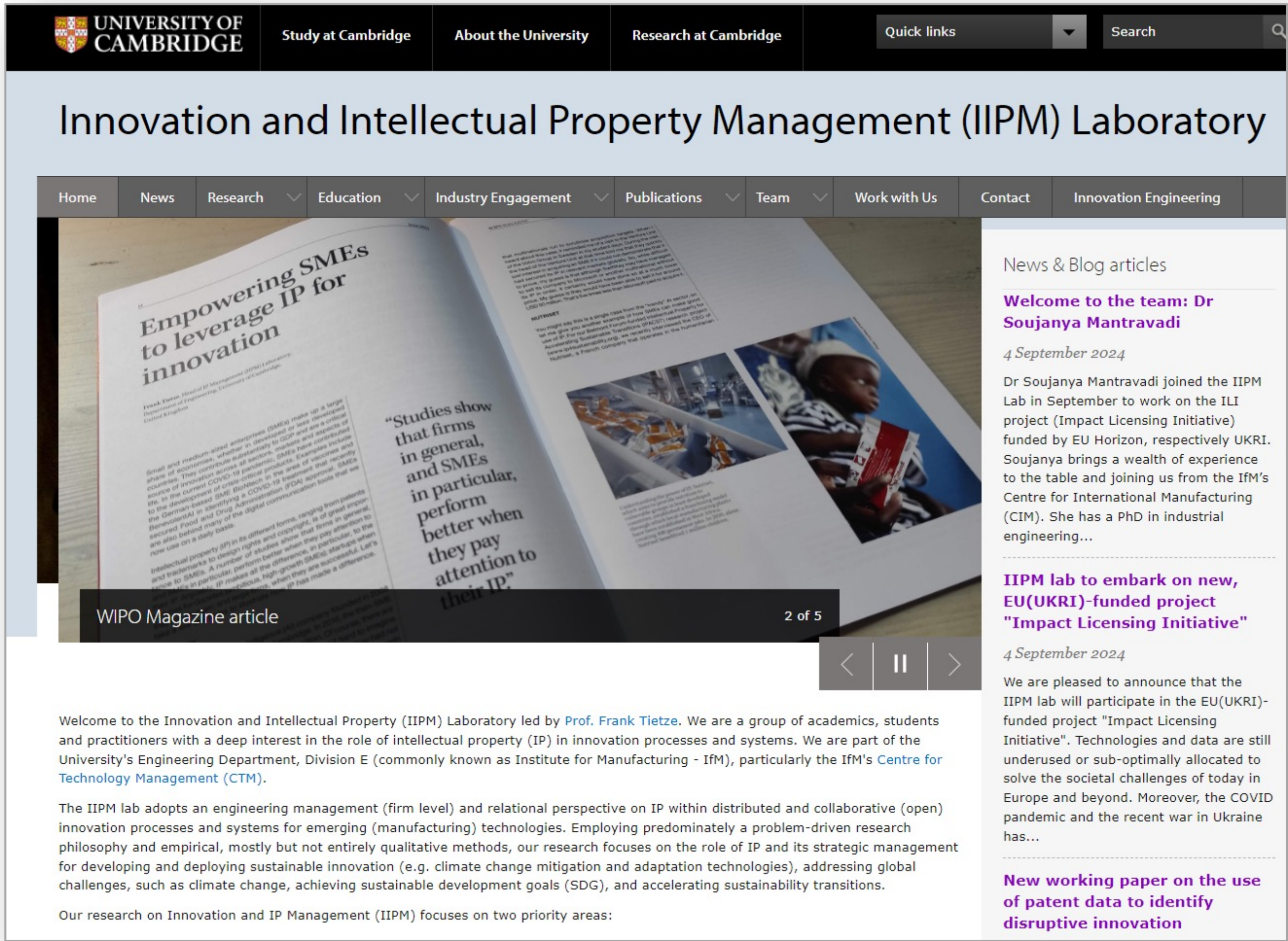
Frank Tietze
Professor of Innovation Engineering
frank.tietze@eng.cam.ac.uk

Head, Innovation and IP Management (IIPM) Laboratory
Centre for Technology Management (CTM)
Department of Engineering

UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

WWW.IIPM.ENG.CAM.AC.UK



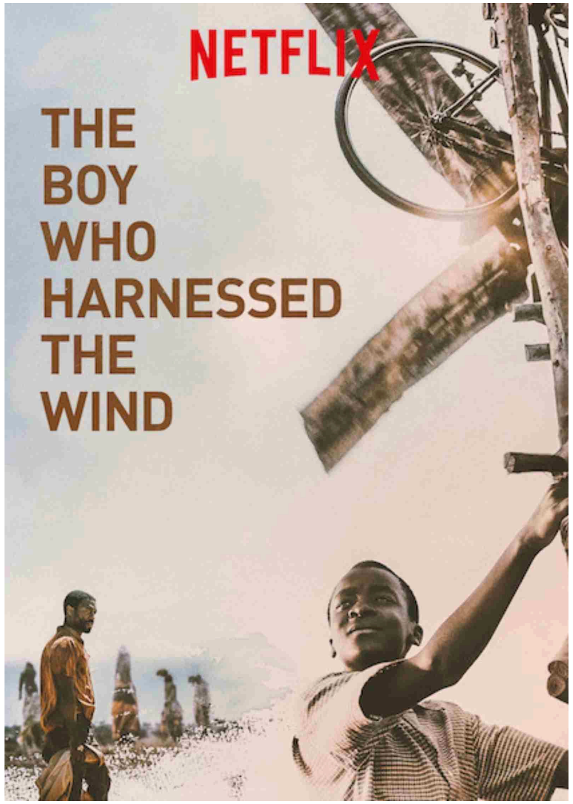
UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

CHIEF ECONOMISTS PANEL AT RECENT EUROPEAN POLICY FOR INTELLECTUAL PROPERTY (EPIP) CONFERENCE

© F. Tietze (2024)

REMINDER: THE ORIGINAL PURPOSE OF PATENT SYSTEMS

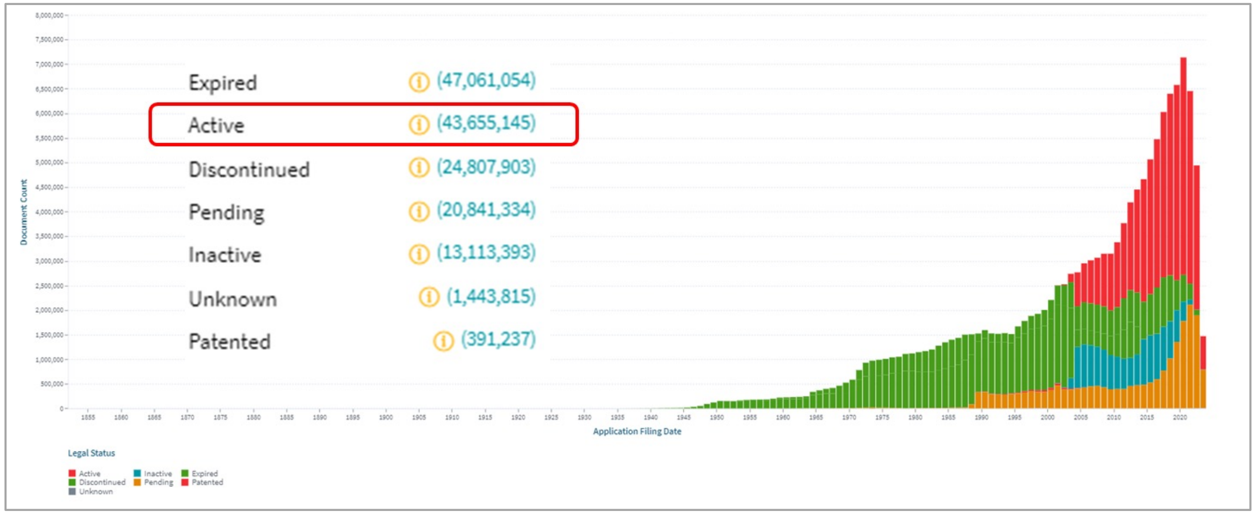


Imagine what William could have done with access to patent data…

PATENT DATABASES ARE THE OLDEST AND WORLD'S LARGEST OPEN DATA REPOSITORIES FOR TECHNICAL SOLUTION KNOWLEDGE

Image generated with https://chatgpt.com

85,872,443 patent families on file [1]



UNIVERSITY OF CAMBRIDGE
Department of Engineering

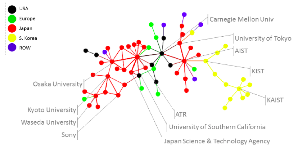
[1] www.Lens.org (30.10.2023)

IfM Innovation and IP Management Lab

PATENT DATA PROVIDED BY THE OFFICES AND DIGITIZATION CREATED AN INDUSTRY

Global patent analytics market size valued at $1.00bn (2023), projected to grow to $3.02bn (2032) [1]

- High search costs
- Focus on individual patents
- Patent data as tool for large firms
- Large value digitising raw data
- 1984: BACON project by the trilateral offices (US, EPO, JP)



- Patent data accessible in a cloud
- Surge in innovative new entrants
- Lower search costs for accessing patent data
- Shift in value from data possession to analytical capabilities

- LLM based patent generation and analysis
- Technology race between filers and offices
- Decentralization of the IP systems allows ownership transfer tracking

1) Analog, pre-digital patent data era

2) The early digital patent data era

3) Growth explosion and consolidation of patent analytics

4) AI for patent analytics

5) The future, eg Blockchain and AI

<1990 | ~2000s | ~2010s | ~today | t

- Patent documents are mainly digital PDFs, not cloud-based
- Decreasing value in data-driven business models
- Predominance of established patent data software providers
- Limited large-scale patent data analysis capabilities

- Patent data fully democratized
- focus on large-scale analysis
- Significant reduction in patent data search costs via free tools
- Patent analytics accessible to non-IP experts
- New AI-compatible architectures by entrants enhance data analysis

2005: My very first patent analysis ;)

See full slide deck: Tietze, F. (2021) A short history and outlook for the patent data and analytics industry, EPO Patent Knowledge Week www.linkedin.com/feed/update/urn:li:activity:6861593777119735808/

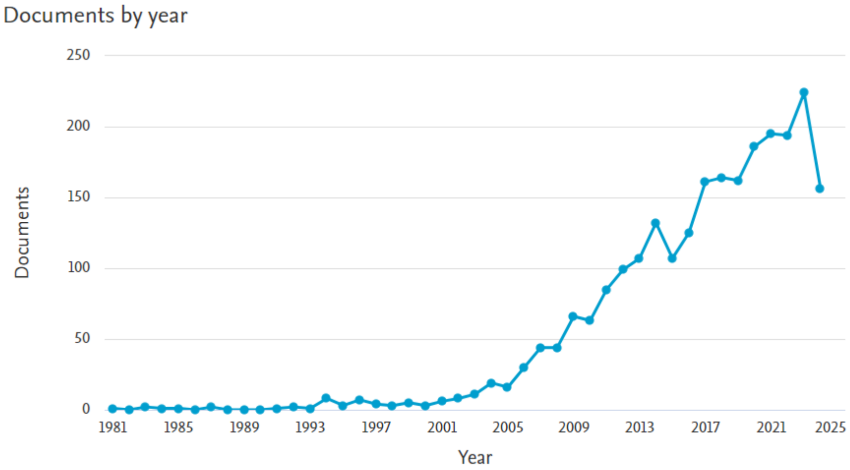
[1] Patent Analytics Market Size, Global Growth Report 2032 www.fortunebusinessinsights.com/patent-analytics-market-102774

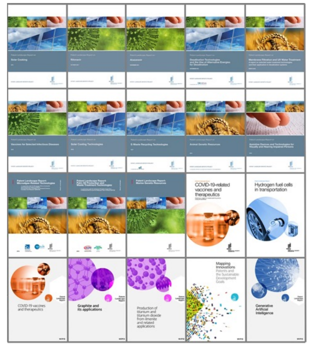
UNIVERSITY OF CAMBRIDGE Department of Engineering

IfM Innovation and IP Management Lab

DEMOCRATIZING PATENT ANALYSIS

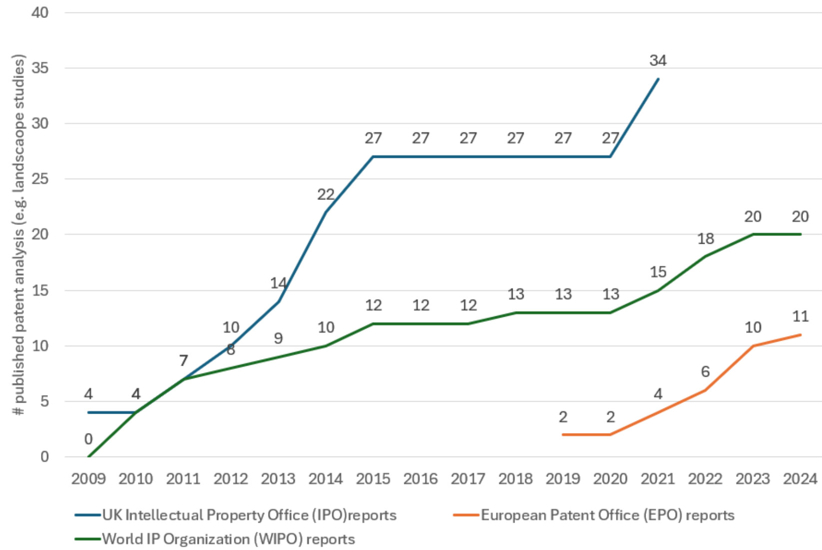
2,425 academic publications [1]





https://www.epo.org/searching-for-patents/business/patent-insight-reports.html

Patent office patent studies



UNIVERSITY OF CAMBRIDGE
Department of Engineering

[1] Source: Scopus.com, publications with "patent analysis" in key, title or abstract (extracted: 15.09.24)

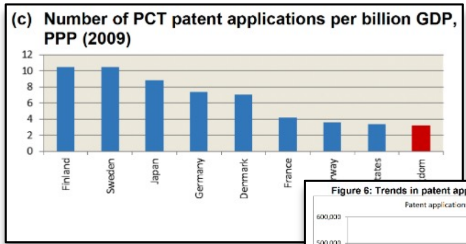
IfM Innovation and IP Management Lab

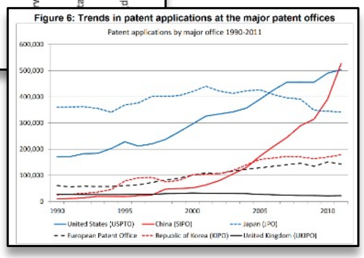
HOW TO MAXIMISE THE SOCIAL VALUE OF PATENT DATA?

Image generated with https://chatgpt.com

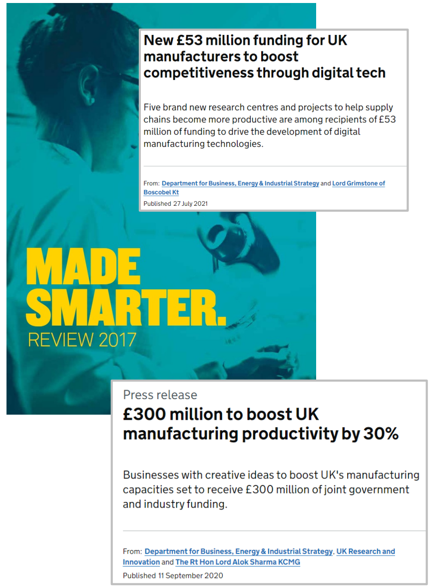
INDIRECT USE OF PATENT ANALYSIS FOR POLICY MAKING

Out of 37 background studies, 6 used patent data





UK industrial and innovation policy



UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

IMPACT OF A TRADEMARK ANALYSIS ON IP LAW

WIPO Economics & Statistics Series

September

2014

Economic Research Working Paper No. 22

Trademarks Squatters: Evidence from Chile

Carsten Fink
Christian Helmers
Carlos Ponce

WIPO
WORLD INTELLECTUAL PROPERTY ORGANIZATION

See also: Fink, C., Helmers, C., Ponce, C.J., 2018. Trademark squatters: Theory and evidence from Chile. International Journal of Industrial Organization 59, 340–371. https://doi.org/10.1016/j.ijindorg.2018.04.004

European Commission

Search

IP Helpdesk

Home | Services | Regional helpdesks | IP management and resources | News & Events | About

Home > News & Events > News > New IP Law in Chile (II): main changes in the trade mark system

NEWS BLOG | 7 October 2022 | 10 min read

New IP Law in Chile (II): main changes in the trade mark system

4) Inclusion of cancellation procedures

As already mentioned under point 2, a new Article (27 bis, item A) states that a trade mark will be cancelled under certain circumstances, if:

1. the trade mark is not used by its owner for a period of five years prior to the date of filing of the cancellation action; or
2. the registrant has tolerated that the trade mark be considered as the usual designation for a determined product or service, losing its distinctiveness. This does not apply if the registrant has consistently used the usual indications of a registered trade mark.

https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/new-ip-law-chile-ii-main-changes-trade-mark-system-2022-10-07_en

THE USE OF PATENT ANALYSIS IN THE DUE DILIGENCE OF MERGERS

Dow and DuPont (2017)

EC carried out patent analysis which confirmed:

1. High importance of both merging parties, and in particular one merging party, as innovators;
2. High degree of concentration in research for new AIs (discovery stage);
3. Significant combined share of research for new AIs accounted by the merging parties, notably in selective herbicides and insecticides; and
4. Closeness between the merging parties in term of innovation efforts.

- Divesting a large part of DuPont's herbicide and insecticide businesses + DuPont's global R&D organisation [1]

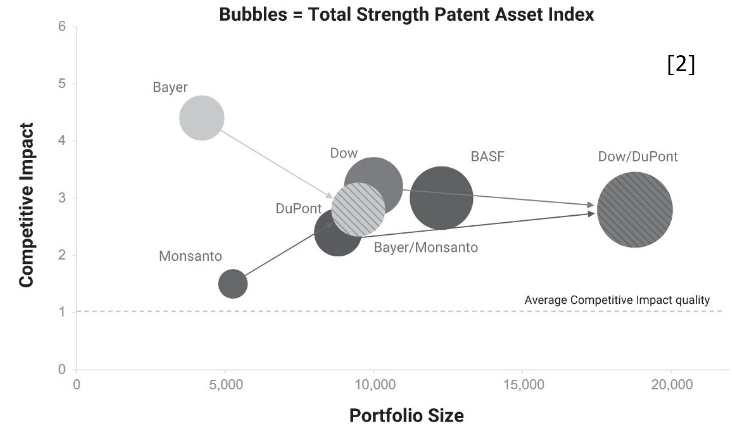


Figure 12.2 Process of concentration in the chemical industry
Source: Own illustration, data provided by PatentSight GmbH, a LexisNexis company (2018).

[1] Buehler, B., Coublucq, D., Hariton, C., Langus, G., Valletti, T., 2017. Recent Developments at DG Competition: 2016/2017. Rev Ind Organ 51, 397–422.

[2] Illustration for Bayer/ Monsanto merger (M.8084) by Ernst, H., Guderian, C.C., Richter, M., (2022) The Innovation Environment and Knowledge Diffusion: Improving Policy Decisions through Patent Analytics, in: Taubman, A., Watal, J. (Eds.), Trade in Knowledge. Cambridge University Press, pp. 376–402.

With thanks to R. Veugelers. See also: (2019) "Innovation/IP in Competition Policy", EPIP Conference, ETH, Zurich

LEVELS OF PATENT ANALYSIS AND EXAMPLE QUESTIONS

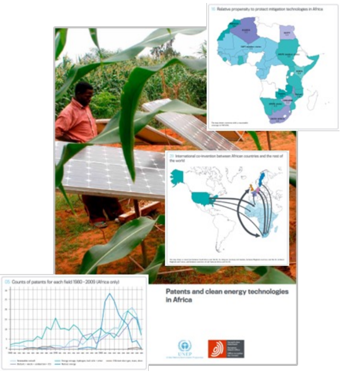
- **Country and regions - geographical (via NUTS codes)**
  (Where is development of a certain technology happening?)
- **Industry (or group of organisations)**
  (Who are the major players in an industry? Which directions are they going?)
- **Technology (IPC/CPC via concordance tables)**
  (Who are key developers for a certain technology?
- **Individual entity level (Individual organisations e.g. universities, corporates, business; inventors and inventor teams)**
  (What is the portfolio, inventive activity of certain entities? Is this a valuable acquisition target?)
- **Portfolio**
  (Which patents are renewed? What is a portfolio worth? Which are the valuable patents?)
- **Product**
  (Which patents / what IP is used in which product?)
- **Patent / IP right level**
  (What are the most relevant patents in a technology? What is the value of a patent? Who would want to license this patent?)

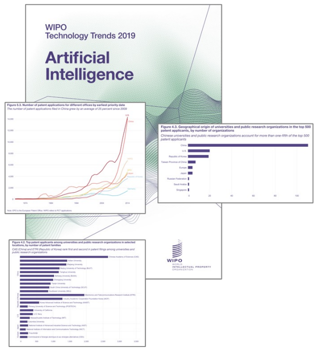
DIFFERENT TYPES OF PATENT STUDIES

- Full (dedicated) patent analytics reports vs. partial background (economic) studies (e.g. as input to a larger study)
- Descriptive patent studies (often gov reports) vs. econometric, often academic studies using patent data as one variable correlated with other variables
- “Pure” patent vs. multi-data studies (incl. patent data)
- Actor studies vs. “relational” studies (e.g. network analysis)
- Published vs. unpublished (hidden) patent analysis
- Commissioned (e.g. by EC) vs. “free” studies
- Single country vs. comparative international studies
- …

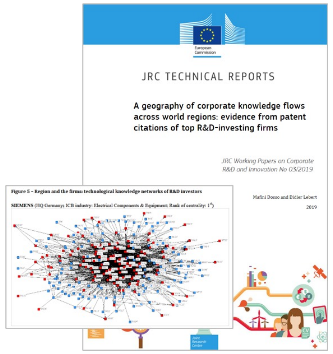
WHERE DOES PATENT ANALYTICS START AND END?
WHERE DOES THE COP WANT TO GET INVOLVED?

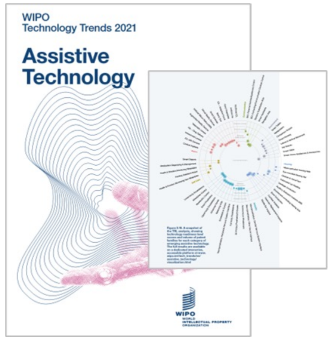
Landscape and trend studies tend to present descriptive results





More advanced analysis, e.g. knowledge flows, tech transfer, TRL estimation

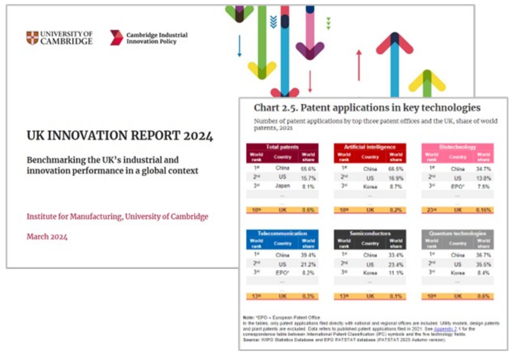




Patent data as part of compound indicators

Studies with combined data that require dataset matching

The many innovation policy/ foresight reports that include some form of patent analysis

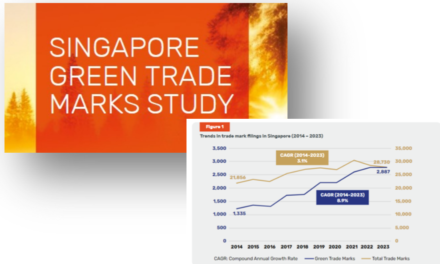


UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

INCREASINGLY DATA IS AVAILABLE FOR OTHER IP RIGHTS THAT COULD BE ANALYSED

Trademarks




GREEN EU TRADE MARKS – 2023 UPDATE

EUIPO

Green EU trade marks - 2023 update

April 2024

Geographical indications

	2020	2012
Italy	*297*	*248*
France	*249*	*192*
Spain	*192*	*161*
Portugal	*138*	*118*
Greece	*104*	*97*
Germany	*91*	*89*
UK	*69*	*46*
Poland	*31*	*35*
Czech Republic	*29*	*28*
Croatia	*24*	*0*
Slovenia	*22*	*16*
Austria	*15*	*14*
Belgium	*15*	*13*
Hungary	*14*	*12*
Slovakia	*12*	*10*
the Netherlands	*11*	*9*
Denmark	*8*	*5*
Finland	*7*	*8*
Ireland	*7*	*4*
Lithuania	*7*	*2*
Romania	*7*	*1*
Sweden	*6*	*6*
Cyprus	*5*	*2*
Luxembourg	*4*	*4*
Bulgaria	*3*	*1*
Latvia	*3*	*0*
Andorra	*1*	*0*

Note(s): Blocks represented by colour, South, North, Central East

Joosse, S., Olders, P., Boonstra, W.J., 2021. Why are geographical indications unevenly distributed over Europe? British Food Journal 123, 490–510.

Design rights

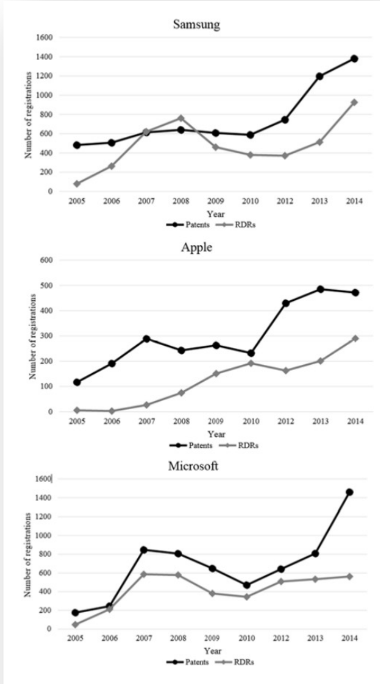
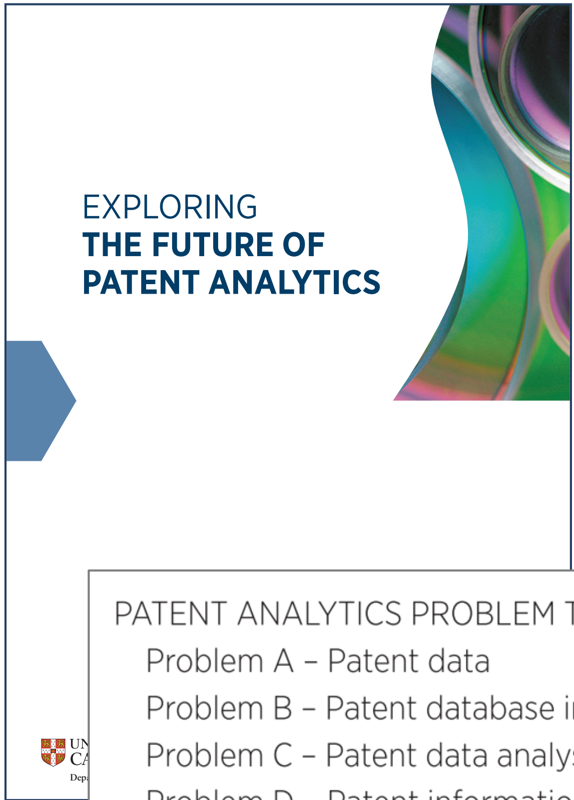


Figure 1: Samsung, Apple, Microsoft RDR and utility patent filings

Wolf, P., Tietze, F., Schweisfurth, T., Moultrie, J. (2017). Registered Design Rights as Innovation Indicators. R&D Management Conference. Leuven, Belgium.

REFLECTING ON THE FIVE PROBLEMS WE IDENTIFIED IN 2017



PATENT ANALYTICS PROBLEM THEMES AND DISCUSSION

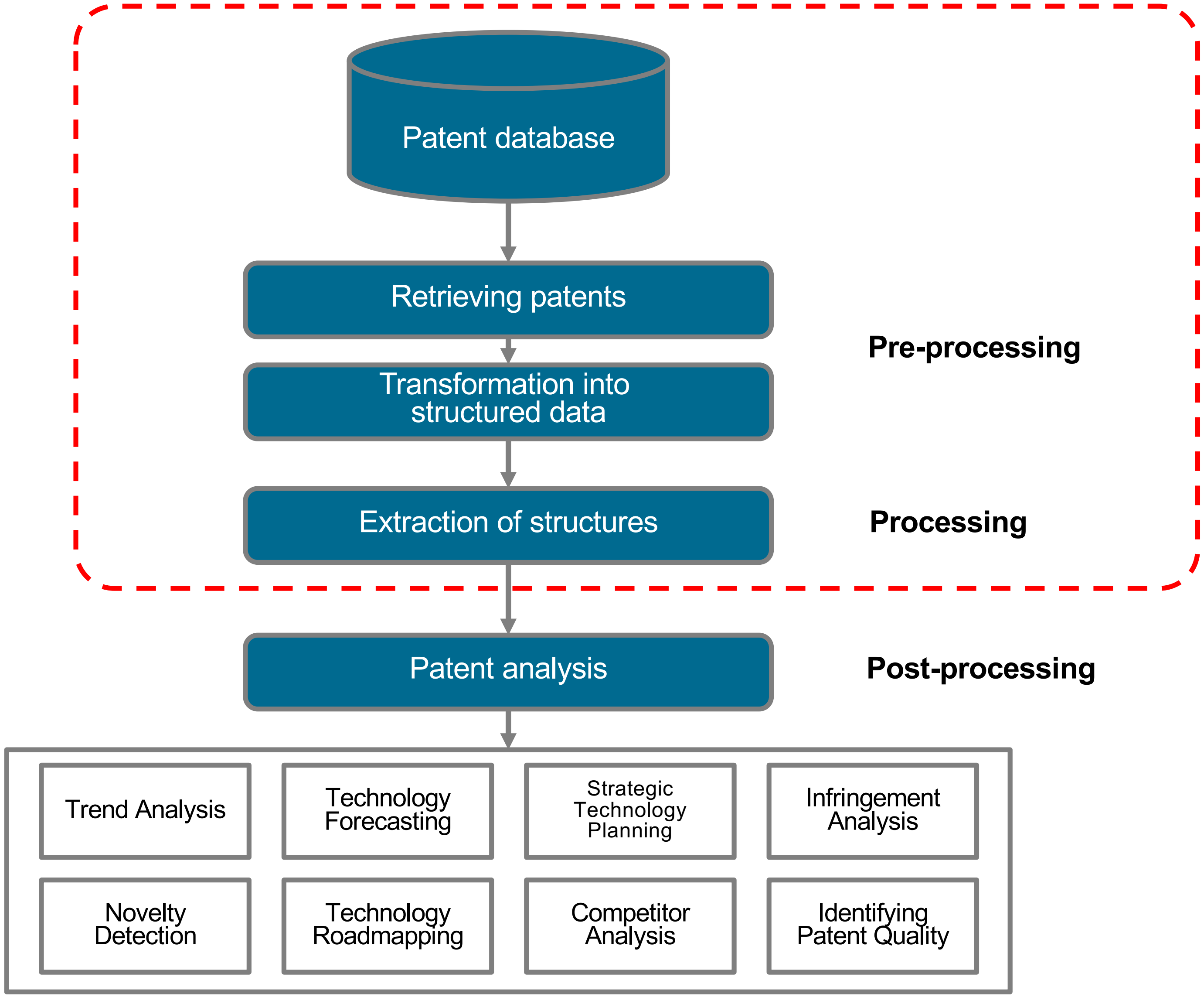
- Problem A – Patent data
- Problem B – Patent database interconnectedness
- Problem C – Patent data analysis
- Problem D – Patent information visualisation
- Problem E – Patent quality

Aristodemou L., and Tietze F., (2017) Exploring the Future of Patent Analytics: A technology roadmapping approach, Institute for Manufacturing, University of Cambridge.
www.ifm.eng.cam.ac.uk/insights/innovation-and-ip-management/exploring-the-future-of-patent-analytics/

UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

THE PATENT ANALYSIS WORK FLOW



Source: Abbas, A., Zhang, L. & Khan, S.U., 2014. A literature review on the state-of-the-art in patent analysis. World Patent Information, 37, pp.3–13.

DATA QUALITY AND THE NEED FOR CLEANSING – EXAMPLE APPLICANT NAME

Nowadays, databases typically employ "corporate tree" functions

Table 8.4 - Examples of consolidated assignee names

Original Assignee Name	Homogenized / Consolidated Original Assignee Name	OATY*
Li Ming Chiang, US	Ming-Chiang Li	1
Li; Ming-Chiang	Ming-Chiang Li	1
Min-Chiang Li	Ming-Chiang Li	1
Arco Chem Tech, US CPC International Inc, US	Unilever	4
Arco Chem Tech, US	Unilever	4
CPC International Inc.	Unilever	4
CPC International Inc, US Arco Chem Tech, US	Unilever	4
CPC International, Inc. Arco Chemical Technology, Inc.	Unilever	4
Bestfoods	Unilever	4

Note: * 1 (individual inventor); 4 (large firm)

Tietze, F. (2012). Technology Market Transactions - Auctions, Intermediaries and Innovation. Cheltenham Edgar Elgar Publishing.

Table 12: Top 10 patenting organizations with patent count and ranking before and after harmonization.

Harmonized Name (after second round of Harmonizing)	After harmonization		Before harmonization		Improvement
	# patents	Rank	# patents	Rank	
MATSUSHITA ELECTRIC INDUSTRIAL COMPANY	442.211	1	326.425	1	35,47%
NEC CORPORATION	347.687	2	184.195	7	88,76%
HITACHI	342.476	3	260.455	2	31,49%
TOSHIBA CORPORATION	336.649	4	236.744	3	42,20%
CANON	334.891	5	202.820	4	65,12%
MITSUBISHI ELECTRIC CORPORATION	305.575	6	187.569	6	62,91%
SAMSUNG ELECTRONICS COMPANY	274.666	7	201.932	5	36,02%
FUJITSU	270.722	8	158.045	8	71,29%
SONY CORPORATION	258.811	9	144.891	9	78,62%
SIEMENS	256.874	10	104.848	15	145,00%

Peeters, B., Song, X., Callaert, J., Grouwels, J., & Van Looy, B. (2009). Harmonizing harmonized patentee names: an exploratory assessment of top patentees. Luxembourg: EUROSTAT working paper and Studies.

BUT: We still do not know who owns which patents (and which are traded/ licensed), but only who registered them in the first place!

DEVELOPING EFFECTIVE SEARCH STRATEGIES AND COMPILING RELIABLE DATASETS REMAINS COSTLY

UNIVERSITY OF CAMBRIDGE

IfM MANAGEMENT TECHNOLOGY POLICY

Additive Manufacturing Innovations by UK Organisations

A review of patents and literature published between 2006 and 2015

Publications

Searches were carried out between 27 March and 24 April 2017. A high-precision search string was used for the paper search across 23 databases, consistent with the approach taken for the patent search. An 'all fields search' ((additive P/3 (manufact* or fabricat*)) OR (3D P/3 PRINT*)) AND af(UK or GB or Britain or England or Scotland or Wales)) resulted in 3416 documents, which was then reduced to a 'title' search only (TI((additive AND (manufact* or fabricat*)) or (3D AND print*)) AND AF(UK or GB or Britain or England or Scotland or Wales)). This resulted in 589 publications after removal of duplicates. We excluded the 2016 publications to ensure we covered the same period as for patent documents. This resulted in a set of 382 publications. The dataset includes dissertations, which were not specifically searched for, but were not deliberately excluded if the search picked them up. Papers were downloaded from ProQuest by publication year for analysis.

EPSRC

ESRC

BITBYBIT

The Geostrategic Race for Leadership in

Future Electric Vehicle Battery Technologies

André Hemmelder[1*], Frank Tietze[2], Simon Lux[1,3], Jens Leker[1,4] and Stephan von Delft[1,5]

[1] University of Münster, Institute of Business Administration at the Department of Chemistry and Pharmacy, Leonardo Campus 1, 48149 Münster, Germany.

[2] University of Cambridge, Intellectual Property and Innovation Management (IPIM) Laboratory, Centre for Technology Management (CTM), Institut...

17 Charles Babbage Road, Cambridge CB3...

[3] Fraunhofer Research Institution for Battery...

[4] Helmholtz Institute Münster, IMD-4, Fors...

Germany.

[5] University of Glasgow, Adam Smith Busine...

* Corresponding author, E-Mail address: ...

3. Supplementary Note 3, Figures and Tables

Abstract

Leadership in electric vehicle battery tech...

analysing a dataset of 22,457 patent fami...

as policy documents, we assess geostrate...

battery technology landscape. The findi...

battery technologies, whereas the US lead...

However, depending on the technology, ...

focus on expanding their domestic state-...

Korea already invest heavily in future b...

has surpassed first movers Japan and South Korea in four out of six future battery technologies.

Patent Landscape Report

Hydrogen fuel cells in transportation

WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION

Fuel cells in transport

Part 1

Part 2

Part 3

World Patent Information 61 (2020) 101964

Contents lists available at ScienceDirect

World Patent Information

journal homepage: http://www.elsevier.com/locate/worpatin

Blockchain patent landscaping: An expert based methodology and search query

Nigel S. Clarke, Björn Jürgens, Victor Herrero-Solana

ARTICLE INFO

ABSTRACT

1. Introduction

Table 3

Blockchain sub-queries Q1-Q4 and full query in GPI query format.

Sub-Query 1 (Q1):

Sub-Query 2 (Q2):

Sub-Query 3 (Q3):

Full Query ((Q1 OR Q2 OR Q3):

UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

INCOMPLETENESS OF DATA REMAINS A PROBLEM

- Y02 data based study of climate adaptation technology contributions
- Patstat … based on inventor residency data we aggregated the data into four groups, i.e. high-income, upper-middle income, lower-middle income, and low-income countries … World Bank country classification
- “inventor residence is not always reported for all patents, resulting in a bias towards countries and patent offices that report inventor residence data more accurately.”

Table 2: Inventor country patenting activity in adaptation technologies between 1980 and 2019

No	Global	Country	Region	Coastal (Y02A10) [%]	Water Y02A20 [%]	Infrastructure Y02A30 [%]	Agriculture Y02A40 [%]	Health Y02A50 [%]	Indirect Y02A90 [%]	Total Patents	Patents / population[1]	Patents / GDP[2]
				High-income countries								
High-income overall				**2.48**	**14.53**	**8.32**	**24.15**	**44.59**	**5.93**	**57,781**	**5.22**	**0.90**
1	1	United States	North America	0.75	6.93	5.31	17.28	58.06	11.68	19,310	5.83	0.92
2	2	Germany	Europe & Central Asia	1.19	9.54	13.11	16.34	56.22	3.60	6,468	7.78	1.66
3	3	Japan	East Asia & Pacific	1.40	10.18	15.25	18.08	49.58	5.51	6,444	5.10	1.28
4	4	France	Europe & Central Asia	1.28	6.41	9.56	20.71	55.39	6.64	3,356	4.97	1.27
5	6	United Kingdom	Europe & Central Asia	1.13	9.32	6.84	16.26	59.16	7.29	3,100	4.62	1.15
6	7	Korea, Rep.	East Asia & Pacific	3.85	11.69	11.00	24.49	45.01	3.96	2,728	5.26	1.66
7	8	Canada	North America	1.48	10.20	5.10	25.65	49.62	7.96	1,961	5.16	1.19
8	9	Netherlands	Europe & Central Asia	1.84	8.01	8.50	44.93	32.74	3.97	1,411	8.09	1.55
9	10	Switzerland	Europe & Central Asia	1.73	6.68	10.58	17.70	59.04	4.28	1,333	15.43	1.80
10	11	Italy	Europe & Central Asia	2.34	11.19	11.41	21.16	51.17	2.72	1,323	2.23	0.70
Remaining high-income countries (68 in total)				2.62	15.48	8.08	24.47	43.39	5.96	10,347	5.00	0.82
				Upper-middle income countries								
Upper-middle income overall				**1.71**	**15.11**	**4.67**	**26.10**	**49.63**	**2.78**	**4,926**	**0.17**	**0.28**
1	5	China	East Asia & Pacific	2.35	12.32	11.74	26.17	42.51	4.92	3,110	0.22	0.21
2	23	Brazil	Latin America & Caribbean	0.79	12.37	2.89	28.42	51.32	4.21	380	0.18	0.26
3	25	Russian Federation	Europe & Central Asia	1.20	21.26	8.38	22.46	38.62	8.08	334	0.23	0.22
4	28	Mexico	Latin America & Caribbean	2.86	15.43	10.29	24.00	44.57	2.86	175	0.14	0.16
5	29	South Africa	Sub-Saharan Africa	3.43	18.86	6.86	25.14	42.29	3.43	175	0.30	0.52
6	33	Turkey	Europe & Central Asia	0.88	17.54	10.53	33.33	34.21	3.51	114	0.14	0.16
7	36	Malaysia	East Asia & Pacific	1.82	8.18	1.82	35.45	50.00	2.73	110	0.33	0.33
8	40	Argentina	Latin America & Caribbean	2.41	7.23	3.61	32.53	49.40	4.82	83	0.18	0.22
9	41	Colombia	Latin America & Caribbean	-	13.70	5.48	19.18	57.53	4.11	73	0.14	0.27
10	42	Thailand	East Asia & Pacific	1.39	4.17	1.39	22.22	69.44	1.39	72	0.10	0.14
Remaining upper-middle income countries (42 in total)				1.71	15.74	4.16	25.86	50.14	2.39	300	0.16	0.28
				Lower-middle income countries								
Lower-middle income overall				**1.58**	**11.39**	**4.03**	**16.84**	**64.72**	**1.43**	**1,396**	**0.05**	**0.19**
1	15	India	South Asia	0.30	7.09	3.54	17.03	67.22	4.82	1,016	0.07	0.38
2	43	Ukraine	Europe & Central Asia	1.61	9.68	4.84	40.32	40.32	3.23	62	0.14	0.40
3	44	Egypt, Arab Rep.	Middle East & North Africa	3.39	38.98	1.69	6.78	47.46	1.69	59	0.05	0.16
4	47	Morocco	Middle East & North Africa	-	30.77	2.56	30.77	33.33	2.56	39	0.11	0.32
5	55	Philippines	East Asia & Pacific	13.79	-	-	51.72	31.03	3.45	29	0.03	0.08
6	62	Vietnam	East Asia & Pacific	0.00	19.05	9.52	14.29	57.14	-	21	0.02	0.06
7	63	Sri Lanka	South Asia	5.26	-	5.26	21.05	68.42	-	19	0.09	0.22
8	66	Tunisia	Middle East & North Africa	-	5.88	11.76	17.65	64.71	-	17	0.14	0.40
9	70	Ghana	Sub-Saharan Africa	-	-	-	13.33	80.00	6.67	15	0.05	0.21
10	72	Bangladesh	South Asia	-	-	6.67	13.33	80.00	-	15	0.01	0.04
Remaining lower-middle income countries (35 in total)				1.24	11.49	3.81	14.53	67.82	1.11	104	0.04	0.17
				Low-income countries								
Low-income overall				**-**	**5.97**	**-**	**19.01**	**75.03**	**-**	**53**	**0.02**	**0.35**
1	85	Sudan	Sub-Saharan Africa	-	27.27	-	-	72.73	-	11	0.02	0.41
2	89	Korea, DPR.	East Asia & Pacific	-	11.11	-	11.11	77.78	-	9	0.03	n/a
3	111	Madagascar	Sub-Saharan Africa	-	25.00	-	50.00	25.00	-	4	0.01	0.31
4	112	Ethiopia	Sub-Saharan Africa	-	-	-	-	100.00	-	4	0.00	0.04
5	113	Gambia, The	Sub-Saharan Africa	-	-	-	-	100.00	-	4	0.16	2.21
6	120	Somalia	Sub-Saharan Africa	-	-	-	-	100.00	-	3	0.02	0.44
7	123	Central African Republic	Sub-Saharan Africa	-	-	-	-	100.00	-	2	0.04	0.86
8	126	Syrian Arab Republic	Middle East & North Africa	-	50.00	-	-	50.00	-	2	0.01	0.18
9	127	Uganda	Sub-Saharan Africa	-	-	-	50.00	50.00	-	2	0.00	0.05
10	128	Eritrea	Sub-Saharan Africa	-	-	-	100.00	-	-	2	0.06	n/a
Remaining low-income countries (19 in total)				-	-	-	16.67	83.33	-	10	0.01	0.16

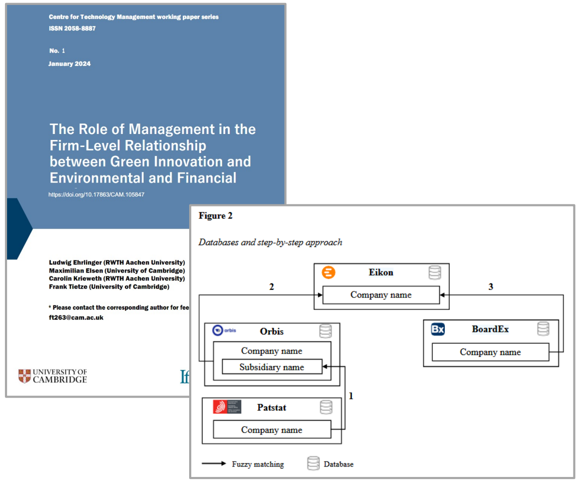
Notes: Underlined values depict values that largely differ from country-group average.
[1] Population calculation is based on high-value adaptation inventions per 100,000 people; population data was retrieved from World Bank for the year 2020.
[2] GDP Calculation is based on high-value adaptation inventions per 1 billion GDP; GDP data was retrieved from World Bank for the year 2020.

Elsen, Maximilian, and Frank Tietze. (forthcoming) ‘Contributions from Low- and Middle-Income Countries to the Development of Climate Change Adaptation Technologies: A Patent Analysis’. Technological Forecasting and Social Change.

MATCHING PATENT DATA WITH OTHER DATASETS REMAINS CHALLENGING

- Classical approach in econ studies: NBER dataset that matches US publicly listed firms with patent data from the USPTO between 1980 and 2015 [1]

- Text based matching approaches:
  - Levenshtein ratios of string similarity (Edit Distance) [2]
  - Jaro Winkler distance [3]
  - Python fuzzy-matching algorithm 'FuzzyWuzzy' using term frequency-inverse document frequency (TF-IDF) [4]



"This matching process accounts for a significant part of the work for this study since it is both time-consuming and fundamental for having an insightful dataset." [4]

[1] Arora, A., Belenzon, S., & Sheer, L. (2021). Matching patents to compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures. Research Policy, 50(5), 104217.

[2] Levenshtein, Vladimir I. (February 1966). "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady. 10 (8): 707–710. Example: e.g. Neuhäusler, Peter, Rainer Frietsch, Carolin Mund, and Verena Eckl. (2016) 'Identifying the Technology Profiles of R&D Performing Firms — A Matching of R&D and Patent Data'. International Journal of Innovation and Technology Management

[3] Jaro, Matthew A. (1 June 1989). "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida". Journal of the American Statistical Association. pp. 414–420.

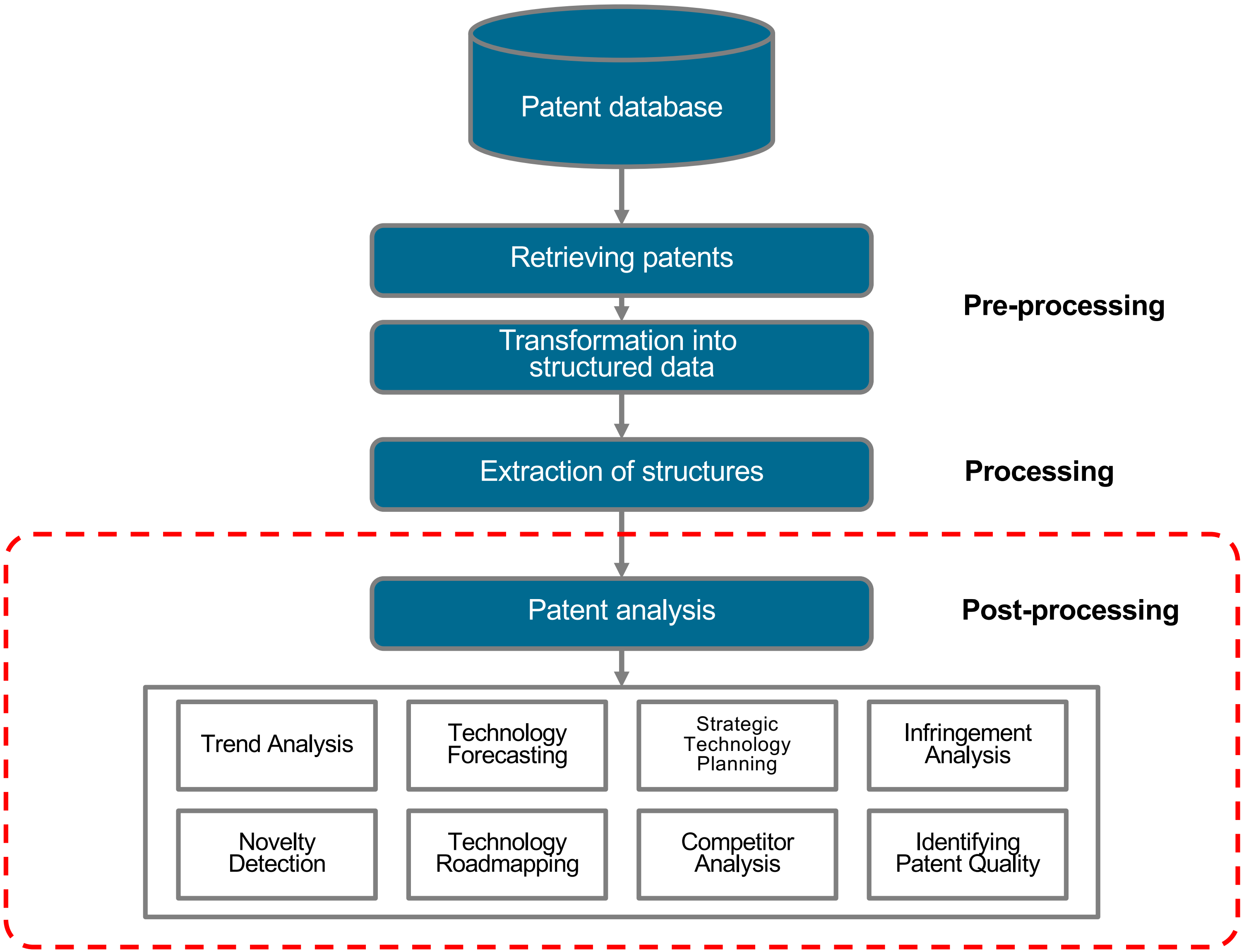
[4] Ehrlinger, Ludwig, Maximilian Elsen, Carolin Krieweth, and Frank Tietze. (2024) 'The Role of Management in the Firm-Level Relationship between Green Innovation and Environmental and Financial Performance'. https://www.repository.cam.ac.uk/handle/1810/364028.

See further: 'String Similarity Metrics – Edit Distance | Baeldung on Computer Science', 14 November 2020. https://www.baeldung.com/cs/string-similarity-edit-distance/.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab

THE PATENT ANALYSIS WORK FLOW



Source: Abbas, A., Zhang, L. & Khan, S.U., 2014. A literature review on the state-of-the-art in patent analysis. World Patent Information, 37, pp.3–13.

THE PATENT INDICATOR "JUNGLE": NUMEROUS (TOO MANY?) PATENT INDICATORS HAVE BEEN DEVELOPED

> 1,000 citations (Google Scholar)

Patent indicator	Definition	Meaning
Patent activity (PA_{iF})	Patent applications (PA) of firm i in technological field (TF) F	Extent of R&D expenditures of firm i in TF F (interest of firm i in TF F)
Technology share (based on patent applications)	PA_{iF}/PA of all competitors in TF F	Competitive technological position of firm i in TF F (quantitative)
R&D emphasis	PA_{iF}/Number of firm's (i) total patent applications	Importance of technological field F for firm i (R&D emphasis)
Co-operation intensity	Number of joint patent applications with partners in TF F/PA_{iF}	Access of firm i to external knowledge (and identification of partners)
Share of granted patents (Q_1)	Granted patents of firm i in TF F/PA_{iF}	Technological quality of firm i's patent applications
Technological scope (Q_2)	Diversity and number of IPC classes in firm i's patent applications (PA_{iF})	Technological quality of firm i's patent applications
International scope (Q_3)	Size of patent family and share of triad (US, JP and EPO) patents of PA_{iF}	Economic quality of firm i's patent applications
Citation frequency (Q_4)	Average citation frequency of PA_{iF}	Economic quality of firm i's patent applications
Average patent quality (PQ_{iF})	Sum of all indicators of patent quality (Q_1–Q_4)	Average total quality of all patent applications of firm i in TF F
Patent strength (PS_{iF})	Product of average patent quality (PQ_{iF}) and patent activity (PA_{iF})	Technological strength of firm i in TF F
Technology share (based on patent strength)	PS_{iF}/PS of all competitors in TF F	Competitive technological position of firm i in TF F (qualitative)
Relative technology share	PS_{iF}/Max. patent strength of a firm in TF F	Distance of firm i to the technological leader in TF F

Ernst, H. (2003). Patent information for strategic technology management. World Patent Information, 25(3), 233-242.

See also:

Lanjouw, J. O. & Schankerman, M. Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators. The Economic Journal 114, 441–465.

OECD Science, Technology and Industry Working Papers. Measuring Patent Quality: Indicators of Technological and Economic Value, 2013.

Table 1
Citation-based measures - patent level.

Measure	Description	Source	Used by	Equation
Citation Index (CI)	CI is the count of the citations received by a company's patents from subsequent patents. It is used to evaluate the technological impact of patents. The number of forward citations mirrors the technological importance of the patent for the development of subsequent technologies, and also reflects the economic value of inventions [24,34]. Patents with high values of CI are often important inventions or fundamental to future inventions, making these inventions valuable [24] due to the cumulative nature of the process.	[10,11,15,23,25,34,54,56,57]	[24,37,41,43,50,53,58]	$CI = \sum_t \sum_1^n Citations\ citing\ patent_{tn}$ where t is the year of patent publication and tn is the publication year of subsequent patents citing the original patent.
Forward Citation Frequency (FCF)	FCF is defined as the number of forward citations received by a patent per year. This is an indication of the impact of a company's patents. FCF can only be compared within a technological area for a particular year, since the number of citations changes per year.	[24,33,34,39,59,60]	[27,54,61]	$FCF = \frac{CI}{Patent\ Age}$
Generality	The generality of a target patent indicates the diversity of citing patents, i.e. the patents that cite the target patent. The index is defined between zero and one, and the measure is high if subsequent patents belonging to a wide range of fields cite a patent. If most citations are concentrated in a few fields the generality index is low [62]. state the numerical range of the index with 0–0.44 (low), 0.45–0.65 (mid) and 0.66–1 (high).	[25,63]	[13,24,30,39,64]	$Generality = 1 - \sum_{J \in Sf} \left(\frac{Number\ of\ Citing\ patents\ belonging}{Number\ of\ Citing\ Patents}\right)^2$ where Sf is the set of classes of citing patents
Influence	The Influence is defined as the number of forward citations a patent received from subsequent patents in the first years since its publication. The higher influence index suggests that a patent has influenced and impacted the technology scope of subsequent patents on publication.	[8]	[8,65]	$Influence_{i,t} = 1 + (\sum_t^{t+5} Cit_{i,t})$ Cit(i,t) is the number of forward citations the i patent received from following patents up to five years after the publication date, where t is the year.

Table 2
Citation-based measures - portfolio level.

Measure	Description	Source	Used by	Equation
Current Impact Index (CII)	CII is the number of times a company's previous 5 years of patents are cited in the current year, relative to all patents in the U.S. patent system. It can measure the influence of a company in the last 5 years, and indicates patent portfolio quality. CII is a synchronous indicator, which looks back from the current year to the previous five years. In general, patents with higher CII values often represent stronger technological ability. It has the advantage to foresee the development of a technology.	[66]	[41,51,52,54,67]	$CII = \frac{100C_i / \sum_i C_i}{100K_i / \sum_i K_i}$ Ci represents the number of times a patent of company I has been cited in a certain year, from previous 5 years. Ki is the number of patents, company i produced the past 5 years.
Herfindal-Hirschman Index of Patents	HHI describes the concentration of patents across patent classes, and is used to measure the concentration level of a firm's technological capability. HHI is described as a patent quality indicator and is used to explore the relationship between patent quality and market value. According to [68], a HHI index constructed from a small number of counts will generally be biased downwards. To avoid the bias, we also use the HHI index adjusted for Bias measure [68].	[24,25,47,68]	[24,34,39,47,60]	For a set of N patents falling into n classes, with Ni patents in each class (Ni ≥ 0, i = 1 … i), the HHI is calculated $HHI\ of\ patents = \sum_{i=1}^n \left(\frac{N_i}{N}\right)^2, 0 \le HHI\ of\ patents \le 1 Adjusted\ bias\ measure = \eta = \frac{N.HHI - 1}{N - 1}$
Hindrance Index (HI)	The measure of hindrance, of the assignee of the target patent is defined as the ratio of 'X' and 'Y' citations received from competitors to all 'X' and 'Y' citations received. If the firms received 'X' or 'Y' references from future patents they can be seen to hinder or block subsequent patents.	[13]	[13]	$HI = \frac{\sum 'X'\ and\ 'Y'\ forward\ citations\ from\ competitors'}{\sum All\ 'X' and\ 'Y'\ forward\ citations\ received}$ Forward citations are split into 'citations by competitors (the no. of patents with a different assignee than the focus patent)' and 'total citations by all other non-competitve companies (the no. of patents developed conjointly between competitors)'.
Relative Patent Position (RPP)	RPP of a company in its most important technological field means the patent counts owned by the company in its technological field where it has more patents than in others divided by the patent counts of the leader in the technological field. RPP is used to measure the degree of leading of the company in the technological field; the higher the value, the more leading position the assignee of that patent has in the field.	[69]	[26,52,70]	$RPP = \frac{\sum N_{p(same\ assignee)}}{\sum N_{p(leader\ in\ the\ field)}}$ The leader in a technological field is defined as the company with the largest amount of patents granted in the field.
Technology Strength (TS)	TS indicator is used to measure the scale of a company in a specific technological field, indicating patent portfolio strength.	[54,71]	[51,54,55,67]	$TS = Number\ of\ Patents\ x\ CII$ It makes the assumption that the number of patents of a company in a specific field is of similar quality and impact as the company's recent patents.

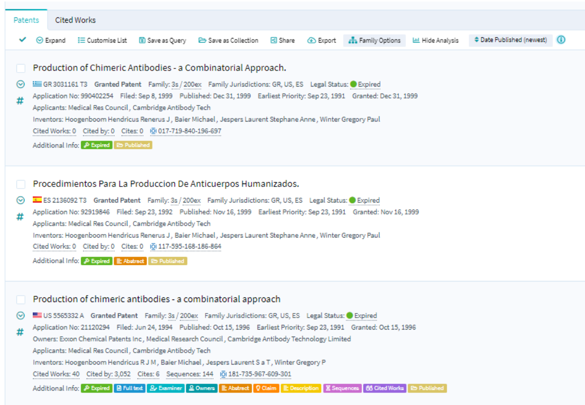
Table 1 in Aristodemou, L. and F. Tietze (2018). "Citations as a measure of technological impact: A review of forward citation-based measures." World Patent Information 53: 39-44.

REMAINING AMBIGUITY IN PATENT FAMILY DEFINITIONS

Simple Family

A simple patent family is a group of patent documents that stem from the same initial document, called the priority document. For example, an applicant might file a patent application in one country, then file other applications in other countries. The simple family covers one single invention.[1]

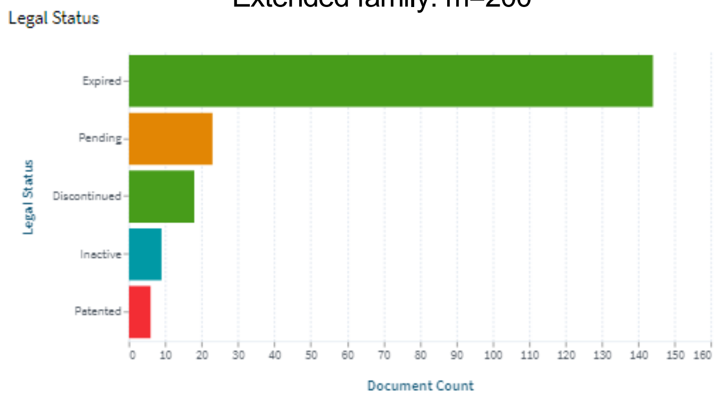
Simple family: n=3



Extended Family

An extended patent family is a collection of patent documents covering a technology – more than one single invention. The technical content covered by the applications is similar, but not necessarily the same. Members of an extended patent family will have at least one priority in common with at least one other member - either directly or indirectly.[2]

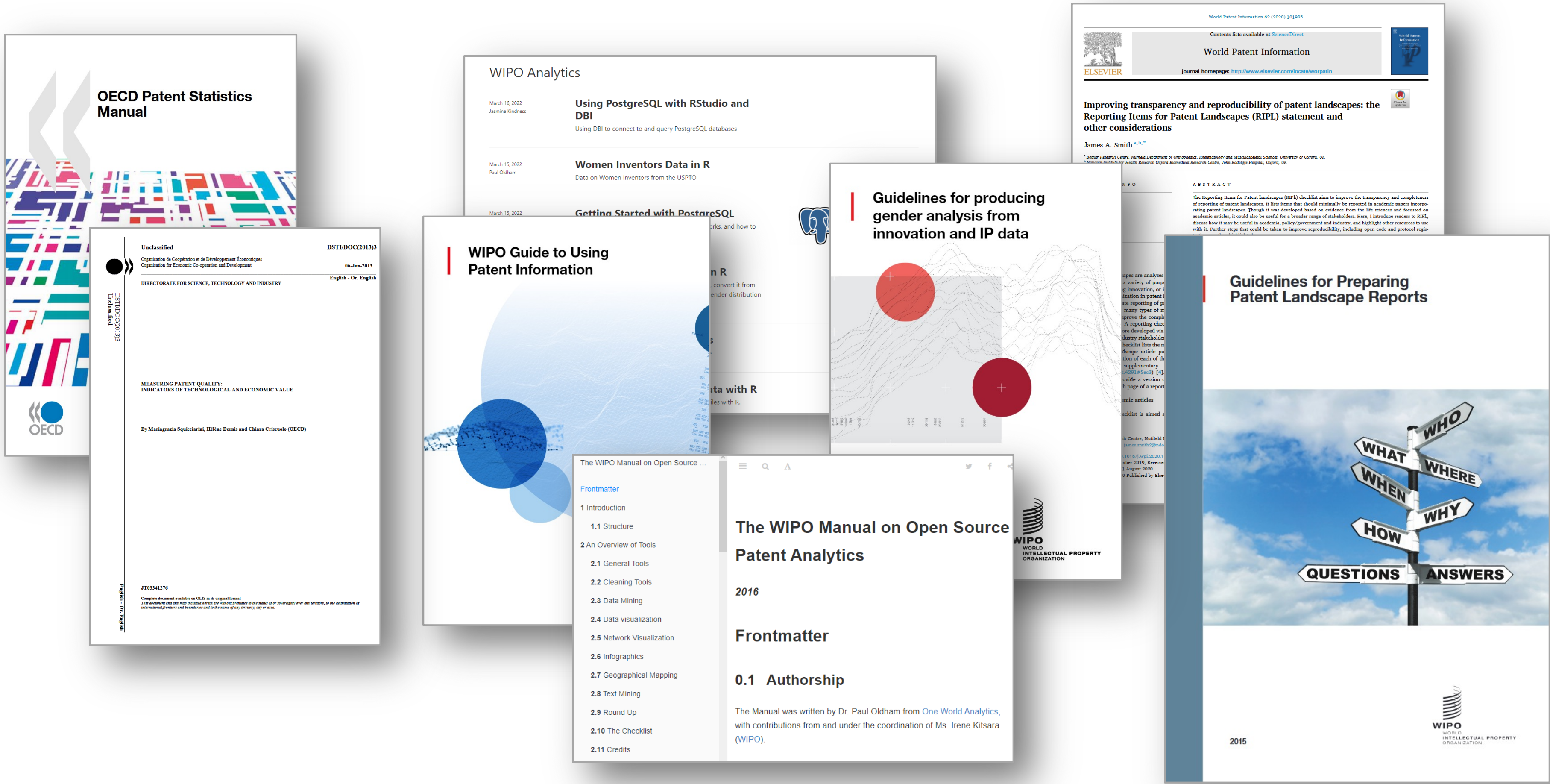
Extended family: m=200



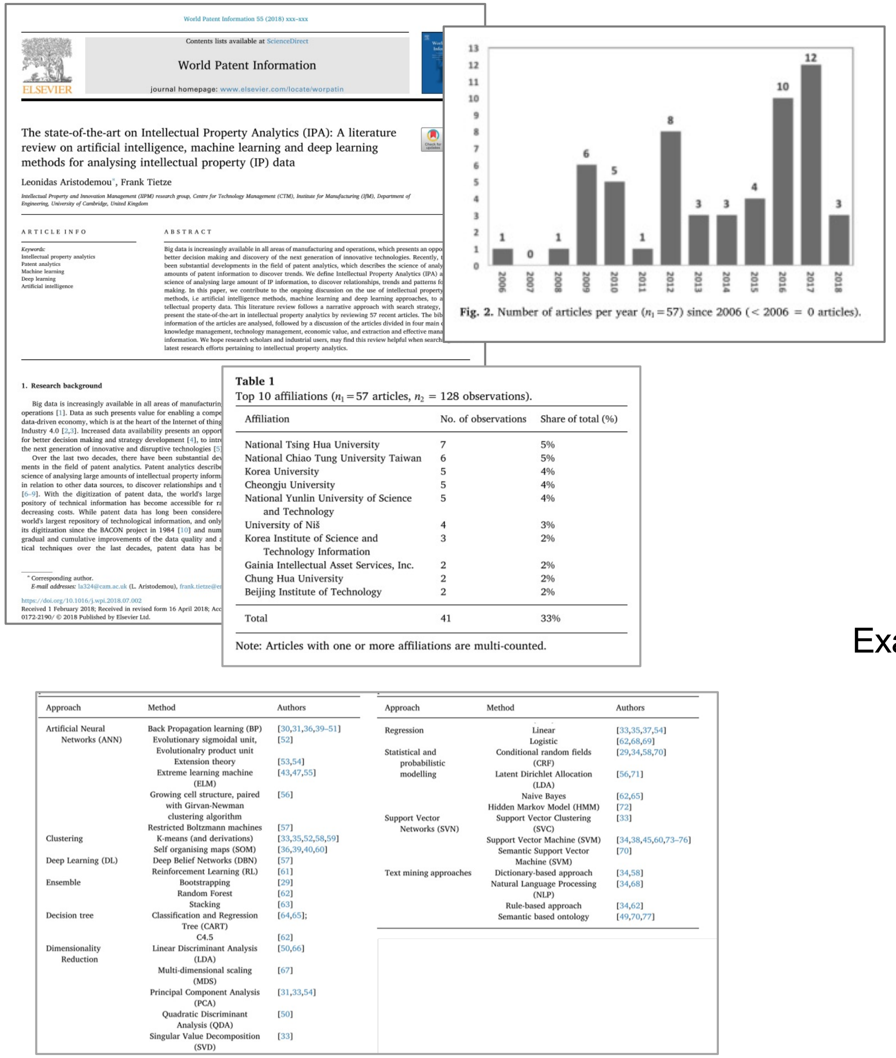
[1] See also: https://www.epo.org/en/searching-for-patents/helpful-resources/first-time-here/patent-families/docdb

[2] See also: https://www.epo.org/en/searching-for-patents/helpful-resources/first-time-here/patent-families/inpadoc

NEED FOR EVEN MORE GUIDANCE AND STANDARDIZATION



INCREASINGLY STUDIES USE AI ALSO FOR UNSTRUCTURED TEXT DATA



World Patent Information 55 (2018) xxx–xxx

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data

Leonidas Aristodemou*, Frank Tietze

Fig. 2. Number of articles per year ($n_1 = 57$) since 2006 (< 2006 = 0 articles).

Table 1
Top 10 affiliations ($n_1 = 57$ articles, $n_2 = 128$ observations).

Affiliation	No. of observations	Share of total (%)
National Tsing Hua University	7	5%
National Chiao Tung University Taiwan	6	5%
Korea University	5	4%
Cheongju University	5	4%
National Yunlin University of Science and Technology	5	4%
University of Niš	4	3%
Korea Institute of Science and Technology Information	3	2%
Gainia Intellectual Asset Services, Inc.	2	2%
Chung Hua University	2	2%
Beijing Institute of Technology	2	2%
Total	41	33%

Note: Articles with one or more affiliations are multi-counted.

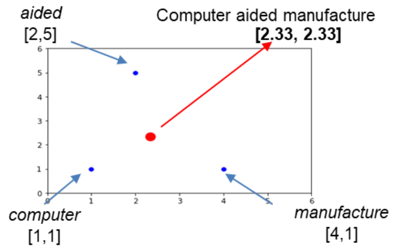
Approach	Method	Authors
Artificial Neural Networks (ANN)	Back Propagation learning (BP)	[30,31,36,39–51]
	Evolutionary sigmoidal unit, Evolutionalry product unit	[52]
	Extension theory	[53,54]
	Extreme learning machine (ELM)	[43,47,55]
	Growing cell structure, paired with Girvan-Newman clustering algorithm	[56]
	Restricted Boltzmann machines	[57]
Clustering	K-means (and derivations)	[33,35,52,58,59]
	Self organising maps (SOM)	[36,39,40,60]
Deep Learning (DL)	Deep Belief Networks (DBN)	[57]
	Reinforcement Learning (RL)	[61]
Ensemble	Bootstrapping	[29]
	Random Forest	[62]
	Stacking	[63]
Decision tree	Classification and Regression Tree (CART)	[64,65];
	C4.5	[62]
Dimensionality Reduction	Linear Discriminant Analysis (LDA)	[50,66]
	Multi-dimensional scaling (MDS)	[67]
	Principal Component Analysis (PCA)	[31,33,54]
	Quadratic Discriminant Analysis (QDA)	[50]
	Singular Value Decomposition (SVD)	[33]

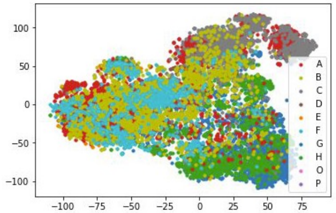
Approach	Method	Authors
Regression	Linear	[33,35,37,54]
	Logistic	[62,68,69]
Statistical and probabilistic modelling	Conditional random fields (CRF)	[29,34,58,70]
	Latent Dirichlet Allocation (LDA)	[56,71]
	Naive Bayes	[62,65]
	Hidden Markov Model (HMM)	[72]
Support Vector Networks (SVN)	Support Vector Clustering (SVC)	[33]
	Support Vector Machine (SVM)	[34,38,45,60,73–76]
	Semantic Support Vector Machine (SVM)	[70]
Text mining approaches	Dictionary-based approach	[34,58]
	Natural Language Processing (NLP)	[34,68]
	Rule-based approach	[34,62]
	Semantic based ontology	[49,70,77]

https://doi.org/10.1016/j.wpi.2018.07.002

0172-2190/ © 2018 Published by Elsevier Ltd.

Text data preparation remains a challenge





First studies experiment with LLMs to analyse patent text [1]

Example from current own work: Text-based classifications of Patent Litigation cases using OpenAi GPT-4o model

UNIVERSITY OF CAMBRIDGE
Department of Engineering

[1] also e.g. Kiiru, E., (2024) Innovative Nonprofits: Contributions to the Patent Corpus. Presented at the European Policy for Intellectual Property (EPIP) conference, Pisa, Italy.

IfM Innovation and IP Management Lab

THE PATENT ANALYSIS WORKFLOW IS CHANGING

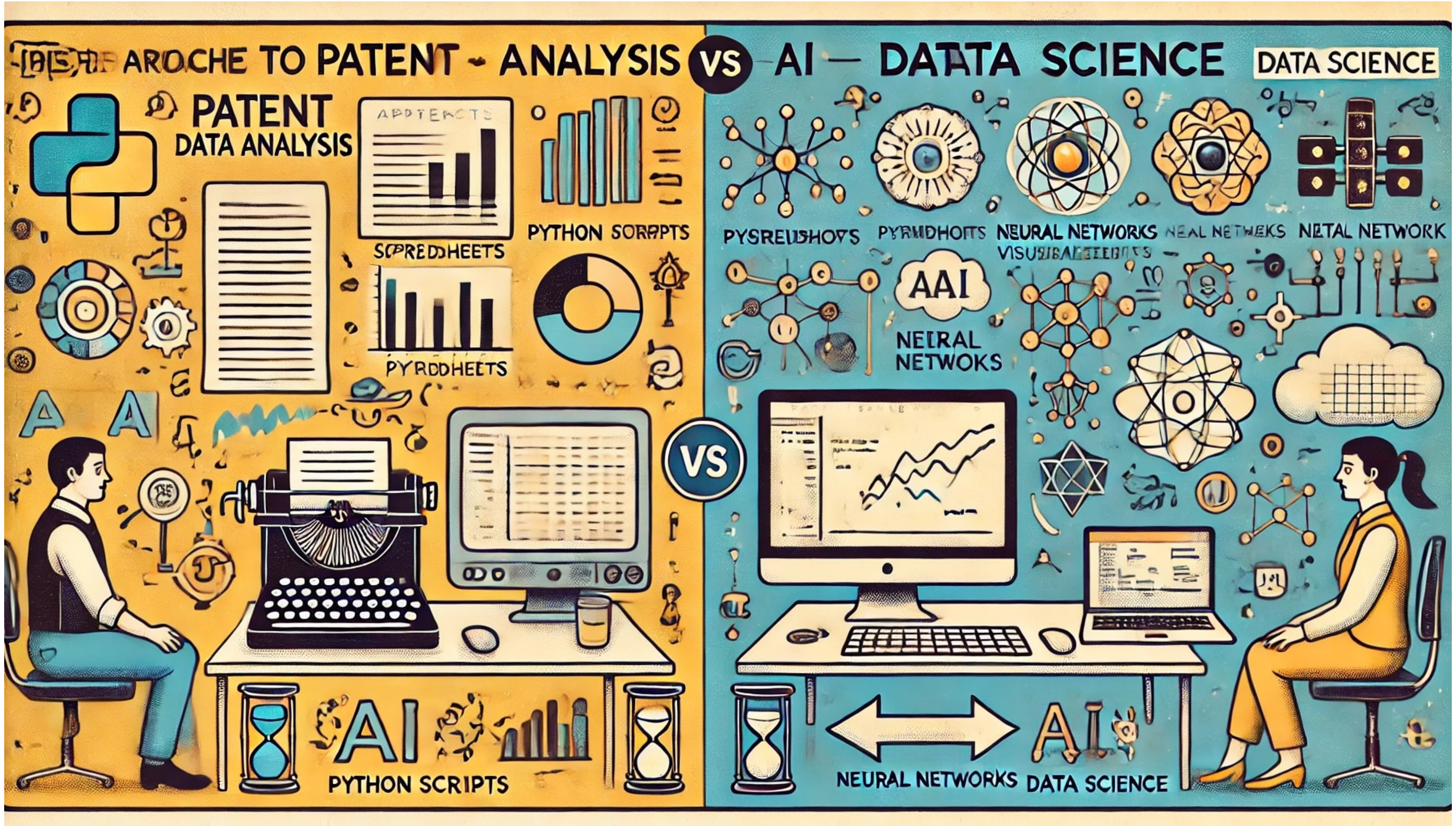


Image generated with https://chatgpt.com

Deploying AI for patent analysis requires new data science skillsets and tools

SOME ADDITIONAL CONSIDERATIONS: WHAT ELSE FOR COP?

- Raising the acceptance of patent studies by showcasing the impact of patent studies, e.g. collecting impact case studies

- Empowering policy makers to better understand patent data, e.g. training

- To what extent to engage with related communities, e.g. academics to address technical challenges, such as indicators/metrics, dataset matching, automated text analysis

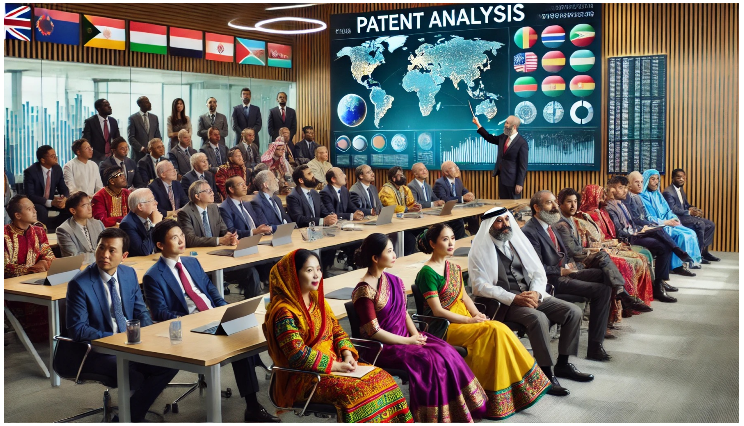


Image generated with https://chatgpt.com

CONCLUSIONS: WHAT COULD THE COP DO?

- Defining its remit and positioning of the COP, i.e. what scope of studies to consider getting involved
- Work together to increase quality of “raw” data, collaborate with academic researchers
- Collaborate on establishing best practices, e.g. for analysis, indicator usage, dataset matching, but also visualisation and dataset sharing
- Help CoP members to develop data science/AI skill sets for analysing data, e.g. develop training material
- Educate policy makers about the use of patent data, e.g. demonstrate impact by showcasing examples
- Work to further reduce access barriers to patent data users, such as in LMIC
- Consider how CoP and patent analytics can best support addressing urgent global challenges, such as climate change, SDGs (e.g. improved Y02 classification)

Image generated with https://chatgpt.com

THANK YOU

Prof. Dr. Frank Tietze
Innovation and IP Management (IIPM) Laboratory
www.iipm.eng.cam.ac.uk
frank.tietze@eng.cam.ac.uk

Department of Engineering
Institute for Manufacturing (IfM)
Centre for Technology Management

17 Charles Babbage Road
Cambridge, CB3 0FS
United Kingdom

UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM Innovation and IP Management Lab