The Patent Lawyer

GLOBAL REACH, LOCAL KNOWLEDGE

www.patentlawyermagazine.com

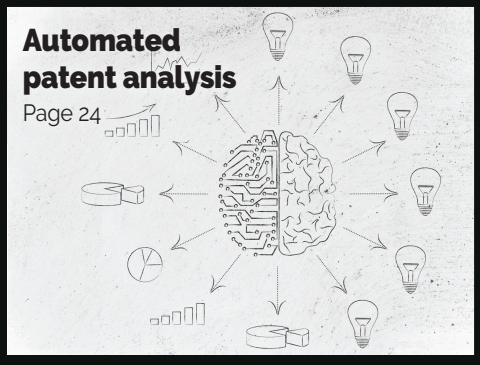
May / June 2021

Augmenting your IP portfolio is virtually the only way to compete in AR/VR

SOUTH AMERICA Law firm RANKINGS

Finnegan, Henderson, Farabow, Garrett & Dunner, LLP experts Christopher Howes, Zachery Olah, Forrest Jones, and Karthik Kumar, discuss the developments in the augmented and virtual reality sphere with advice for protecting innovation.

Ismat Levin, Synamedia
Page 12



Automated patent analysis
Page 24

Indefiniteness
Page 60

CTC Legal Media

Why automated patent analysis can be wrong, even when it's right

Axel Contreras-Alvarez, former IPR Commercialization Manager at Ericsson, evaluates the reliability of software and automated analysis for patent valuation, considering the factors used by algorithms, and with an action plan for those wishing to use such platforms.

Connectivity is changing our communication dynamics. Long gone are the days when the only function of a phone was to make calls. Nowadays, we can share pictures and videos with our friends, calculate our time of arrival to any destination, order food and get it delivered to our door, join business conferences, etc. Connectivity is also spreading to all kinds of devices, not only our phones are connected but also the lights, the TV, the coffee machine, and the cars[1], to name a few. These complex fast-developing cutting-edge technologies enabling connectivity are the result of years of work and heavy R&D investments in innovation. This is why they are typically protected by patents, which may be licensed. We live in an era of technology licensing.

Both licensors and licensees understand that a timely valuation of the technologies is crucial, in particular for the negotiation of the licenses. With the recent boom of artificial intelligence (AI), attention has been drawn to analytics and other automated methods to help in the generation of said information. This has created a market for computer platforms which offer automated patent analysis and valuation.

Axel Contreras-Alvarez

Résumé

Hector Axel Contreras is former IPR Commercialization Manager at Ericsson

Hector is currently focused in technology standardization and patent licensing in the telecommunications sector. Has a background in informatics and a master's degree in Intellectual Property and Competition Law at the Munich Intellectual Property Law Center, where he is now a tutor for the current students. Also acquired experience as entrepreneur, patent agent and lecturer. Passionate for new technologies, digital transformation, and cycling.

> **We live in an era of technology licensing.**

Automated patent valuation platforms can provide quick results, and the process is also less expensive compared to human expert analysis, but its reliability is questionable. In fact, although the software for automated analysis has experienced a fast-paced development in the recent years, its implementation in the legal field has been limited, mainly because this requires legal knowledge and other forms of advanced abstract thinking which the algorithms are not yet capable of.

Nevertheless, such platforms are already being used in the patent licensing environment, a context which demands both technical and legal proficiency. Moreover, the information that these platforms generate is accessible not only to industry stakeholders but also to other relevant factors such as policy makers. For these reasons, it is important that both users and service-providers are aware of the benefits, the capabilities, and the limitations of these tools[2]. Equally relevant would be to understand how the information provided by the platforms should be interpreted. In order to answer this, we need to look at how the results are being obtained.

Patent factorization

For the first step, we need to ask the following: *what data do these platforms analyse?* We often hear terms such as analytics, text mining, and

[1] https://www.bbc.com/news/av/business-30786714

[2] https://papers.ssrn.com/abstract=3658631

big data. In a broad sense, they all refer to the acquisition of measurable information from the patents. Patents are publicly accessible, but a computer cannot understand them as they are. It is necessary to break down the patent into a group of less complex pieces of information which can be individually assessed. I refer to these pieces of information as factors.

Factors are not new to the field; they have been used to benchmark patents even before the automated platforms appeared (for instance, by Larry M. Goldstein). But nowadays, factors are used to simplify the patents in order for a computer to be able to analyse them. It is worth noting that not all platforms use the same factors. Some platforms use more or less factors than others, and some even come up with new factors by further subdivision or combination of known factors. In all cases, it is important for the users to pay attention to which factors are used by the platform at hand. An example of a factor is the number of 'backward citations' in a patent, that is, the number of other patents that are cited in a specific patent. When referring to the number of other patents which cite the patent in question, the factor is called 'forward citations'. Another factor is the so-called 'grant lag', which refers to the number of days elapsed since the application date of the patent until the grant date. There are many examples of factors, the most popular ones are listed right.

Computers can quickly analyse these factors from a multitude of patents. Sums, comparisons, averages, and trends can be obtained from portfolios or even entire technical fields in a matter of seconds. Errors in the computation are close to non-existent thanks to the power of computers. All the obtained information can be useful and, as previously mentioned, the factors have been used as benchmarks for several years already. The recent increased efficiency of software and hardware makes it possible to extract those factors from multiple patents in a shorter time and then analyse them according to the platforms' algorithms. However, the analysis is not about entire patents with context and details, but about the simplified factors. The idea behind is that those factors may be able to tell us whether the patent in question is valuable or not.

Factor	Definition as frequently used by the platforms
Backward citations	Number of patents cited as prior art in a patent document.
Citations to non-patent literature	Number of scientific sources (non-patents) cited as prior art in a patent document.
Forward citations	Number of citations received from later patents.
Claims	Number of claims in a granted patent.
Essentiality Declarations	Identifies whether the patent has been declared as potentially essential towards a Standards Development Organization (SDO).
Family size	Number of granted or pending applications which share a common priority application (usually in different countries).
Grant lag	Number of days elapsed between the filing and the granting of a patent.
Legal status	Specifies whether the patent is pending, granted, lapsed, or abandoned.
Licensing	Whether the patent has been previously licensed.
Opposition or Litigation	Indicates if a patent has survived opposition or litigation procedures.
Ownership	Rating of a patent based on who the holder is and whether there have been ownership changes.
Term	Number of days left until the patent's 20-year lifecycle lapses.
Technological scope	Number of distinct four-digit IPC subclasses in which a patent was classified by its corresponding patent office.

Searching for value

After identifying the factors, the second important question is: *are those factors truly indicators of patent value?* To answer this, we need to look at how they are interpreted. One obstacle is that, in some cases, it is challenging to know how the platforms are processing the information because it is protected as a trade secret, thus the description of the process is not accessible to users.

Even when the company providing the platform is aiming for transparency, they frequently present the factors as direct indicators of value (or lack of it) based on simplified, and sometimes-undisclosed, assumptions. As an example, when analysing the backward citations some IP specialists argue that a high number may reflect that the owner of the patent possesses expertise in the field. It may also reflect that the invention is an incremental innovation of a market-proven

technology, which may make those inventions valuable. However, other IP experts argue that a high number of backward citations may also mean that the patented invention was created in a field where there is more competition and this would result in less market opportunities, making the patent less valuable. In real world, as these examples show, the relation between the factors and the real value of the patent is usually more complex than what the platforms typically assume.

The discrepancies in how the factors are interpreted affect the results calculated by the platforms. In other words, even when the computing is correctly done, the theoretical assumptions behind it might be inaccurate. This is evident with simple numerical factors, as is the case of citations, and further issues arise with more complex factors such as essentiality for standard essential patents (SEPs), which no platform has been able to address. Current platforms that try to use essentiality as a factor, rely on declarations of potential essentiality made by the patent owners towards standard development organizations. These self-proclaimed declarations reflect only potentially essential patents and patent applications[3], which do not necessarily result in a SEP. A genuine essentiality check involves over 40 hours of expert analysis per patent, and current technology has not been able to perform it. Thus, automated platforms offering data and analysis on 5G standards[4] are likely to provide inaccurate results.

“Thus, automated platforms offering data and analysis on 5G standards are likely to provide inaccurate results.”

[3] https://www.iam-media.com/frandseps/transparency-iot-licensing

[4] https://www.iam-media.com/frandseps/5g-decision-making-full-facts

Some of the factors can be directly obtained from patent offices or the patent documents themselves, but other factors require external information which is not always accessible. The availability of this external information represents another obstacle in the correct assessment of the patents because there may be time - or quantity - constraints. For instance, in the case of 'forward citations', a patent that was published a long time ago will be more often cited than a more recently granted patent. As a result, the score of earlier-published patents would be overinflated, while the value of most recent innovations would not be accurately reflected. Another example is the 'opposition or litigation' factor. On the one hand, the fact that a patent survived opposition/litigation can be a reliable indicator of value. On the other hand, there is only a limited amount of data related to this factor because, in practice, very few patents are opposed or litigated. Therefore, the impact of these factors on the results is limited.

Machine-learning approaches have been equally unable to provide consistently favorable results because of two main reasons. First, there is not enough information publicly available to create a reliable dataset about patent value, the information is scarce and difficult to gather. The second issue is that patents are, by definition, technically unique and they are shaped by the peculiarities of jurisdiction-specific laws and regulations. The results which can be obtained

Factor	Usability as value indicator	Availability of information
Backward citations	No	Good
Citations to non-patent literature	No	Limited
Forward citations	Context-specific	Delayed
Claims	No	Good
Essentiality Declarations	Weak, as it only considers a declaration of potential essentiality towards an SDO	Very Limited
Family size	Strong (But relies on previous assessment by the owner)	Good
Grant lag	No	Good
Legal status	Weak (But works as a filter)	Good but requires frequent updates
Licensing	Strong	Very limited
Opposition or Litigation	Strong	Limited and difficult to extract
Ownership	No	Good
Term	Weak	Good
Technological scope	No	Good

from processing such a heterogeneous source will likely be impacted by the parameters and assumptions that are applied to the data in order to analyse it. For example, when more and more parameters for analysis are added, there is a risk to produce an algorithm that corresponds precisely to the data set which is being used as the training model, but which does not fit other data sets. An algorithm is typically developed (or trained) with the intention to generate accurate results over a wider data population, and its performance is measured in a test with a control data set. Eventually, the algorithm can accurately pinpoint the "correct" patents within the control data set, but when faced with a different group of patents it may fail to identify any "correct" patents at all. This happens because the program was created to identify details and nuances specific to the control test, but which are not general rules for the identification of correctness. This phenomenon is known as overfitting and is often observed when analysing data sets which were created with objectives different than software analysis, such as patent databases. The opposite effect, underfitting, can also happen causing the algorithm to fail in the obtention of the desired results. It is usually a result of an oversimplified model which leaves out important parameters in the analysis.

Conclusions

In sum, patent value is complex and depends on the specific situation of each negotiation. Even the same patent can be perceived as having different value by different individuals in different circumstances. Automated platforms frequently present the factors as direct indicators of value (or lack of it) based on simplified, and sometimes-undisclosed, assumptions. Although these platforms can correctly process some factors in a quick manner and generate metadata which could be very useful for the industry (e.g. by identifying patents which survived nullity actions), some other results provided by these platforms should be taken with prudence and caution.

> **First, there is not enough information publicly available to create a reliable dataset about patent value, the information is scarce and difficult to gather.**

Unfortunately, there is no factor that is reliable for all cases. Even the ones with the strongest connection to value, suffer from issues that prevent their applicability in some circumstances. This variation in the reliability of the factors makes human involvement necessary for a trustable result of the process. Moreover, users should know that a platform applying a higher number of factors does not necessarily provide the most accurate results. For example, if weight is placed on unreliable factors or the wrong algorithm is used, the results will be wrong.

The users must pay attention to the factors analysed and to how the information is processed, so they can evaluate whether the information is relevant for their specific scenario and needs. Fortunately, the platforms can be useful for certain cases (e.g., a general analysis before a merger) even if the results are not sharply precise. In other scenarios however, a more careful assessment may be required. The more complex the assessment is (e.g., in essentiality determination where technical and legal knowledge are involved) the less reliable an automated system will be.

Action plan

- Automated platforms aim to assess patent value by breaking down the information from patents into less-complex pieces called factors.
- Factors are presented as direct indicators of value (or lack of it) based on simplified assumptions. These assumptions can be inaccurate. In the real world, patent value is complex and context-dependant.
- With current technology, automated platforms are insufficient for a more complex assessment, such as determining the essentiality of a patent.
- Caution and scrutiny are crucial for users when using information from automated platforms, especially if the information is going to be used in negotiations, policy-making or other practices with wide-reaching repercussions.
- Transparency from service providers is fundamental, the details about what information is processed and how it is interpreted allows users to make better use of this kind of tools.

The views expressed herein are those of the author alone and do not necessarily represent Ericsson's views.