Università degli
Studi di Bergamo

Università degli
Studi di Pavia

Doctoral Programme in
Economy and Management of Technology (DREAMT)

Curriculum
Innovation and Technology Management

STRATEGIES FOR REDUCING RISK IN PATENT APPLICATIONS' ANALYSIS

Doctoral Dissertation of

Paolo Carrara

Supervisor:
Prof. Caterina Rizzi

Tutor:
Prof. Davide Russo

PhD Programme Coordinator:
Prof. Gianmaria Martini

2018 – XXXI PhD Cycle

Abstract

Patents are a unique and exhaustive source of technological knowledge. Technical information we can find in them is not possible to obtain in other ways, such as market and economic analysis, voice of customer, etc. Patent databases also have high accessibility (since they are free available on the web) and a high level of format uniformity. This lets patent data can be electronically searched individually or together. These special features make patents a strategic source for supporting CEOs in decision making activities.

At present, worldwide patent database contains over 100 million of documents. Over the last decades, the number of patent applications per year is globally raising. The global growth in patent activity can be understood as an effect due to the shift of the economy towards the knowledge-based economy paradigm. According to this, the outcomes generated by knowledge, like patents, are business products or productive assets, which can be exploited as economical goods. In such a framework, it is crucial for patent owners knowing the value of held patents to adopt the best exploitation strategy.

For whom works in the patents' environment, the main difficulty relates to the proceeding the application for patent is subjected, which generally is long and complex.

The application filing is the first step in an ‘obstacle course’. Dozens of events and scenarios can affect the likelihood that the application reaches the grant, some of which might cause the unavoidable fall of the application itself. The first effect of this contest is the lack of certainties and the need to adopt work strategies and assessment criteria that take the risk into account.

The surge of patent filings had drastically increased the uncertainty status of patent literature.

The tools and methods currently available for patent experts are not designed to manage the risk due to this uncertain scenario. IP offices of firms, patent valuation experts of banks and other expert-in-the-field people must take the risk and manage it through their own professional expertise: a difficult job which this work addresses to.

Despite the high relevance and practical consequences of the uncertainty and risk related to the procedural aspects of patent applications, only few works paid attention to them. They did not give suggestions about tools or methods able to prevent or assess

the level of uncertainty in patent proceeding, neither to support the applicant carrying out patent analyses in presence of high share of patent applications.

This thesis is a sort of full immersion in the uncertainty of the patent application environment. From the coarsest errors anyone might do, to suggestions about most up-to-date sources of information, tools and strategies available to limit the uncertainty risk, up to an analytical system to compute the impact of procedural events on the success likelihood of the application for patent. It is a journey into the complex world of patent seen from a non-common point of view that can give useful insight to anyone working in the field.

Chapter 1 presents an overview on the currently available valuation methods for patents and the limitation they have in working with uncertainty due to patent applications.

Chapter 2 is an in-depth discussion about issues related to the transformations the text of patent application may undergo during the PCT and EPC proceedings.

Chapter 3 expounds a wide analysis that carried out in EP patent register to make an infographic about the success-rate of EP applications in grant and post grant proceedings.

Chapter 4 gives operative indications about building a business intelligence to assess the background into which positioning a patent application.

Finally, the Chapter 5 deals with the extraction of information about the market structure from patent data. In presence of patent thicket, dominant positions of main incumbent competitors might hindrance the access to the market of new entrant

Acknowledgements

Any PhD is a route. As any route, it passes through different places and brings you to meet several people. It doesn't matter about the reason they are in that place, or why they travel next to you for a stretch of road. They are there and they walk with you. After the end of that route, you will may remember some places or special happenings; above all, you will remember people with which you shared part of your way.

And then, a new path starts. It could follow new directions than previous, and your travel companions might change. So, it is crucial to greet the current ones and thanks them for walk, words, opinions, work, successes, projects, activity, laughers, time, daily life you shared.

Here I would like to thank the people who contributed to my PhD route, starting from Federica who pushed me to take this unknown path and who made possible the other amazing happenings of the last three years.

I would like to express my sincere gratitude to Prof. Davide Russo, my PhD tutor, for his continuous support, motivation, enthusiastic approach to the work and for his great humanity. His guidance helped me in all the time of research and writing of this thesis.

I would like to thank my PhD supervisor, Prof. Caterina Rizzi, for her encouragement, patience, and ironic comments. Her recommendations drove me to write the thesis in a better way.

My sincere thanks also go to Prof. Raffaella Manzini and Prof. Monica Giulietti, referees of my thesis, for their insightful questions and comments which enhanced the clearness of the thesis.

Thanks to the PhD labmates, Antonio, Christian and Giancarlo who made fun the work and gave different points of view on my research.

I would like to express my gratitude to the management of Warrant Innovation Lab, especially its director Matteo Marusi, who decided to fund this PhD project.

I would like to thank also my PhD-mates, Mariangela, Chiara, Alice and Dario, for their willingness and congeniality. They were an essential source of information and discussion about the administrative aspects (and others) of the PhD.

And thanks to all of the people who had been walking with me in the last three years. Have a nice continuation.

Index

List of Figures

List of tables

1 Introduction

Patents are a unique and exhaustive source of technological knowledge. Technical information we can find in them is not possible to obtain in other ways, such as market and economic analysis, voice of customer, etc. Patent databases also have high accessibility (since they are free available on the web) and a high level of format uniformity. This lets patent data can be electronically searched individually or together. These special features make patents a strategic source for supporting CEOs in decision making activities.

At present, worldwide patent database (e.g. Espacenet) contains over 100 million of documents[1]. Over the last decades, the number of patent applications per year is globally raising. The only exception have been happened in the first years of the recent economic crisis, but after those years, the acceleration in patenting activity increased, up to reach 3 million of applications per year in the 2016 (WIPO, 2017).

The global growth in patent activity can be understood as an effect due to the shift of the economy towards the knowledge-based economy paradigm (Drucker, 1969; Powell and Snellman, 2004). According to this, the outcomes generated by knowledge, like patents, are business products or productive assets, which can be exploited as economical goods. In such a framework, it is crucial for patent owners knowing the value of held patents in order to adopt the best exploitation strategy (Munari and Oriani, 2011).

Unfortunately for whom works in the patents' environment, patent proceedings are source of a large amount of uncertainty. The main difficulty relates to the proceeding the application for patent is subjected, which generally is long and complex. More over the final result is difficult to be foresaw. Patent professionals who have to express opinions about patent searches manages a larger and larger uncertainty risk, due to application lifecycle.

This thesis aims to give methods and tool able to support patent professionals in managing the risk due to the application uncertainty in granting and post-grant patent proceedings. It indicates most up-to-date sources of information, tools and strategies available to limit the uncertainty risk and prevent the coarsest errors anyone might do. It also introduces an analytical system to compute the impact of most relevant procedural events on the success likelihood of the application for patent. It is a journey into

[1] https://worldwide.espacenet.com/?locale=en_EP

the complex world of patent seen from a non-common point of view that can give useful insight to anyone working in the field.

The application filing is the first step in an 'obstacle course'. Dozens of events and scenarios can affect the likelihood that the application reaches the grant, some of which might cause the unavoidable fall of the application itself. Events occur within a period of years, even decade, in supranational proceedings before the national entry (up to 48 months in PCT[2] and 52 and 74 months on average in EPC and Euro-PCT respectively (Beatty, 2017)). The first effect of this contest is the lack of certainties and the need to adopt work strategies and assessment criteria that take the risk into account.

The surge of patent filings had drastically increased the uncertainty status of patent literature. Currently, the share of the documents having a non-definitive status in European patent DB is about 48% (see Table 6) and their likelihood of success is about 61% (see Figure 23).

The tools and methods currently available for patent experts are not designed to manage the risk due to this uncertain scenario. IP offices of firms, patent valuation experts of banks and other expert-in-the-field people must take the risk and manage it through their own professional expertise: a difficult job which this work addresses to.

Despite the high relevance and practical consequences of the uncertainty and risk related to the procedural aspects of patent applications, only few works paid attention to them. (Lemley, 2001) analysed the actual socioeconomical opportunity in reducing the uncertainty in US system related to litigations and oppositions. (Sternitzke, 2009) used examination reports from EPO examiners to assess which the causes of the failure are for PCT applications entered in EP phase in chemical, pharmaceutical and biotechnology filed. He points out the novelty is often anticipated by *available* patent literature, while *known* non-patent literature mainly discloses the inventive step. To reduce the uncertainty, he suggests performing an in-depth patent and non-patent search ex ante to identify potential priority citations.

Other articles point to the uncertainty due to the final result of the patent proceeding. (Guellec and van Pottelsberghe de la Potterie, 2000) studied the correlation between some economical parameters and the success rate of patent applications at the European Patent Office, considering, among other indicators, the filing route. (Frietsch, Neuhäusler and Rothengatter, 2013) raises the question whether the choice of the filing route influences the likelihood of a patent filing being granted, withdrawn or refused after the examination process at the EPO. (Harhoff and Reitzig, 2004; Jerak and

[2] http://www.wipo.int/pct/en/texts/time_limits.html

Wagner, 2006) suggest indicators to forecast the occurrence of opposition. Nevertheless, they did not give suggestions about tools or methods able to prevent or assess the level of uncertainty in patent proceeding, neither to support the applicant carrying out state-of-the-art patent analyses in presence of high share of patent applications.

Lee (Lee, 2017) proposes a Natural Language Processing (NLP) tool to perform a similarity check on US patent literature to retrieve potential invalidating patents in order to prevent the unsuccessful filing and/or forecast the invalidation of grant. Although it is an interesting development direction, the author himself stated the limited effect of the automated algorithm and highlighted the need of a supervised method to assess the relevance of results.

The following part of Chapter 1 presents an overview on the currently available valuation methods for patents and the limitation they have in working with uncertainty due to patent applications.

Chapter 2 collects in-depth discussion about issues related to the transformations the text of patent application may undergo during the PCT and EPC proceedings.

To quantify the risk of failure for a European patent application, Chapter 3 expounds a wide analysis has been performed in the EP patent register. An infographic is proposed as success-rate map to easy support the decision makers in choose the actions to be taken to progress in patent proceedings.

Chapter 4 gives operative indications about building a business intelligence to assess the background into which positioning a patent application. Knowing the relevant state-of-the-art and its business history, the analyst can easier assess the importance of the application itself.

Finally, the Chapter 5 deals with the extraction of information about the market structure from patent data. In presence of patent thicket, dominant positions of main incumbent competitors might hindrance the access to the market of new entrants, independently from the novelty and quality of their patented ideas.

1.1 Patent valuation reasons

The range of strategies the companies adopt in exploiting the Intellectual Property (IP), then patents, is very large. Despite they might be very different from a company to another (Cohen *et al.*, 2002), (Otsuyama, 2003) depicted an evolutionary pattern through them (see Figure 1). He linked the level of exploitation to the value of patent and the importance for the company to adopt a valuation strategy. As the level of direct exploitation of IPs increases, the importance of IP valuation grows too.

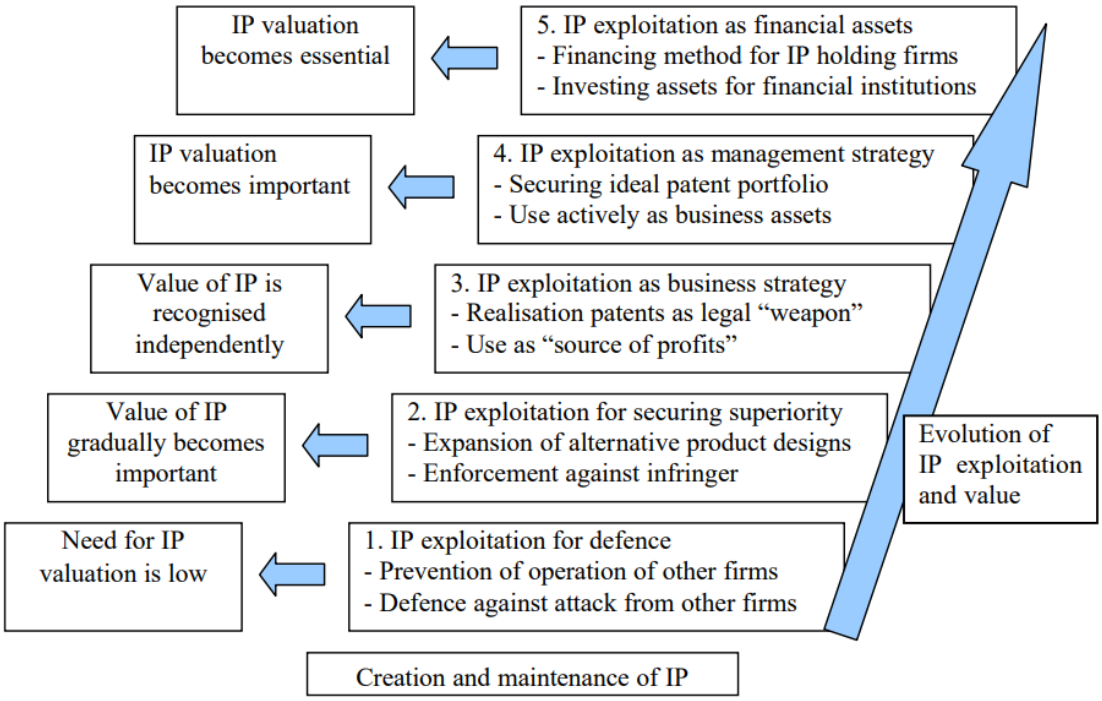


Figure 1: evolutionary pattern of the reasons to exploiting IPs as business assets and related importance level of valuating them (Otsuyama, 2003)

The exploitation trend shown in Figure 1 is the result of some pushing evolutionary pattern, like the need of external sources of technology in order to support their innovation process (Chatterji and Manuel, 1993; Chatterji, 1996; Jones *et al.*, 2002) and obtain Freedom-to-operate (Grindley and Teece, 1997; Kryder, Kowalski and Krattiger, 2000; Lichtenthaler, 2007), the 'market for technology' (Arora, Fosfuri and Gambardella, 2001; Roberts, 2001) and the leverage on knowledge in asset exchanges (Haour, 2004).

Consequently, the demand for valuation of patents is rising and the growth of patent activity worsen the situation. This pushes the patent specialists toward working in more and more difficult context, characterised by bigger patent pools than in the past, which have higher share of patent applications: in other words, higher level of uncertainty.

1.2 Patent valuation techniques

There are two main families of patent valuation methods for business purposes: quantitative methods and qualitative methods (Kamiyama, Sheehan and Martinez, 2006). The first ones aim to define a value of the patent relying on measurable data. Due to

this feature, they can be considered objective and reproducible methods. The output of this kind of approach is a value, or range of values, expressed in monetary units. On the other hand, the qualitative methods do not rely on measurable and objective data. They focus the analysis on the intended use of the patent and the environment in which they will exploited. The outcomes of such analysis are scores and/or rates according to some declared criteria[3].

Among authors involved in valuation of intellectual property (e.g. (Austin, 1993; Deng, Lev and Narin, 1999; Ernst, Leptien and Vitt, 2000; Baglieri *et al.*, 2001; Neifeld, 2001; Anson, 2002; Harhoff and Reitzig, 2004) and many others), (Pitkethly, 1997) focused on the uncertainty nature of patents and its consequences on the usage of quantitative, or monetary, valuation methods. Beyond highlighting the primary role of the evolution of the risk parameter in a patent lifecycle, he proposed an interesting classification of valuating methods (see Figure 2) according to the increasing amount of information they need and the related reliability of the resulting value.

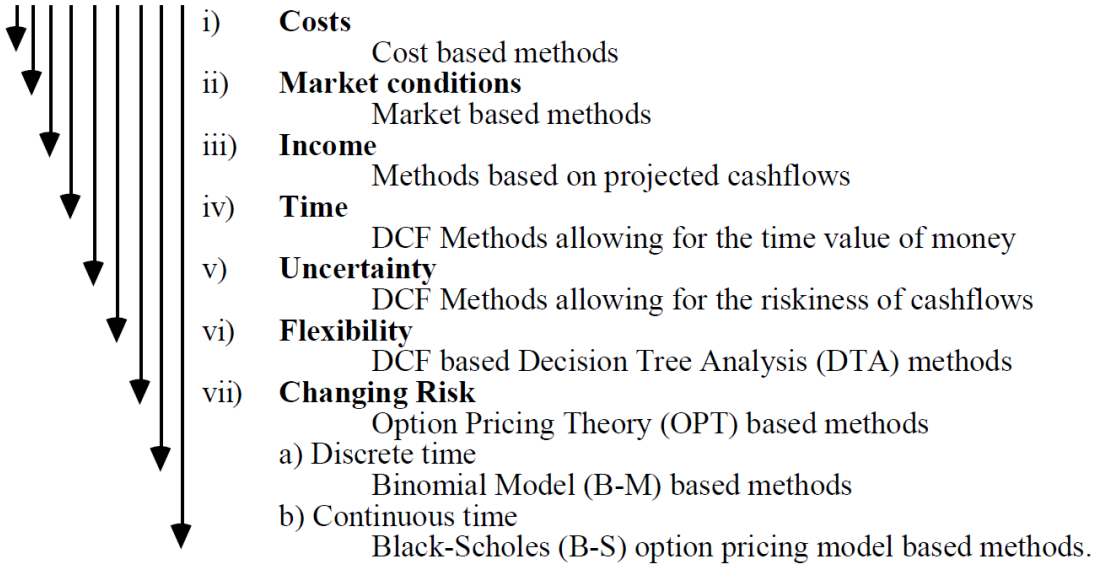


Figure 2: classification of patent valuation methods according to (Pitkethly, 1997)

Other than methods, also appraisal processes are object of research, both in scientific literature (Reilly and Schweihs, 1998; Park and Park, 2004; Chiesa, Gilardoni and Manzini, 2005; Chiesa *et al.*, 2008) and in patents (US2012310847, KR20140013473,

[3] http://www.ip4inno.eu/index.php?id=108

CN104574150, KR101625124, KR20170128008, WO2017132450). Difficulties related to the uncertainty of the patent lifecycle make disappointing the methods and the appraisal processes. (Chiesa *et al.*, 2007) addressed the reason of such a limited operability to the need of monetizing any parameter that might influence the patent value. This consideration agrees with the fact that in the monetary methods, there are no indications about the way to measure the risk related to the patent proceeding.

To improve the reliability of valuation methods, (Chiesa *et al.*, 2007) introduced a framework which arranges different kinds of qualitative variables able to affect the patent value in a structured classification. In the pool of variables, some of them can be easily linked to the feature of patent system, i.e. the novelty of the technology, the difficulty of invent around, the availability of technological alternatives and their feasibility, the technology family lifecycle and the development stage of the asset. Nonetheless, these variables do not deal with some pitfalls typical of the patent applications, where the most misleading is the uncertainty about the technological content (Carrara and Russo, 2017).

The framework is substantially recalled by IPscore[4], the IP scoring tool provided by the EPO, and the "Patent Valuation Grid" of the Italian patent office[5]. They are operational tools that guide the appraiser in patent valuation task.

A comparison between quantitative and qualitative methods is available in (van Zeebroeck, 2011) which provides the guidelines for the choice of the most suitable valuation method according to the reason for which the IP value is needed

1.3 Valuation parameters

Qualitative methods could consider different kinds of parameters and indicators, e.g. among others (Harhoff, Scherer and Vopel, 2003; Hung and Tseng, 2010; Ernst and Omland, 2011; van Zeebroeck, 2011; van Zeebroeck and van Pottelsberghe de la Potterie, 2011; Hsieh, 2013; Fischer and Leidinger, 2014; Grimaldi *et al.*, 2015; Frattini, Bianchi and Franzò, 2018). (Munari, 2012) collected some of them, mainly used in framework in order to automatically assess at least a part of patent value (see Figure 3).

(Reitzig, 2004), in reviewing the most interesting variables, classify them into three 'generation':

[4] https://www.epo.org/searching-for-patents/business/ipscore.html#tab-1
[5] http://www.uibm.gov.it/index.php/brevetti/utilita-brevetti/griglia-di-valutazione-economica/259-proprieta-industriale/2006171-griglia-brevetti

- The first one collects patent data concerning the *general* economy of the patent, like backward and forward citations, family size and ownership;
- Parameters belonging to the second generation take into account more sophisticated feature of the patent, like the number of 4-digit IPC codes, the priority country and the kind of backward. Nonetheless, they exploit just the first page data and the search report of patent;
- The third generation of parameters relies on the availability of patent DBs collecting the full-text data of patents. It is the most promising in revealing the patent value due to the opportunity to extract the actual content of the patent.

Topic	Common indicators	Rationale	Empirical support
Legal status	Grant Y/N; Pending Y/N;	Valid patents strongly discourage competitors to use the invention.	Strong
International scope	Number of countries in which patents on the invention have been applied for; Triadic patent Y/N	Each patent is valid for a certain territory only. Covering all markets by patents requires significant investment.	Strong
Forward citations	Number of citations received from later patents, corrected for time-dependency	Further investment into related developments made; Invention contained useful aspects; Relevance of invention in later technology space	Strong
Opposition and Litigation	Survived opposition Y/N; Survived annulment Y/N; Infringement lawsuit Y/N	If competitors invest money in order to challenge the patent (or if they illegally use the technology), then this means the patent is valuable if it is upheld.	Strong
Technological Scope	Number of 4-digit IPC classes assigned to the patent	A broader technological scope could mean the market for the invention is larger.	Limited (contradictory findings)
Claims	Number of claims	The breadth of the claims defines the scope and effectiveness of protection from imitation	Limited (few tests)
Patent Filing Strategy	Choice of PCT system Y/N	Choices in the application process may reveal patent value as perceived by the applicant.	Weak (one indicator falsified, others essentially untested)
Inventors	No. of inventors	Number of inventors related to size of R&D investment; Patents of key inventors are more likely to be valuable;	Limited (few tests)

Figure 3: table of most commonly used parameters to ranking patent documents as presented by (Munari, 2012)

Information about the technological content generally concerns to one expert-in-the-field who must perform an accurate analysis to assess the quality of the invention. When the valuation object is a patent application, this analysis would be made at any procedural step, indeed the main uncertainty about the application related to its technological content. Furthermore, the analyst expert-in-the-field must consider, each time, the updated state of the art, which collects in turn new applications.

The thesis aims to integrate the current parameters for patent analyses, among which the patent valuation, with other coming from patent procedural DBs and focused patent data analysis tools. They allow the user to get useful indication about the risk of failure of a patent application and relevant technological background.

2 Working on patent applications: the most common mistakes in patent analyses

What seen in previous chapter is suitable for patent documents having a definitive and definite purpose. When the analysis is performed in presence of high amount of patent applications, it must be considered that their content evolves in time. Therefore, when a patent opinion is request, it is crucial paying attention to this kind of uncertainty.

This chapter collects a set of common mistakes faced by patent professionals in writing patent searches opinion in presence of a pool of document having a high number of applications.

The reasons for which a patent professional leads a patent search can be very numerous; for simplicity they can be summarized in three main categories:

- the first collects all kinds of search dealing with patent intelligence that means transforming patent data into technical, business and legal knowledge as the monitoring and survey of a specific technology or product, the survey of competitors patents, technology transfer, identification of emerging technology and technological trends;
- in the second group there are searches related to the state of the art for evaluating the patentability of a new innovative idea, writing a new patent, avoiding potential infringements with other patents, preparing legal action in order to protect our business/patents, determining the residual life of a competitor protection;
- the last group deals with due diligence for investment or transfers/acquisitions.

Each of these activities needs a specific patent search having its own peculiarity, specific strategies, techniques and search tools.

In this chapter, we focused on activities collected in the second group, which request to analyse a big number of patent applications and express an opinion before their grant. The group aggregate different kinds of search, like e.g. freedom-to-operate, clearance search, infringement search, right-to-use, etc[6].

[6] Unfortunately, the definition of cited searches lacks in precision and shared explanation. For example, WIPO in its guidelines edited by (Trippe, 2015) and (Alberts *et al.*, 2011) consider clearance, freedom-to-operate, infringement and right-to-use, as synonyms, whereas (Hunt, Nguyen and Rodgers, 2012) discriminate between infringement and clearance, while not distinguish right-to-use and freedom-to-

Although these searches differ each other, they share the evaluation of the legal bounds claimed by a pool of patents. Especially, they focus on the claims of a patent document in order to define which is the matter bound by them and what is free from restriction.

Despite the little interest given to the topic by the literature, the importance of search opinion uncertainty is increased due to the great growth of patenting activity in the last years, as shown by main Patent Offices reports: EPO[7], WIPO[8], USPTO[9].

Below in Table 1 and Figure 4, the data about the growth of patent activity in the world since 2000.

Table 1: patent document publications in the world. The growth rate refers to total published documents. Source: WIPO Intellectual Property Statistics[8]

Year	*Applications filed*	*Patents granted*	*Total published*	*Growth rate*
2000	1,377,400	517,600	1,895,000	-
2001	1,456,600	538,100	1,994,700	5%
2002	1,444,200	561,600	2,005,800	1%
2003	1,484,200	620,600	2,104,800	5%
2004	1,574,300	626,400	2,200,700	5%
2005	1,703,200	633,100	2,336,300	6%
2006	1,791,700	754,500	2,546,200	9%
2007	1,875,000	776,300	2,651,300	4%
2008	1,930,100	781,700	2,711,800	2%
2009	1,855,900	814,400	2,670,300	-2%
2010	1,997,500	914,800	2,912,300	9%
2011	2,158,400	1,001,700	3,160,100	9%
2012	2,356,500	1,137,700	3,494,200	11%
2013	2,564,500	1,169,700	3,734200	7%
2014	2,680,700	1,173,900	3,854,600	3%
2015	2,886,700	1,240,200	4,126,900	7%
2016	3,125,100	1,352,300	4,477,400	8%
2017	3,168,900	1,404,600	4,573,500	2%

operate from clearance. We also found many IP specialists that introduce specific definitions for clearance and for freedom-to-operate exchanging their meanings (http://www.filament.com.au/patent/search/clearance-search.aspx).

[7] www.epo.org/about-us/annual-reports-statistics/annual-report.html
[8] www.wipo.int/ipstats/en
[9] www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm

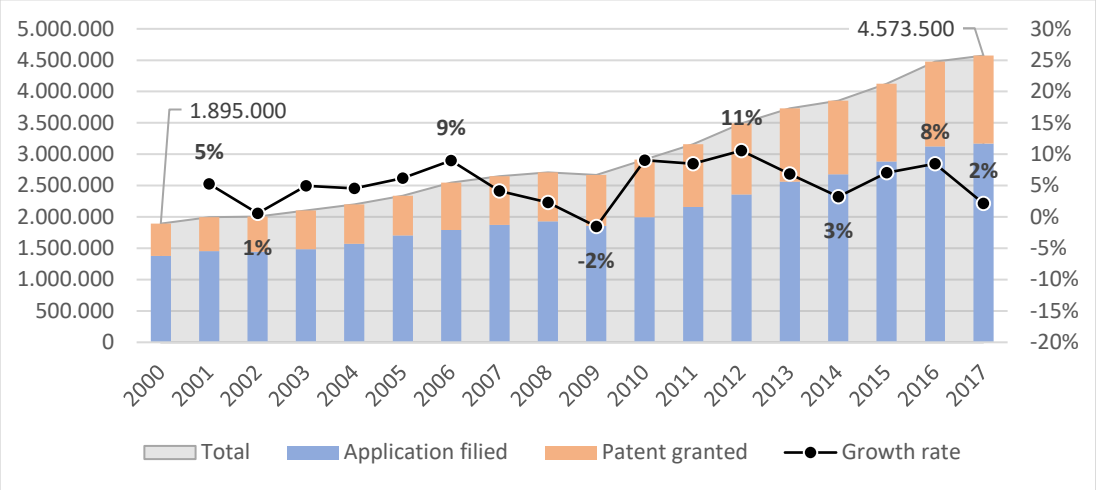


Figure 4: world patenting activity behaviour and growth rate. Source of data: WIPO Intellectual Property Statistics[8]

A great number of patent applications causes that the opinion cannot pinpoint a boundary in which there are no legal risks, but, rather, it gives a rough idea about which are the technological features having an uncertain protection. Thus, it is crucial be aware of the risks the applications carry with them and consider all the sources of information able to limit them.

In writing a patent search opinion, one of the potential consequences of neglecting the uncertainty can be, for example, the launch of an R&D project that lead to the filing of a patent application, which infringes the right of a prior one. This can lead to many alternative scenarios as function of numerous factors. For the sake of simplicity, we report the potential procedural events resulting from an unaware EP application filing:

- *application withdrawal*: as a result of the Preliminary Search, the Patent Authority publishes a Search Report highlighting prior documents claiming the same topic. The applicant can decide to withdraw the application, assuming the risk of unprotected investment.;
- *filing of amendments*: in same situation of a negative Search Report, to continue with examination, the applicant can change the text of the application, paying the unexpected fees;
- *refusal of patent*: if the examination has been requested, the examiner can refuse the application, e.g. because it lacks in novelty or originality. Again, there is the risk of unprotected investment worsened by the examination fees;
- *opposition filing*: in case of patent grant, a third party can file with the Patent Authority an opposition to fight the issue. If it reaches the goal, the patent will

be revoked. Once again, the effect is an unprotected investment worsened by unexpected fee cost.

To understand the impact of amendments and withdrawals, we collect procedural information from PATSTAT database, the procedural raw-data DB published by EPO. Table 2 and Figure 5 show the behaviour in time of the two events in European grant-proceeding.

Table 2: behaviour of EP application filed, amended and withdrawn per year elaborated from PASTSTAT

Year	*App. filed*	*Procedural event*		*Event share*	
		Amendment	Withdrawal	Amendment	Withdrawal
2005	128,709	24,387	40,585	19%	32%
2006	135,399	29,872	42,198	22%	31%
2007	141,231	31,844	44,136	23%	31%
2008	146,244	42,861	52,459	29%	36%
2009	134,511	49,566	59,742	37%	44%
2010	151,015	63,526	63,945	42%	42%
2011	142,822	92,293	56,245	65%	39%
2012	148,562	108,353	53,565	73%	36%
2013	148,027	112,508	57,488	76%	39%
2014	152,703	117,582	58,931	77%	39%
2015	160,004	146,320	61,178	91%	38%

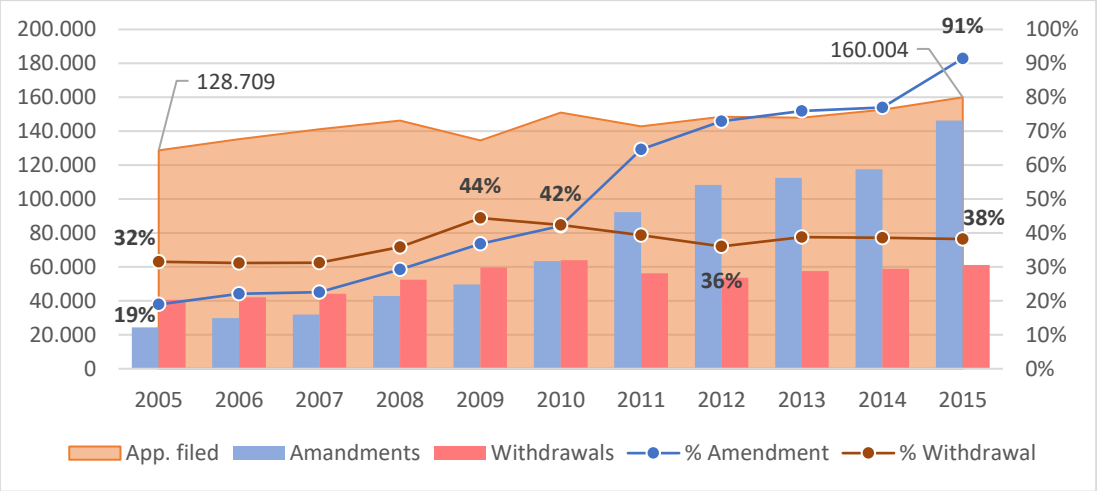


Figure 5: behaviour of EP application filed, amended and withdrawn per year

The share detailed in Table 2 and Figure 5 refers to the number of applications filed, despite amendments and withdrawals occur, generally, some months after the filing. Nonetheless, we consider the ratios representative enough to address the impact of the two events.

The effect of the refusal must be measured in reference to the examinations carried out by the patent office in the target year. We assume that decision taken after examination is binary: grant or refusal. For uniformity, in Table 3 and Figure 6 we consider European patent data about examination provided and patent granted, collected by EPO annual reports[7].

Table 3: behaviour of EP application filed, amended and withdrawn per year

Year	Examination	Decision		Decision share	
		Grant	Refusal	Grant	Refusal
2005	84,719	53,251	31,468	63%	37%
2006	96,422	62,777	33,645	65%	35%
2007	90,310	54,700	35,610	61%	39%
2008	99,053	59,800	39,253	60%	40%
2009	102,178	51,952	50,226	51%	49%
2010	114,991	58,117	56,874	51%	49%
2011	110,331	62,108	48,223	56%	44%
2012	111,860	65,655	46,205	59%	41%
2013	116,820	66,707	50,113	57%	43%
2014	96,062	64,613	31,449	67%	33%
2015	113,586	68,419	45,167	60%	40%

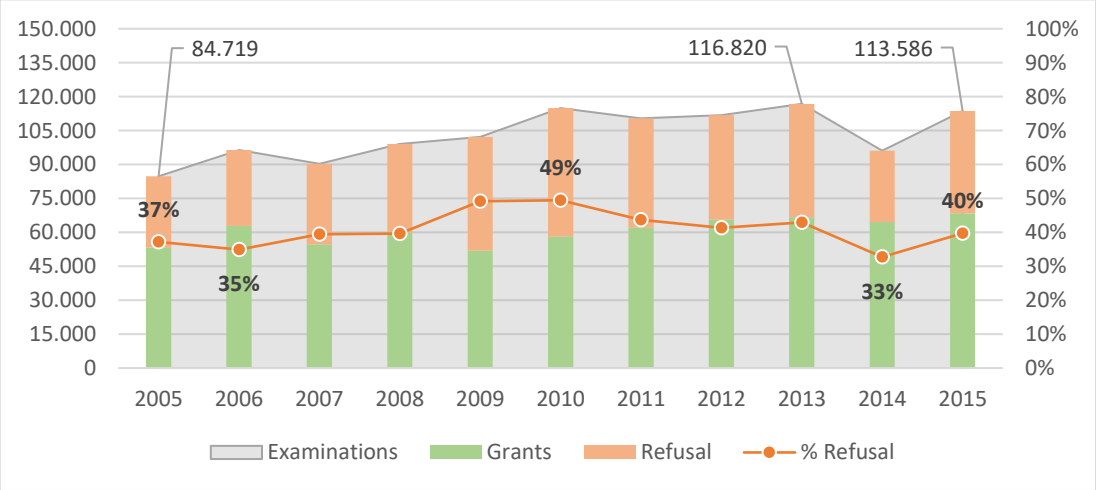


Figure 6: behaviour of EP examinations, granted patents and refusals per year

According to what depicted in Table 3 and Figure 6, across 2005-2015 the share of refusal fluctuated between one third and nearly a half of examination decisions. Part of the refusals can be referred to the uncertainty due to applications in pre-filing patent searches. Reducing the uncertainty, it should decrease the refusal ratio.

Finally, the opposition is the chance a third party has in order to contest a patent grant issued to the applicant. If the opposition were accepted, the patent would revoke or at least amended. To assess the impact of the opposition on the granted patent, we considered the data published by EPO in annual reports[7].

Table 4: behaviour of EP patent granted and opposed per year

Year	*Patents granted*	*Oppositions filed*	*Opposition share*
2005	53,251	3,126	5.9%
2006	62,777	3,000	4.8%
2007	54,700	3,293	6.0%
2008	59,800	2,840	4.7%
2009	51,952	2,695	5.2%
2010	58,117	2,766	4.8%
2011	62,108	2,945	4.7%
2012	65,655	2,994	4.6%
2013	66,707	2,963	4.4%
2014	64,613	3,063	4.7%
2015	68,419	2,898	4.2%

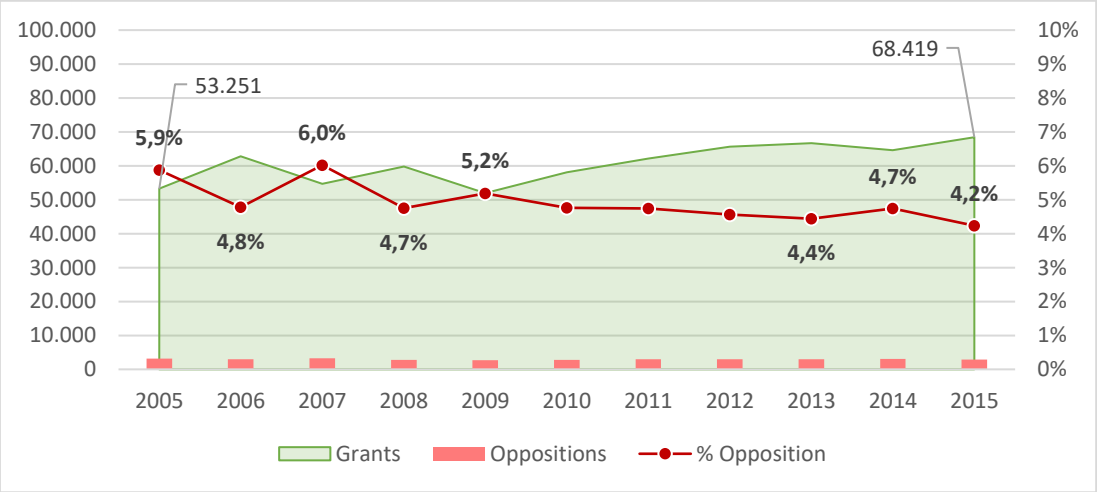


Figure 7: behaviour of EP patent granted and opposed per year

Although the limited impact of oppositions on patent literature, they are interesting due to two reasons. First, they affect granted patent, revoking or amending them. Second, to support the opposition, the third party must give evidence of the presence of known prior art to which the contested patent refers. Collecting information to reduce the uncertainty should reduce the chance of opposition filing.

Application amendment, withdrawal, refusal and patent opposition are examples of possible consequences of neglecting the uncertainty in patent searches carried out in presence of a big number of patent applications. Many other scenarios can occur, all of them entail unexpected costs, risk of unprotected investment and sometimes litigations.

The main factor related to patent applications that raises complexity of patent searches and worsen effort and risk taken on writing an opinion is the content of the document, that is not definitive and may change during the procedure.

As shown in Figure 5, the amendments on applications took more than 90% of EP patent applications in 2015. The filing of amendments is the way an applicant has in order to modify the text of an application, for example after receiving the Search Report. It can modify the title, the abstract, the description and the claims. Furthermore, amendments can be filed more than once.

Actually, the patent text may also change after that the patent application has been granted, but fortunately, the impact of the number of oppositions on the number of granted patents is very low and almost constant (see Table 4).

Chapter 2.1 offers a survey of the issues about the content of patent applications, which are the most common sources of risk, and where the patent specialist can collect information about it in order to limit their negative influence.

2.1 Ultimate version of a patent text

One of the major problems faced by patent examiners is the fact that during the progress of the procedure, the patent application has many occasions to change its contents. This causes uncertainty in its interpretation. Below, we report some events that might happen and some indications about the sources of information in order to know whether they occur and eventually, which modifications they cause.

2.1.1 Amendments in PCT procedure

The PCT procedure allows the applicant to amend its PCT application at specific steps[10], which are:

- PCT Article 19 Amendments
- PCT Article 34 Amendments

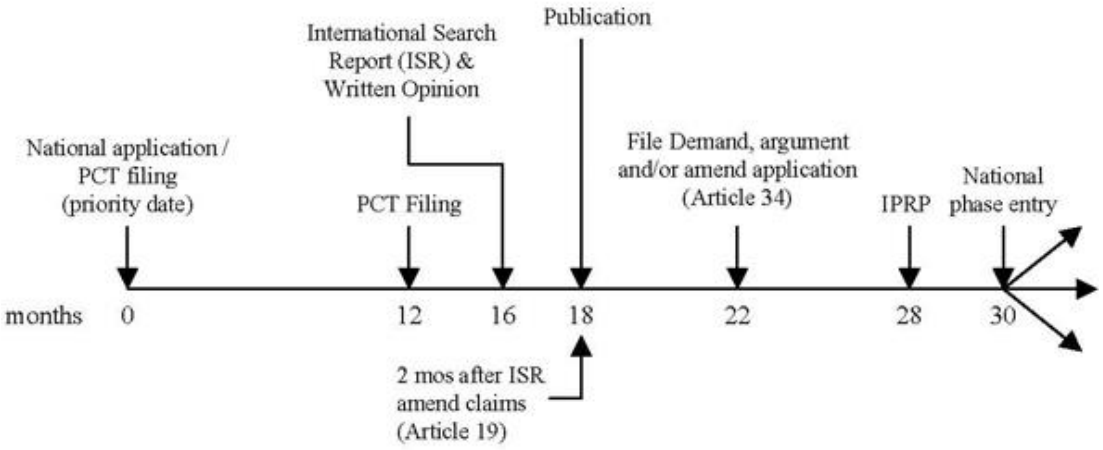


Figure 8: PCT application proceeding timeline. Source: WIPO[11]

PCT Article 19 amendments

First amendments may occur within 16 months from priority date[12] or, if later, 2 months after the transmittal of the International Search Report (ISR) by the International Searching Authority (ISA) to the applicant[13]. The available publications of the PCT application can have different kind codes, which gives partial information about Article 19 amendments progress. The patent analyst can perform simple checks to better understand whether the applicant filed amendments, still has time in order to file them or did not file them:

[10] http://www.wipo.int/pct/en/faqs/faqs.htm
[11] https://www.wipo.int/export/sites/www/pct/en/seminar/basic_1/timeline.pdf
[12] This is a theoretical time limit only, indeed "Amendments received by the International Bureau after the time limit are still accepted if they have been received before the technical preparations for international publication have been completed"
(http://www.wipo.int/pct/en/faqs/amendments_19_and_34.html)
[13] http://www.wipo.int/pct/en/faqs/amendments_19_and_34.html

A1 code: application with ISR

- **in amended form** (e.g. WO2014059450), in which the amended sheets show "AMENDED SHEET (ARTICLE 19)" stamp. In case the amendment is only a cancellation of entire sheets, the "*Letter accompanying Amended claims*" record, in PATENTSCOPE, proves the exploitation of Article 19;
- **with subsequent A4 publication** (e.g. WO2016038606), that is the official publication of the amendments;
- **having remaining time to be amended**. This is a risky situation because the patent search is done within the time available to the applicant to exploit the Article 19 amendments, hence no information about them might be find. The risk lasts for some time after the two-month-long window, because there is technical delay in publication on PATENTSCOPE of amendments. Some indications are available in paragrph 2.1.4.
- **not amended** (e.g. WO2014172463). The analyst can state the application was not amended if he/she does not find available documentation about amendments in PATENTSCOPE during the search made beyond the reasonable time-limit of 15 weeks from the date of the ISR transmittal, which is reported in the "*Date of mailing of the international search report*" field in ISR. In such a case, the reference application is the A1 publication.

A2 and A3 codes: application without ISR (A2) and later ISR publication (A3)

- **with subsequent A4 publication** (e.g. WO2016050173), that is the official publication of the amendments;
- **having remaining time to amend**. Like above, this is a risky situation for the same reasons. Notice that A3 publication (the ISR) will ever published, but the delay can be very long (e.g. for WO2015090938), so also the time before ISR transmission can be considered a risky period. Some indications are available in paragraph 2.1.4.
- **not amended** (e.g. WO2015196121). Again, there is a reasonable time window of 15 weeks from the ISR transmittal beyond which the analyst can state the Article 19 was not exploited. In such a case, the reference application is the A2 publication.

PCT Article 34 amendments

As intrinsically suggested by WIPO[14], the applicant who wants to hide as long as possible its definitive application does not exploit the Article 19 amendments, but files for Demand for Preliminary Examination[15]. This gives him/her the chance to amend the application under Article 34, hiding the changes which are confidential till to the 30th month from priority date. The demand automatically indicates that the results of examination will be used in all PCT countries (*PCT Applicant's Guide 10.001*). The only available information about Article 34 procedure is the filing date of the demand, available, if filing was done, from the "*International Report on Patentability (IPRP) Chapter II of the PCT"* field of International Application Status Report (IASR), on PATENTSCOPE. The applicant can file the demand up to 22 months from priority date or, if later, 3 months from the ISR transmittal date. Three countries, Luxemburg, Uganda and Tanzania, have shorter time limit for filing, i.e. 19 months from priority date, because they not yet adapted their national law to the *PCT Article 22*[16]. If the applicant files the Demand within this shorter time limit, the procedure in these countries pursues like other ones, otherwise they start the national phase, upon fee payment, within the 21th month. The potential scenarios are:

- demand filed
    - **within 19 months from priority date**, then the international application will undergo potential amendments and be examined for all PCT countries;
    - **between 19 and 22 months from priority date** (or, if later, 3 months from ISR transmission), then the international application will be examined and will undergo potential amendments for all PCT countries, less Luxemburg, Uganda and Tanzania;
- demand not yet filed and search done
    - **after 22 months from the priority date** (or if later, 3 months from ISR transmission), then the international application can enter in national phase in its published form (A1, A2 or A4);
    - **between 19 and 22 months from priority date** (or if later, 3 months from ISR transmission), then the applicant still have chance to amend under Article 34 for all PCT countries less Luxemburg, Uganda and Tanzania;

[14] http://www.wipo.int/pct/en/faqs/amendments_19_and_34.html
[15] The Demand filing is always possible, regardless the exploitation of Article 19
[16] http://www.wipo.int/pct/en/texts/time_limits.html

- **before 19 months from priority date**, then the analyst has no data to reduce the risk about amendments under Art. 34.

Amendments under *PCT Article 34* can be filed together with the demand or after, up to the international preliminary examination report is established. Anyway, the patent analyst cannot get them before the end of international phase, that occurs at the 30th month from priority date. At that time the International Bureau (IB) publishes all documents about the international application. The IPRP[17] is the final document of the international phase.

National or regional entry amendments

The national or regional phases are independent each other, indeed each jurisdiction complies with its own national/regional patent law. The applicant can start them before the end of the international phase. Furthermore, the entries may be done in different times. For the countries with prior entry, the subsequent amendments on the international application not need to be take into account[18]. Contrary for other countries in which the entry occurs after the amendments filing, the reference application is the amended one. Nevertheless, at the entry in national phase each national/regional application can be further amended, worsening the risk. Unless the US Continuation-In-Part procedure, which allows the applicant to add matter to the prior disclosure (*US MPEP 201.08*), the amendments cannot add new matter to the first application.

2.1.2 Amendments in EPC procedure

Even EPC procedure allows the applicant to amend its application in description, claims or drawings, but only after the receiving of the Search Report (*EPC Rule 137*). Even in this case, the amendments cannot add new matter to the first disclosure. If a European application coming from a PCT procedure, the amendments can be done in response to the International Search Report after the entry in regional phase (e.g. EP2793232). Information about the amendments can be retrieve from the EP register.

[17] Chapter II if the demand was filed, Chapter I otherwise
[18] https://xepc.eu/node/pa_062010

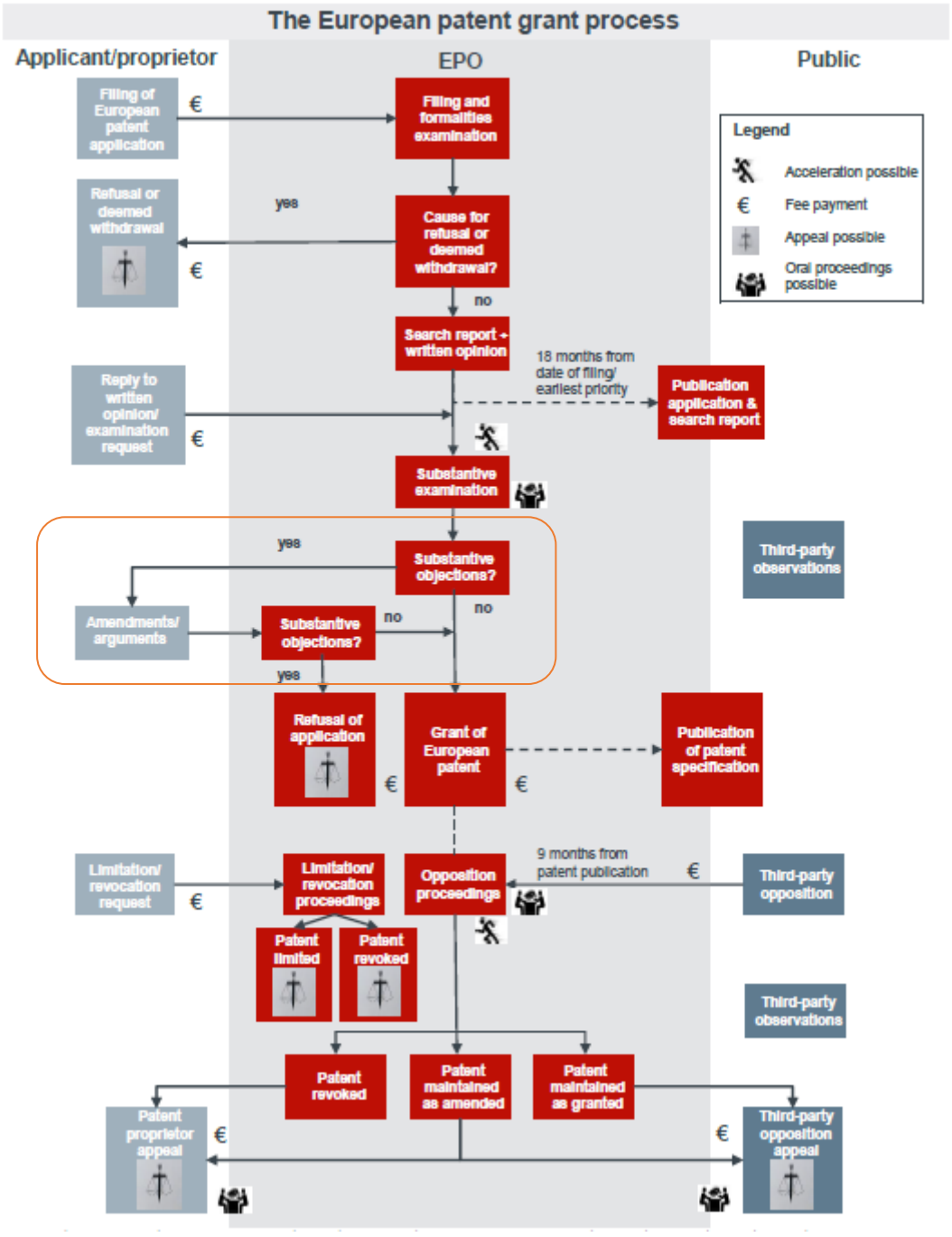


Figure 9: EPC grant-proceeding timeline. Source: EPO

2.1.3 Amendments due to opposition decision

The opposition proceeding is an objection made by third party about the grant of a patent. Decision in opposition cases may amend the text of the granted patent. The opposition mechanism is defined by each jurisdiction complying with its own patent law. Its procedures differ, inter alia, in filing time limits. The EPO lets the opponent to file within nine months[19] from the publication of the grant, like the USPTO which calls this kind of procedure Post-Grant Review[20]. The limit for the JPO is shorter, 6 months[21]. In China the opposition procedure was abrogated in 2001, thus the only current procedure allowing third party to oppose a grant is the invalidity procedure[22], which does not have time limit (likewise other countries). The EPO publishes the documents about opposition to EP patents in the EP Register, USPTO collects them in the Patent Review Processing System (PRPS)[23] and JPO publishes the decisions in J-PlatPat[24].

2.1.4 Potential amendments prediction based on search report and written opinions

When the applicant still has time to amend its application, the search reports and written opinions may help the patent specialist to evaluate what is the likelihood with which the text can change during the procedural steps left. Search reports from some authorities are more interesting than other ones, e.g. the US because USPTO examiners usually cite US document only for X and Y categories (Michel and Bettels, 2001; List, 2010). Very useful are the search report of Asian patent office, indeed they suggest precisely some documents otherwise not considered by patent specialist not skilled in Asian languages. Especially the JPO is considered the most eminent Asian office thanks to its efforts in accelerate the examination and stabilize the granted patents (*JPO Annual Report 2014*). Likewise, some offices are considered stricter than others, e.g. DPMA (Germany) and again JPO[25] (de Rassenfosse, Jaffe and Webster, 2016). Granted patents in those offices are stronger than in many other countries and their search reports contains information hard to reach. The search reports are available on Espacenet in the EP register/Global Dossier or in national patent offices.

---

[19] https://www.epo.org/about-us/jobs/examiners/what/opposition.html
[20] https://www.uspto.gov/patents-application-process/appealing-patent-decisions/trials/post-grant-review
[21] https://www.jpo.go.jp/english/ip-rights/appeals-trials/pdf/opposition/operation_of_system.pdf
[22] http://english.sipo.gov.cn/FAQ/200904/t20090408_449718.html
[23] ptabtrials.uspto.gov
[24] www.j-platpat.inpit.go.jp/web/all/top/BTmTopEnglishPage
[25] http://www.fosspatents.com/2011/08/preliminary-injunction-granted-by.html

Reliability of written opinion in PCT procedure when a withdrawal of priority claim occurs

Time limits of PCT procedure are function of the priority date. Generally, this is the date of the first national filing, but the applicant might decide to withdraw the priority during the procedure (e.g. WO2015053620 and WO2013188948). All time limits based on the priority date, not yet expired, will be postponed on the basis of the new priority, up to the international filing (*PCT Rule 90bis*). In such a case, ISR will remain valid, indeed it collects the "relevant prior art" literature published before the international filing date (WIPO, 2004). Instead, the written opinion of the ISA provides a detailed explanation of the relevance of the references published before the priority date (WIPO, 2004), and how they affect the patentability of the invention. The difference about the two dates could be very important, even more than 1 year (e.g. WO2013188948)! The patent specialist can reach the request of withdrawal on PATENTSCOPE some days after the receiving by IB.

2.2 Conclusions

The rapid growth of patent applications tends to increase the level of uncertainty in the patent context. This leads to take on greater risks in writing a patent search opinion than in the past. An improvement to the methods aiming to the risks reduction and reliability increasing of a patent search opinion is increasingly felt.

Referring to patent applications, the evolution of their content during granting proceeding is a crucial origin of uncertainty. We listed some specific and common cases in which the content of the application may change. The large amount of opportunity to do it in grant and post-grant proceedings highlights how difficult is tracking the content evolution of an application. Even more when a pool of documents includes many of them.

This work indicates the sources of information available to retrieve the last version of an application content. It is useful for reducing the risk in patent application analysis and for ameliorating already existent tools for patent searches, implementing new modules for warn the user about the revised versions of patent text and get an overview of the patent evolution along the procedural time.

3 A risk assessment tool

The use of patent applications has turned out to be pivotal for the evaluation of intellectual property right's quality, showing an increasing path of two main factors. First, the number of patent applications per year is globally raising, up to reach 3 million of applications per year in the 2016 (WIPO, 2017). Second, although the attempts made by patent offices, the time taken by patent granting procedure is still very long[26,27] (Schultz and Madigan, 2016; Beatty, 2017).

As discussed in Chapter 2, applications carry with them uncertainty about claimed matter. This Chapter focuses on most interesting steps in granting and post-grant European proceedings to assess the likelihood an EP application has in order to be granted. It aims to support the attorneys in estimation of the success rate for an EP application.

Figuring the final result of patent proceedings out is not a trivial work. Actually, the failures of applications are around 40%, as shown in Figure 6 and Table 3 and detailed in this Chapter.

Patent applications influence different kinds of patent analysis: risk management in patent opinions (Bergmann *et al.*, 2008; Fenton, 2016; Carrara and Russo, 2017), valuation methods (Austin, 1993; Pitkethly, 1997; Reitzig, 2004; Chiesa, Gilardoni and Manzini, 2005; Chiesa *et al.*, 2007; Chiu and Chen, 2007; Suzuki, 2011), patent indicators (Baglieri *et al.*, 2001; Reitzig, 2004; Chiesa, Gilardoni and Manzini, 2005; Gans and Hsu, 2008; Nagaoka, Motohashi and Goto, 2010; van Zeebroeck, 2011; Thompson, 2016), market for innovation (Chiesa, Manzini and Pizzurno, 2008; Harhoff and Wagner, 2009), competitive intelligence (van der Drift, 1988; Zanasi, 1998; Rouach and Santi, 2001; Grandjean *et al.*, 2005; Shih, Liu and Hsu, 2008, 2010), patent office performances[26,27] (Quillen and Webster, 2001; McAleer and Slottje, 2005; Yang, 2008; Frietsch, Neuhäusler and Rothengatter, 2013; Schultz and Madigan, 2016; WIPO, 2017), etc.

Anyhow, independently from the reason for which patent applications are interesting, a crucial parameter about them is the likelihood they have in reaching a grant. Nonetheless, the attention given to their *success rate* is very limited

[26] https://www.ic.gc.ca/eic/site/cipointernet-internetopic.nsf/eng/h_wr04112.html
[27] https://patentlyo.com/patent/2015/06/patent-pendency-redux.html

3.1 Success Rate of a patent application

To avoid possible confusion, we define the *success rate* (*SRate*) of a patent application as its likelihood to be granted. This index differ from what considered by (van der Drift, 1988; Schankerman, 1998; Quillen and Webster, 2001; Yang, 2008; van Zeebroeck, 2011; Schultz and Madigan, 2016; WIPO, 2017), named *grant ratio* (*GR*), that is the number of yearly granted patents over the number of applications filed per year. Both *SRate* and *GR* refer to a specific Patent Office. Below, the comparison of *grant ratio* and *success rate* formulas.

$$GR^{PO,Y} = \frac{\sum_i g_i^{PO,Y}}{\sum_j f_j^{PO,Y}}$$

$$SRate^{PO,Y_1} = \mathcal{L}\left(f_k^{PO,Y_1} \middle| g_k^{PO,Y_n}\right)$$

Where *GR* and *SRate* are computed for Patent Office *PO* and year *Y*, where *f* means "application filing" and *g* means "patent granted".

GR is an index useful to measure the annual performances of a patent office using a "fast" and easy to compute value. Indeed, number of applications filed f_i and patent granted g_j per year are available at the closing of fiscal year, whereas the grant g_k of an application f_k might occur many years after the filing.

About *success rate*, a first study has carried out by (Guellec and van Pottelsberghe de la Potterie, 2000). They studied the correlation between some economical parameters and the success rate of patent applications at the European Patent Office. Within the results, the most interesting one is the correlation between the filing route and the success rate of the application. This specific topic was deepen by (Frietsch, Neuhäusler and Rothengatter, 2013), who raises the question whether the choice of the filing route influences the likelihood of a patent filing being granted, withdrawn or refused after the examination process at the EPO. Both the cited works found the filing route could be a strategic variable to consider when file for an EP patent application.

Although other variables have been studied in these articles, they cannot be used as strategical parameters in order to maximize the granting likelihood, like filing route does. The dimension of the company and the country of the inventor, for example, cannot be modified by the applicant, and designed countries or the patent family dimension could not be modified without changing the patent effectiveness.

No other contributes have been found about the likelihood for an application to become a patent.

Except for what concerns the filing route, available analyses do not consider which path the patent application follows along EP proceedings. This work put the focus on main events that could occur in these examination processes.

If a likelihood in reaching a grant exists, then a risk that the application for patent fails the issue exists too. In the latter case, no right of exclusion could be enforced against potential competitors and the value of the application is dramatically reduced or totally voided in filing country.

Currently, the professionals able to express the most reliable judgement about *SRate* are the patent attorneys., They are skilled in understanding the path followed by an application throughout the granting process. Most interesting events they keep in consideration are:

- The Search Report (SR), published by the Patent Office, whose content is the first official opinion about the patentability of the application;
- The amendment (A), filed by applicant, which shows that it modified the content of the application also considering the objections highlighted in SR;
- The opposition (O), filed by a third party within 9 months from the grant date, that proves the attempt to oppose to the granting.

It is a demanding task, also considering a single application. In fact, some deep analyses must be carried out to retrieve all the information needed and arrange them on a timeline. It requires assessing the structure and the state of the patent family, reading different documents, sometimes in different languages, comparing different content versions modified by amendments, reading the search report and opinion and compare them with the attorney answer, etc. Carrara and Russo discussed about some drawbacks related to this work (Carrara and Russo, 2017). If an assessment on the value of a large patent portfolio is needed, the judgement of the success rate of patent applications becomes a prohibitive task, both in time consumption and in costs.

In a knowledge-based economy, where the application surge pushes toward ever larger patent pools (Powell and Snellman, 2004; van Zeebroeck, 2011), populated mainly by applications, it is crucial to develop a method able to systematically assess the grant likelihood an application has and use it as a risk indicator for other analyses, like valuation, infringement risk, freedom-to-operate, etc.

As a first attempt, this chapter introduces a statistical description of common procedural scenarios that considers the most relevant and frequent events in a granting procedure, to understand which their impact on the granting likelihood is. The data are

extracted from the Worldwide Patent Statistical Database (PATSTAT), mainly from the EP Register section.

3.2 PATSTAT Register

The European Patent Office collects and publishes publications (applications, search reports and granted patents) from other Offices in the world. Twice a year, the EPO export all the data collected in PATSTAT, a patent raw-data database specifically dedicated to statistical usage and prepared by EPO on behalf of the OECD Taskforce on Patent Statistics (Kang and Tarasconi, 2016; European Patent Office, 2017a).

In EP Register and Global Dossier[28], EPO publishes also procedural documents from main Patent Offices: CIPO, CNIPA, JPO, KIPO, USPTO and WIPO, other than EPO itself. PATSTAT has a set of tables dedicated to procedural data, PATSTAT Register, but only to the European ones. It is the most complete and up-to-date procedural information on European patent applications and grants[29]. It collects data since 1978, up to the week of its publication that occur twice a year.

Due to the completeness and the long history of data collected by PATSTAT Register, this is the most reliable and suitable data source from which extract the procedural data to analyse and compute the alternative paths followed by applications along the granting and post-grant processes. No other database is so rich in patent proceeding data like this. This means the analysis cannot consider applications other than European ones. The work presented here used the 2017 Spring Edition, version 3.04, published in April 1st, 2017.

Unfortunately, not always the data collected in PATSTAT are easy to be read. For example, the procedural event codes listed in table '*REG301_EVENT_DATA'* are difficult to be interpreted and associated to the related event. As function of the law changes, also some codes might change although they refer to the same event. There are also some codes used differently among examiners. These issues must be faced in order to obtain reliable data on which build the statistical model for the granting likelihood of patent applications.

PATSTAT Register arranges information as shown in Figure 10, using a structure based on relational DB. The rectangles represent sets of tables grouped according to the kind of information they gather. Each row in the rectangles starts with the code of

[28] https://worldwide.espacenet.com/help?locale=en_EP&topic=globaldossier&method=handleHelpTopic

[29] https://www.epo.org/searching-for-patents/legal/register.html#tab-1

a table, e.g. *REG101_APPLN*, followed by a brief description of the content. Deeper details are available in (European Patent Office, 2017b).

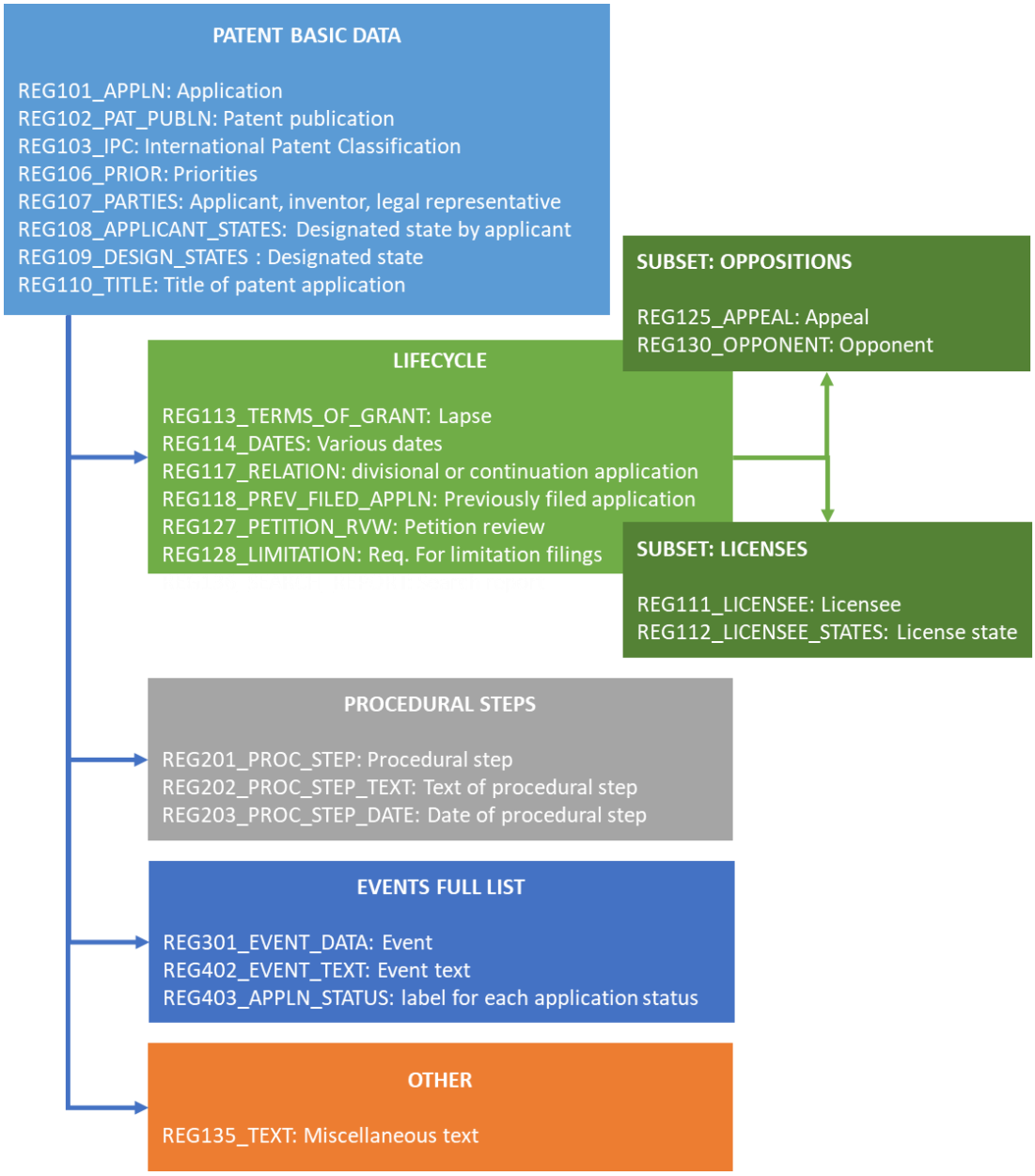


Figure 10: Main structure and information contained in PATSTAT Register 2017 Spring Edition. Source: (Tarasconi, 2018)

EPO classifies the content of PATSTAT Register in two main type of applications:

- EP direct applications, which group
    - EP first filing applications;

    - EP filings under the Paris Convention[30] from previous national filings;
- PCT international applications, which are either
    - EURO-PCT applications: European patent applications that have entered the European procedure via a PCT international application;
    - PCT-applications which have not (yet) entered the European procedure

PATSTAT defines application types using the available combinations of *EP application identifier (EP ID* and *PCT international application identifier (PCT ID)* which are the univocal identification codes provided to retrieve a specific application.

EP ID does not discriminate between first filing and Paris Convention filing, but in combination with *PCT ID* it allows the user to separate EURO-PCT from PCT applications. Table 5 and Figure 11 summarize the share taken by each type of application in PATSTAT Register 2017 Spring Edition.

Table 5: types of application collected by PATSTAT Register and their share

Application Type	*EP ID*	*PCT ID*	*#*	*Share*
EP direct	> 0	= 0	1.605.399	35%
EURO-PCT	> 0	> 0	1.508.958	33%
PCT	= 0	> 0	1.448.381	32%
???	= 0	= 0	72	0%
Total			*4.562.810*	*100%*

They show a fourth unclassified type of applications, which has not EP nor PCT identifier. Since PATSTAT does not give details about this class and its very low impact (around 10^{-5}), it has been neglected in this analysis.

Despite the presence in PATSTAT of PCT-applications not yet entered in European phase, no useful data are available about their proceedings, indeed Europe is only a designed country at this stage.

At the time of publication, PATSTAT Register classifies the procedures according to their progress in patent proceeding lifecycle. A list of 18 statuses resumes the progress in EP proceedings of a patent application. Figure 12 show the flowchart of the statuses.

[30] http://www.wipo.int/treaties/en/text.jsp?file_id=288514

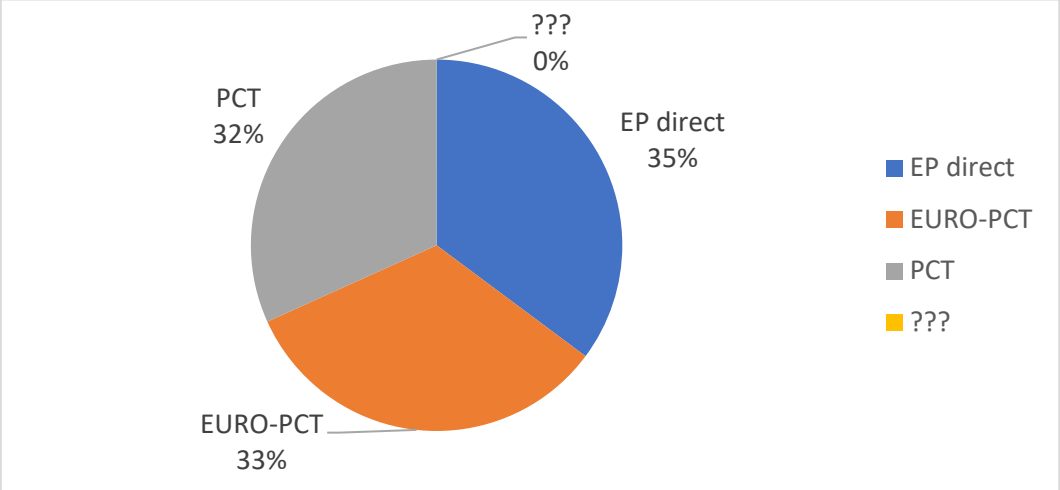


Figure 11: share of application types in PATSTAT Register 2017 Spring Edition

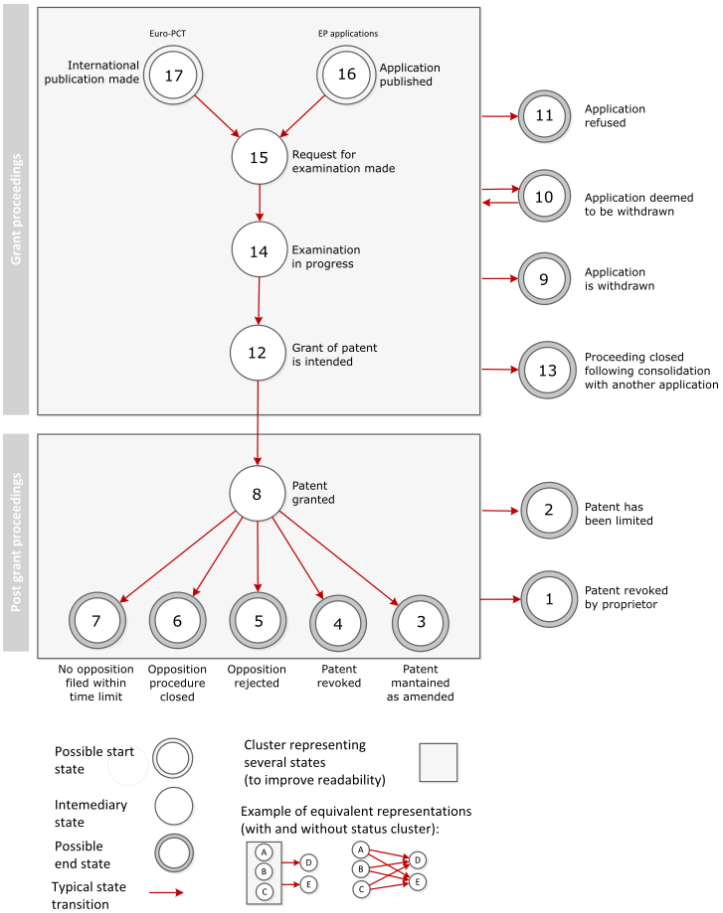


Figure 12: status transition diagram, from Data catalogue PATSTAT Register (European Patent Office, 2017b)

The diagram shows the flow through the granting and post-grant proceedings of the applications. The number assigned by the EPO to each status is reverse in reference to the chronological sequence, so the first status of an application could be 16 or 17 and the status after grant, opposition and decision could be 1, 2, 3 or 4.

During the lifecycle of the patent, an application might change its end status. For example, an *application deemed to be withdrawn* could be recover by the applicant paying an additional fee. In this case, in next edition of PATSTAT register, the same application would appear with another status, for example *patent granted*. Although the uncertainty about the end status exists, we assume that it is limited and does not modify considerably our results.

Table 6: amount and shares of the statuses of applications in grant and post-grant proceedings collected in PATSTAT Register 2017 Spring Edition

Status	*Description*	*# of documents*	*Share*
1	Patent revoked by proprietor	67	0.0%
2	Patent limited	326	0.0%
3	Patent maintained in amended form	21,450	0.5%
4	Patent revoked	24,449	0.5%
5	Opposition rejected	15,701	0.3%
6	Opposition procedure closed	5,707	0.1%
7	No opposition filed within timelimit	1,305,421	28.6%
8	Patent granted	113,378	2.5%
9	Application withdrawn	185,385	4.1%
10	Application deemed to be withdrawn	770,046	16.9%
11	Application refused	83,773	1.8%
12	Grant of patent intended	41,859	0.9%
13	Proceedings closed following consolidation with another application	211	0.0%
14	Examination in progress	179,963	3.9%
15	Examination requested	332,871	7.3%
16	Application published	33,750	0.7%
17	International application published	1,448,381	31.7%
18	Unknown	72	0.0%
	Total	*4,562,810*	*100%*

Applications and patents are assigned to statuses at the time of PATSTAT publication. Their distribution composes a sort of photography about the situation at that time. It is a simpler way to assess the progress of documents in proceedings then retrieving all data needed to understand it.

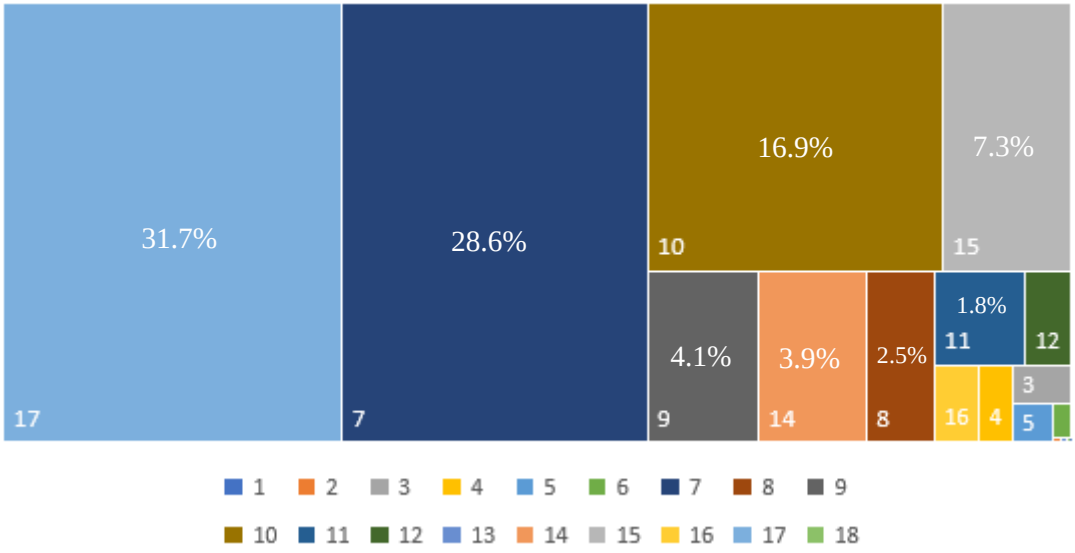


Figure 13: shares of the states of applications in grant and post-grant proceedings collected in PATSTAT Register 2017 Spring Edition

According to what said about PCT applications not yet entered in EP proceedings, the number of documents of statuses 17 and 18 in Table 6 are the same of the ones of PCT and unknown class in Table 5.

3.3 Dataset

This work focuses on procedural events that may happen along the grant and post-grant proceedings at the European Patent Office. Consequently, only those applications filed with the EPO, both directly or through PCT, have been considered. According to what shown in Table 5, only application having a positive *EP ID* are useful in order to make up the dataset on which develop the study.

Furthermore, as listed in Table 6, the statuses from 14 to 18 and 12 refer to the intermediate steps of the grant proceedings. Applications belonging to these cannot give information about their actual success or failure result, thus, it has been set up a bound on the filing year, on or before 2001, to filter out 'Temporary' statuses from the dataset. This limit reduces under 1% the impact of those applications which have not reached at least one of the possible end statuses, for both EP direct filing and EURO-PCT docs separately.

Table 7 summarises the Boolean filters used to define the documents included in dataset according to what just described. Filters are presented sequentially, they work on results of the row above, defining a subset of them.

Table 7: sequence of Boolean filters used to define the dataset and number of documents retrieved

Filter	*# of documents*
PATSTAT EP Register	4,562,810
EP ID > 0	3,114,357
Filing date ≤ 2001-12-31	1,339,018
Total in dataset	*1,339,018*

To focus the analysis on *success rate*, statuses have been grouped in classes according to proceeding result they assess. Descriptions in Table 6 give information about the successful or failure result of proceedings:

- **Success**: it collects the statuses related to applications that reach the grant and grant opposed which maintain at least part of the IP right;
- **Failure**: it groups the statuses related to applications what not reach the grant or granted patent revoked for any reason;
- **Temporary**: it gathers the statuses that cannot give definitive information about the IP right.

Table 8 summarizes the content of dataset. It groups the statuses according to main classes and shows the share of single group and the one of the classes. As a first result, the reader can see that EP application filed on or before 2001 were successful in 61% of the cases.

Figure 14 shows the share of each status in dataset. As evident, the main status is the *'no opposition filed within time limit'*, which means the application reached the grant and this is not opposed by third parties, thus the exit from the EP proceedings was successful.

Figure 15 depicts the distribution of documents among classes. It highlights that the successful proceedings take the majority.

Table 8: content of the dataset

Status	*Class*	*Description*	*# of docs*	*Share*
2	Success	Patent limited	140	0.0%
3	Success	Patent maintained in amended form	17,091	1.3%
5	Success	Opposition rejected	12,555	0.9%
7	Success	No opposition filed within timelimit	787,373	58.8%
		Success subtotal	*817,159*	*61.0%*
1	Failure	Patent revoked by proprietor	23	0.0%
4	Failure	Patent revoked	19,130	1.4%
9	Failure	Application withdrawn	90,987	6.8%
10	Failure	Application deemed to be withdrawn	342,703	25.6%
11	Failure	Application refused	48,567	3.6%
13	Failure	Proceedings closed following consol-idation with another application	211	0.0%
		Failure subtotal	*501,621*	*37.4%*
6	Temporary	Opposition procedure closed	4,108	0.3%
8	Temporary	Patent granted	7,928	0.6%
12	Temporary	Grant of patent intended	785	0.1%
14	Temporary	Examination in progress	4,972	0.4%
15	Temporary	Examination requested	2,008	0.1%
16	Temporary	Application published	437	0.0%
17	Temporary	International application published	0	0.0%
18	Temporary	Unknown	0	0.0%
		Temporary subtotal	*20,238*	*1.5%*
		Total	***1,339,018***	***100%***

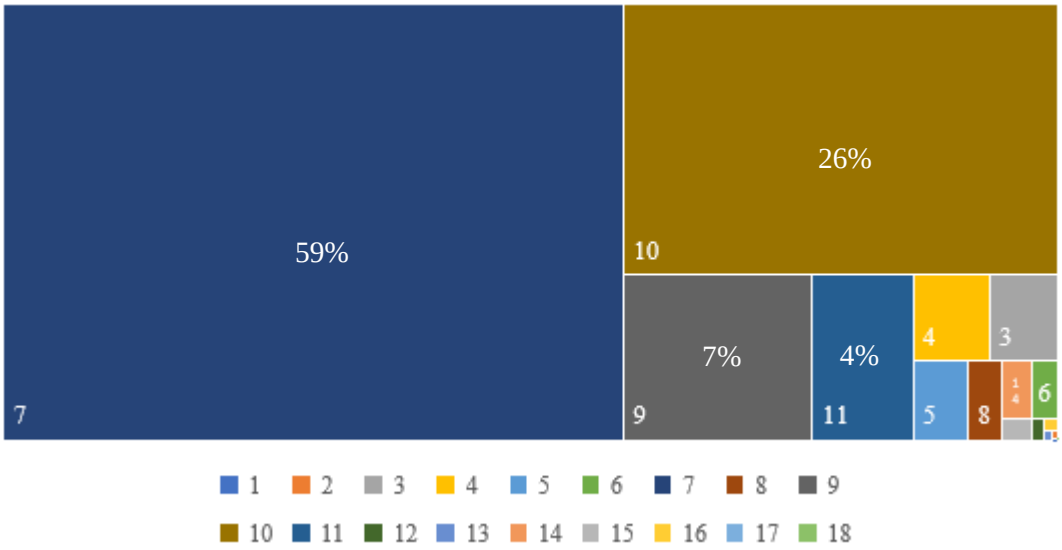


Figure 14: shares of the statuses in applications filed with the EPO on or before 2001

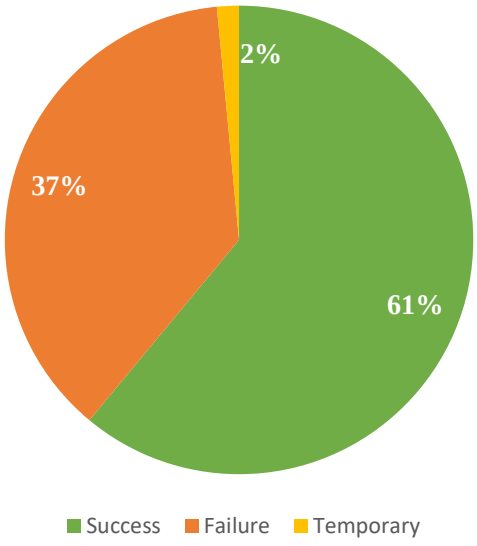


Figure 15: pie chart of shares of success-related classes for applications in dataset

The statuses of the applications can be used as reference for the final result of the proceedings, but they cannot give information about the history of the procedure. They are like a photography that shows the state of the applications in the time of the database publication. In order to gather information about the history of the applications, deeper analyses have to be performed on those steps that are typical of the patent lifecycle.

3.4 Filing Route

According to what found by (Guellec and van Pottelsberghe de la Potterie, 2000; Frietsch, Neuhäusler and Rothengatter, 2013), in EP proceedings, the filing route has been considered as a parameter influencing the *success rate*.

The filing route is the way an application enters in the European patent proceeding. It can belong to one of three types:

1. National filing with EP extension under Paris Convention;
2. Filing directly with the EPO;
3. Entry in European regional phase under PCT.

As mentioned in section 3.2, PATSTAT Register distinguishes only between EP direct filings, which aggregates both first and second of cases listed above, and EURO-PCT applications, the third case. To discriminate between them is used the *PCT ID* that is 0 for EP direct cases and positive for EURO-PCT ones.

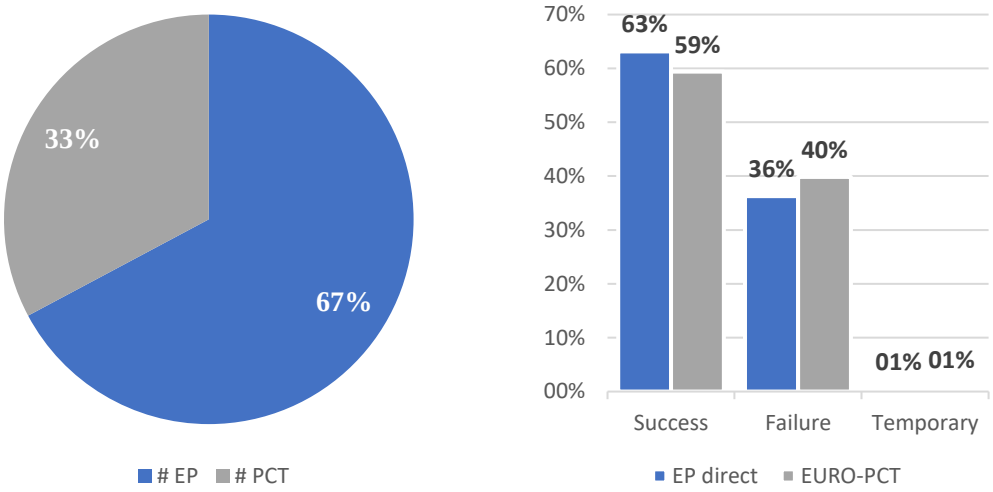
Table 9 summarize the distribution of the applications according to the mentioned classes and the route of filing (R) with the EPO.

Table 9: comparison of the success-related classes between direct-EP route of filing and EURO-PCT one in applications filed with the EPO on or before 2001

Class	*EP direct*		*EURO-PCT*	
Success	568,463	63.2%	260,732	59.4%
Failure	326,624	36.3%	174,997	39.9%
Temporary	4,840	0.5%	3,362	0.8%
Total	*899,927*	*100%*	*439,091*	*100%*

Figure 16 A shows the share of routes of filings in the dataset. The first evidence coming from the analysis is that the applications filed EPO directly are twice than EURO-PCT ones. Figure 16 B compares the success rate of the routes. It shows a difference

around the 4% between them. According to this, in average a direct filing with the EPO has more chances to be successful than a EURO-PCT one.



A B

Figure 16: A) share of filing route at the EPO in dataset; B) comparison of success-related classes for filing routes

3.5 Search report

The Search Report (SR) is the first formal response given by the Search Authority (SA) to the applicant. It can give crucial information about the content of an application generally concerning its patentability. This led to consider information in SR as interesting parameters influencing the *success rate*.

According to the filing route described in section 3.4, the first SR is published by different SAs. In case of an EP direct filing, it is the EPO. Otherwise, in case of EURO-PCT application it is the International Search Authority (ISA) (WIPO, 2015), which might be the EPO or other Patent Offices like, for example, JPO or USPTO[31].

There are several kinds of SRs, each of which collected in PATSTAT. For example, PCT proceeding could have a SR from Chapter II option (WIPO, 2015), an 'advanced stage' SR. They are issued in distinct steps of the proceeding, by different SAs. Here, it is interesting the first answer about patentability, i.e. the first SR got, because it

[31] http://www.wipo.int/pct/en/access/isa_ipea_agreements.html

contains the information on which the successive decisions about the continuation of proceeding rely. Thus, the two kinds of SR interesting for this work are:

- the International Search Report (ISR) for EURO-PCT applications;
- the European Search Report (SEA in PATSTAT) for EP direct filings.

The data collected in SR are references to patent or scientific literature building the state of the art of the claimed invention[32]. They could be specified by the applicant in order to underline an inventive step up from the previous knowledge, or marked by the SA to warn the applicant about previous documentation that might disclose and anticipate, at least a part of, the invention. In the latter case, they are called backward citations.

Backward citations indicated by SA might represent an obstacle in the granting proceeding, because they contest novelty and/or originality satisfaction of the application. As function of the contested requirement, they are coded with letters X or Y in the SR, regardless it is ISR or SEA:

- X: it indicates a single document that discloses the content of marked claims. It underlines a lack in novelty of the application for patent;
- Y: it indicates at least two documents which anticipate the claiming of the marked claims if considered together. It means the patent application lacks in originality.

They have been studied by many authors (among others (Pitkethly, 1997; van Zeebroeck, 2011; Grimpe and Hussinger, 2014; Liu *et al.*, 2017) and US20110161089) to assess their correlation with the value of patents. Generally, they found a weak or no correlation with the patent value. Nevertheless, this work focuses on the likelihood to obtain a grant, then it is not an assessment about the value of the patent, but on the actual chances to access to that value. Furthermore, despite such a little or null contribute to the value, it must keep in consideration that the backward citations are the first step in the grant proceeding and, later, there will be opportunities to modify the applications according to what contested by the SA. Such opportunities are described in detail in the next section.

Although Y citation category needs at least a couples of documents, PATSTAT has some records with single Y citation (e.g. WO9834831 A3). A manual check has been done on these applications, verifying that the problem is about the way with which the SR is written, indeed, these documents have misalignment between the type label (Y)

[32] https://www.wipo.int/edocs/mdocs/aspac/en/wipo_ip_kul_11/wipo_ip_kul_11_ref_t20.pdf

and the document reference or a double label for the same document (see Figure 40 in Appendix A.1). Those ways to write SR generate errors in its automatic reading. For simplicity Applications having a single Y citation have been grouped with the 2Y ones.

Table 10 and Figure 17 collect and summarise the distribution of documents receiving backward citations. They consider X and Y kinds separately for both EP-direct filings and EURO-PCT applications.

Table 10: distributions of X and Y citations for EP-direct and EURO-PCT

EP-direct

# of X	*# of docs*	*share*
0	529409	59%
1	172255	19%
2	91833	10%
3	49882	6%
4	26490	3%
5	13559	2%
6	7006	1%
7	3792	0%
8	2101	0%
9	1168	0%
10	805	0%
>10 up to 94	*1627*	*0%*

A

EURO-PCT

# of X	*# of docs*	*share*
0	223745	51%
1	85861	20%
2	49728	11%
3	31056	7%
4	18671	4%
5	11027	3%
6	6614	2%
7	4094	1%
8	2488	1%
9	1701	0%
10	1097	0%
>10 up to 105	*3009*	*1%*

B

# of Y	*# of docs*	*share*
0	658007	73%
*1**	-	-
2	165710	18%
3	40329	4%
4	19967	2%
5	7866	1%
6	4021	0%
7	1733	0%
8	1020	0%
9	538	0%
10	296	0%
>10 up to 42	440	0%

C

# of Y	*# of docs*	*share*
0	294064	67%
*1**	-	-
2	76686	38%
3	27400	6%
4	17003	4%
5	9494	2%
6	5553	1%
7	3170	1%
8	2058	0%
9	1214	0%
10	781	0%
>10 up to 51	1668	0%

D

**) grouped in category 2*

Although the number of citations in SR can be very high, up to 105 X and 51 Y (separately) in EURO-PCT proceeding, the main part of SRs is in categories having no citations or few of them.

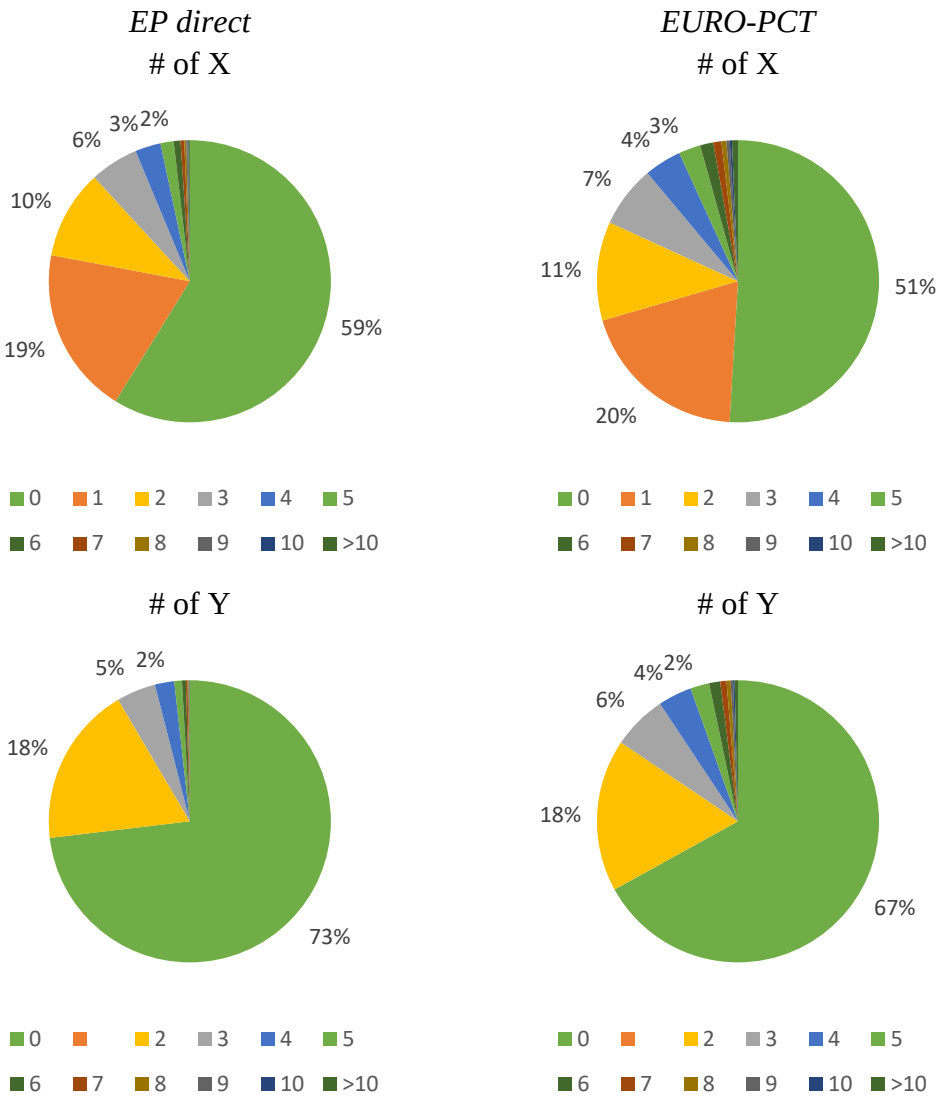


Figure 17: distributions of number of X and Y citations (separately) for EP-direct filings and EURO-PCT applications

Although information provided by Table 10 and Figure 17 gives an idea on the relevance of backward citation in patent literature, it is not exhaustive. Indeed, a SR can indicate both kind of citations for a single application. Thus, the analysis of the combinations of X and Y citations.

Table 11 and Figure 18 collects data about combination of X and Y citation in SRs. For the sake of simplicity, only first cases are detailed.

Table 11: patent applications amount and share for the first combinations of Search Report citations for both EP direct and EURO-PCT applications

EP				PCT			
X	*Y*	*# of docs*	*%*	*X*	*Y*	*# of docs*	*%*
0	0	406633	45%	0	0	157842	36%
0	2	87764	10%	0	2	32876	7%
0	3	19452	2%	0	3	12759	3%
1	0	111724	12%	1	0	50162	11%
1	2	41597	5%	1	2	20902	5%
1	3	10521	1%	1	3	6677	2%
2	0	63357	7%	2	0	31535	7%
2	2	18471	2%	2	2	10210	2%
2	3	5224	1%	2	3	3520	1%
3	0	35172	4%	3	0	20579	5%
3	2	9231	1%	3	2	5598	1%
3	3	2556	0%	3	3	1957	0%
other		88225	10%	other		84474	19%

The interest of this work focuses on the main paths followed by applications throughout grant and post-grant proceeding, which involve only a limited part of X and Y combinations. Then, only first combinations on citations have been studied, whereas other cases could be analysed in further works.

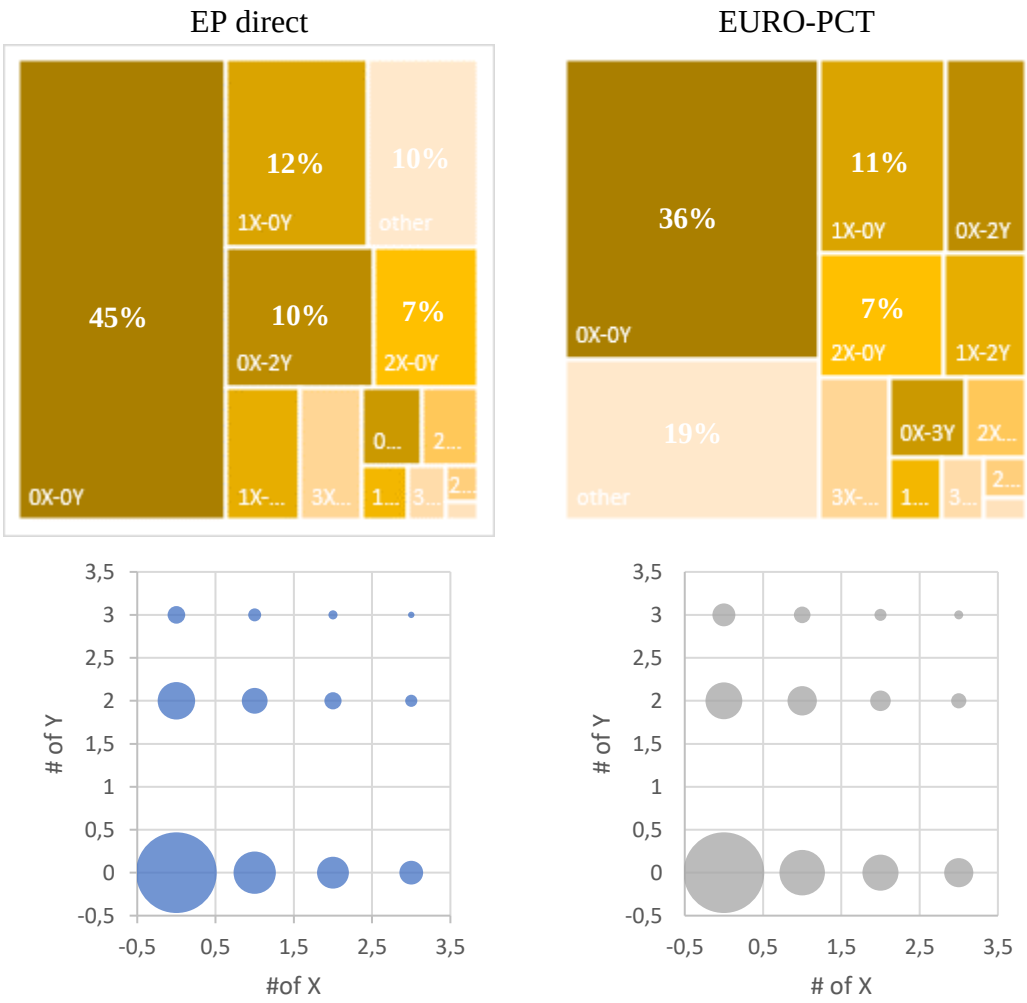


Figure 18: share of applications for combinations of X and Y citations and their distribution for both EP direct and EURO-PCT applications

Figure 19 shows the behaviour of the success rate as function of the citations highlighted in the SR. For EP direct filings, the effect on the success rate is higher for the X citation than Y citations. Otherwise, EURO-PCT applications, the difference between the effects of the two type of citations is not so clear.

Table 12: success and failure rates for citation combinations in SR for both EP direct and EURO-PCT applications

EP direct

X	*Y*	*Success*	*Failure*	*Temp.*	*%S*	*%F*	*%T*	*Total*
0	0	281,583	120,043	5,008	69%	30%	0%	406,634
0	2	55,196	31,454	1,114	63%	36%	0%	87,764
0	3	12,153	6,972	327	62%	36%	1%	19,452
1	0	66,130	43,816	1,778	59%	39%	1%	111,724
1	2	23,499	17,516	582	56%	42%	1%	41,597
1	3	5,801	4,548	172	55%	43%	1%	10,521
2	0	34,558	27,731	1,068	55%	44%	1%	63,357
2	2	9,852	8,341	278	53%	45%	1%	18,471
2	3	2,678	2,434	112	51%	47%	1%	5,224
3	0	18,106	16,451	640	51%	47%	1%	35,197
3	2	4,764	4,305	169	52%	47%	1%	9,238
3	3	1,238	1,271	47	48%	50%	1%	2,556
Other		45,530	44,652	41,742	51%	47%	2%	88,192
Total		*560,210*	*326,624*	*13,093*	*62%*	*36%*	*1%*	*899,927*

EURO-PCT

X	*Y*	*Success*	*Failure*	*Temp.*	*%S*	*%F*	*%T*	*Total*
0	0	101,099	54,173	2,570	64%	34%	2%	157,842
0	2	19,790	12,586	500	60%	38%	2%	32,876
0	3	6,894	5,650	215	54%	44%	2%	12,759
1	0	30,168	19,267	727	60%	38%	1%	50,162
1	2	11,684	8,892	326	56%	43%	2%	20,902
1	3	3,722	2,839	116	56%	43%	2%	6,677
2	0	18,081	12,968	486	57%	41%	2%	31,535
2	2	5,502	4,533	175	54%	44%	2%	10,210
2	3	1,883	1,586	51	53%	45%	1%	3,520
3	0	11,342	8,916	323	55%	43%	2%	20,581
3	2	2,908	2,575	115	52%	46%	2%	5,598
3	3	1,017	899	41	52%	46%	2%	1,957
Other		42,859	40,113	1,500	51%	47%	2%	84,472
Total		*256,949*	*174,997*	*7,145*	*59%*	*40%*	*2%*	*439,091*

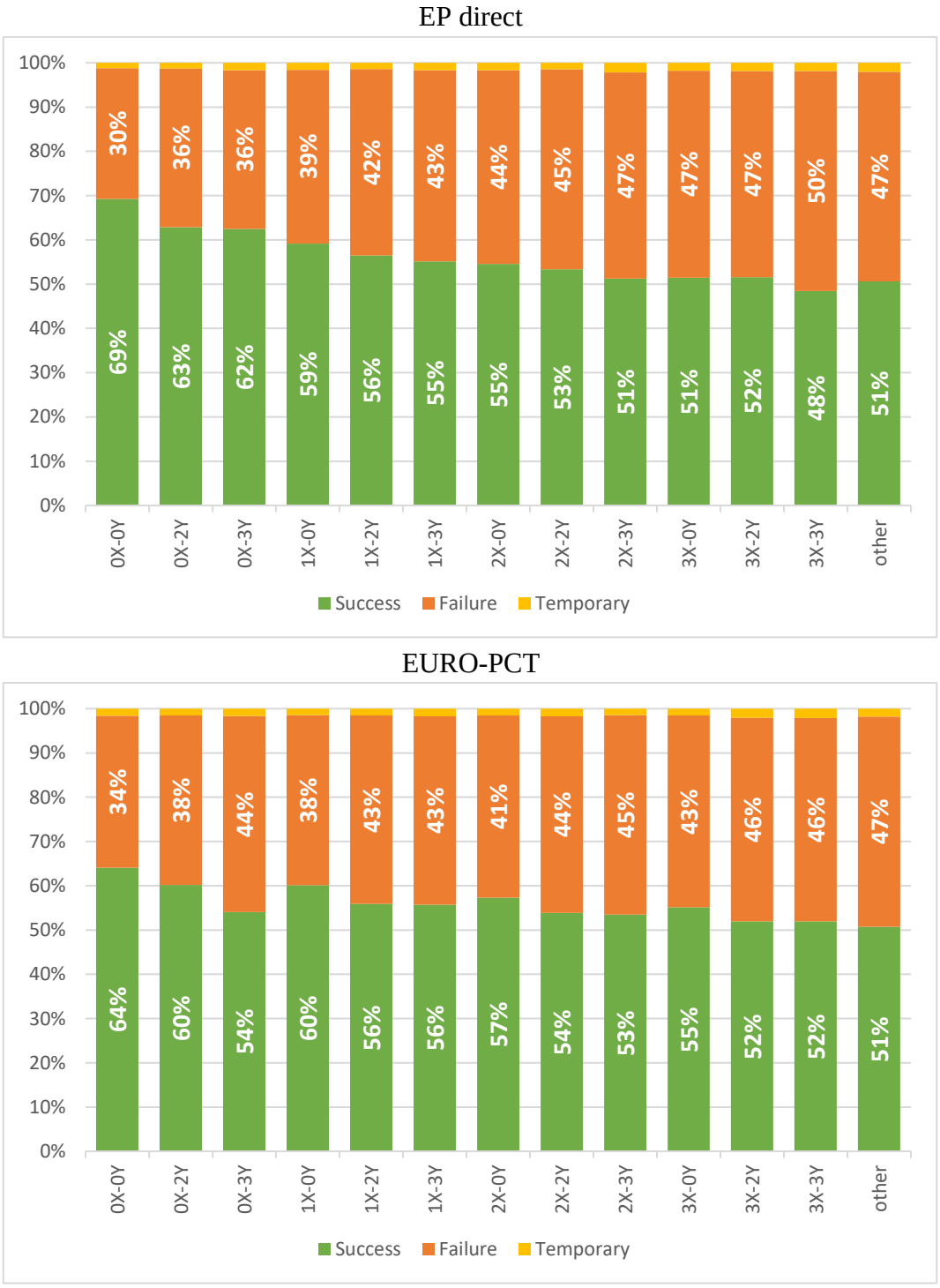


Figure 19: graph of success and failure rates for first SR citation combinations for both EP direct and EURO-PCT applications

3.6 Events

As mentioned in section 3.2, PATSTAT Register collects the data about the procedural events both for grant and post-grant European proceedings, up to the opposition period (European Patent Office, 2017b).

Referring to what shown in Figure 8 and Figure 9, there are some events in application timelines that can be addressed as pivotal for the success or failure on an application:

- Amendments filed by applicant in response to the SR;
- Application deemed to be withdrawn from the EPO;
- Opposition filed by third party.

The amendments (A) are the tool the applicant has whether it wants to modify the content of a filed application when the SA underlines content anteriority, lacks in description or other problems related to the content of the application. They may be filed many times during the proceeding and can modify different sections of the application: title, description, abstract, drawings and claims (refer to Chapter 2 for details).

When the applicant does not take expected actions within the time limit, usually, EPO send to the applicant a remainder. If after that the applicant does not matter about the application, the EPO publish the '*Application deemed to be withdrawn'* communication (D). Despite it is not the definitive withdrawal of the application, but a statement that the final withdrawal is imminent, in many cases it is the last official communication before the final withdrawal, as could be observed by the status 10 in Table 6 and Table 8.

If the application passes the examination and reaches the grant, a third party who considers the granting incorrect may file for an opposition with the EPO. In this case the EPO publishes the '*Opposition filed*' event (O). It is the first formal event that discloses the interest of a third party to the content of the patent. It represents the attempt to tear the IP right down, a clear signal of the importance the invention has in a true technological market. Economics literature largely treated the opposition filing. It has a great importance as an indicator of patent value (among others (Harhoff, Scherer and Vopel, 2003; Lemley and Shapiro, 2005; Chiesa *et al.*, 2008; Suzuki, 2011; van Zeebroeck, 2011; Liu *et al.*, 2017)).

Considering these pivotal events in proceedings, a subset of event data could be extracted from PATSTAT Register in order to assess the success rate of a European application.

3.6.1 Amendments

Since the IP right arises from patent claims, the most interesting amendments (A) are the ones acting on claims. PATSTAT Register indexes them with a specific event code, distinguishing the ones voluntary filed by applicant from the ones filed in reply to an action of the office. The latter are the ones considered for the analysis.

The code is the same for both EP direct filings and EURO-PCT applications, but it must be considered that the EURO-PCT route allows the user to modify the content of the application at the entry in regional phase without any declaration. There are no data about this kind of 'hidden' amendments. Thus, an error must be expected on the amendment impact on EURO-PCT route

Table 13: success and failure rates of patent applications as function of combinations of Route of filing and Amendment options

Route	*A*	*Success*	*Failure*	*Temporary*	*%S*	*%F*	*%T*	*Total*
EP	No	492,550	297,882	8,571	62%	37%	1%	799,003
EP	Yes	67,660	28,742	4,522	67%	28%	4%	100,924
PCT	No	187,672	140,176	3,848	57%	42%	1%	331,696
PCT	Yes	69,277	34,821	3,297	65%	32%	3%	107,395

Data show that, although the chance to change the application at the entry in Europe regional phase, the EURO-PCT route has more than twice amendments than EP direct filings (24% vs 11%). Furthermore, the effect of amendments in EURO-PCT route have about 2% better effect than the ones on EP direct filings, indeed, amendments in EURO-PCT increase the success rate of about 9% (66%-57%); otherwise, the EP direct amendments raise the success rate of about 7% (69%-62%).

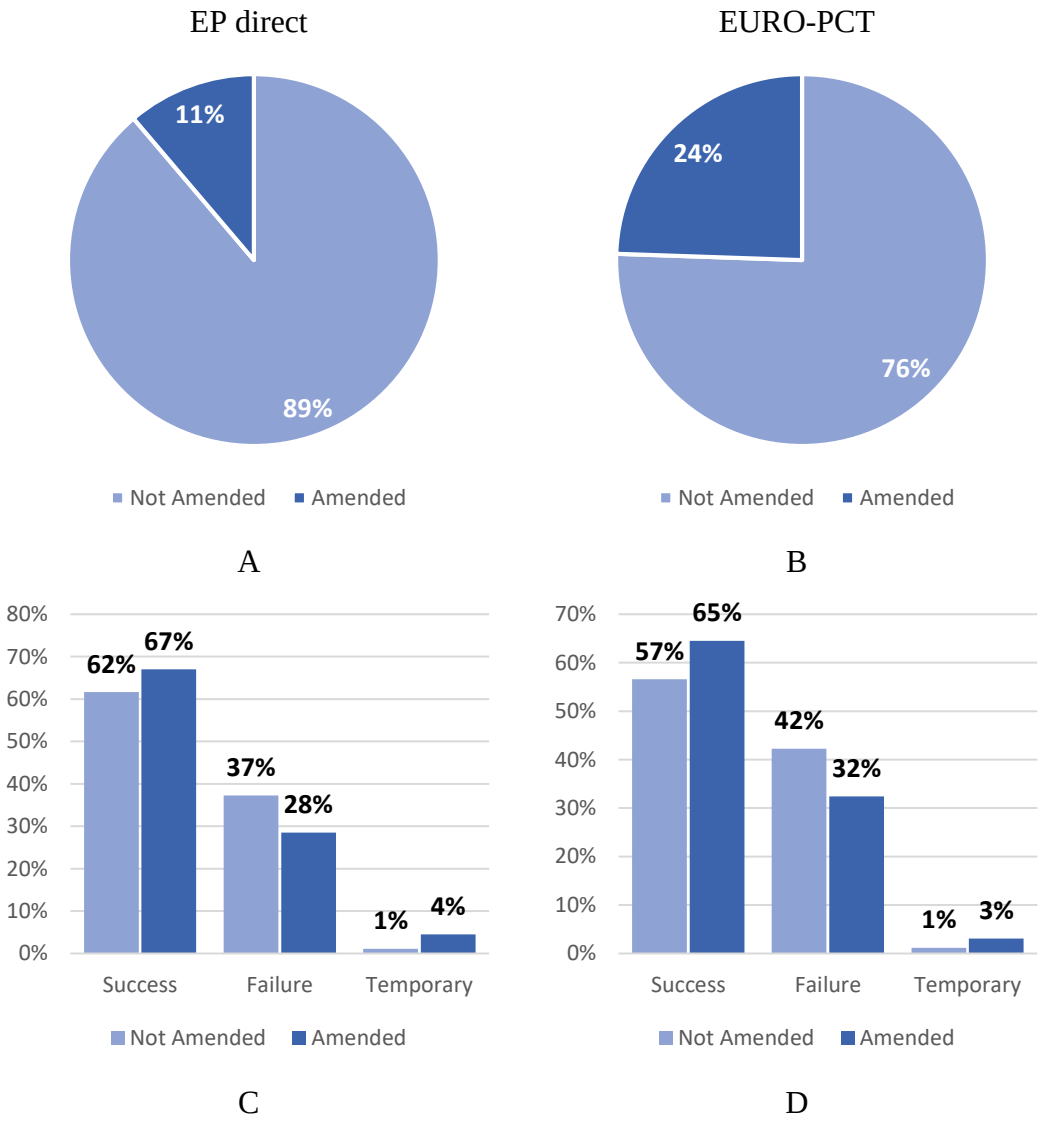


Figure 20: share of amended and not amended applications for A) EP direct and B) EURO-PCT routes and comparison of success and failure rates between amended and not amended application for C) EP direct and D) EURO-PCT routes

3.6.2 Application deemed to be withdrawn

Such an event is a public communication from the EPO and could be understood as an indicator about the intention to not proceed of the applicant. Although it collects some different causes that could lead towards the final withdrawn, for the aim of the analysis, the interest focuses on the procedural hindrances occurred in granting steps. They could be missed payments or missed documentation filings or other missed procedural fulfilments.

Table 14 summarises the effect of the procedural event '*Application deemed to be withdrawn'* (D) on the EP-direct filings and EURO-PCT applications.

Table 14: success and failure rates of patent applications as function of combinations of Route of filing and "Application deemed to be withdrawn" event occurrence

Route	*D*	*Success*	*Failure*	*Temporary*	*%S*	*%F*	*%T*	*Total*
EP	No	560,745	267,339	3,678	67%	32%	0%	831,762
EP	Yes	7,718	59,285	1,162	11%	87%	2%	68,165
PCT	No	249,532	107,264	2,541	69%	30%	1%	359,337
PCT	Yes	11,200	67,733	821	14%	85%	1%	79,754

Figure 21 C and D show clearly the effect the event '*Application Deemed to be Withdrawn'* has on the success rate of the application. Like above, the occurrence is higher (more than twice) in EURO-PCT than EP direct route. Differently from what observed in Amendments, the effect on the success rate of this event is the same for both the filing routes: an increment of about 55% in application failure (87%-32% for EP direct and 85%-30% for EURO-PCT).

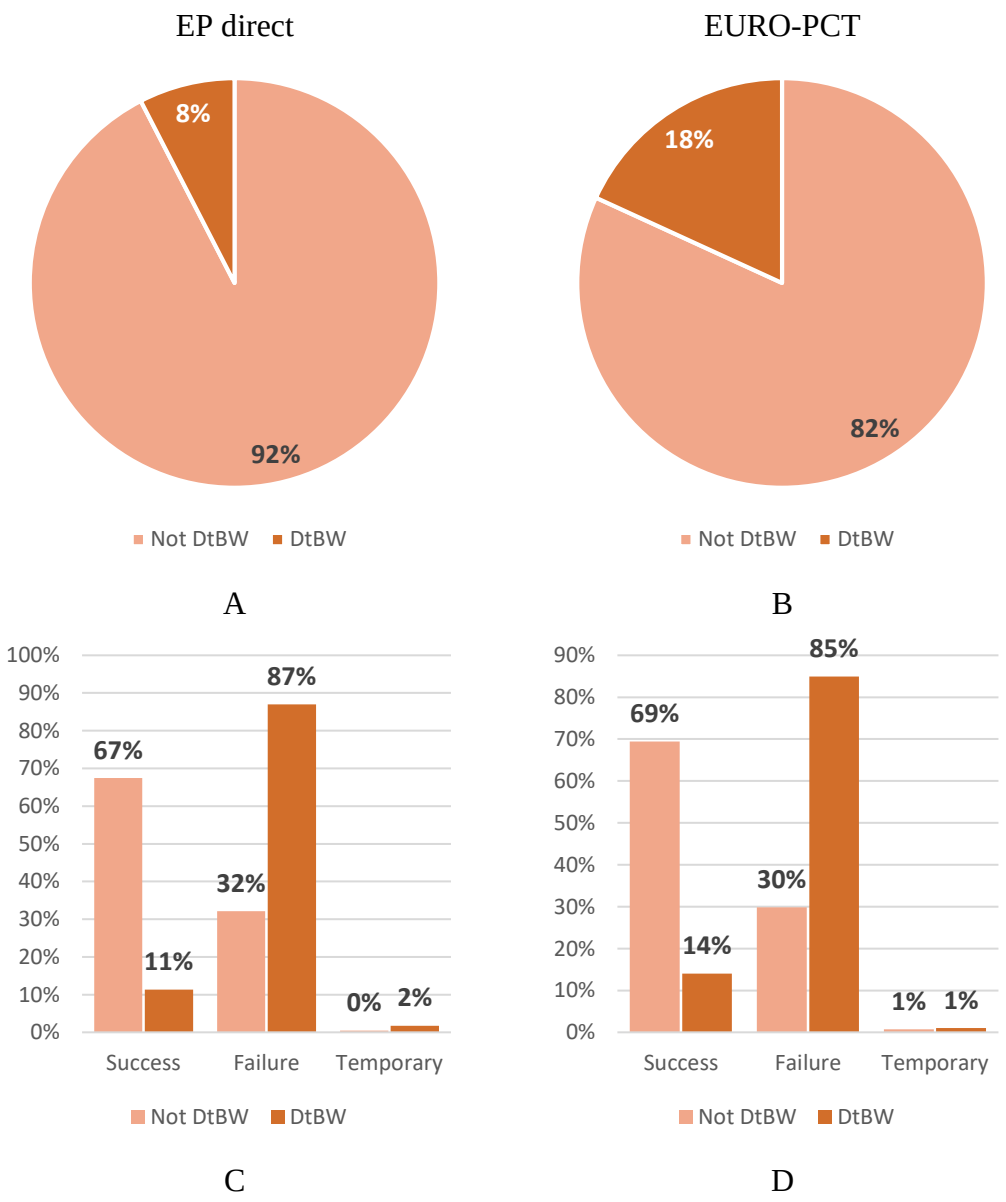


Figure 21: share applications with and without D occurrence for A) EP direct and B) EURO-PCT routes and comparison of success and failure rates between application with and without D occurrence for C) EP direct and D) EURO-PCT routes

3.6.3 Opposition

The opposition is here interesting due to its consequences on the IP rights. After an opposition filing (O), a specific proceeding starts, which can lead towards different results, going from the rejection of the opposition up to the revocation of the patent (see status 1-5 in Table 6).

The opposition is an objection to a patent granting, thus only when an application reaches the grant a third party can file for it. To understand which the effect of the opposition on the *success rate* is, the dataset has been modified, considering only the number of applications that reached the grant.

Table 15 shows shares referred to the new dataset. To define which is the likelihood for O to nullify the IP rights, the analysis considers the output of the whole proceeding successful ff the patent maintains at least a part of the rights. When patent grants is still in the opposition window or opposition phase has closed but the decision is not yet available, it is assigned to Temporary class. For the sake of completeness, the number of applications failed before the grant has been shown.

Table 15: success and failure rates of patent applications as function of combinations of Route of filing and opposition filing

Route	*O*	*Success*	*Failure*	*Temporary*	*%S*	*%F*	*%T*	*Total*	*Failed before grant*
EP	No	539007	8	4477	99%	0%	1%	543492	318358
EP	Yes	21203	13095	3776	56%	34%	10%	38074	3
PCT	No	248477	19	1870	99%	0%	1%	250366	172308
PCT	Yes	8472	6031	1913	52%	37%	12%	16416	1

Figure 22 A and B show that the share of opposition on EP grants is the same for both the filing routes. However, Figure 22 C and D underline a difference in the effect of the oppositions on success rate. Granted patents filed through the EURO-PCT route are more sensitive (37% vs 34%) than the ones filed directly with the EPO.

Figure 22: share applications with and without opposition for A) EP direct and B) EURO-PCT routes and comparison of success and failure rates between application with and without opposition for C) EP direct and D) EURO-PCT routes

3.7 Procedural map

We discussed about the effect of a single parameter on the success rate of a European application for patent, considering them independent each other. Actually, they are not unconnected. For example, as mentioned, the amendments are the answer to the SR form the applicant who wants to enhance the success rate of its application. So, the amendments are a sort of counterbalancing to a 'negative' SR. The main interest of the study is about the effect of combination of terms onto the success rate.

Collecting data about the procedural path of an application is useful to understand its position into the proceeding. The applicant can exploit this information in order to figure out which potential scenarios can be expected and what is the success rate for each of them. By this way, it could enhance the success rate of the application following the best available path or it might decide to withdraw the application due to the too high risks related to the continuation in proceeding.

A first procedural map can be built form the parameters considered so far. Using it the applicant can find the potential scenarios, according to the data considered.

Figure 23 shows the success and failure rates at different steps of the EP proceedings. The central ring diagram illustrates the rates for the whole applications in EP phase. Moving away from the centre, the map displays the partition by Route of filing and the sharing of the SR results respectively, each of which with the rates shown in related bar. The details about the events are available only for the first four responses of the SR, 0X-0Y, 0X-2Y, 1X-0Y and 1X-2Y. From the centre towards borders (right side for EP direct filings and left side for EURO-PCT applications) the map combines sequentially:

- Application amendments (light blue-blue column);
- 'Application deemed to be withdrawn' events (pink-red column);
- Opposition filings (grey-light yellow-yellow column)
- The bar charts aside each partitioning column illustrate the success-failure rates of each combination not worrying about the vertical scale. In 'Opposition filed' area, the grey parts indicate applications failed before reaching the grant, for which the opposition cannot be filed, indeed their success rates are 0%.

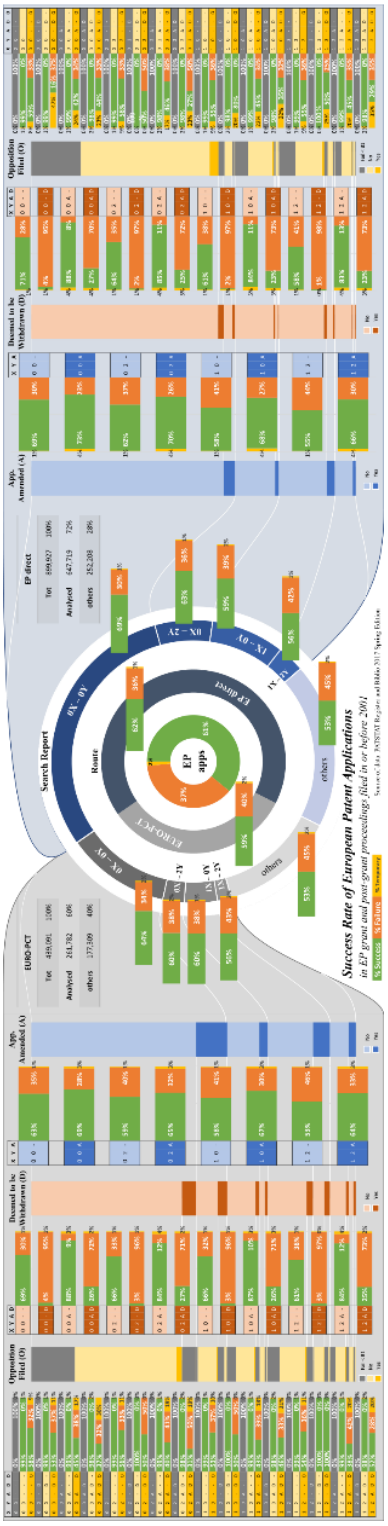


Figure 23: infographic about the Success Rate of European Patent Application in EP grant and post-grant proceedings filed in or before 2011 (for a full size view, see Figure 41, Figure 42 and Figure 43 in Appendix A.2)

The map points out an intriguing situation: all of the possible paths have been followed for both the filing routes, although in some cases, only few applications followed a specific path (see Table 16)

Table 16: procedural paths with less than 10 occurrences in EP proceedings

# of App.	*Route*	*X*	*Y*	*A*	*D*	*O*	*Success*	*Failure*	*Temporary*
1	EURO-PCT	1	2	No	Yes	Yes	1	0	0
2	EURO-PCT	0	2	No	Yes	Yes	1	1	0
2	EP direct	0	2	No	No	Yes	1	1	0
4	EP direct	1	0	No	Yes	Yes	2	1	1
5	EP direct	1	2	No	Yes	Yes	4	0	1
6	EURO-PCT	1	0	No	Yes	Yes	3	3	0

3.8 Effect of terms on success rate

The sequences of terms considered describe the procedural path followed by the application throughout the EP proceedings. The effect of the single term (Route, SR citations or event) is affected by the variance due to the other terms. Thus, it has been analysed the range of impact due to a single term, considering the whole path of the application.

The analysis has been split into two parts, before and after grant, because the opposition filing is available only after the grant. So, the database on which evaluate its effect cannot be the same of the other terms, which act on the application. For both parts, the success rate refers to the final result in EP proceedings, as described in Table 8, not to the grant.

For the first part, the terms considered are:

- **R**: the choice of EURO-PCT (1) route instead of EP direct filing option (0);
- **X**: 1 X citation occurrence in SR, considering only the cases of 0 and 1 X;
- **Y**: 2 Y citation occurrence in SR, considering only the cases of 0 and 2 Y;
- **A**: Amended claims filing, at least one;
- **D**: Application deemed to be withdrawn.

Table 17 summarise the dataset referred to the granting proceeding, useful to evaluate the impact of above terms.

Table 17: Success Rate reference dataset for granting proceeding

	R	*X*	*Y*	*A*	*D*	*Success*	*Failure*	*Total*	*SRate*
	0	0	0	0	0	268036	104269	372305	0.720
	0	0	0	0	1	438	10563	11001	0.040
	0	0	0	1	0	15327	1384	16711	0.917
	0	0	0	1	1	1505	3827	5332	0.282
	0	0	2	0	0	50503	26735	77238	0.654
	0	0	2	0	1	49	2751	2800	0.018
	0	0	2	1	0	4947	643	5590	0.885
	0	0	2	1	1	464	1325	1789	0.259
	0	1	0	0	0	55776	33817	89593	0.623
	0	1	0	0	1	98	5557	5655	0.017
	0	1	0	1	0	10230	1292	11522	0.888
	0	1	0	1	1	1037	3150	4187	0.248
	0	1	2	0	0	19850	13777	33627	0.590
	0	1	2	0	1	26	1992	2018	0.013
	0	1	2	1	0	3580	533	4113	0.870
	0	1	2	1	1	399	1214	1613	0.247
	1	0	0	0	0	84341	36280	120621	0.699
	1	0	0	0	1	449	10871	11320	0.040
	1	0	0	1	0	15533	1601	17134	0.907
	1	0	0	1	1	2021	5421	7442	0.272
	1	0	2	0	0	15454	7481	22935	0.674
	1	0	2	0	1	93	2892	2985	0.031
	1	0	2	1	0	3896	545	4441	0.877
	1	0	2	1	1	641	1668	2309	0.278
	1	1	0	0	0	21159	10201	31360	0.675
	1	1	0	0	1	150	4930	5080	0.030
	1	1	0	1	0	8129	936	9065	0.897
	1	1	0	1	1	1193	3200	4393	0.272
	1	1	2	0	0	7975	4808	12783	0.624
	1	1	2	0	1	58	2153	2211	0.026
	1	1	2	1	0	3323	481	3804	0.874
	1	1	2	1	1	516	1450	1966	0.262

	Success Rate stats for fixed values					***General Success Rate stats***	
Value	**0**	**0**	**0**	**0**	**0**	**Min**	0.013
Min	0.013	0.018	0.017	0.013	0.590	**Q1**	0.195
Median	0.436	0.468	0.452	0.315	0.795	**Median**	0.436
Max	0.917	0.917	0.917	0.720	0.917	**Q3**	0.758
						Max	0.917
Value	**1**	**1**	**2**	**1**	**1**		
Min	0.026	0.013	0.013	0.247	0.013		
Median	0.451	0.431	0.434	0.576	0.144		
Max	0.907	0.897	0.885	0.917	0.282		
Delta	**1-0**	**1-0**	**2-0**	**1-0**	**1-0**		
Min	-0.021	-0.097	-0.066	0.197	-0.680		
Median	0.013	-0.015	-0.023	0.235	-0.630		
Max	0.052	0.000	0.006	0.280	-0.577		

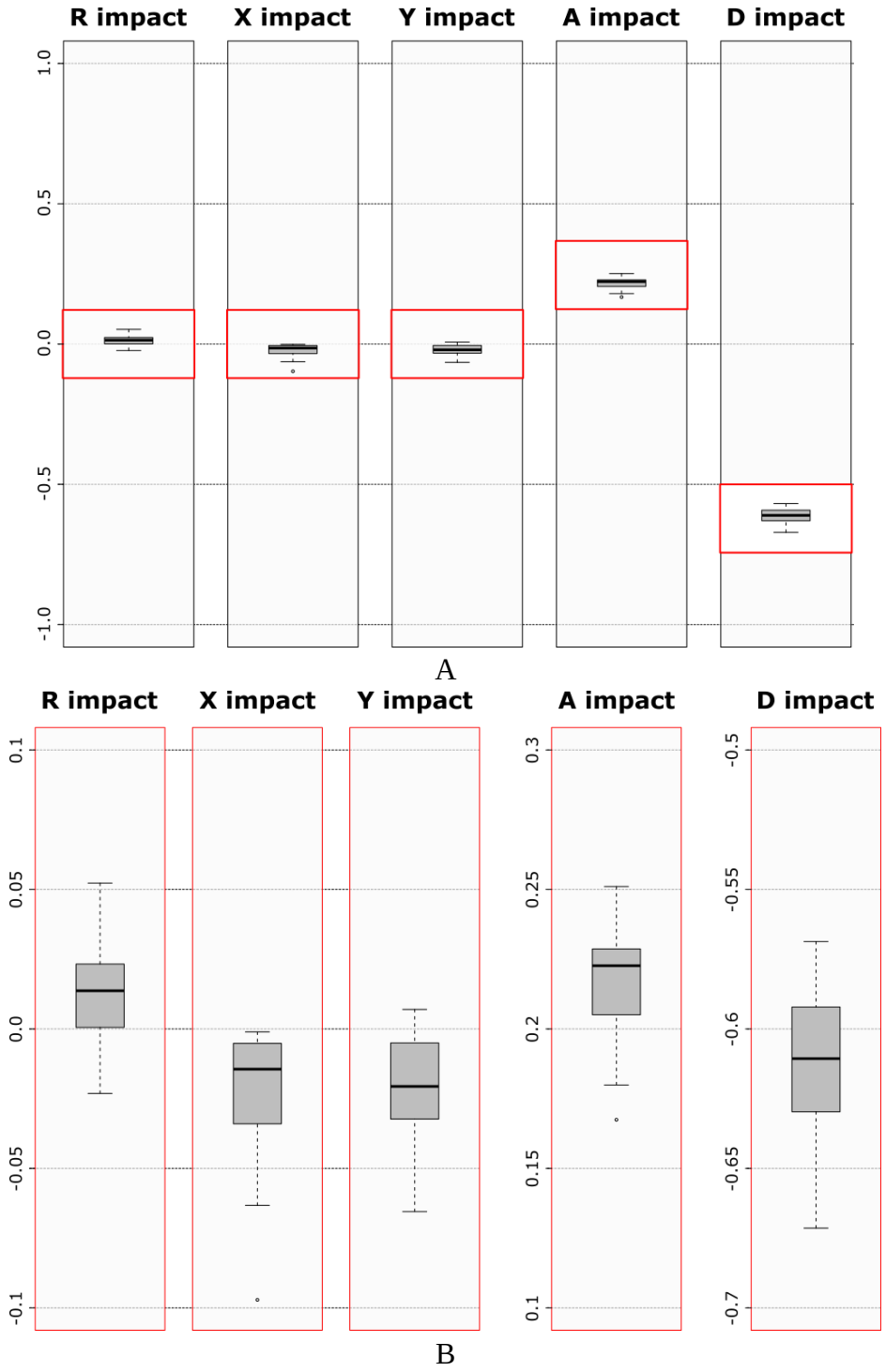


Figure 24: A) comparison of impacts due to considered terms on success rate for patent applications (grant proceeding); B) zoom of boxes highlighted in A

Table 17 and Figure 24 show that the filing Route has a weak effect on the success rate, indeed it increase of about 1.4% choosing the international route instead a direct filing with the EPO. However, the rage of deviation due to the other terms is such that the first quartile is negative.

As for the Route, also the effects of the Search Report citations, X and Y, are weak. The presence of an X or a couple of Y decreases the success rate of about 1.4% and 2.1% respectively. There is a difference between the two effects. While the effect of X is always negative (from -0.1% to -6.5%), the range of deviation of Y causes an excess in positive range (up to +0.7%).

Differently from the mentioned, the effect of the amendments on the success rate is relevant. It increases the chance to have success of about 22.2%, with a range from 16.7% up to 25.1%.

Also ‘deemed to be withdrawn’ event is relevant, but, contrary to the amendments, it causes a fall in success rate of about 61.1% with a range from -67.1% to -56.9%.

To enhance the success rate of an application for European patent, the applicant must see about the procedural fulfilments in order to avoid the application deem to be withdrawn. Even the amendments are crucial to increase the chances of success. Conversely, it is not important which filing route the applicant takes to file the application, while the search report citation (limited to 1 X and 2 Y) are small hindrances.

The second part of the analysis of the term effects considers also the opposition filing from a third party (**O**), at least one. As mentioned, the reference documents are granted patents, so all applications failed before the grant are not taken into account, then also the effects of the other terms can change.

Table 18: Success Rate reference dataset for post-grant proceeding

	R	X	Y	A	D	O	Success	Failure	Total	SRate
	0	0	0	0	0	0	256384	4	256388	1.000
	0	0	0	0	0	1	11652	5699	17351	0.672
	0	0	0	0	1	0	426	0	426	1.000
	0	0	0	0	1	1	12	7	19	0.632
	0	0	0	2	0	0	14899	1	14900	1.000
	0	0	0	2	0	1	428	220	648	0.660
	0	0	0	2	1	0	1458	0	1458	1.000
	0	0	0	2	1	1	47	23	70	0.671
	0	0	1	0	0	0	48273	0	48273	1.000
	0	0	1	0	0	1	2230	1118	3348	0.666
	0	0	1	0	1	0	48	0	48	1.000
	0	0	1	0	1	1	1	1	2	0.500
	0	0	1	2	0	0	4809	0	4809	1.000
	0	0	1	2	0	1	138	83	221	0.624
	0	0	1	2	1	0	453	0	453	1.000
	0	0	1	2	1	1	11	11	22	0.500
	0	1	0	0	0	0	53217	32375	85592	0.622
	0	1	0	0	0	1	2559	1442	4001	0.640
	0	1	0	0	1	0	93	5557	5650	0.016
	0	1	0	0	1	1	5	0	5	1.000
	0	1	0	2	0	0	9927	1138	11065	0.897
	0	1	0	2	0	1	303	154	457	0.663
	0	1	0	2	1	0	991	3128	4119	0.241
	0	1	0	2	1	1	46	22	68	0.676
	0	1	1	0	0	0	18922	13261	32183	0.588
	0	1	1	0	0	1	928	516	1444	0.643
	0	1	1	0	1	0	23	1991	2014	0.011
	0	1	1	0	1	1	3	1	4	0.750
	0	1	1	2	0	0	3483	480	3963	0.879
	0	1	1	2	0	1	97	53	150	0.647
	0	1	1	2	1	0	388	1208	1596	0.243
	0	1	1	2	1	1	11	6	17	0.647
	1	0	0	0	0	0	81298	34826	116124	0.700
	1	0	0	0	0	1	3043	1454	4497	0.677
	1	0	0	0	1	0	437	10864	11301	0.039
	1	0	0	0	1	1	12	7	19	0.632
	1	0	0	2	0	0	15117	1351	16468	0.918
	1	0	0	2	0	1	416	250	666	0.625
	1	0	0	2	1	0	1954	5389	7343	0.266
	1	0	0	2	1	1	67	32	99	0.677
	1	0	1	0	0	0	14809	7162	21971	0.674
	1	0	1	0	0	1	645	319	964	0.669
	1	0	1	0	1	0	92	2891	2983	0.031
	1	0	1	0	1	1	1	1	2	0.500
	1	0	1	2	0	0	3788	469	4257	0.890
	1	0	1	2	0	1	108	76	184	0.587

	1	0	1	2	1	0	628	1655	2283	0.275
	1	0	1	2	1	1	13	13	26	0.500
	1	1	0	0	0	0	20244	9661	29905	0.677
	1	1	0	0	0	1	915	540	1455	0.629
	1	1	0	0	1	0	147	4927	5074	0.029
	1	1	0	0	1	1	3	3	6	0.500
	1	1	0	2	0	0	7865	766	8631	0.911
	1	1	0	2	0	1	264	170	434	0.608
	1	1	0	2	1	0	1142	3175	4317	0.265
	1	1	0	2	1	1	51	25	76	0.671
	1	1	1	0	0	0	7620	4611	12231	0.623
	1	1	1	0	0	1	355	197	552	0.643
	1	1	1	0	1	0	57	2153	2210	0.026
	1	1	1	0	1	1	1	0	1	1.000
	1	1	1	2	0	0	3226	410	3636	0.887
	1	1	1	2	0	1	97	71	168	0.577
	1	1	1	2	1	0	498	1443	1941	0.257
	1	1	1	2	1	1	18	7	25	0.720

	Success Rate stats for fixed values							***General Success Rate stats***	
Value	**0**	**0**	**0**	**0**	**0**	**0**		**Min**	0.011
Min	0.011	0.031	0.016	0.011	0.577	0.011		**Q1**	0.500
Median	0.665	0.670	0.667	0.641	0.668	0.675		**Median**	0.436
Max	1.000	1.000	1.000	1.000	1.000	1.000		**Q3**	0.782
								Max	1.000
Value	**1**	**1**	**2**	**1**	**1**	**1**			
Min	0.026	0.011	0.011	0.241	0.011	0.500			
Median	0.627	0.641	0.643	0.662	0.500	0.645			
Max	1.000	1.000	1.000	1.000	1.000	1.000			
Delta	**1-0**	**1-0**	**2-0**	**1-0**	**1-0**	**1-0**			
Min	-0.969	-0.989	-0.250	-0.324	-0.661	-0.500			
Median	0.000	-0.017	-0.012	0.014	-0.106	-0.014			
Max	0.250	0.500	0.500	0.291	0.360	0.984			

In post-grant proceeding, all terms seem to have undefined effect on *success rate*. A more refined analysis can be done.

Considering the relative weight of each combination, Figure 25 shows that term O is the most important parameter to assess the *success rate* in post-grant proceedings, since it has a great negative effect. This is what was expected in this proceeding. O term has a remarkable negative effect, from –96.8% to -77.8% with a median of -91.9%.

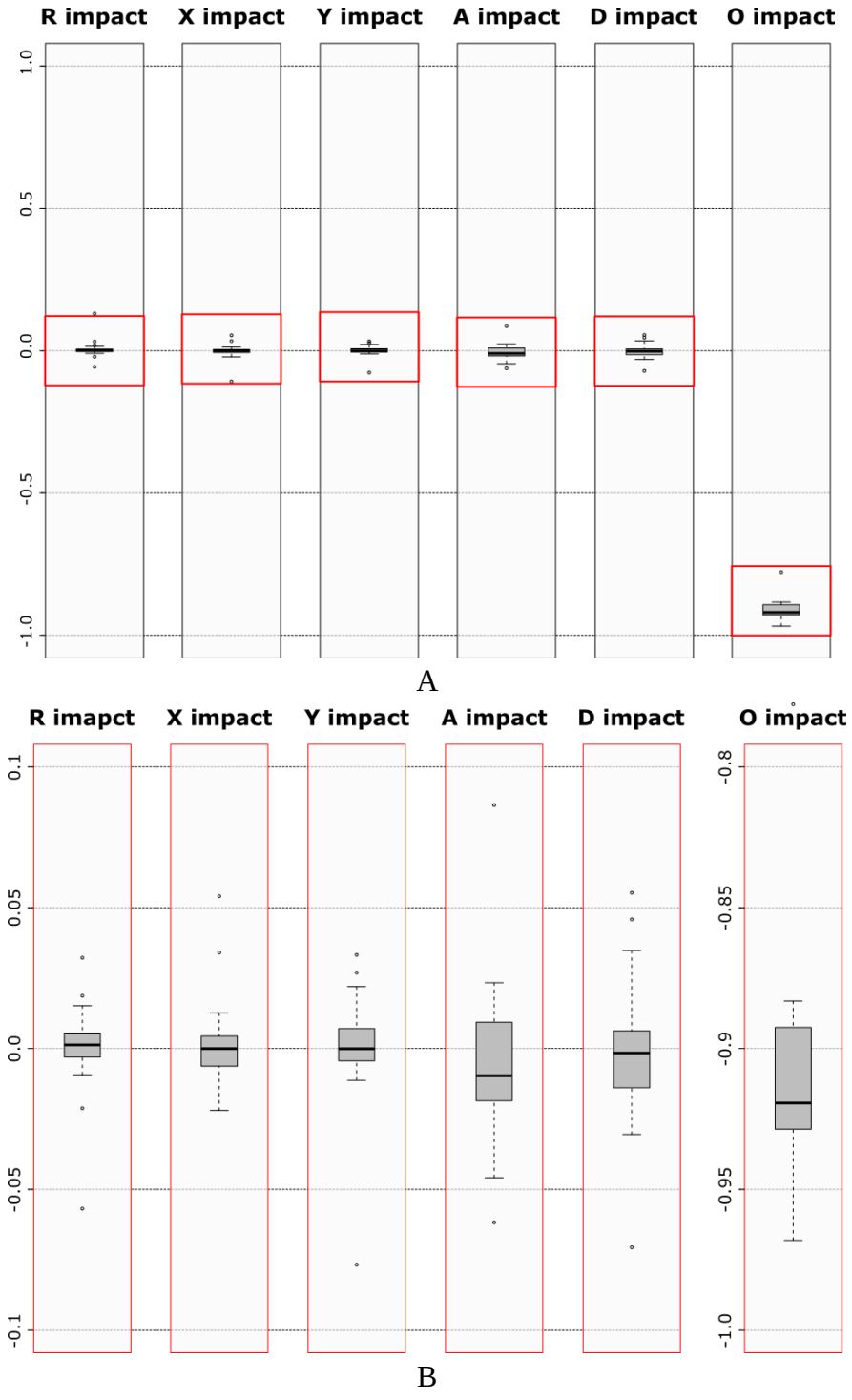


Figure 25 A) comparison of impacts due to considered terms on success rate for granted patent (post-grant proceeding); B) zoom of boxes highlighted in A

3.8.1 Binary regression

The data collected and analysed are suitable for the study of a binary multivariate regression to compute the success likelihood for a European patent application.

As for the term effects computation, also the regression models considered work on the two datasets listed in Table 17 and Table 18.

Due to the existence of procedural dependencies through the terms, the regression model considers the whole set of interaction terms. For granting proceeding the binary model proposed is:

$$y = \alpha + \sum_{i} \beta_i z_i + \sum_{i \neq j} \gamma_{ij} z_i z_j + \sum_{i \neq j \neq k} \delta_{ijk} z_i z_j z_k +$$

$$+ \sum_{i \neq j \neq k \neq l} \varepsilon_{ijkl} z_i z_j z_k z_l + \zeta z_1 z_2 z_3 z_4 z_5$$

The data have been analysed with *logit* and *probit* models.

Table 19: granting proceeding coefficient value and relevance for logit and probit regression models

	logit			*probit*		
(Intercept)	0.94415 ***	0.94407 ***	0.94402 ***	0.58265 ***	0.58291 ***	0.58327 ***
	(0.0036)	(0.0036)	(0.0036)	(0.0022)	(0.0022)	(0.0022)
R	-0.10055 ***	-0.09865 ***	-0.09876 ***	-0.06049 ***	-0.06147 ***	-0.06015 ***
	(0.0073)	(0.0070)	(0.0070)	(0.0044)	(0.0042)	(0.0042)
X	-0.44377 ***	-0.44340 ***	-0.44172 ***	-0.27047 ***	-0.27084 ***	-0.27162 ***
	(0.0078)	(0.0077)	(0.0077)	(0.0048)	(0.0047)	(0.0047)
Y	-0.15404 ***	-0.15450 ***	-0.15451 ***	-0.09344 ***	-0.09390 ***	-0.09372 ***
	(0.0042)	(0.0041)	(0.0041)	(0.0026)	(0.0025)	(0.0025)
A	1.46049 ***	1.42990 ***	1.43285 ***	0.80370 ***	0.80038 ***	0.78355 ***
	(0.0283)	(0.0172)	(0.0172)	(0.0141)	(0.0089)	(0.0078)
D	-4.12704 ***	-4.09237 ***	-4.09219 ***	-2.33550 ***	-2.33333 ***	-2.35586 ***
	(0.0489)	(0.0338)	(0.0338)	(0.0218)	(0.0170)	(0.0141)
RX	0.32975 ***	0.32254 ***	0.31559 ***	0.20127 ***	0.20288 ***	0.19965 ***
	(0.0157)	(0.0153)	(0.0149)	(0.0096)	(0.0093)	(0.0091)
RY	0.09499 ***	0.09229 ***	0.09331 ***	0.05760 ***	0.05934 ***	0.05255 ***
	(0.0088)	(0.0085)	(0.0085)	(0.0053)	(0.0053)	(0.0049)
RA	-0.03175			-0.00600		
	(0.0391)			(0.0198)		

RD	0.09661						0.05873	.	0.05493	***	0.04374	**
	(0.0689)						(0.0308)		(0.0145)		(0.0143)	
XY	0.08646	***	0.08640	***	0.08642	***	0.05151	***	0.05151	***	0.05124	***
	(0.0078)		(0.0075)		(0.0075)		(0.0048)		(0.0046)		(0.0046)	
XA	0.10826	**	0.13047	***	0.09655	***	0.09938	***	0.10691	***	0.13229	***
	(0.0415)		(0.0309)		(0.0256)		(0.0214)		(0.0154)		(0.0134)	
XD	-0.41119	***	-0.41833	***	-0.41767	***	-0.08900	.	-0.07520	*		
	(0.1132)		(0.0965)		(0.0965)		(0.0462)		(0.0311)			
YA	-0.02808						0.00038					
	(0.0256)						(0.0133)					
YD	-0.26846	***	-0.15208	***	-0.15869	***	-0.08432	**	-0.04582	*		
	(0.0762)		(0.0417)		(0.0417)		(0.0308)		(0.0195)			
AD	0.78911	***	0.81331	***	0.81030	***	0.37300	***	0.35917	***	0.41698	***
	(0.0641)		(0.0422)		(0.0422)		(0.0317)		(0.0213)		(0.0165)	
RXY	-0.13918	***	-0.12766	***	-0.12772	***	-0.08431	***	-0.08479	***	-0.07602	***
	(0.0155)		(0.0141)		(0.0141)		(0.0095)		(0.0093)		(0.0085)	
RXA	-0.10500	.	-0.08319	*			-0.08682	**	-0.10431	***	-0.09128	***
	(0.0615)		(0.0416)				(0.0319)		(0.0200)		(0.0177)	
RXD	-0.06558	.	0.34550	**	0.34801	**	0.02501					
	(0.0375)		(0.1161)		(0.1161)		(0.0626)					
RYA	0.20237	*	-0.06937	***	-0.08187	***	-0.04372	*	-0.04136	***		
	(0.0960)		(0.0193)		(0.0183)		(0.0197)		(0.0120)			
RYD	-0.01771		0.09552	**	0.10538	***	0.06542					
	(0.0885)		(0.0309)		(0.0305)		(0.0400)					
RAD	0.01340						-0.02418					
	(0.0381)						(0.0436)					
XYA	0.18557						-0.00191					
	(0.1348)						(0.0201)					
XYD	0.56891	***					0.06756		0.06665	*		
	(0.1292)						(0.0529)		(0.0327)			
XAD	0.39258	***	0.52608	***	0.55706	***	0.15440	**	0.13863	***		
	(0.0860)		(0.1067)		(0.1055)		(0.0578)		(0.0389)			
YAD	0.21958		0.21793	***	0.22547	***	0.14280	***	0.11988	***	0.05207	***
	(0.1490)		(0.0417)		(0.0415)		(0.0381)		(0.0238)		(0.0093)	
RXYA	0.07762						0.05396	.	0.05060	**		
	(0.0567)						(0.0302)		(0.0194)			
RXYD	-0.06857						-0.00578					
	(0.1669)						(0.0677)					
RXAD	-0.26653		-0.36980	**	-0.44748	***	-0.03377					
	(0.1725)		(0.1284)		(0.1224)		(0.0792)					
RYAD	-0.15862						-0.03566					

	(0.1105)			(0.0505)		
XYAD	-0.22825			-0.08307	-0.09753 *	
	(0.1472)			(0.0626)	(0.0386)	
RXYAD	0.03452			-0.02087		
	(0.1860)			(0.0821)		
n	904943	904943	904943	904943	904943	904943
AIC	1031835	1031830	1031832	1031835	1031823	1031836
AIC/n	1.14022	1.4022	1.14022	1.14022	1.14021	1.14022
R^2_{McF}	0.11076	0.11075	0.11075	0.11076	0.11076	0.11073

*coefficient value, (standard error), relevance * p<1%, ** p<0.1%, *** p<0.01%*

As shown in Table 19, the regression considers all of the combinations possible. By this way, the potential influence of earlier terms on the decision about one of the following can be catch.

At the limit of 1% of asymptotic relevance, the logit model takes into account the fourth level of interaction by the RXAD coefficient, while probit model stops at the third level with RXY, RXA and YAD coefficients. Anyway, the performance of the regression is limited, indeed the *Pseudo-R^2*, computed with *McFadden* approach, is small. Nonetheless, the regression computed with this set of regressing terms, explains about 11% of the variance.

The binary model proposed for the post-grant proceeding is:

$$y = \alpha + \sum_i \beta_i z_i + \sum_{i \neq j} \gamma_{ij} z_i z_j + \sum_{i \neq j \neq k} \delta_{ijk} z_i z_j z_k +$$

$$+ \sum_{i \neq j \neq k \neq l} \varepsilon_{ijkl} z_i z_j z_k z_l + \sum_{i \neq j \neq k \neq l \neq m} \zeta_{ijklm} z_i z_j z_k z_l z_m + \eta z_1 z_2 z_3 z_4 z_5 z_6$$

The data have been analysed with *logit* and *probit* models.

Table 20: post grant proceeding coefficient value and relevance for logit and probit regression models

	logit						*probit*					
(Intercept)	11.0681	***	11.0832	***	11.0831	***	4.1645	***	4.1677	***	4.1678	***
	(0.5000)		(0.4471)		(0.4471)		(0.1141)		(0.1017)		(0.1017)	
R	-10.2204	***	-10.2367	***	-10.2362	***	-3.6398	***	-3.6431	***	-3.6440	***
	(0.5000)		(0.4471)		(0.4471)		(0.1142)		(0.1018)		(0.1018)	
X	-10.5711	***	-10.5876	***	-10.5891	***	-3.8544	***	-3.8588	***	-3.8586	***
	(0.5001)		(0.4471)		(0.4471)		(0.1142)		(0.1018)		(0.1018)	
Y	3.2490						0.7057					
	(9.0066)						(2.2161)					
A	-1.4591						-0.3459					
	(1.1181)						(0.2718)					
D	6.4974						1.4113					
	(191.6962)						(47.1643)					
O	-10.3530	***	-10.3754	***	-10.3754	***	-3.7203	***	-3.7275	***	-3.7286	***
	(0.5003)		(0.4473)		(0.4473)		(0.1145)		(0.1021)		(0.1021)	
RX	10.4632	***	10.4824	***	10.4820	***	3.7889	***	3.7959	***	3.7954	***
	(0.5002)		(0.4473)		(0.4473)		(0.1145)		(0.1020)		(0.1020)	
RY	-3.3096		-0.0555	***	-0.0564	***	-0.7425		-0.0361	***	-0.0335	***
	(9.0066)		(0.0065)		(0.0065)		(2.2161)		(0.0038)		(0.0036)	
RA	3.0263	**	1.5378	***	1.5360	***	1.2128	***	0.8424	***	0.8427	***
	(1.1184)		(0.0264)		(0.0264)		(0.2722)		(0.0123)		(0.0123)	
RD	-10.5584		-4.1100	***	-4.1170	***	-3.7024		-2.3158	***	-2.3027	***
	(191.6962)		(0.0404)		(0.0403)		(47.1643)		(0.0189)		(0.0181)	
RO	10.2437	***	10.2814	***	10.2813	***	3.6541	***	3.6691	***	3.6701	***
	(0.5013)		(0.4483)		(0.4483)		(0.1162)		(0.1037)		(0.1037)	
XY	-3.3197		-0.0682	***	-0.0655	***	-0.7496		-0.0420	***	-0.0424	***
	(9.0066)		(0.0058)		(0.0057)		(2.2161)		(0.0035)		(0.0035)	
XA	3.1281	**	1.6566	***	1.6566	***	1.3013	***	0.9543	***	0.9543	***
	(1.1185)		(0.0270)		(0.0270)		(0.2723)		(0.0141)		(0.0141)	
XD	-11.0846		-4.6370	***	-4.6367	***	-3.8545		-2.4470	***	-2.4467	***
	(191.6962)		(0.0938)		(0.0938)		(47.1644)		(0.0362)		(0.0362)	
XO	10.4295	***	10.4446	***	10.4455	***	3.7676	***	3.7677	***	3.7730	***
	(0.5014)		(0.4485)		(0.4485)		(0.1164)		(0.1040)		(0.1040)	
YA	0.7295						0.1730					
	(29.9168)						(7.3617)					
YD	-3.2501						-0.7055					

	(300.9523)						(74.1264)					
YO	-3.2613						-0.7132					
	(9.0066)						(2.2162)					
AD	1.4598		1.3434	***	1.3762	***	0.3460					
	(217.9070)		(0.1032)		(0.1022)		(53.6150)					
AO	1.4094						0.3156		0.4791	***	0.4708	***
	(1.1213)						(0.2767)		(0.0433)		(0.0432)	
DO	-6.6736						-1.5195					
	(191.6968)						(47.1653)					
RXY	3.2616						0.7135					
	(9.0066)						(2.2162)					
RXA	-3.1061	**	-1.5566	***	-1.5509	***	-1.2788	***	-0.9003	***	-0.9019	***
	(1.1196)		(0.0464		(0.0463)		(0.2735)		(0.0237)		(0.0237)	
RXD	10.8938		4.5956	***	4.6191	***	3.7902		2.4299	***	2.4288	***
	(191.6963)		(0.1047		(0.1041)		(47.1644)		(0.0432)		(0.0432)	
RXO	-10.5327	***	-10.5318	***	-10.5314	***	-3.8317	***	-3.8254	***	-3.8263	***
	(0.5055)		(0.4517)		(0.4517)		(0.1229)		(0.1092)		(0.1092)	
RYA	-0.8319		-0.0619	**	-0.0625	**	-0.2191					
	(29.9169)		(0.0212)		(0.0212)		(7.3617)					
RYD	3.1936						0.6912		0.0297	*		
	(300.9524)						(74.1264)		(0.0119)			
RYO	3.3047						0.7395					
	(9.0067)						(2.2163)					
RAD	-0.8282		-0.6165	***	-0.6510	***	-0.0711		-0.1420	**	-0.1340	**
	(217.9070)		(0.1143)		(0.1132)		(53.6151)		(0.0481)		(0.0480)	
RAO	-3.2059	**	-1.7652	***	-1.7632	***	-1.3231	***	-0.9669	***	-0.9891	***
	1.1249		0.0799		0.0798		0.2821		0.0496		0.0483	
RDO	10.5351		3.8017	***	3.8025	***	3.6881		2.4736	***	2.4816	***
	191.6974		0.3964		0.3964		47.1662		0.1544		0.1543	
XYA	-0.7509						-0.1771					
	29.9169						7.3617					
XYD	3.1355						0.6779					
	300.9524						74.1265					
XYO	3.3387		0.0939	***	0.0923	***	0.7612		0.0634	***	0.0552	***
	9.0067		0.0257		0.0256		2.2163		0.0163		0.0158	
XAD	-0.1880						0.1273					
	217.9070						53.6151					
XAO	-2.9752	**	-1.5863	***	-1.5859	***	-1.2076	***	-0.8878	***	-0.9129	***
	1.1266		0.0935		0.0935		0.2845		0.0580		0.0566	
XDO	28.2548		5.0034	***	5.0228	***	9.1811		2.5557	***	2.5710	***

	1779.5482		0.3153		0.3154		437.8582		0.1801		0.1800	
YAD	-0.7283						-0.1732					
	320.4805						78.9271					
YAO	-0.7957						-0.2137		-0.0501	*		
	29.9170						7.3618		0.0255			
YDO	2.9930						0.5450					
	300.9533						74.1279					
ADO	-1.2344		-1.5866	***	-1.6177	***	-0.2078		-0.5068	***	-0.5169	***
	217.9077		0.2313		0.2310		53.6161		0.1225		0.1223	
RXYA	0.8388		1.6112	***	1.6057	***	0.2279					
	29.9169		0.1575		0.1575		7.3617					
RXYD	-3.0201						-0.6159					
	300.9524						74.1265					
RXYO	-3.2327						-0.6957					
	9.0069						2.2166					
RXAD	0.4567						-0.0248					
	217.9071						53.6151					
RXAO	3.0953	**					1.2719	***	0.9072	***	0.9278	***
	1.1365						0.2987		0.0964		0.0959	
RXDO	-28.3918		-4.7645	***	-4.7832	***	-9.3233		-2.3832	***	-2.3986	***
	1779.5485		0.4293		0.4293		437.8586		0.2485		0.2485	
RYAD	0.9705		0.1227	***	0.1490	***	0.2840					
	320.4805		0.0309		0.0286		78.9271					
RYAO	0.8364						0.2214					
	29.9171						7.3621					
RYDO	-3.1887						-0.6882					
	300.9542						74.1294					
RADO	1.0321		1.2251	*	1.2543	**	0.1963					
	217.9084		0.4810		0.4809		53.6171					
XYAD	0.9418						0.2529		0.0429	**	0.0571	***
	320.4805						78.9271		0.0149		0.0138	
XYAO	0.7742						0.1916					
	29.9171						7.3621					
XYDO	-11.1195						-2.9722					
	934.3898						229.9323					
XADO	-16.9705						-5.4468					
	1782.5622						438.6000					
YADO	0.7067						0.1601					
	320.4815						78.9287					
RXYAD	-1.1306						-0.3553					
	320.4806						78.9272					

RXYAO	-0.8953			-0.2627		
	29.9175			7.3627		
RXYDO	20.0086			5.8367		
	2187.6705			538.3372		
RXADO	17.0727			5.5777		
	1782.5625			438.6005		
RYADO	-0.9872			-0.2937		
	320.4825			78.9302		
XYADO	7.2914			2.1970		
	940.8612			231.5247		
RXYADO	-15.6368			-4.7308		
	2190.4424			539.0193		
n	791511	791511	791511	791511	791511	791511
AIC	490658	490661	490672	490694	490661	490667
AIC/n	0.61995	0.61990	0.61990	0.61995	0.61990	0.61991
R^2_{McF}	0.57720	0.57718	0.57717	0.57720	0.57717	0.57716

*coefficient value, (standard error), relevance * p<1%, ** p<0.1%, *** p<0.01%*

Table 20 shows the logit and probit regression results for the post-grant proceeding modelling considering potential interaction between terms. Both the models take into account more than one fourth level coefficient, at an asymptotic relevance of 1%. As in the case of the granting proceeding, the difference in performance through the models is small, but, differently from above, the significance level is greater. The *Pseudo-R^2* of McFadden indicate a variance explanation of about 58%.

3.9 Conclusion

The applications for patent are taking increasingly an important role in many kinds of patent analysis. The crucial parameter about them is the likelihood they have in reaching a grant, which can drastically modify the impact of an application onto the result of the analysis. The attention to this topic is unexpectedly low. The works which concerned it dealt the success rate with the economical point of view, considering as terms of influence some economical parameters and filing route.

Currently, the most reliable judgement about the success rate of an application is the patent attorney's opinion, who is skilled in understanding the path followed by the application through the events of the granting process, but this is a time-consuming

task. The surge of applications for patent request to the attorneys to reduce the time needed for the patent analysis.

This work proposes a set of parameters, proper of the granting and post grant proceeding of an EP application, to support the attorneys in the assessment of the success rate for an EP application.

Data about the success rate of patent applications have been extracted from PATSTAT Register and cover a time span from 1978 up to 2001, as reference filing year. According to the status definitions of PATSTAT, as successful applications, it has been considered those applications which left the granting and post-grant EP proceedings with an active exclusive right, while as failed applications, the study considers those ones which did not reached the grant or whose granted patents have been revoked.

The statuses do not explain the history of the application throughout the EP proceedings. Some terms have been considered to collect information about the most important steps of the application lifecycle: the filing route chosen by the applicant, the search report citations coming from the EPO examiner, the amendments filed by applicant in response to the search report, the 'deemed to be withdrawn' event and the opposition filing event.

For each term, an analysis of the mean impact on the success rate has been done. Furthermore, it is proposed an infographic map that collects the mean impacts related to the path followed by applications throughout the granting and post-grant European proceedings.

To delimit the variance of contribution of each term, an analysis about the effect of each term has been carried out. It shows that the success rate does not matter about the filing route and the search report citations, limited to the 1 X and 2 Y, are small hindrances to the success of the application. Contrariwise, the effect of claims amendments, deemed to be withdrawn event and opposition filing have big relevance.

In the EP granting proceeding, the amended claim filing increases the success rate from 16.7% up to 25.1% while the 'deem to be withdrawn' event causes a fall from -67.1% to -56.9%.

In the EP post-grant proceeding, previous terms loss their impact. The lone term relevant is the opposition filing from third parties that causes a decreasing in success rate from -96.8% to -77.8%.

A regression of success rate on the analyzed terms has been computed with *logit* and *probit* models. The results in granting proceedings cover about 11% of the variance

considering linear terms and their interaction, while in post-grant proceeding, the coverage rises up to 57.7%.

Thus, the terms considered are not enough to explain the variance of the success rate and are not able to predict its value. Nonetheless, the information about their relative effect on the success rate are reliable. Further studies could be carried out in order to identify other terms of regression to enhance the performance of the predictive model.

Although a reliable predicting model is not available, the information about the effects of the terms considered can help patent attorneys in retrieve objective information in a systematic and fast way in order to get useful indicators that support them in assessing the impact of applications on the target of the patent analysis performed.

4 A method to set a Business Intelligence up assessing patent application background

The patent valuation activity, and other kinds of patent analysis (Alberts *et al.*, 2011), must consider the proposal of invention into the relevant technological background, which is the collection of patent documents dealing with related matter. In other words, it makes up the state of the art of the invention.

Currently, the patent analyst builds the state of the art using search software that are expressly developed for the search task in patent literature. Based on data retrieved from those tools, the analyst leads a Patent Intelligence to identify the most relevant information for the specific target of the analysis. It compares the parameters of the target patent document and the aggregated data of the state of the art to understand what the position of the target document is, and which the trends in the relevant technological domain are.

In the last 25 years, most part of retrieval techniques implemented by patent search software have been only restricted on the Boolean combination of keywords and classification codes (IPC or CPC) (Khode and Jambhorkar, 2017). The complexity of both used language and classification system and the presence of sparse class assignments to the patent make this approach too general, limiting the quality of the patent search and its results.

Furthermore, the search gives a list of documents, each of which is the representative of its family. A patent family is a collection of patent documents, applications and/or grants, which are linked by the priority application (Martinez, 2010; Martínez, Martinez and Martínez, 2011; Carrara and Russo, 2017; European Patent Office, 2017c). Usually the family members are patent documents filed in different countries claiming the same invention. Unfortunately, each country has its own patent law system, then each application may have different stories.

Aggregated data might mislead the analyst in retrieving the information about the actual situation of the state of the art, like, for example, patent content and legal status.

The analysis of patent families is useful in order to understand which patent are more representative than others for the aim of the BI. The family analysis is useful, for example, to learn latent characteristics of patents of a third party (Harhoff and Wagner, 2009), to estimate whether a granted patent will be opposed (Reitzig, 2004), to esteem a patent value (Harhoff, Scherer and Vopel, 2003; Reitzig, 2004; Harhoff and Wagner, 2009), also using particular methods (US2002178029, CA2504580). Figure 25 shows the behaviour of the number of members of a family along time.

Although the mean size of patent families[33] is decreasing over time since the first years of 2000s due to the growth of single patents, there is an increase of the number of families that has almost reached 300.000 per year. Especially there is a great increasing in families having between 2 and 6 members, while the larger ones are decreasing. The larger family set includes more than 100 members, 6 of which with more than 300 members, up to 472 (US7309763B2).

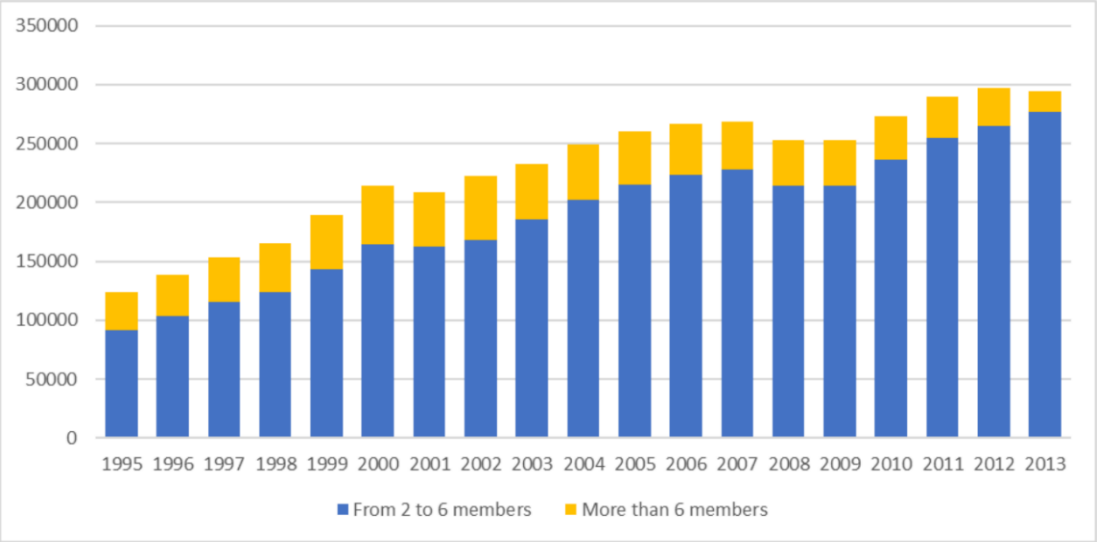


Figure 26: Count of families per earliest filing year grouped by family size.

In this section, families are defined according to the DOCDB definition, which define them as the set of patent document which has the same priority numbers.

To overcome the drawbacks of the analysis of patent families' list obtained from current keyword-based patent search tools, a three-stepped method is suggested. The first one corresponds with the current search but using a simpler keyword approach focused on recall increasing. The second step involves *syntactic parsing* and *patent ontology* to identify and select only those patent families related to the state of the art. It seeks for a set of syntactic dependency patterns specific for the description of the technical problem faced by the invention in order to increase the precision. The third step elaborates the data of selected families to disaggregates them and extract information about each family member.

[33] We considered the DOCDB patent family definition

In this chapter, we will overview the riskiest issues due to patent families and legal status that might mislead the analyst in making the final opinion. Then we introduce the parsing technique and present the suggested method for BI.

4.1 Family-related issues

The search for patents on the most diffused international repositories does not collect a list of single patents, but a pool of patent families. Usually, in a patent family there are applications and/or granted patents claiming rights for the same invention in different countries. In some families, however there might be applications and patents claiming rights for different inventions in the same country or, again claiming the same invention in the same country with different set of claims.

4.1.1 Choice of correct document(s)

The population of a family depends on the rules of the database (Simmons, 2009), so, depending on the database on which the patent specialist lead the search, he/she obtain different results. The rules about the building of family regard also the choice of representative document of it. It is essential for the patent specialist to understand which is/are the patent(s) to take into account inside a family, indeed there is high risk of mistakes if he/she rely on the reference document only.

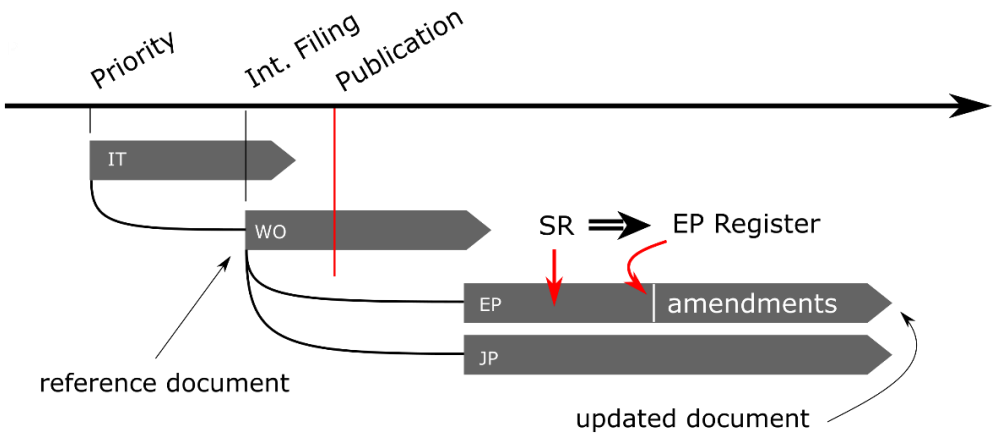


Figure 27: choice of the correct document in a patent family. Depiction of an example of a process of selection of the most interesting document into a patent family

4.1.2 Lack of unity of invention: Divisional applications

The applicant has to limit its application to only one invention. Unless the PCT procedure, which requires additional searches and fees whether further inventions have to be examined (*PCT Applicant's Guide 7.016*) (ISR and Written Opinion indicate the claims not examined (e.g. WO2015114601)), the examiner who finds a "*Lack of Unity of Invention*" requires the applicant to reduce its application to only one invention ("*Restriction Requirement*" in US (*37 CFR 1.142*) and "*Lack of Unity Objection*" at EPO (*Guidelines for Examination 8.1*)) and possibly allows the filing of applications claiming the other invention(s) which will have the same priority and filing dates. Anyway, no new matter can be claimed in reference to the parent disclosure. Each country comply with its own law and time limits about divisional applications (*US MPEP 201.06*)[34,35] (*Patent Information News*, 2010). E.g. EPO lets the applicant to file (so-called *mandatory* (*Patent Information News*, 2010)) divisional application up to 24 months after the *Lack of Unity Objection* was notified (*Patent Information News*, 2010), while US lets to file a divisional application before the parent will issued, rejected or abandoned. The patent analyst has to check whether in the patent family exist some divisional applications in order to choose the most relevant one. US applications must declare whether they are divisional on the first sheet (e.g. US7075825), while EP Register shows the information about EP parent/divisional applications (e.g. EP2892287).

4.1.3 "Continuation of" and Voluntary Divisional applications

If an application claims are too strict compared to the disclosure/description, or after a claim cancellation by the examiner, or for other reasons, the applicant can file a new application for the same invention claiming for same prior and filing dates. No new matter is allowed. US procedure calls it "Continuation of" application[36] , EP procedure refers to it as (*Voluntary*) "Divisional" (*Patent Information News*, 2010). Similar aspects can be found in Japanese divisional procedure[37]. Even in this case the patent specialist must check for divisional and continuations in order to consider the most interesting application. US *Continuation* shows its relationship with prior filed application in its first sheet (e.g. US20150195331), and is allowed up to the issuance, rejection or abandonment of the parent application (*US MPEP 201.06*). EP *voluntary*

[34] https://www.epo.org/searching-for-patents/helpful-resources/asian/china/faq.html
[35] https://www.jpaa.or.jp/old/english/whatsnew/pdf/green_technology_patent.pdf
[36] In the past the continuation practice caused great uncertainty on patent term (Simmons and Spahl, 2000; Simmons, 2009). Currently it is a source of interpretation risk only, indeed it allows the applicant to change the claim set (*US MPEP 201.07*)
[37] https://www.jpo.go.jp/tetuzuki_e/t_tokkyo_e/files_guidelines_e/06_0101_e.pdf

divisional can be filed up to 24 months after the *First communication examination* was notified (*Patent Information News*, 2010).

4.1.4 "Continuation-in-Part" applications

The USA let the applicant to file a "Continuation-in-part" (CIP) application, which repeats some substantial portions of or all earlier application and adds new matter not disclosed before. The filing is available till to the issuance, rejection or abandonment of the earlier application and the new one has to declare on first sheet its CIP relation (e.g. US20110249806). The expiration term of the patent based on the CIP application is 20 years from the filing date of earliest application for which the benefit is claimed (*US MPEP 2701*). Notice that foreign applications, amended at filing can be considered a sort of CIP (Simmons, 2009).

4.1.5 National route of the first national filing

Often, it occurs the first national filing proceeds on its own route independently from the PCT procedure (WO2014086620). The content of resulting applications may be different due to different proceedings, moreover the result of the procedures might be different.

4.1.6 Double patenting in the same country

Generally, the possibility to patent the same invention by the same applicant is prevent by the patent law of any country. WIPO, EPO, USPTO, JPO and SIPO rule it with different approaches and results[38] (WIPO, 2004; Tegernsee Experts Group, 2012) (*EPO Guidelines for Examination 5.4*, *US MPEP 804*). The general rule says an applicant having more than one co-pending applications about the same invention, claimed by the substantially same claim set, with the same effective filing date, cannot receive more than one patent about that invention. Notice that whether the descriptions/disclosures are the same, but the claims sets are different (like in parent-divisional cases) the co-pending applications are considered different. If the patent specialist finds a potential double patenting, he/she has to check in detail the correspondence of the claim sets.

4.1.7 Simultaneous deposit in different countries

Applicant may apply for patent in different country directly. Furthermore, he/she can file for the same invention, in the same day, in different countries without declare the

[38] http://english.sipo.gov.cn/news/iprspecial/919037.htm

reference between the applications. As a result, the applications will belong to different patent families. There are no tools or standard methods to retrieve the link between applications, so it is very difficult for the patent specialist to find it. The current opportunity to highlight it is through a foreign extension citing the applications (e.g. WO2013074032). In such a case the patent specialist can use the INPADOC patent family tool in Espacenet to retrace the simultaneous deposits.

4.1.8 Early publication in some countries

In some cases, the applicant needs to anticipate the publishing of the application for patent (e.g. WO2016080555) in order to earlier exploit the provisional protection. As said before, each national application is independent from others, so might occurs that the applicant publishes its application in some "urgent" countries and follows the ordinary route in other countries or procedures. In such a case the patent search reaches only a part of the family, while the rest will still hide up to the national/regional entry. If the early publication occurs for a PCT application, the EP register offers useful information related to the entry in national and regional phases.

4.2 Legal-status-related issues

As mentioned above, a patent search into international databases will get a list of patent families. Not necessarily a family has uniform legal status among its members. The patent specialist can refer to dedicated databases (e.g. INPADOC and PAtLegal) in order to check the legal status of each member of the family.

4.2.1 Database updates delays

The update of these databases is not in real time, so, for those patent searches for which it is important to know the legal status of the individual document, the patent analyst must check for information updates in the website of the national offices, when available, in order to avoid mistakes.

4.2.2 National entry deadlines in PCT procedure

The PCT procedure automatically designates all PCT countries (*PCT Rule 4.9(a)*) at the international application and allows the applicant to delay the choice of countries in which entry in national or regional phase. The time limit to do it changes from country to country going from 19 months from the priority date, for Luxemburg, Uganda and Tanzania, up to 48 months for Singapore[39]. Germany, Japan and Republic

[39] http://www.wipo.int/pct/en/texts/time_limits.html

of Korea can be excluded from designation (*PCT Rule 4.9(b)*), and the request will be accessible at the date of first publication in "*(RO/101) Request form*" record in PATENTSCOPE. Notice that the withdrawal of designation of Germany is binding only for PCT procedure, indeed if the patent application enters in the EP regional phase, the Germany will still a designated state from EPC procedure.

4.2.3 Reliability of maintenance payment deadlines

Patenting procedures have some "flexible" deadlines that might be sources of risk for the patent specialist. As an example, USPTO offers a 6-month after due date, called "grace period", within which it is still possible to pay the maintenance fee with a surcharge (*US MPEP 2506*) in order to avoid expiration of the patent. Furthermore, may occur some problems in the payment notice (*US MPEP 2530*) that slow down the update of the legal status. Moreover, the applicant may challenge the Decision of expiration of a patent within two months from the decision (*US MPEP 2580* and *2590*).

4.2.4 Loss of rights and their re-establishment in EPC procedure

If a loss of rights occurs, the applicant may file for a decision about the matter within two months from it (*Rule 112 EPC*). The potential re-establishment of rights competes to the department who took the decision about the loss (*Rule 136 EPC*). The event and document concerning the loss and re-establishment of rights can be find in the EP register (EP1043016)

4.2.5 Revocation due to opposition decision

Decision in opposition case may revoke the granted patent (e.g. EP0993241). It definitely changes the legal status of the patent in jurisdiction in which the opposition has been filed.

4.2.6 Changes due to decision of Board of Appeal

Every national law system let the applicant to request for review of a decision. An appeal against a decision on patent application at EPO has a suspensive effect (*EPC Article 106*) as long as the appeal proceeding. The appellant can appeal within 2 months of notification of the decision (*EPC Article 108*). The patent specialist can find the documentation about the appeal in the Register and about decision in the Board of Appeal Decisions Database[40]. In USA, an applicant may appeal against a decision to the Patent Trial and Appeal Board within the six months (*US 37 CFR 1.134*). The

[40] www.epo.org/law-practice/case-law-appeals/advanced-search.html

appeal is available after the second rejection of the claims, also whether the claim belongs to different applications[41]. The proceedings are available for searches in the PRPS. The appeal against "Decision of Refusal" can be filed with the JPO within 3 months (4 months for overseas residents) from the date on which a certified copy of the examiner's decision has been transmitted[42]. The decisions taken by the JPO are available at the dedicated web page[43]. Also in China the time limit is 3 months from the date of receipt of the notification of rejection[44].

4.2.7 Uncertainty due to "Grace Period"

In some patent law systems (e.g. in USA since September 16, 2011 and China up to 2009) there is a so-called "grace period", within which the inventor of a publicly disclosed invention can still file for a patent on the same invention (*US AIA 35 U.S.C. 102(b)(1)(A)*). The Patent Office allows the applicant to exploit a *grace period* only in defined situations[45]. Thus, within the *grace period*, the public disclosure of an invention might not constitute prior art. The only chance in order to get information about this kind of uncertainty is to search into non-patent-literature published within the *grace period*.

4.2.8 Changes of expiration date of a patent

Generally, the life of patent lasts 20 years, assuming the applicant pays for all maintenance fees, but as function of some events, the term may change (*EPC Article 63*). USPTO makes available to the users a calculator, although having no legal value[46].

4.3 Parsing Procedure

Syntactic parsing is a method that takes a text as input data and builds a hierarchical syntactic structure according to the role of each word inside the proposition (e.g. subject, action, object, etc.). It offers high precision, in fact, it is employed to simplify the complex sentences of a patent claim in smaller and more readable utterances (Wang, Lu and Loh, 2015) or to enhance the performance of patent search engines (Sakthika and Kogilavani, 2015).

---

41 https://www.uspto.gov/patents-application-process/appealing-patent-decisions/patent-trial-and-appeal-board-ptab-faqs
42 https://www.jpo.go.jp/english/faqs/patent.html
43 www.jpo.go.jp/torikumi_e/t_torikumi_e/decisions.htm
44 http://english.sipo.gov.cn/faq/948972.htm
45 http://www.apta.org/HealthCareReform/MakingSense/
46 *www.uspto.gov/patent/laws-and-regulations/patent-term-calculator*

The topic '*parsing tool'*, usually refers to a text analysis software that contribute to ease the extraction of useful information from an input text corpus. It computes the relationships among the tokens composing a sentence and encodes a dependency graph of the sentence structure (see Figure 28)

The dependencies are the binary-syntactic relationships (*nsbj*, *pobj*, *prep*, *det*, etc.) between lemma couples. They are limited in number and changes according to the software used.

In order to enhance this parsing task on patents, the authors tested *spaCy* package[47]. It is a probabilistic parser based on Stanford dependency approach (de Marneffe and Manning, 2008) and integrates the CoNLL dependency one (Johansson, 2008), to enrich important relations like object predicates and to minimize unclassified dependencies (Choi and Palmer, 2012). It was selected because able to work on long-distance dependency elaboration, very typical in patent text.

In this work, the authors used spaCy for setting a pool of Syntactic Dependency Pattern (SDP), useful for identifying the syntactic features related to technical problem definition in patent documents.

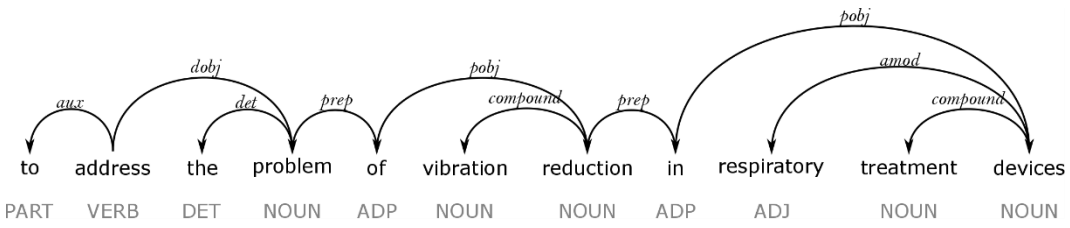


Figure 28 An example of a dependency graph labelled with spaCy dependency.

In Figure 28, the word-line shows the tokenization of the sentence. Under each token, in gray, it puts related PoS-tag. Arrows define the binary dependencies, from the "head" to the "dependent" lemmas.

4.4 Supervised state of the art analysis for Business Intelligence

4.4.1 First step: recall increment with keyword-based patent search

As mentioned above in section 4, the first step of the method does not modify the usual approach to the patent search of attorneys. The focus of this step is to enlarge the coverage of the patent pool by semantically expand the keyword set, thus enlarge the patent recall.

[47] http://spacy.io

To find additional keyword related to the searched topic there are several strategies (Lupu *et al.*, 2011; Hunt, Nguyen and Rodgers, 2012; Khode and Jambhorkar, 2017), within which the use of specialised thesaura and query expansion tools (Montecchi and Russo, 2011, 2015; Russo, Montecchi and Duci, 2015).

In this step, the analyst does not matter about the precision. Thus, the query composition could be simpler than the current approach, which considers also precision issue (Salton and McGill, 1983).

4.4.2 Second step: supervised identification of state of the art

An interesting branch of research that works on overcoming the drawbacks of the Boolean search deals with the combination of the syntactic parsing tools with patent ontologies.

Patent ontologies define patent-specific fundamental concepts and relations. It contains information about the structure of the document (i.e. first page, claim section, description, etc.), basic parameters (i.e. patent number, inventor, title, etc.…) and technical features (design aim, working principle, problem to be solved, application field, materials, functions, etc.) (Wang, Lin and Yang, 2013).

Setting up a system based on the syntactic parsing and specific patterns extraction, we are able to generate the list of problems that lets the user to get automatically the state-of-the-art of a pool of patents and extracts the documents related to them.

Nowadays does not exist any tool able to support this kind of analysis, what needs time and expertise from the user, who has to know all kind of problems in advance in order to translate them in search queries.

According to the proposed method, the authors show how such an approach is able to extract the relevant sentences from a patent (or a patent text corpus) and organize them in a concise and easy-to-be-read list useful for a first, coarse classification, unsupervised by the user who has to define the state-of-the-art.

In order to identify a problem, syntactic parser is only a part of the method. It needs also linguistic patterns and a strict ontology for understanding what a problem is. Furthermore, according to FOS logic, we want to extract in automatic a list of problems and assess a raw state-of-the-art about the inventive problems the inventors dealt with.

The design entity that has to be automatically identified in this paper is only the *problem*. It describes the need or the goal to which the claimed invention must fulfil.

In natural language, a problem can be described in many different ways, which combinations are potentially infinite in number. However, the requirement that a patent must declare what is the problem the invention claims to overcome limits the variance of its description. Furthermore, according to what said by (Wang, Berant and Liang, 2015), there are a limited amount of cases that a simple general grammar cannot reach in a domain specific corpus.

The main idea is to retrieve the pertinent patents, using only few (SDP), as shown in Figure 29. It highlights the SDP using black dependency arrows, whereas the grey ones refer to dependencies related to the context of the problem.

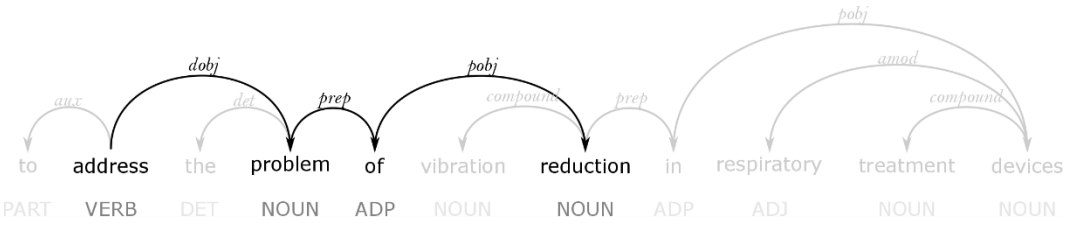


Figure 29 Example of use of a SDP for the extraction of 'problem' feature.

Each SDP contains syntactic patterns related to the problem (i.e. for ***solving*** the ***problem of, prevent from, caused by, etc.),*** as shown in Table 21.

Table 21 List of dependency patterns for retrieving problem description

1	[**solve**] *dobj* **[problem]** *prep* of *pobj (problem description)*
i.e. a fine platform **for solving the problem** of (*abnormal noise from the coolant pump bearings*)	
2	**[action]** *prep* **without** *pobj (problem description)*
i.e. A ring could be **injected without** *(perfect alignment of the two halves of the retaining ring cavity)*	
3	**[prevent]** *prep* **from** *pobj* (*problem description*)
i.e. *edge gated nozzle may be **prevented from** thermal expansion*	
4	**[cause]** *agent* **by** *pobj* (*problem description*)
i.e. *corrosion **caused by** the gases released during the moulding*	
SDP Description *dobj* - Direct object	

Prep- Prepositional modifier *pobj* -Object of a preposition *Agent*- agent

Linguistic patterns are combined with syntactic parser for reaching a high precision. For ameliorating the recall a list of lemmas dealing with the action of solving and with synonyms of "problem" is proposed as shown in Table 22.

Table 22. List of lemmas about 'solving and 'problem'

Solve		
Eliminate	Overcome	Avoid
Deal with	Mitigate	Address
Prevent	Cause	…
Problem		
Issue	Drawback	Challenge
Trouble	Limitation	Deficiency
Disadvantage	Anomaly	…

By this method only sentences that suggests which are the problems are extracted from the patent pool. No domain-specific indications (less than functions) need, which means that also a user not expert-in-the-field can retrieve domain-specific problems.

The list of sentences is a concise and easy-to-be-read result that summarizes the 'problem environment' described in the patent pool. Such a result fulfils the requirement of the state-of-the-art to find at least one of documents related to each kind of problem faced by the patent pool analysed.

Case study

In order to explain the method problems extraction in the field of the *injection molding* is proposed. No specific topics was defined to limit the number of documents analysed.

Building of the patent pool

The authors used *Orbit Intelligence*[48] to querying the worldwide patent DB. The aim of such a first extraction is collecting a quite large number of patents document able to gather the most of state-of-the-art about the problems of the target topic. The search

[48] www.orbit.com

has been based on keywords *injection* and *molding* (Table 23 lists the result of each search step).

The first step was querying the title (TI) and abstract (AB) patent fields with keywords linked by '*AND*' Boolean relation (step 1). To enhance the precision of results, we reduced the distance between keywords using '*nD'* operator, where *n* sets the maximum number of words between keywords (step 2).

Table 23. List of queries and related results

#	Query	# of results
1	(injection **AND** molding)/TI/AB	121359
2	(injection **2D** molding)/TI/AB	106976
3	(**inject+** AND **mold+**)/TI/AB	134605
4	(inject+ AND **mo?ld+**)/TI/AB	171829
5	4 AND (US OR EP)/EAPC	12959
6	5 AND EPRD>=1995	7542

'Injection molding' states an action, '*molding'*, performed in a specific way, '*injection*'. Both, the action and the way, could be referred to by different words, e.g. the conjugation of the relative verb or plural of the corresponding noun, and so on. In order to match this kind of variations we can use the truncation wild-chart '+' (step 3). 'Molding' is a word of US English, but in UK English its equivalent is 'mo*u*lding'. Another wild-chart, '?', allows the user to generalize the query in order to match such a kind of variance (step 4). For simple explanation, we do not consider other query expansion techniques.

At this step, using keywords or Boolean relation, we could not have reduced more the patent pool, indeed additional keywords or stricter Booleans might have been reduce the technology coverage of the pool. For the same reason, no filter on IPC or CPC classifications was take into account.

We are interested mainly on English texts to test the method. To get original English texts we filtered the patents on US or EP as first application country (EAPC) (step 5). Last, we gathered patents whose first-priority application has been filed since 1995 (EPRD) (step 6).

The patent corpus, considered for this study, is a collection of more than 7500 patents belonging to different application fields. For each patent, the authors extracted only free-text fields: Title, Abstract, Description and Claims.

Extraction of problems-related sentences

Table 24 shows the resulting sentences gathered by the methodology for the case study. It reports the most important part of each sentence to simplify its reading and corresponding patent number in which they occur.

The last step of the method is to summarize the history of most interesting problems giving the user the evolutive perspective of such an issue. Using suitable graphs, e.g. bubble graph, the patents involved in one of target problems can give at same time the information about which technology field are interested and which is the behavior in time of the patent applications.

The search tool built on syntactic patterns and based on FOS approach is very sharp, indeed the sentences it got are very close to the target. Differently, considering a keyword-based search, the list of results tends to bring noise in results. E.g. looking for *'solv+ 2D problem? 4D mix+'* in the final pool of Table 23. List of queries and related results, we got sentences like *'The **mixing** tool 36 referenced above **solves** that **problem** by incorporating a stripping element 37 into the said tool 36'* that is far from the target.

Table 24: List of sentences describing 'problems' in injection molding patent pool

[solve] *dobj* [problem] *prep* of *pobj* (NOUN)	
SOLVE-PROBLEM	
US20160108919	[…] a fine platform for ***solving*** the ***problem of*** abnormal *noise* from the coolant pump bearings
EP2527125	[…] while ***solving*** the environmental ***problem of*** Cr *plating*
EP2508555	Therefore, to ***solve*** the ***problem of*** non-uniform *mixing* and right *dosage* of the raw materials […]
US8182723	[…] are ***solved*** the ***problems of*** the time-consuming fabrication *process* and poor lens *alignment* […]
EP2294133	[…] ***solves*** the ***problem of*** premature *vulcanization* associated with fast curing rubber compounds, […]
…	…
OVERCOME-PROBLEM	
WO201474760	To ***overcome*** the ***problem of*** *freeze off* , […]

US9414895	[…] ***overcomes*** the ***problems of*** flowable *composite* such as the increase shrinkage [..], the lower polishability […], the lower wear resistance […], and the lower strength of floable composite
WO200248753	[…] ***overcomes*** the ***problems of*** poor mold surface *replication* and residual part *stresses* […]
WO200074922	To ***overcome*** the ***problems of*** poor mold sunace *replication* bv […]
EP-972714	This ***overcame*** the ***problem of*** thread *distortion* when stripping […] thread with less than two full turns
…	…
ADDRESS-PROBLEM	
US20160340494	[…] embodiments […] can ***address problems of*** high flow viscosity melted *thermoplastics* […]
WO2016141047	[…] does not ***address*** the ***problem of*** production *downtime* related to material change over, […]
US9452569	The present methods also ***address*** the ***problem of*** high flow *viscosities* […]
US7931249	[…] considered in ***addressing*** the ***problem of*** mold *sticking* to a substrate on which it is heat treated
US7131833	This heating configuration attempts to ***address*** the ***problem of*** uneven *heating*
…	…
[action] *prep* without *pobj* (NOUN)	
INJECT	
US20170030455	[…] ring could be ***injected without*** perfect *alignment* of the two halves of the retaining ring cavity
US9414895	[…] paste composite to allow it to be ***injected without*** the *use* of lesser filled resins such as flowable resin
US7381455	[…], the material being ***injected without*** irremediable *degradation* by exceeding the limit shearing speed
US5876765	[…] different volumes can be ***injected without*** *changes* in the actual device for supplying material
EP1683739	[…] the material […] ***injected*** thereinto ***without*** the *gate* interfering with the feeding of the material […]
…	…
MOLD	
WO2017185076	[…] all of the features […] can be ***molded without*** *slides*
WO200775202	Such mixed glass/thermoplastic bundles may be shipped and molded without any additional additives […]
WO9638283	[…] small quantities of different parts may be molded without shutdown of the mold machine
US8951454	[…] the present invention allows for two piece balls to be compression molded without the need to provide pre-formed cups which require the more difficult, expensive and labor-heavy injection molding

US20090077802	The plastic airfoil portion 24 is molded without the need for fiber reinforcement, […]
…	…
[prevent] *prep* from *pobj* (NOUN)	
US20170340172	The structure of the cutting board is firm in one piece so as to ***prevent from*** *peeling*
EP2487358	[…] reinforcement plate […] is prevented from decoupling from the seaming cap unless force is applied
EP2228194	[…] conventional edge gated nozzle may be ***prevented from*** thermal *expansion* in a longitudinal direction
US20100038811	[…] the sliding element is ***prevented from*** *displacement* via the pressure of the polymer melt injected […]
EP1854611	In order to ***prevent from*** a *disarrangement* of the nail 110 with respect to the broken bone […]
…	…
[cause] *agent* by *pobj* (NOUN)	
EP3216578	[…] *movement* of the laminate within the mold ***caused by*** an injection *flow* of an overmold
WO201740527	[…] reduced solidification *rate* […], typically ***caused by*** the outer *layer* of molding material, […]
WO201715573	The *displacement* ***caused by*** the *addition* of the opening force to the closing force […]
US20170015033	[…] *oxidation*, *wear*, high *temperatures* and *corrosion* ***caused by*** the *gases* released during the moulding process, […]
US20170001344	Volatile organic compound (VOC) *emissions* are ***caused by*** low molecular weight *compounds*
…	…

Thus, the method suggested can drastically enhance the *precision* of the patent search. Contrary its *recall* feature is very limited in comparison with keyword-based tools. The use suggested by authors of dependency patterns is about the state-of-the-art. Thus, the need is not the accurate collection of pertinent patents dealing with a target matter, but it is the big picture of the technology or application field. Therefore, it requires to give the qualitative behaviour in time of patents related to a problem.

We found that patent trends coming from the two approaches are very similar when both pools have only relevant patents. Thus, the qualitative information is retrieved faster using the proposed method than the keyword-based one. In order to get an accurate state-of-the-art, e.g. for a patentability search, the user has to do a deeper and more focused search.

4.4.3 Third step: member's data extraction

Companies' decision makers use business information to plan the operations and pursue the business aims of companies. To simplify the reading of a large amount of data, they organize them with a Business Intelligence (BI) method.

The BI is a collection of strategies and tools for the data analysis of business information (Dedić and Stanier, 2016). Many kinds of functions concur to the arrangement of data in order to obtain a comprehensive and easy-to-be-read dashboard of crucial information.

BI aims to ease the interpretation of business information and extract from them insights able to suggest new business opportunities or strategies in order to assure a competitive market advantage and long-term stability (Rud, 2009).

When applied to patent data, the BI is also called Patent Intelligence. Current software tools of patent search offer different patent distribution analyses to visualize their behaviour over time, country of filing or publication, technology field, applicant and owner. The most advanced tools allow the user to graph the co-applicant network in order to retrieve R&D collaborations in the target technological field.

As mentioned in sections 4 and 4.1, current search tools work on patent families. Despite such a method simplify the searching task, it aggregates patent data that could be more useful in an elementary form when analyzed.

A selection of simple tools to ameliorate the patent intelligence is presented. Starting from the export of current search tools, especially Orbit Intelligence(orbit.com), the package of scripts allows the user to arrange the data in the most useful way.

Applicant aggregation

In patent literature the applicant field contains the business name of the fist owner(s) of the application. Unfortunately, often it is written neglecting the legal form (e.g. Ltd., AG., S.p.A., GmbH, etc.) or in a wrong way (e.g. Siemes instead of Siemens in US2007083658 on Esp@cenet). This causes a wrong grouping of patent portfolios.

Commercial tools for patent search allow the user to retrieve company names from databases of corporate trees, but these data do not cover dismissed business names and do not consider errors in writing. Nonetheless, they retrieve corporate trees from financial DBs, thus are able to catch potential company merges; in this case is important knowing when occurred the last update.

To aggregate patents per applicant, a BI tool can help the user in building the specific thesaurus of alternative names of most interesting companies. Then a simple data indexing aggregates the data to collect the right amount of patent document for each competitor (see Figure 30).

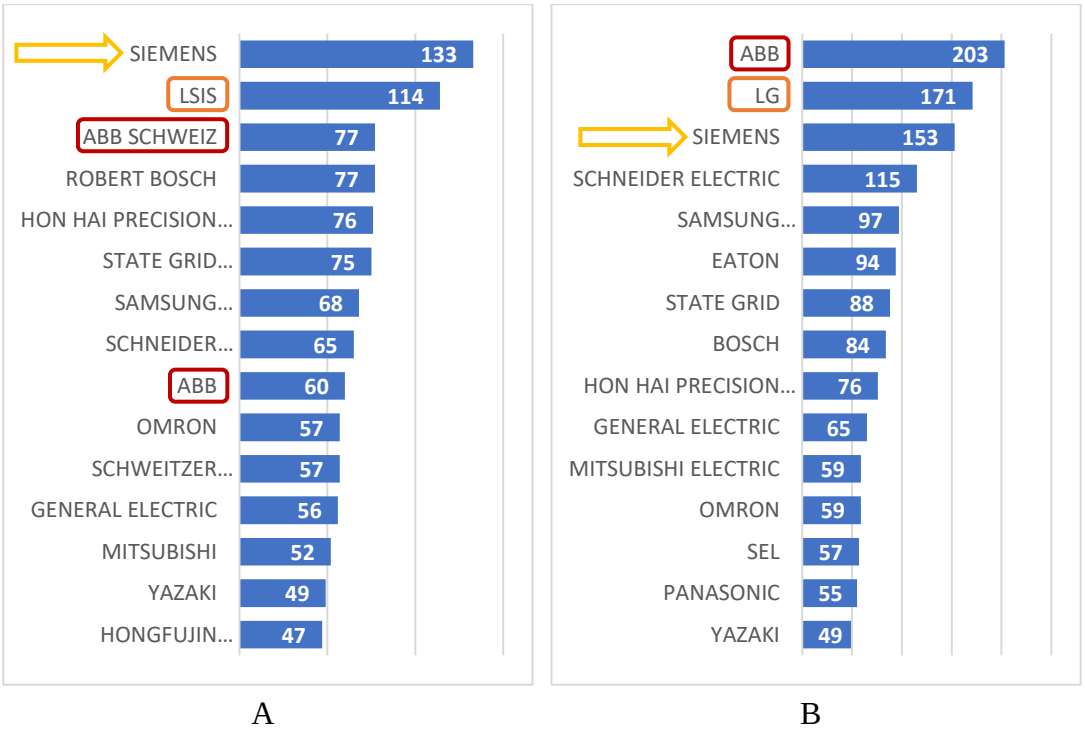


Figure 30: comparison between the TOP 15 applicant in MV relays before (A) and after (B) aggregation by applicant

In Figure 30, red rectangles highlight the simple aggregation by name similarity, whereas orange ones indicate the aggregation of applicant using corporate tree. The yellow arrow points to the most important effect of aggregation, i.e. the adjustment of the rank of applicant, in this case Siemens from the 1st position falls to 3rd one

Data extraction of family members

The export from commercial search tools can include the list of patent numbers and bibliographic data of each family member. The BI extraction tool employs a data mining technique able to read the list and separate it in different records.

According to the raw data available, the BI data mining tool can retrieve a wide amount of information. Table 25 summarizes what could be extracted for each patent family and its members.

Table 25: description of patent family data extractable by BI data mining tool from a patent data family record exported by a commercial patent search tool

Label	*Meaning*
ID	Univocal identification number of the resulting record
Family ID	Univocal identification number of the family in search tool database
Nr of Application Countries	Number of countries/jurisdictions with which an application filing occurs. WO applications does not matter about designating states. EP applications count as different from nationalizations.
Application Countries	Country label of each application country according to the Espacenet country codes[49]
First Priority Date	According to the need of the analysis, the time ordering can be arranged per year, trimester, month or single day
Nr of Inventor	Number of inventors of patented matter
Inventor names	Names of inventors of patented matter
Nr of current applicant	Number of owners of the patent document at the moment of export
Current Applicant Names	Unified name of current applicants according to aggregation thesaura
Nr of reassignement	Number of times a member of the patent family has been changed its owner
Assignees history	Chronological sequence of assignees of a patent family member
Dates of ownership change	List of dates in which a change of ownership occurred
Nr of Technological Domains	Number of technological domains related to the patent document according to a specific classification[50]

[49] https://worldwide.espacenet.com/help?locale=en_EP&method=handleHelpTopic&topic=countrycodes

[50] Different kinds of technological classification are available. According to the BI aim, the user can choose the suitable one: ISIC (OECD, 2011), NACE (http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=EN), ISI-OST-INPI (Schmoch, 2008), NAICS (https://www.naics.com/search), etc.

Technological Domains	List of technological domains for each family member
Average Nr of cited patents documents	Average of patent document cited by members in patent family. Source of citation (applicant, examiner,opposition…) does not matter
Average Nr of literatures cited by examiner	Average number of scientific papers cited by the examiner in patent family
Average nr of patents citing by examiner	Average number of patent document citing a family member by examiner in patent family
Nr of claims	Number of claims of the priority application
Nr of independent claims	Number of indipendent claims of the priority application
Legal status	Legal status of each patent family member

For each of data described in Table 25, a deeper analysis can be done.

Below some examples, coming from Smart Grid application field, are discussed briefly to explain their contribution and risks they might hide.

Patent timeline

When a patent pool is consistent with a specific topic, its timeline explains how it is important in the reference market. A linear growth may suggest an increasing interest of the actors which invest in R&D to increase their knowledge around the topic, while an exponential rise highlights a new technology in which many applicants might be new entrant motivated to exploit a disruptive technology. Contrariwise, a descending trend generally means the topic is obsolete.

Figure 31 depicts the patent filing timeline for MV relay technology, in which a moderate acceleration can be find around 2010.

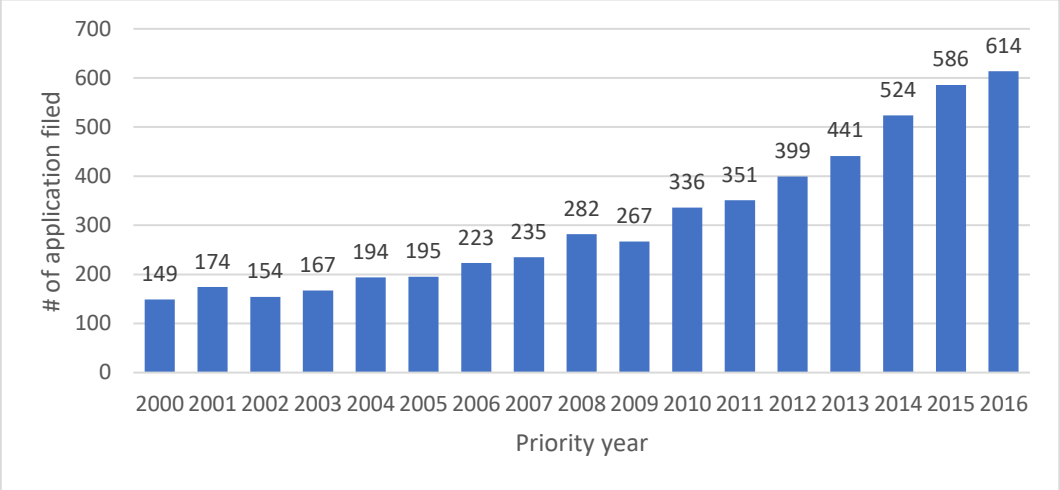


Figure 31: the patent filing timeline for MV relay technology.

Geographical distribution

The patent country coverage can be easily viewed the with a geographical distribution map, as showed in Figure 32.

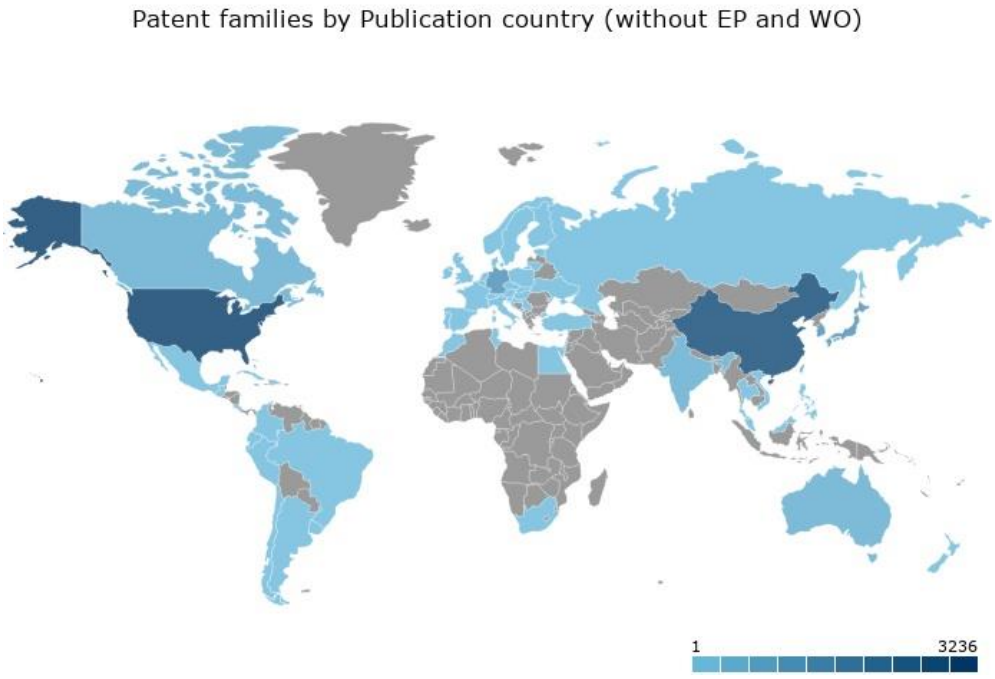


Figure 32: patent country coverage for MV relays patent pool

Opinion relying on this kind of diagram lacks in crucial information about the future nationalizations of WO and EP applications, as stated in title of Figure 32. In fact, the

two proceedings let the applicant to postpone the decision at the end of the proceedings themselves. When geographical information is strategically important for the analysis, the map of filing and/or publication might mislead the user.

Citation network

Another interesting example for patent analysis, is the citation network. In a patent pool could be present a document disclosing a pivotal invention, from which many other patents (at least in part) come from. That document would be cited by all of them and it will at least in part limit their claims. A specific analysis to highlight most cited patent documents is important to verify the potential claim overlap of a new application.

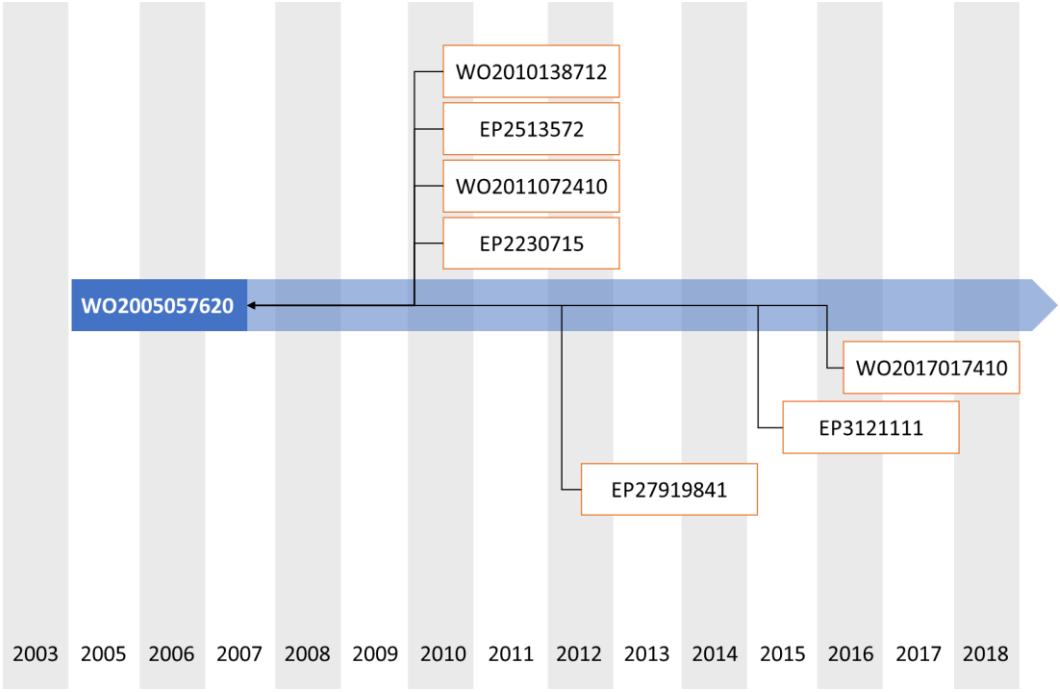


Figure 33:citing patent application for WO2005057620 in patent citation network. Only X and Y citation are shown

Re-assignment

In the patent valuation task, a very important information is the reassignment of a patent document. The reassignment means a transfer of the ownership of the patent, then of the exclusive right. This event is an indicator of the economic interest of the buyer on the patent document. Due to the territorial feature of patents, the new owner of a patent member could limit it acquisition to the country in which it operates. In this case, the value of the patent family member increases, but not necessarily the value of the other members does the same.

5 Market structure and patent application risk management

So far, the work has been focused on a single patent application, extracting from it useful information to assess its positioning in relation to relevant the state of the art and procedural path.

This chapter concerns the computation of an indicator based on data about the network of patents linked to the target application. The indicator points to assess the type of market structure in order to identify potential hindrances due to dominant positions of main incumbent competitors.

The market *for* innovation is particularly characterized by the use of licensing and sale of patents which negotiations could both either promote or prevent the diffusion of a specific technology, affecting firms' incentive to invest in further innovation (Arora, 2001). Patent data can help the decision makers in identification of tricky market structures that hinder the company to access to the market itself, for example building a patent thicket (Shapiro, 2001).

The patent thicket is "a dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology" (Shapiro, 2001). It affects the structure of the market, usually characterized by cumulative or complementary technologies (Von Graevenitz, Wagner and Harhoff, 2013), defending against competitors designing around a single patent (Rubinfeld and Maness, 2004) and/or building hindrances to the innovation (Bawa, Bawa and Maebius, 2005; Sabety, 2005; Clarkson and DeKorte, 2006; D'Silva, 2009; Galasso and Schankerman, 2010a, 2010b; Hargreaves, 2011).

It arises when, according to the claims made in the patents about both physical and methodological components of the patented product or process, it is likely that the ownership of the exclusive intellectual property right can overlap across different economics agents (i.e. firms), preventing their use and creating a proper barrier, especially for the follow-on innovation (Shapiro, 2001). It is a peculiar characteristic of complex industries.

A potential new entrant in a specific technological market might be discouraged by the presence of a thicket, simply because of the high cost related to the bargaining process, which is requested to advance further the innovation. This is even more evident when the market is characterized by the presence of few big players that may have a dominant position by owning a large pool of patent, then conditioning the access to the market. According to the complexity of the patent thickets, the new entrant can be

discouraged by the sticky and costly process, and as a result it would not undertake any further investment in the advancement of the new technology (Hall, Helmers and von Graevenitz, 2015). Alternatively, whether the investment takes place by licensing agreements and a new technology is proposed, due to the initial high cost its diffusion will be limited, preventing future advancement in the innovation process. (Shapiro, 2001). Galasso and Schankerman (Galasso and Schankerman, 2010a) show how the presence of patent thickets obstruct follow-on research especially in the complex industries (for example in information and communications technology, electrical–electronics and medical instruments). Once we accept that innovation depends mainly on the paradigm which sees it as a cumulative process upon which new ideas are generated, in Gallini (Gallini, 2017) we could find a comprehensive review of the literature which highlights the issues yielded by the presence of patent thickets.

An indicator based on this kind of consideration could give to patent analysts an additional information about success of patented technology.

5.1 Patent thickets as indicator of hindrances to the market access

Business Intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information, in order to identify new business opportunities with a competitive market advantage (Marchand and Raymond, 2008; Brannon, 2010; Alaskar and Poulis, 2015; Walsh, Lee and Jung, 2016). Firms make use of BI strategies to support a wide range of business decisions, which include both operational (i.e. production, and distribution) and technical (i.e. analysis of data) procedure able to improve their own efficiency. In the most recent times, it has been adopted to support the product innovation process, also exploiting IP marketing, and to minimize the risk of R&D management decisions.

Knowing whether or not there is a patent thicket, before starting an innovation activity, is pivotal, as it allows to anticipate and update the BI strategy before the launch of the product will occur into the market. It help also the R&D activity in ranking of the list of product requirement with highest market potential (Livotov, 2015).

Currently there are no tools able to automatically extract the data necessary to unveil the risk to meet with patent thicket; however, there are strategies based on backward and forward citations with which it is possible to obtain indications regarding any priority relations between patent pairs (or applicant pairs) (Von Graevenitz, Wagner and Harhoff, 2011). Other useful information can be derived from the co-applicant and co-inventors maps, which show collaboration and relationships even between different companies (Ijichi, Yoda and HIRASAWA, 1994) that are assumed to be competition within the same market, unless R&D cooperation agreements have been signed, and

then emerging from the co-patenting measurement. A simple indication of the presence of patent thickets it is not informative *per se*, and for that reason the approach we propose would instead suggest the major players in the market, highlighting also the relative structure of the market.

In order to take into account the market structure, while ranking the market potential of product/process requirements, we present a new method able to incorporate both characteristics. We investigate a large pool of patents extracted from Machine Learning field in order to identify the presence of possible patent thickets to pinpoint the major market players and all possible constrains faced by a new innovator entering into the market.

5.2 A method for identifying patent thicket

The strategy to measure the density of a patent thicket proposed by (Von Graevenitz, Wagner and Harhoff, 2011) relies on the patent literature cited by the examiner in the search report having kind X or Y. He uses the triple (see Figure 34) of applicants involved in mutual blocking citation relationships as a unity of measure of the thicket density. Higher the number of triples, worse the hacking through the thicket.

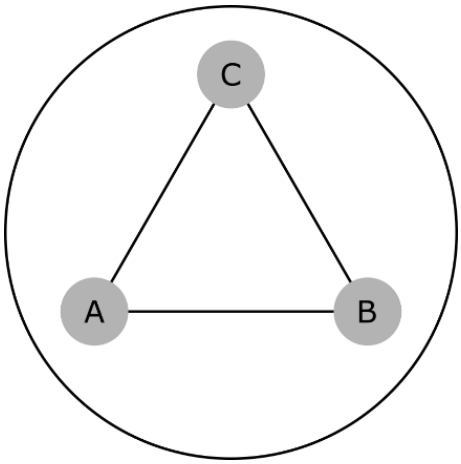


Figure 34: Schematic presentation of the structure of triple. Source (Von Graevenitz, Wagner and Harhoff, 2011)*)*

In Figure 34, A, B and C circles identifies the firms involved in triple; the straight lines represent the mutual blocking relationships between firm couples

Given the complexity of such a type of dimension, it finds some limitation for application to BI analysis. The required skills ranging from IT field, especially SQL language to querying PATSTAT, the statistical DB published by EPO[51], to patent procedural knowledge. This makes hard to reproduce the analyses available in literature.

Furthermore, the method has been proposed in an economical context. It tend to make an economic analysis of the market for the main classes of OST-INPI/FhG-ISI technology nomenclature (OECD, 1994). The typical dimension of the patent pools used by von Graevenitz is excessively large in order to give an information exploitable by decision makers concerning the reference market structure of a specific technical solution.

Moreover, although the triples count is an interesting method to measure the thicket density, it does not take into account the inner balance (or imbalance) of the patent portfolios, which explains the effective polarization toward one or two competitors in the triple.

In this article, the authors introduce an algorithm able to automatize the triples extraction process in a delimited technology environment. The output is a navigable network of citation links, in which the user can identify the main players, taking into account the contribution of the balance/imbalance information.

5.3 Triples extraction method for identify patent thicket

To extract the information about the possible presence of patent thicket in a technological domain related to a patent application, we suggest a modified triples evaluation algorithm.

It works in a patent pool selected in a more refined way than OST-INPI/FhG-ISI (or NACE) classification, considering the reference application field only. Thereby the measure of thicket density is 'local' and the approach can index the main players involved in thickening.

The algorithm gives three different indexes about the triple inner imbalance.

First it shows the number of citations for each couple in both directions (see Figure 35). This might unveil that one (or two) of the player involved in the triple is not effectively disturbing the other two because its blocking patents are limited in number.

[51] www.epo.org\searching-for-patents\business\patstat.html

Therefore, the other two applicants might not actually worry about the triple and control the market as a duopoly (or monopoly).

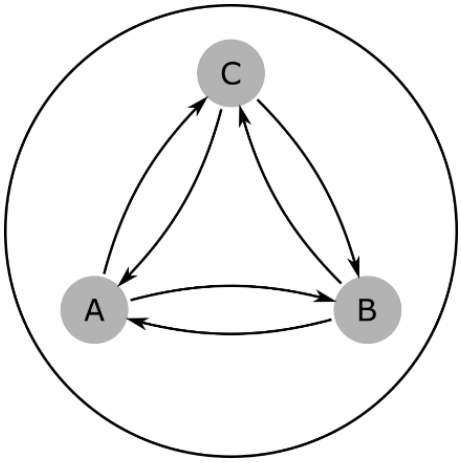


Figure 35: a generic triple schema. Arrows start from the cited player and go to the citing one (blocking direction).

The second index refers to the inner relative strength due to the portfolios size comparison of the triple players. The triples with an important imbalance due to this reason may suffer the effect of dominant positioning of one (or two) player. Thus, the actual configuration of the thicket tends to become a monopoly (or duopoly).

The last index measures the ratio between the target-technology-related portfolio of a firm in the triple and its whole patent portfolio. It indexes which is the effective interest of a single player involved in the triple to the target technology, and its relative market. A high ratio means the player considers strategic the target technology and the involving in triples could be a great fail risk.

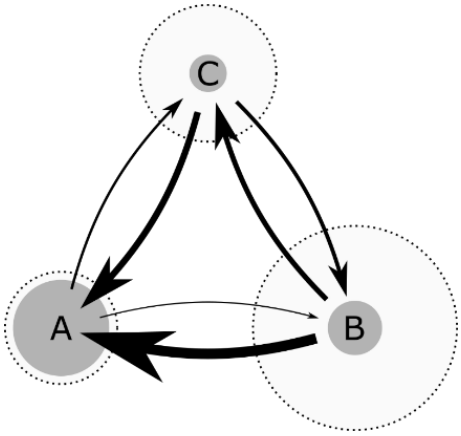


Figure 36: a complete visualization of the inside imbalances of a triple.

In Figure 36, the arrow thickness indexes the weight to the bilateral citation relationship, starting from the blocking player and pointing to the blocked one. The dotted circles indicate the dimension of the patent portfolios, related to the target technology. The number of the documents involved in the triple is proportional to the area of the grey circle. The second index gives us the information about the inner imbalance between the players in a triple, while the third index shows the importance/interest of the target technology for each player (see Figure 36).

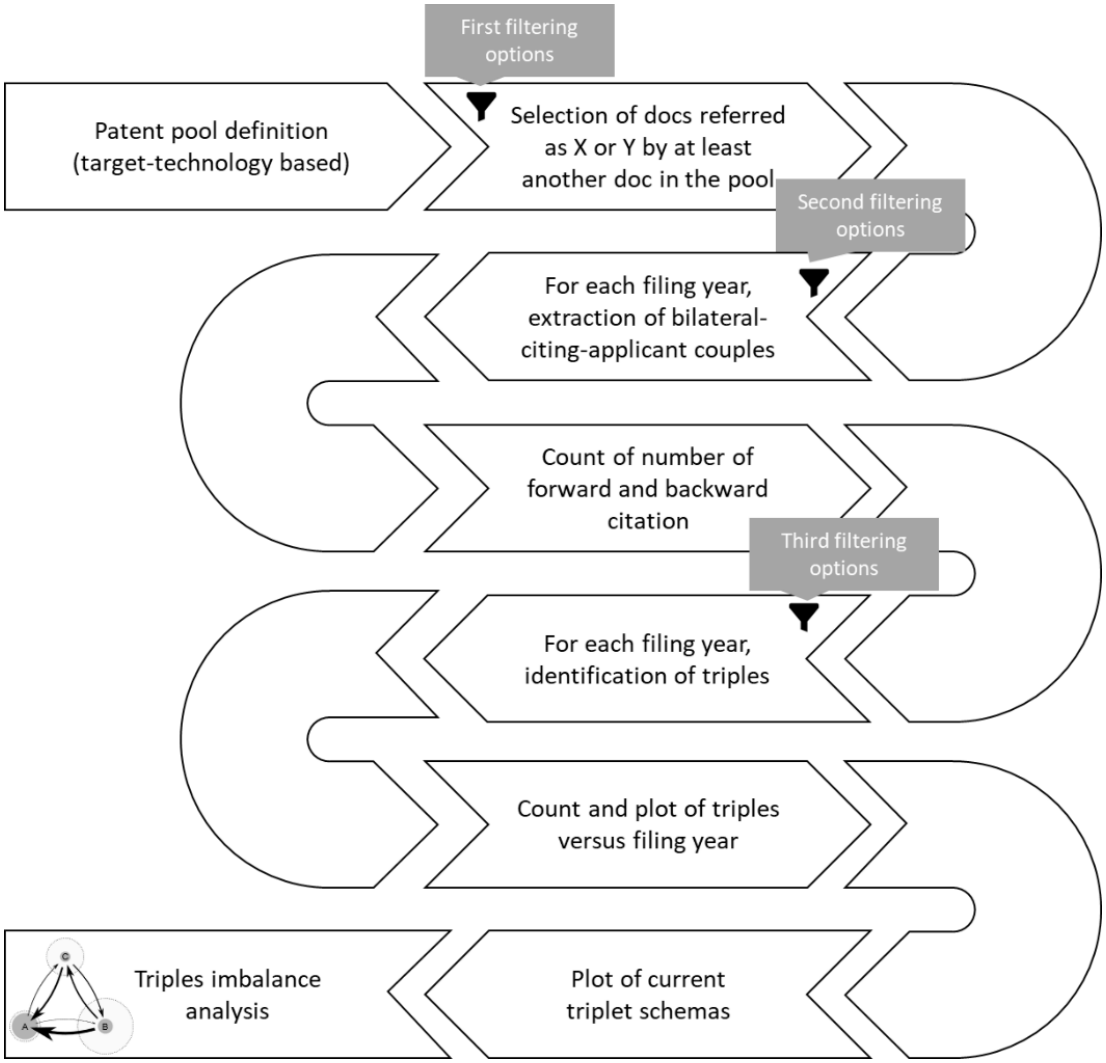


Figure 37: the algorithm for the extraction from a technology-based patent pool of triples and related data.

The assessment on potential imbalance in triples is a useful indication that could be integrated into the BI tools in order to unveil the actual structure of the market to the decision makers and let them to make choices in a more informed way.

Figure 37 shows the algorithm used to compute the triples, considering filters in the choice of patent documents, application activity of the firms and inner imbalance in the triples. The first filtering option filters documents by filing year and applicants by minimum number of applications *(Von Graevenitz, Wagner and Harhoff, 2013)*. Second filter acts on the lifetime (in years) of a bilateral citation *(Von Graevenitz, Wagner and Harhoff, 2013)*. Filter 3 considers the imbalance parameters (relative portfolios dimension, documents involved in triple and number of bilateral citations)

As an example of the application of our proposed method we present the case of a pool of patent extract from Machine Learning patent field, and due to possible time lags in the patent office register update, we censor the last two years as suggested by the patent literature.

We extracted all patents from Google, IBM and Microsoft focusing on machine learning and artificial intelligence. From Figure 38 we can observe this market segment is dominated by the presence of three major players, Google, IBM, and Microsoft Technology. We use citation as a measure of a patent market value (Hall, Jaffe and Trajtenberg, 2005), and in particular number of co-citation to identify the extent of the patent thickets among the major players. It emerges clearly that both Google and IBM tend to interact and make citation in a reciprocal relationship mainly with Microsoft Technology Licensing. Whereas the co-co-citation numbers between Google and IBM drop by almost 30 per cent compare to the same measure they have with Microsoft Technology. From this scenario it looks like that Microsoft Technology Licensing play a role as a leader controlling the number of citations which the other two players. This very preliminary analysis would suggest an oligopoly given the presence of three firms (with other two very small), where the one which owns the larger number of patents might play an important role in controlling the market entrants. Of course, a furthermore details analysis is needed to be able to understand if any illegal behaviour is in place (i.e. collusion). However, from this simple picture would it be already enough to confirm the presence of high barriers to entry which could discourage the progress of this specific technology, leaving if in the hands of few actors.

5.4 Conclusions

This article introduces a method that aims to extract quantitative data from patents to identify the structure of the reference market for a target technology field. It aims to enhance the innovation strategy for a firm that proposes a new technology, especially when it operates in a complex market.

The method extracts citations data from patent database by which computing triples and measuring their inner balance.

It has been tested in machine learning field and it shown a dominant positioning of Microsoft, IBM and Google. The experimentation proves and highlights the risk for an unaware entrant to incur in unexpected extra charge due to the hard licensing bargaining.

The method is valid for any technical domain.

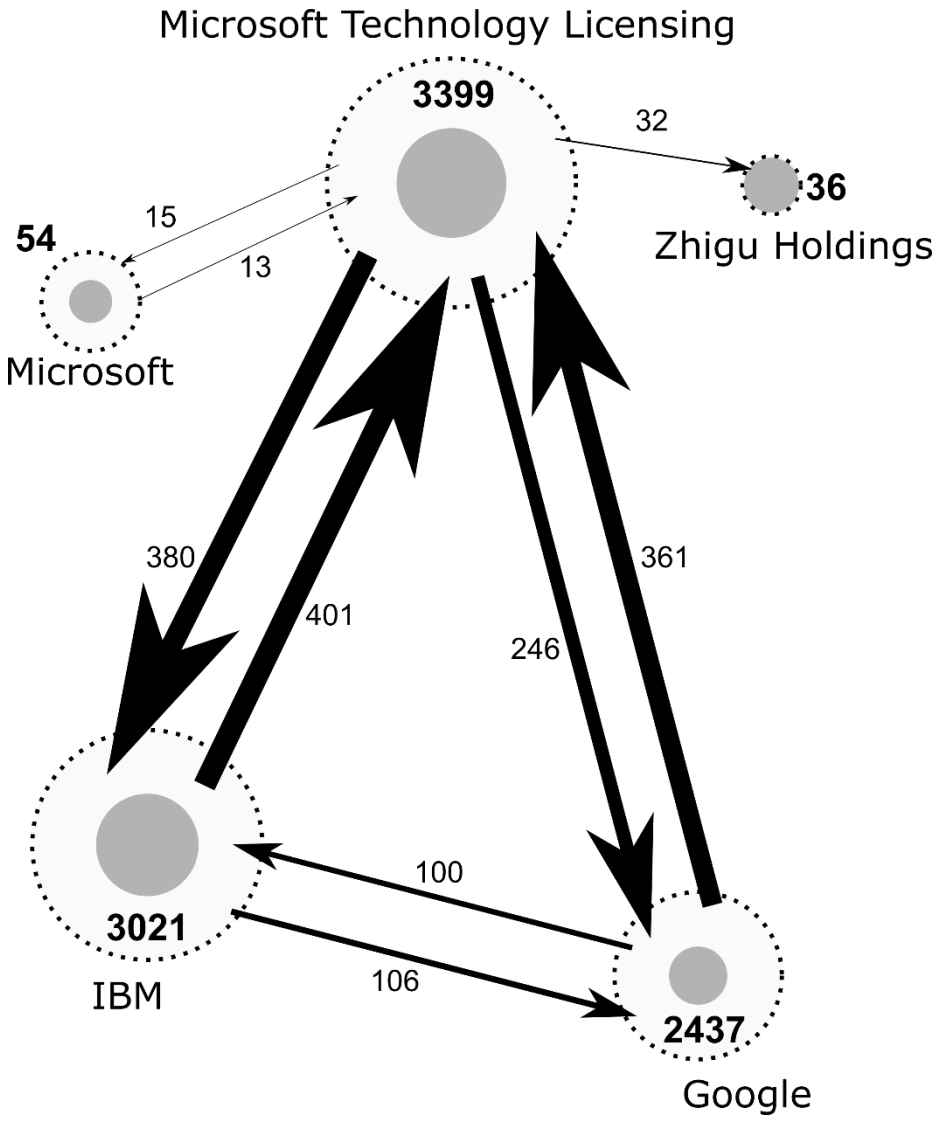


Figure 38. Most interesting triple isolated by the algorithm in Machine Learning sector involving Google, IBM and Microsoft.

Conclusions

Patents are a unique and exhaustive source of technological knowledge. Technical information we can find in them is not achievable in other ways, such as market and economic analysis, voice of customer, etc. This makes patents a strategic source for supporting CEOs in decision making activities. At present, worldwide patent database contains over 100 million of documents. The number of patent applications reached 3 million per year in 2016 and is globally raising.

The main drawback of application surge is the great amount of uncertainty they bring in any kind of patent analysis; indeed, their content is temporary and forecasting the grant and post grant proceedings outcome is difficult.

Despite the high relevance and practical consequences of the uncertainty to the procedural aspects of patent applications, only few works paid attention to this issue. Nevertheless, they did not give suggestions about tools or methods able to prevent or assess the level of uncertainty in patent proceeding, neither to support the applicant carrying out state-of-the-art patent analyses in presence of high share of patent applications.

Knowing that the outcomes generated by knowledge, like patents, are business products or productive assets, which can be exploited as economical goods, it is crucial for patent owners knowing the value of held patents to adopt the best exploitation strategy. Unfortunately, the tools and methods currently available for patent valuation experts are not designed to manage the risk due to this uncertain scenario. IP offices of firms, patent valuation experts of banks and other expert-in-the-field people work in a riskier environment than in the past. The main pitfalls are: the time consumption, the need of deep expertise of the appraiser and the request of reiteration at each content change. A methods improvement which aims to reduce the risk and increase the reliability of a patent search opinion and valuation is increasingly required.

Firstly, in this thesis specific and common cases of patent application content change are listed. For each, the available sources of information related to the risk reduction in writing an opinion or valuation are also specified. Furthermore, these suggestions can be useful to improve already existent tools for patent searches, e.g. by implementing new modules for warn the user about the revised versions of patent text and/or get an overview of the patent evolution along the procedural timeline.

Aside the content change, the uncertainty that a patent application brings in assessment of its value also refers to the likelihood of failure in reaching the grant. Currently, the most reliable judgement about the success rate of an application is the patent attorney’s

opinion, who is skilled in understanding the path followed by the application through the events of the granting process.

To improve the reliability of success rate judgement, this work develops the impact analysis of a set of parameters, proper of the grant and post grant proceedings of EP applications. The analysis has been carried out in grant and post grant scenarios separately.

In general, the results show that the success rate does not matter about the filing route chosen by applicant in filing the application with the EPO. In the grant proceeding, the Search Report citations, limited to the 1 X and 2 Y, are small hindrances to the success of the application. Otherwise, the effect of claims amendments and '*application deemed to be withdrawn'* event have big relevance. Relying on these results, the two events can be used as binary indicators of success rate shifts.

Contrariwise, in a post-grant proceeding scenario, previous terms lose their impact. The lone term relevant is the opposition filing from third parties that causes a dramatic decreasing in success rate.

To easily extract the information about the success rate of a patent application, the big amount of dataset was depicted using an infographic map. For each procedural path followed by EP applications, the map shows its average success rate.

A regression of success rate on the analyzed terms shows that they are not enough to explain the variance of the success rate and are not able to predict its value.

The patent valuation activity, and other kinds of patent analysis, must consider the proposal of invention into the relevant technological background, which is the collection of patent documents dealing with related matter. Knowing the relevant state-of-the-art and its business history, the analyst can easier assess the importance of the application itself and its value.

The complexity of both used language and classification system and the presence of sparse class assignments to the patent make the classical keyword-based approach too general, limiting the quality of the patent search and its results.

To enhance the relevance of patent search results, this thesis proposes a Business Intelligence based on a supervised method able to maximise both retrieval recall and precision. The core of the approach is composed by syntactic-semantic modules, called Syntactic Dependency Patterns (SDPs), which have been developed using a rigorous patent ontology. They allow the user to retrieve the most pertinent sentences introducing the *technical problem* described by patents. The choice of relevant sentences is

referred to the user. By this way, the patent families composing the state-of-the-art can be easier and faster collected and their relevance is ensured by the analyst who need to read only pertinent sentences. The SDPs retrieval approach was tested in injection moulding field with positive results.

Furthermore, the method proposed by this thesis build a patent-based Business Intelligence considering the elementary data about the patent family members, instead the single representative of a family. This can unveil detailed information that classical patent-based BIs overlook.

Finally, this thesis introduces a method that aims to extract quantitative data from patents to identify the structure of the reference market for a target technology field. It aims to enhance the innovation strategy for a firm that proposes a new technology, especially when it operates in a complex market.

The method extracts citations data from patent database by which computing triples and measuring their inner balance. These information are useful to assess the complexity of a technological market and unveil potential dominant positioning of players.

It has been tested in machine learning field and it shown a dominant positioning of Microsoft, IBM and Google. The experimentation proves and highlights the risk for an unaware entrant to incur in unexpected extra charge due to the hard licensing bargaining.

To sum up, this thesis analysis many hurdles on the course which anyone, also expert-in-the-field people, might meet working with patent applications instead of granted patents. To overcome the hurdles, strategies are suggested and tools able to compute risk indicators are proposed. This approach provides quantitative indicators able to support the professionals in assessing the actual reliability level of patent analysis carried out in high uncertainty fields.

This work can be a good starting point to improve patent valuation methods by including original risk indicators based on specific patent application features.

References

Alaskar, T. and Poulis, E. (2015) 'Business Intelligence Capabilities and Implementation Strategies', *International Journal of Global Business*, 8(1), pp. 34–45.

Alberts, D. *et al.* (2011) 'Introduction to Patent Searching-Practical Experience and Requirements for Searching the Patent Space', in Lupu, M. et al. (eds) *Current Challenges in Patent Information Retrieval*. Springer Berlin Heidelberg, pp. 3–43. doi: 10.1007/978-3-642-19231-9_1.

Anson, W. (2002) 'Valuing and monetizing intellectual property in bankruptcy', *Secured Lender*, 58(3), pp. 8–82.

Arora, A. (2001) 'Markets for Technology and their Implications for Corporate Strategy', *Industrial and Corporate Change*. Oxford University Press, 10(2), pp. 419–451. doi: 10.1093/icc/10.2.419.

Arora, A., Fosfuri, A. and Gambardella, A. (2001) 'Markets for Technology and their Implications for Corporate Strategy', *Industrial and Corporate Change*. Oxford University Press, 10(2), pp. 419–451. doi: 10.1093/icc/10.2.419.

Austin, D. H. (1993) 'An Event-Study Approach to Measuring Innovative Output: The Case of Biotechnology', *The American Economic Review*, 83(2), pp. 253–258.

Baglieri, E. *et al.* (2001) 'Evaluating Intangible Assets: The Measurement of R&D Performance', *SSRN Electronic Journal*. doi: 10.2139/ssrn.278260.

Bawa, R., Bawa, S. R. and Maebius, S. B. (2005) 'The nanotechnology patent "gold rush"', *Journal of Intellectual Property Rights*, 10(September), pp. 426–433.

Beatty, J. (2017) 'Speed of processing at the EPO'. Munich: European Patent Office.

Bergmann, I. *et al.* (2008) 'Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips', *R and D Management*. Blackwell Publishing Ltd, 38(5), pp. 550–562. doi: 10.1111/j.1467-9310.2008.00533.x.

Brannon, N. (2010) 'Business Intelligence and E-Discovery.', *Intellectual Property & Technology Law Journal*, 22(7), pp. 1–5.

Carrara, P. and Russo, D. (2017) 'Patent searches opinion: How to minimize the risk when reviewing patent applications', *World Patent Information*. Elsevier Ltd, 49, pp. 43–51. doi: 10.1016/j.wpi.2017.05.003.

Chatterji, D. (1996) 'Accessing External Sources of Technology', *Research-Technology Management*, 39(2), pp. 48–56. doi: 10.1080/08956308.1996.11671051.

Chatterji, D. and Manuel, T. A. (1993) 'Benefiting from External Sources of Technology', *Research-Technology Management*, 36(6), pp. 21–26. doi:

10.1080/08956308.1993.11670937.

Chiesa, V. *et al.* (2007) 'Searching for factors influencing technological asset value', *European Journal of Innovation Management*, 10(4), pp. 467–488. doi: 10.1108/14601060710828781.

Chiesa, V. *et al.* (2008) 'DETERMINING THE VALUE OF INTANGIBLE ASSETS — A STUDY AND AN EMPIRICAL APPLICATION', *International Journal of Innovation and Technology Management*. World Scientific Publishing Co. Pte. Ltd., 05(01), pp. 123–147. doi: 10.1142/S0219877008001278.

Chiesa, V., Gilardoni, E. and Manzini, R. (2005) 'The valuation of technology in buy-cooperate-sell decisions', *European Journal of Innovation Management*. Emerald Group Publishing Limited, 8(1), pp. 5–30. doi: 10.1108/14601060510578556.

Chiesa, V., Manzini, R. and Pizzurno, E. (2008) 'The market for technological intangibles: a conceptual framework for commercial transactions', *International Journal of Learning and Intellectual Capital*, 5(2), p. 186. doi: 10.1504/IJLIC.2008.020151.

Chiu, Y. J. and Chen, Y. W. (2007) 'Using AHP in patent valuation', *Mathematical and Computer Modelling*, 46(7–8), pp. 1054–1062. doi: 10.1016/j.mcm.2007.03.009.

Choi, J. D. and Palmer, M. (2012) 'Guidelines for the CLEAR Style Constituent to Dependency Conversions', *Center for Computational Language and EducAtion Research University of Colorado Boulder Institute of Cognitive Science Technical Report 01-12*, p. 53.

Clarkson, G. and DeKorte, D. (2006) 'The problem of patent thickets in convergent technologies', *Annals of the New York Academy of Sciences*. Wiley/Blackwell (10.1111), 1093(1), pp. 180–200. doi: 10.1196/annals.1382.014.

Cohen, W. M. *et al.* (2002) 'R&D spillovers, patents and the incentives to innovate in Japan and the United States', *Research Policy*, 31(8–9), pp. 1349–1367. doi: 10.1016/S0048-7333(02)00068-9.

D'Silva, J. (2009) 'Pools, Thickets and Open Source Nanotechnology', *European Intellectual Property Review*, p. 300.

Dedić, N. and Stanier, C. (2016) 'Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting', in. Springer, Cham, pp. 225–236. doi: 10.1007/978-3-319-49944-4_17.

Deng, Z., Lev, B. and Narin, F. (1999) 'Science and Technology as Predictors of Stock Performance', *Financial Analysts Journal*, 55(3), pp. 20–32. doi: 10.2469/faj.v55.n3.2269.

van der Drift, J. (1988) 'Statistics of European patents on legal status and granting data', *World Patent Information*. Pergamon, 10(4), pp. 243–249. doi: 10.1016/0172-

2190(88)90275-X.

Drucker, P. F. (1969) 'The Age of Discontinuity; Guidelines to Our Changing Society', *Management Science*. Harper and Row, Publishers, Inc., pp. B153–B154. doi: 10.1016/B978-0-434-90395-5.50019-5.

Ernst, H., Leptien, C. and Vitt, J. (2000) 'Inventors are not alike: the distribution of patenting output among industrial R&D personnel', *IEEE Transactions on Engineering Management*, 47(2), pp. 184–199. doi: 10.1109/17.846786.

Ernst, H. and Omland, N. (2011) 'The Patent Asset Index – A new approach to benchmark patent portfolios', *World Patent Information*. Pergamon, 33(1), pp. 34–41. doi: 10.1016/J.WPI.2010.08.008.

European Patent Office (2017a) 'Data Catalog PATSTAT Biblio'.

European Patent Office (2017b) 'Data Catalog PATSTAT Register', pp. 1–153.

European Patent Office (2017c) *Patent families at the EPO*.

Fenton, B. G. M. (2016) 'Application of Enterprise Risk Management (ERM) Principles To Patent Freedom To Operate (FTO) Analysis: A Novel "IP-RM" System', *Journal of the Licensing Executives Society*, 4(December), pp. 246–255.

Fischer, T. and Leidinger, J. (2014) 'Testing patent value indicators on directly observed patent value—An empirical analysis of Ocean Tomo patent auctions', *Research Policy*. North-Holland, 43(3), pp. 519–529. doi: 10.1016/J.RESPOL.2013.07.013.

Frattini, F., Bianchi, M. and Franzò, S. (2018) 'Identifying Factors Influencing the Price of Technology Licenses: a Framework Grounded in Negotiation Research', *International Journal of Innovation Management*. World Scientific Publishing Company, p. 1950039. doi: 10.1142/s1363919619500397.

Frietsch, R., Neuhäusler, P. and Rothengatter, O. (2013) 'Which road to take? Filing routes to the European Patent Office', *World Patent Information*. Pergamon, 35(1), pp. 8–19. doi: 10.1016/j.wpi.2012.09.011.

Galasso, A. and Schankerman, M. (2010a) 'Patent thickets, courts, and the market for innovation', *RAND Journal of Economics*, 41(3), pp. 472–503. doi: 10.1111/j.1756-2171.2010.00108.x.

Galasso, A. and Schankerman, M. (2010b) 'Patent thickets, courts, and the market for innovation', *The RAND Journal of Economics*, 41(3), pp. 472–503. doi: 10.1111/j.1756-2171.2010.00108.x.

Gallini, N. (2017) 'Do patents work? Thickets, trolls and antibiotic resistance', *Canadian Journal of Economics*, 50(4), pp. 893–926. doi: 10.1111/caje.12312.

Gans, J. S. and Hsu, D. H. (2008) 'The Impact of Uncertain Intellectual Property

Rights on the Market for Ideas: Evidence from Patent Grant Delays', *Management Science*, 54(5), pp. 982–997. doi: 10.1287/mnsc.1070.0814.

Von Graevenitz, G., Wagner, S. and Harhoff, D. (2011) 'How to measure patent thickets-A novel approach', *Economics Letters*. Elsevier B.V., 111(1), pp. 6–9. doi: 10.1016/j.econlet.2010.12.005.

Von Graevenitz, G., Wagner, S. and Harhoff, D. (2013) 'Incidence and growth of patent thickets: The impact of technological opportunities and complexity', *Journal of Industrial Economics*, 61(3), pp. 521–563. doi: 10.1111/joie.12032.

Grandjean, N. *et al.* (2005) 'Competitive intelligence and patent analysis in drug discovery.', *Drug discovery today. Technologies*, 2(3), pp. 211–5. doi: 10.1016/j.ddtec.2005.08.007.

Grimaldi, M. *et al.* (2015) 'The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning', *Technological Forecasting and Social Change*. North-Holland, 94, pp. 286–302. doi: 10.1016/J.TECHFORE.2014.10.013.

Grimpe, C. and Hussinger, K. (2014) 'Pre-empted patents, infringed patents and firms' participation in markets for technology', *Research Policy*, 43(3), pp. 543–554. doi: 10.1016/j.respol.2013.12.001.

Grindley, P. C. and Teece, D. J. (1997) 'Managing Intellectual Capital: Licensing and Cross-Licensing in Semiconductors and Electronics', *California Management Review*. SAGE PublicationsSage CA: Los Angeles, CA, 39(2), pp. 8–41. doi: 10.2307/41165885.

Guellec, D. and van Pottelsberghe de la Potterie, B. (2000) 'Applications, grants and the value of patent', *Economics Letters*, 69(1), pp. 109–114. doi: 10.1016/S0165-1765(00)00265-2.

Hall, B. H., Jaffe, A. B. and Trajtenberg, M. (2005) 'Market Value and Patent Citations', *RAND Journal of Economics*, 36(1), pp. 16–38. doi: 10.1007/s00216-009-2643-x.

Hall, B., Helmers, C. and von Graevenitz, G. (2015) *Technology Entry in the Presence of Patent Thickets*. 21455. Cambridge, MA. doi: 10.3386/w21455.

Haour, G. (2004) *Resolving the Innovation Paradox*. London: Palgrave Macmillan UK. doi: 10.1057/9780230510555.

Hargreaves, I. (2011) *Digital Opportunity - A Review of Intellectual Property and Growth*. doi: 10.1177/0266666907079076.

Harhoff, D. and Reitzig, M. (2004) 'Determinants of opposition against EPO patent grants - The case of biotechnology and pharmaceuticals', *International Journal of Industrial Organization*, 22(4), pp. 443–480. doi: 10.1016/j.ijindorg.2004.01.001.

Harhoff, D., Scherer, F. M. and Vopel, K. (2003) 'Citations, family size, opposition and the value of patent rights', *Research Policy*. North-Holland, 32(8), pp. 1343–1363. doi: 10.1016/S0048-7333(02)00124-5.

Harhoff, D. and Wagner, S. (2009) 'The Duration of Patent Examination at the European Patent Office', *Management Science*, 55(12), pp. 1969–1984. doi: 10.1287/mnsc.1090.1069.

Hsieh, C.-H. (2013) 'Patent value assessment and commercialization strategy', *Technological Forecasting and Social Change*. North-Holland, 80(2), pp. 307–319. doi: 10.1016/J.TECHFORE.2012.09.014.

Hung, S.-W. and Tseng, S.-C. (2010) 'A New Framework Integrating Environmental Effects into Technology Evaluation', *Journal of Business Ethics*. Springer Netherlands, 95(4), pp. 543–556. doi: 10.1007/s10551-010-0439-8.

Hunt, D., Nguyen, L. and Rodgers, M. (2012) *Patent Searching: Tools & Techniques*. Edited by D. Hunt, L. Nguyen, and M. Rodgers. Wiley.

Ijichi, T., Yoda, T. and HIRASAWA, R. (1994) 'Mapping R&D Network Dynamics: Analysis of the Development of Co-author and Co-inventor Relations', *The Journal of Science Policy and Research Management*. Japan Society for Research Policy and Innovation Management, 8(3_4), pp. 263–275. doi: 10.20801/jsrpim.8.3_4_263.

Jerak, A. and Wagner, S. (2006) 'Modeling probabilities of patent oppositions in a Bayesian semiparametric regression framework', *Empirical Economics*. Wiesbaden: DUV, 31(2), pp. 9–32. doi: 10.1007/s00181-005-0047-0.

Johansson, R. (2008) *Dependency-based semantic analysis of natural-language text*. Lund University Department of Computer Science.

Jones, T. *et al.* (2002) 'Strategies for maximizing value from intellectual capital in a technology-driven business', *The Licensing Journal*, (July), pp. 1–6.

Kamiyama, S., Sheehan, J. and Martinez, C. (2006) *Valuation and Exploitation of Intellectual Property*, *OECD Science, Technology and Industry Working Papers*. Paris: OECD Publishing. doi: 10.1787/307034817055.

Kang, B. and Tarasconi, G. (2016) 'PATSTAT revisited: Suggestions for better usage', *World Patent Information*. Pergamon, 46, pp. 56–63. doi: 10.1016/j.wpi.2016.06.001.

Khode, A. and Jambhorkar, S. (2017) 'A Literature Review on Patent Information Retrieval Techniques', *Indian Journal of Science and Technology*, 10(36), pp. 1–13. doi: 10.17485/ijst/2017/v10i37/116435.

Kryder, R. D., Kowalski, S. P. and Krattiger, A. F. (2000) 'The Intellectual and Technical Property Components of pro-Vitamin A Rice (GoldenRiceTM): A Preliminary Freedom-To-Operate Review', *ISAAA Briefs*, (20), p. 56.

Lee, J. H. (2017) *Predicting Bad Patents*. University of California, Berkeley.

Lemley, M. A. (2001) 'Rational Ignorance at the Patent Office', *SSRN Electronic Journal*. doi: 10.2139/ssrn.261400.

Lemley, M. A. and Shapiro, C. (2005) 'Probabilistic Patents', *Journal of Economic Perspectives*, 19(2), pp. 75–98. doi: 10.1257/0895330054048650.

Lichtenthaler, U. (2007) 'The Drivers of Technology Licensing: An Industry Comparison', *California Management Review*, 49(4), pp. 67–89. doi: 10.2307/41166406.

List, J. (2010) 'An A to X of patent citations for searching', *World Patent Information*, 32(4), pp. 306–312. doi: 10.1016/j.wpi.2010.01.004.

Liu, X. *et al.* (2017) 'On Predictive Patent Valuation: Forecasting Patent Citations and Their Types', *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) On*, pp. 1438–1444.

Livotov, P. (2015) 'Using patent information for identification of new product features with high market potential', in *Procedia Engineering*. Elsevier B.V., pp. 1157–1164. doi: 10.1016/j.proeng.2015.12.438.

Lupu, M. *et al.* (2011) *Current Challenges in Patent Information Retrieval*, *Information Retrieval*. Edited by M. Lupu et al. Berlin, Heidelberg: Springer Berlin Heidelberg (The Information Retrieval Series). doi: 10.1007/978-3-642-19231-9.

Marchand, M. and Raymond, L. (2008) 'Researching performance measurement systems', *International Journal of Operations & Production Management*, 28(7), pp. 663–686. doi: 10.1108/01443570810881802.

de Marneffe, M.-C. and Manning, C. D. (2008) 'The Stanford typed dependencies representation', in *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pp. 1–8. doi: 10.3115/1608858.1608859.

Martinez, C. (2010) *Insight into Different Types of Patent Families*, *OECD STI Working Papers*. doi: 10.1787/5kml97dr6ptl-en.

Martínez, C., Martinez, C. and Martínez, C. (2011) 'Patent families: When do different definitions really matter?', *Scientometrics*, 86(1), pp. 39–63. doi: 10.1007/s11192-010-0251-3.

McAleer, M. and Slottje, D. (2005) 'A new measure of innovation: The patent success ratio', *Scientometrics*, 63(3), pp. 421–429. doi: 10.1007/s11192-005-0222-2.

Michel, J. and Bettels, B. (2001) 'Patent citation analysis: A close look at the basic input data from patent research reports', *Scientometrics*. Kluwer Academic Publishers, 51(1), pp. 185–201. doi: 10.1023/A:1010577030871.

Montecchi, T. and Russo, D. (2011) 'FBOS : Function / Behaviour – Oriented Search', *TRIZ Future 2011, Dublin, Ireland*, pp. 165–175.

Montecchi, T. and Russo, D. (2015) 'Dynamic and Semantic Database of Effects for Technology Transfer', *TRIZ Future Conference 2015*, 1(2), pp. 1–5.

Munari, F. (2012) 'La valutazione dei brevetti : metodi ed applicazioni'.

Munari, F. and Oriani, R. (2011) 'When, why and how to value patents? An introduction', in *The Economic Valuation of Patents: Methods and Applications*, pp. 3–24. doi: 10.4337/9780857936516.00007.

Nagaoka, S., Motohashi, K. and Goto, A. (2010) 'Patent statistics as an innovation indicator', *Handbook of the Economics of Innovation*, 2(1), pp. 1083–1127. doi: 10.1016/S0169-7218(10)02009-5.

Neifeld, R. A. (2001) 'A Macro-Economic Model Providing Patent Valuation and Patent Based Company Financial Indicators', *Journal of the Patent and Trademark Office Society*, 83.

OECD (1994) *The Measurement of Scientific and Technological Activities Using Patent Data as Science and Technology Indicators*, *Ocde/Gd(94)114*. OECD Publishing. doi: 10.1787/9789264065574-en.

OECD (2011) *ISIC REV. 3 Technology intensity definition, Economic Analysis*.

Otsuyama, H. (2003) 'Patent valuation and intellectual assets management', in Samejima, M. (ed.) *Patent strategy handbook*. Chuokeizai. Tokyo, p. 175.

Park, Y. and Park, G. (2004) 'A new method for technology valuation in monetary value: procedure and application', *Technovation*. Elsevier, 24(5), pp. 387–394. doi: 10.1016/S0166-4972(02)00099-8.

Patent Information News (2010) 'Divisional applications at the EPO', June, pp. 4–5.

Pitkethly, R. (1997) 'THE VALUATION OF PATENTS : A review of patent valuation methods with consideration of option based methods and the potential for further research', *New Developments in Intellectual Property : Law and Economics*, (March), pp. 1–32.

Powell, W. W. and Snellman, K. (2004) 'The Knowledge Economy', *Annual Review of Sociology*, 30(1), pp. 199–220. doi: 10.1146/annurev.soc.29.010202.100037.

Quillen, C. D. and Webster, O. H. (2001) 'Continuing patent applications and performance of the U.S. Patent and Trademark Office', *Federal Circuit Bar Journal*, 11(1), pp. 1–21.

de Rassenfosse, G., Jaffe, A. and Webster, E. (2016) *Low-quality Patents in the Eye of the Beholder: Evidence from Multiple Examiners*, *NBER WORKING PAPER SERIES*. Cambridge, MA. doi: 10.3386/w22244.

Reilly, R. and Schweihs, R. (1998) *Valuing intangible assets*, *M*. McGraw Hill Professional.

Reitzig, M. (2004) ‘Improving patent valuations for management purposes—validating new indicators by analyzing application rationales’, *Research Policy*. North-Holland, 33(6–7), pp. 939–957. doi: 10.1016/j.respol.2004.02.004.

Roberts, E. B. (2001) ‘Benchmarking Global Strategic Management of Technology’, *Research-Technology Management*, 44(2), pp. 25–36. doi: 10.1080/08956308.2001.11671416.

Rouach, D. and Santi, P. (2001) ‘Competitive Intelligence Adds Value: Five Intelligence Attitudes’, *European Management Journal*, 19(5), pp. 552–559. doi: 10.1016/S0263-2373(01)00069-X.

Rubinfeld, D. and Maness, R. (2004) ‘The strategic use of patents: implications for antitrust’, *Antitrust, patents and copyright - EU and US perspectives*, (January 2004), pp. 85–102.

Rud, O. P. (2009) *Business intelligence success factors: Tools for aligning your business in the global economy., Business intelligence success factors: Tools for aligning your business in the global economy*. Joh Wiley & Sons. doi: 10.1109/MSR.2009.5069473.

Russo, D., Montecchi, T. and Duci, S. (2015) ‘A new benchmark method for the identification and assessment of alternative technologies’, in *Fourth Scientific Meeting EuNetAir*, pp. 59–62. doi: 10.5162/4EuNetAir2015/16.

Sabety, T. (2005) ‘Nanotechnology innovation and the patent thicket: Which IP policies promote growth’, *Albany Law Journal of Science & Technology*, 15, p. 477.

Sakthika, R. I. and Kogilavani, S. V. (2015) ‘An integration of triple parser with predicate based approach for search engine development’, in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*. IEEE, pp. 553–557. doi: 10.1109/ECS.2015.7124969.

Salton, G. and McGill, M. J. (1983) *Introduction to Modern Information Retrieval*. McGraw-Hill.

Schankerman, M. (1998) ‘How Valuable is Patent Protection? Estimates by Technology Field’, *The RAND Journal of Economics*, 29(1), p. 77. doi: 10.2307/2555817.

Schmoch, U. (2008) *Concept of a technology classification for Country comparison*. Karlsruhe: Fraunhofer Institute for Systems and Innovation Research.

Schultz, M. and Madigan, K. (2016) *The Long Wait for Innovation: The Global Patent Pendency Problem*. Center for the Protection of Intellectual Property.

Shapiro, C. (2001) ‘Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting’, in Jaffe, A. B., Lerner, J., and Stern, S. (eds) *Innovation Policy and the Economy*. MIT Press, pp. 119–150. doi: 10.2139/ssrn.273550.

Shih, M.-J., Liu, D.-R. and Hsu, M.-L. (2008) 'Mining Changes in Patent Trends for Competitive', in *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 999–1005. doi: 10.1007/978-3-540-68125-0_105.

Shih, M.-J., Liu, D.-R. and Hsu, M.-L. (2010) 'Discovering competitive intelligence by mining changes in patent trends', *Expert Systems with Applications*, 37(4), pp. 2882–2890. doi: 10.1016/j.eswa.2009.09.001.

Simmons, E. S. (2009) '"Black sheep" in the patent family', *World Patent Information*, 31(1), pp. 11–18. doi: 10.1016/j.wpi.2008.08.005.

Simmons, E. S. and Spahl, B. D. (2000) 'Of submarines and interference: legal status changes following citation of an earlier US patent or patent application under 35 USC §102 (e)', *World Patent Information*, 22(3), pp. 191–203. doi: 10.1016/S0172-2190(00)00046-6.

Sternitzke, C. (2009) 'Reducing uncertainty in the patent application procedure – Insights from invalidating prior art in European patent applications', *World Patent Information*, 31(1), pp. 48–53. doi: 10.1016/j.wpi.2008.04.007.

Suzuki, J. (2011) 'Structural modeling of the value of patent', *Research Policy*. North-Holland, 40(7), pp. 986–1000. doi: 10.1016/j.respol.2011.05.006.

Tarasconi, G. (2018) 'EP register for patent data analysis'.

Tegernsee Experts Group (2012) *Treatment of conflicting applications*. Munich.

Thompson, M. J. (2016) 'Measuring patent quality: A claim and search report approach', *World Patent Information*. Pergamon, 45, pp. 47–54. doi: 10.1016/j.wpi.2016.03.003.

Trippe, A. (2015) *Guidelines for Preparing Patent Landscape Reports*, *Wipo*.

Walsh, J. P., Lee, Y.-N. and Jung, T. (2016) 'Win, lose or draw? The fate of patented inventions', *Research Policy*. North-Holland, 45(7), pp. 1362–1373. doi: 10.1016/j.respol.2016.03.020.

Wang, F., Lin, L. F. and Yang, Z. (2013) 'An Ontology-Based Automatic Semantic Annotation Approach for Patent Document Retrieval in Product Innovation Design', *Applied Mechanics and Materials*, 446–447, pp. 1581–1590. doi: 10.4028/www.scientific.net/AMM.446-447.1581.

Wang, J., Lu, W. F. and Loh, H. T. (2015) 'A two-level parser for patent claim parsing', *Advanced Engineering Informatics*. Elsevier Ltd, 29(3), pp. 431–439. doi: 10.1016/j.aei.2015.01.013.

Wang, Y., Berant, J. and Liang, P. (2015) 'Building a Semantic Parser Overnight', in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language*

Processing (Volume 1: Long Papers), pp. 1332–1342. doi: 10.3115/v1/P15-1129.

WIPO (2004) *PCT International Search and Preliminary Examination Guidelines*.

WIPO (2015) 'Introduction to the Patent Cooperation Treaty Preface', *PCT Distance Learning Course Introduction*. World Intellectual Property Organization.

WIPO (2017) *World Intellectual Property Indicators*, *World Intellectual Property Indicators Series*. Geneva.

Yang, D. (2008) 'Pendency and grant ratios of invention patents: A comparative study of the US and China', *Research Policy*, 37(6–7), pp. 1035–1046. doi: 10.1016/j.respol.2008.03.008.

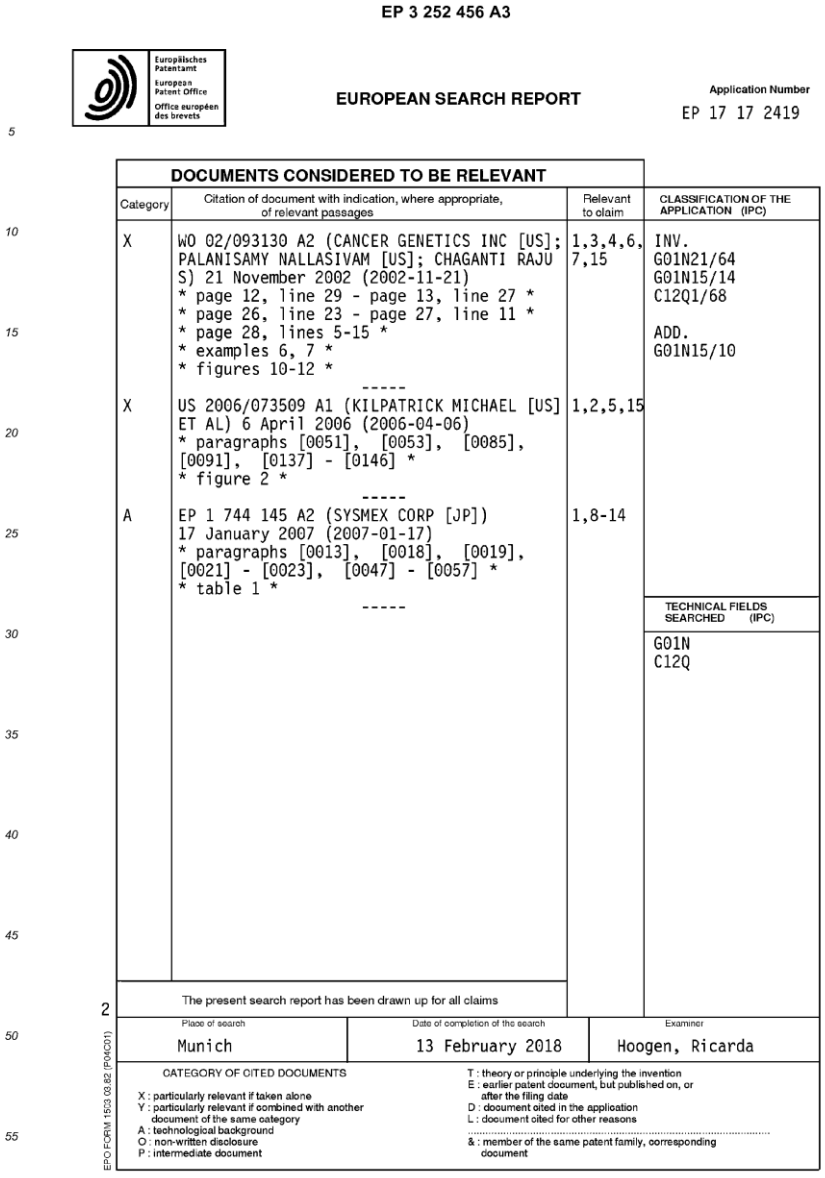
Zanasi, A. (1998) 'Competitive intelligence through data mining public sources', *Competitive Intelligence Review*, 9(1), pp. 44–54. doi: 10.1002/(SICI)1520-6386(199801/03)9:1<44::AID-CIR8>3.0.CO;2-A.

van Zeebroeck, N. (2011) 'The puzzle of patent value indicators', *Economics of Innovation and New Technology*. Routledge , 20(1), pp. 33–62. doi: 10.1080/10438590903038256.

van Zeebroeck, N. and van Pottelsberghe de la Potterie, B. (2011) 'Filing strategies and patent value', *Economics of Innovation and New Technology*. Routledge , 20(6), pp. 539–561. doi: 10.1080/10438591003668646.

APPENDIXES

A.1 Examples of Search Report



EP 3 252 456 A3

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 17 17 2419

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	WO 02/093130 A2 (CANCER GENETICS INC [US]; PALANISAMY NALLASIVAM [US]; CHAGANTI RAJU S) 21 November 2002 (2002-11-21) * page 12, line 29 - page 13, line 27 * * page 26, line 23 - page 27, line 11 * * page 28, lines 5-15 * * examples 6, 7 * * figures 10-12 *	1,3,4,6, 7,15	INV. G01N21/64 G01N15/14 C12Q1/68 ADD. G01N15/10
X	US 2006/073509 A1 (KILPATRICK MICHAEL [US] ET AL) 6 April 2006 (2006-04-06) * paragraphs [0051], [0053], [0085], [0091], [0137] - [0146] * * figure 2 *	1,2,5,15	
A	EP 1 744 145 A2 (SYSMEX CORP [JP]) 17 January 2007 (2007-01-17) * paragraphs [0013], [0018], [0019], [0021] - [0023], [0047] - [0057] * * table 1 *	1,8-14	

TECHNICAL FIELDS SEARCHED (IPC)
G01N
C12Q

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	13 February 2018	Hoogen, Ricarda

CATEGORY OF CITED DOCUMENTS
X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Figure 39: example of European Search Report

Figure 39, on the left, shows the column with the citation codes. Near, in second column, there are the reference to the cited documents (patent or non-patent). The column on the right side indicates the claims for which the citation is relevant. Even on the

right side, outside the main table, there are the classification codes assigned to the application and the classes in which the search has been performed.

INTERNATIONAL SEARCH REPORT

In. ational Application No
PCT/IB 98/00159

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 A63B21/008

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A63B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 123 641 A (ABBOUDI ET AL.) 23 June 1992	1
Y	see column 5, line 55 - column 6, line 50; figures 4-7	2-4
Y	US 5 354 253 A (AWBREY ET AL.) 11 October 1994	2-4
A	see column 4, line 12 - line 41; claims 1,2; figure 1A	5
A	FR 2 347 062 A (METAIS ET AL.) 4 November 1977 see claims; figures	1,5
X,P	EP 0 803 270 A (SCOTT) 29 October 1997 see claim 16; figures	1

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 August 1998	10/08/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Grunfeld, M

1

Form PCT/ISA/210 (second sheet) (July 1992)

Figure 40: example of Search Report with 2 Y citation but listed as single Y citation in PATSTAT

Figure 40 show a Search Report having double citations. The application has backward citation of both kinds X and Y which point to the patent US5123641. Y one, which is no aligned with the patent number was not automatically read, then PATSTAT Register collects the search report with a single Y. Notice that the same situation holds for a code on patent US5354253, but it is not a backward citation.

A.2 Patent success rate infographic

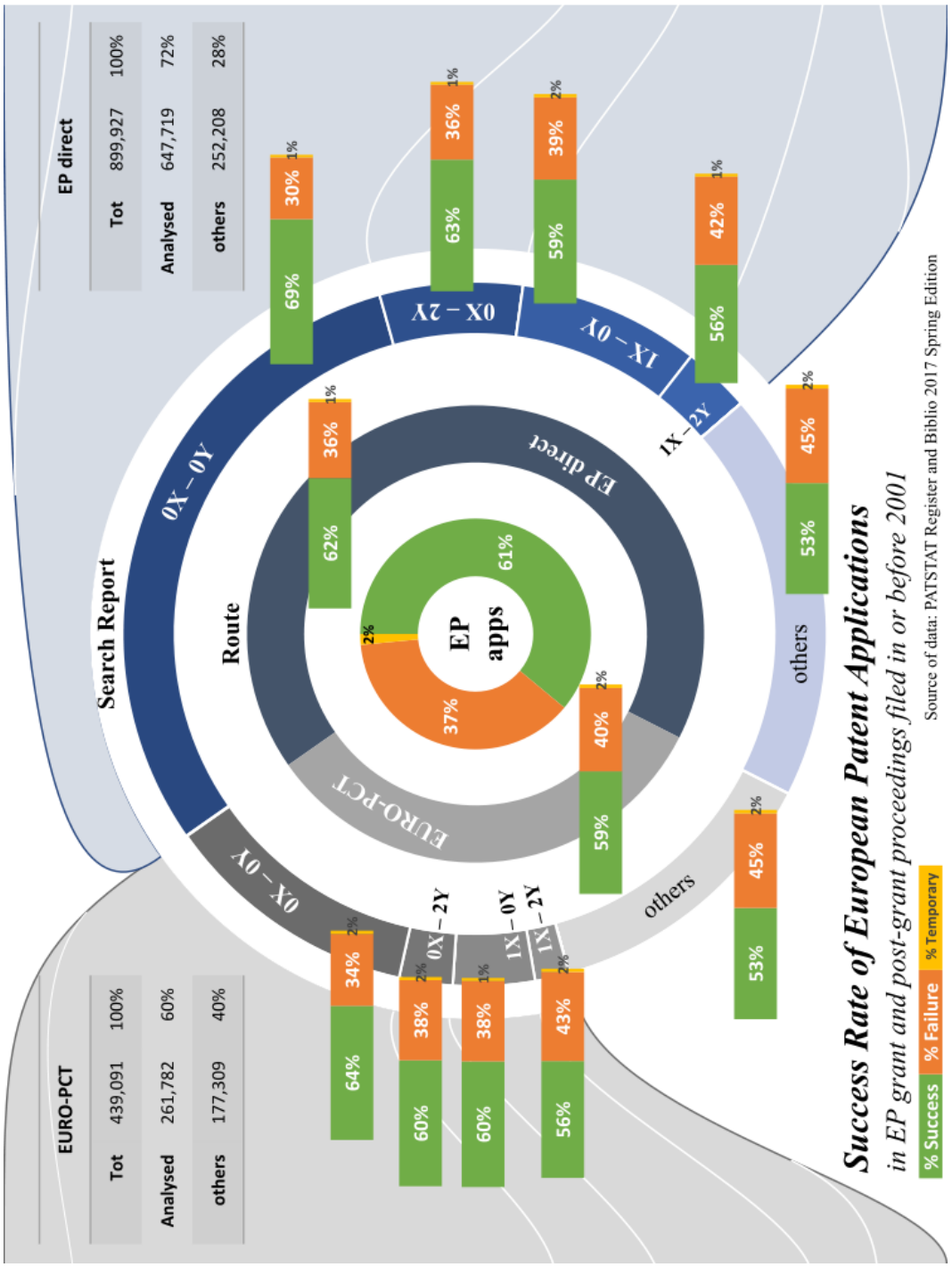


Figure 41: Patent success rate infographic, central section

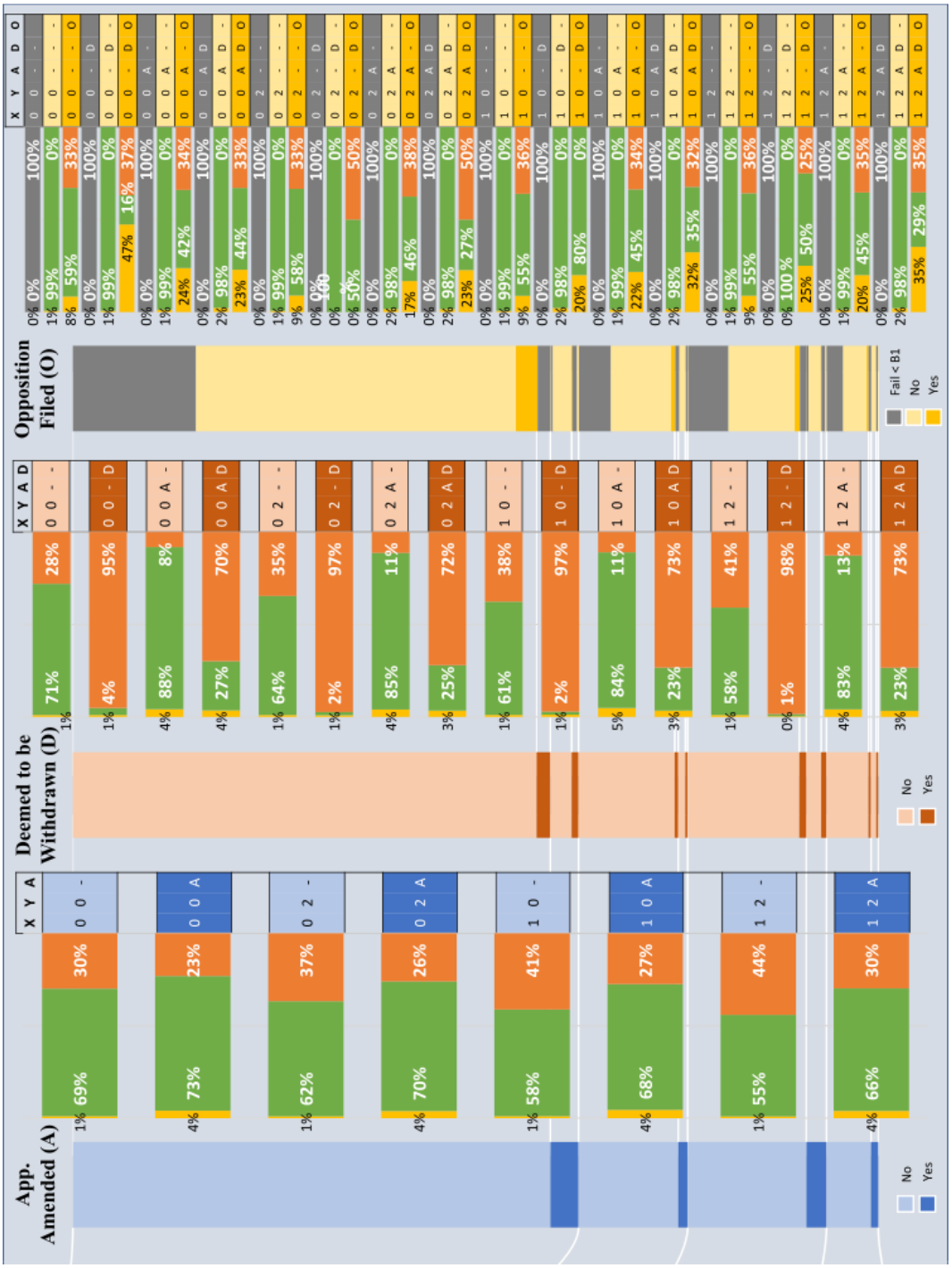


Figure 42: patent success rate infographic, left (EP) section

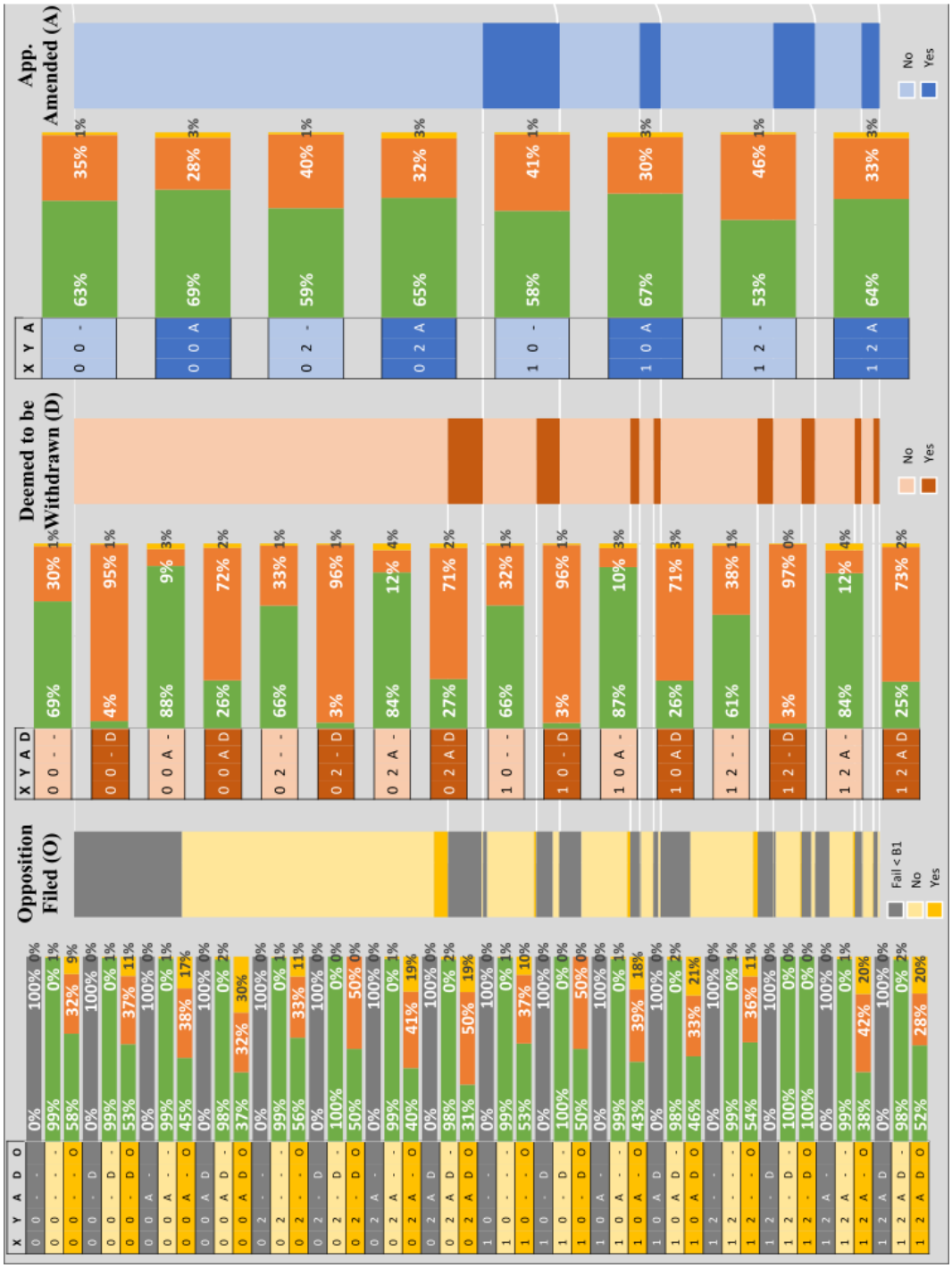


Figure 43: patent success rate, right (PCT) section