

**IPwe**

[info@ipwe.com](mailto:info@ipwe.com)  
[www.ipwe.com](http://www.ipwe.com)

# IPwe Platform Analytics White Paper



# IPwe Platform Analytics



## BACKGROUND:

This IPwe white paper provides more detailed information on how IPwe Analytics functions and its features and capabilities.

IPwe Analytics has been in development for 10 years and is based on pioneering research conducted by Professor George Karypis — Distinguished Professor of Computer Science at the University of Minnesota.<sup>1</sup> IPwe Analytics

represents a total investment of approximately \$20 million over the past ten years and has been utilized in transactions involving over one billion dollars in licensing, acquisition and capital market transactions.

IPwe offers analytics free of charge without any tracking or advertising. An advanced version can be accessed by request by emailing [analytics@ipwe.com](mailto:analytics@ipwe.com).

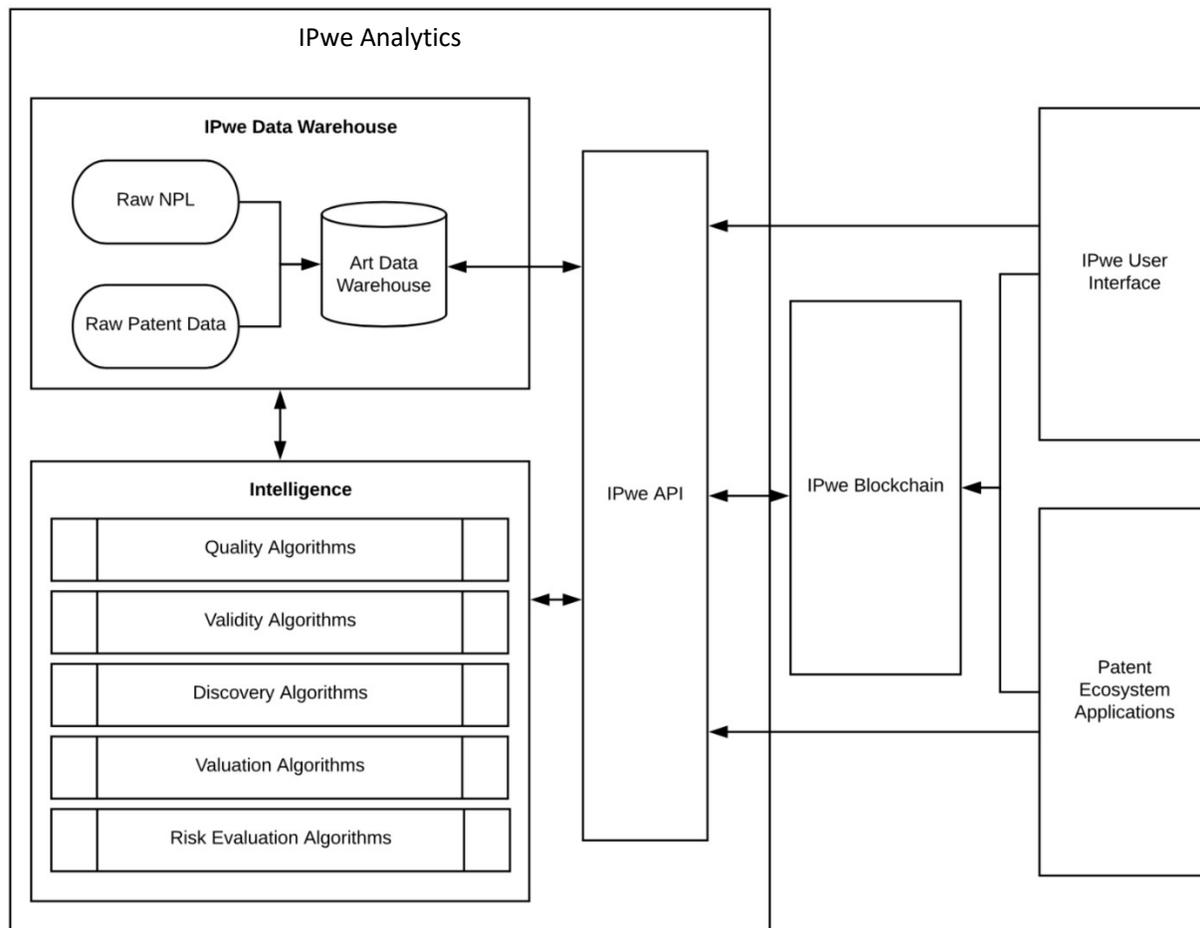
## HOW IPwe ANALYTICS HELPS UNLOCK THE PATENT ASSET CLASS:

IPwe Analytics fundamentally alters the dynamics of the patent market by automating the more mundane and tedious aspects of patent analysis. While patents by their nature are unique assets, few organizations can dedicate the time and resources required to manually review each patent in a large portfolio. With respect to an individual patent, you want to identify the patent that is most likely to be worth your time reviewing and want to be able to quickly identify key attributes about that patent. IPwe Analytics allow just that, freeing up time and financial resources so parties engaged in the patent space can efficiently execute their legal, technical, financial or business analysis.

<sup>1</sup> <https://scholar.google.com/citations?user=ElqwScwAAAAJ&hl=en>

# IPwe Platform Analytics

## IPwe ANALYTICS OVERVIEW:



IPwe Analytics uses a collection of artificial intelligence and data mining techniques that have been developed to solve fundamental and more complex problems related to patents.

We believe that the collection of techniques IPwe Analytics uses are capable of addressing the basic but critical questions regarding patents. There are some very simple questions that IPwe Analytics can address that would otherwise require significant human resources (e.g., identifying patents that address particular technology areas). There are more complex questions IPwe Analytics can address such as “patent quality” and “validity” that, depending on the purpose for which this information is being used, can either provide adequate first level analysis (e.g., is this

# IPwe Platform Analytics

patent likely invalid and worth additional examination) or provide a basis to begin analysis that will benefit from human augmentation (e.g., preparing a submission to a patent office).

One problem encountered every day is that stakeholders in the patent space are presented with a very broad issue that would potentially implicate hundreds of patents (e.g., I plan to launch product X with attributes A and B—are there any patents that I should consider in designing or launching Product X?) and the only choices are to ignore any patent implications or to conduct some level of analysis on potentially relevant patents. In finance, stakeholders are approached with a potential opportunity that includes a patent aspect and have to decide whether the patent should be considered in evaluating the potential opportunity. In research, stakeholders may have an interest in a particular technology area and are attempting to determine what others have done in a particular technology area. While ultimately it may be prudent to retain counsel or other third party advisors in an attempt to answer these questions in more depth, we believe there is obvious benefit to using analytic tools like IPwe Analytics to at least start answering these types of questions. IPwe believes that for a majority of scenarios stakeholders face, IPwe Analytics is an excellent tool to provide initial guidance and to refine and help more precisely define relevant areas for further inquiry.

In addition to the benefits for researchers, inventors and other members of the patent ecosystem, we believe that patent professionals also benefit from IPwe Analytics. We do not believe that tools like IPwe Analytics replace the need for counsel and expert advisors. We believe that IPwe Analytics provides a level of analysis that is more than adequate to gain a high-level understanding of a portfolio or particular patent. Where IPwe Analytics also excels is in work that has traditionally been outsourced to lower labor cost solutions can now be returned to counsel's and expert's laptops and they can then apply their skill and learning to higher value tasks.

For an additional information on the use of AI (as distinguished from other search methods), click [How artificial intelligence outsmarts current patent search tools and will change the patent world.](#)<sup>2</sup>

---

<sup>2</sup> Written by Dan Bork, CTO at IPwe

# IPwe Platform Analytics

## Patent Quality – Q Score:

“Patent quality” is a term often used, but there is no agreed upon definition or metrics to establish patent quality. The USPTO Office of Patent Quality (<https://www.uspto.gov/aboutus/organizational-offices/office-commissioner-patents/office-deputy-commissioner-patent-19>) endeavors to bring definition and metrics to this term, but its definition does not appear to have gained broad acceptance. We believe that the market will ultimately answer this question.

IPwe uses IPwe Analytics to provide an indication of patent quality we refer to as the “Q Score.” IPwe Analytics computes a score for each patent in its collection, which is designed to measure its overall quality relative to the collection of patents. The goal of this single score is to capture various aspects related to how well the invention is described, its pioneering nature, its market/commercial importance and significance, and how relevant it is for technologies that are currently being actively developed and deployed.

The idea behind the approach that IPwe Analytics uses to compute the Q scores is to imitate the process that an expert would follow when asked to identify the set of the most important patents related to one or more patents describing a particular technology. The expert will start from these “seed patents” and expand the set by iteratively collecting the patents that are cited by the seed and the cited patents. After a few iterations of this process, the expert will end up with a set of patents that she will prioritize and rank based on a variety of factors, including how often they were cited in this expansion process, whether they are in the same or related art areas (by using their assigned classification codes), by the size of their respective patent families, by how determined their respective patent owners were in prosecuting them (office actions) and keeping them in force by paying their respective maintenance fees, by how often they changed owners and by how well their claims are supported by the specification. The expert will then select some of the highest ranked patents to iteratively identify additional patents that cite them, and thus allow her to identify patents that have been issued in more recent years, which may have significantly improved on the initial inventions to the point that these patents may make the set of already identified highest-ranked patents obsolete. The expert will proceed to prioritize and rank this new set of patents using similar criteria to those used earlier and combine them with the previously highly ranked patents to either get her final list or use them as a new seed set of patents to repeat the above process.

# IPwe Platform Analytics

IPwe Analytics operationalizes the above process by using a random-walk-with-restart (RWR), which is a stochastic process on networks (graphs), and in its simplest form computes for each network node the steady state probability that a random walker will end up on that node. Nodes with high steady-state probabilities correspond to central nodes and depending on the underlying domain, they correspond to topical authorities (e.g., performing a random walk on the citation network of scientific articles), important web-pages (e.g., performing a random walk on the network corresponding to the hyperlink structure of the web), proteins involved in many biological processes (e.g., performing a random walk in a protein-protein interaction network) and influencers (e.g., performing a random walk in a social network or a follower-follower network).

IPwe Analytics employs a novel RWR model, referred to as *expert-random-walker-with-restart* (ERWR). This patent pending method was specifically developed by Professor Karypis for patent analysis and is rationally designed to model the expert's process using RWR as applied to patents. The network that IPwe Analytics uses corresponds to the patents' citation network that contains directed edges for both the "a *cites* b" and "b *cited-by* a" relations. The *cites* edges are used to model the initial expansion that an expert performs by following the set of patents being cited by the seed patents, whereas the *cited-by* edges are used to perform the second level of expansion that focuses on the recent inventions. The expert's various prioritization criteria are encoded by utilizing a wide range of primary and derived information to assign different weights to the network's edges and to assign non-uniform restart probabilities to the different nodes. As a result, ERWR can also be viewed as an RWR process on a heterogeneous information network, consisting of different types of edges modeling the different prioritization criteria. The relative importance of the different types of information was estimated via extensive experimentation and close collaboration with patent and technical domain experts.

We believe ERWR is a superior method for determining patent quality among a collection of patents. Many of the patent analytic tools available attribute, for example, significant weight to citation count, various aspects of prosecution history or similar factors. While these individual factors are potentially indicative of various attributes, we believe it is essential to deploy additional levels of analysis through ERWR to reach a useful result. Simply put, we believe that ERWR achieves better results as a result of being reflective of the process an expert would use in attempting to analyze a given patent or set of patents.

# IPwe Platform Analytics

We continue to refine “Q scores” and expect that over time we will improve results and permit users to focus in on the individual elements that currently contribute to Q Scores.

## Patent Validity – V Score:

“Patent Validity” is a term that encompasses the various different ways a patent can be rendered unenforceable in a court or administrative proceeding. Patents are most often rendered invalid as a result of prior art (in general, where the invention is disclosed in a printed publication—like another patent or application or article—at least a year prior to the patent filing date), lack of patentable subject matter (there are a number of things that are not eligible to be patented, like laws of nature), that the invention is not “enabled” (someone with the requisite skill must be able to replicate your invention based on the disclosure in the patent) and failure provide a “written description” (failure to describe the invention in the patent specification). There are other ways to challenge a patent – for example, failing to disclose all the inventors, but these are far less common.

Currently, IPwe Analytics focuses in on “prior art” analysis. In particular IPwe Analytics relies on searching over 59 million patent and patent application records dating as far back as the early 1900’s and over two million articles and journals we have legally obtained and indexed and regular downloads from the Internet, which we call our “IPwe Analytics Prior Art Library.” The IPwe Analytics Prior Art Library is constantly being updated and we are working with a number of third parties to establish an even larger library of key prior art references. In the near future, we plan to make the IPwe Analytics Prior Art Library available to any third party that commits to contribute and share in the goal of creating a relevant prior art library accessible by all.

While data search sources like Google, Bing and others are excellent resources that should be consulted in any prior art search, there are limits in their utility for various structural reasons. For example, a prior art reference in the medical device field may not be useful in invalidating a patent in the field of agricultural sciences. In other words, it is not “analogous art.” Through a time consuming process, we call “indexing,” IPwe Analytics endeavors to address analogous art issues. Moreover, in certain analyses, one must demonstrate a “motivation to combine” multiple references to invalidate claims in a patent. Through various analytical techniques, IPwe Analytics is now able to assess and provide results that reflect this “motivation to combine” requirement.

# IPwe Platform Analytics

IPwe Analytics compiles its results on a claim element by claim element basis, again mimicking the process that an expert would follow.

We have begun development and testing of algorithms that we believe will be useful in analyzing other aspects of validity, including lack of patentable subject matter and enablement. These will ultimately be incorporated into IPwe Analytics.

## Prior Art Analysis:

For the purpose of evaluating validity, each claim of a patent is considered separately on an element by element basis. IPwe Analytics conducts its evaluation on a claim by claim basis and a claim element by claim element basis. For example, IPwe Analytics identifies relevant prior art by taking into account the limitations of the claim under consideration, the text of the art, the link structure of the citation network and the patent classification. IPwe Analytics first constructs a network that consists of two types of nodes: (i) the art node (patents and non-patent literature) and (ii) the classes of the patent classification node (currently IPwe Analytics uses the international patent classification known as the “IPC”). Each art node is linked to all the art nodes that it cites and is linked to all the classification nodes to which it belongs. The weights of the first set of edges are determined based on the content similarity of the corresponding art nodes, whereas the weights of the second set of edges are determined based on the classification strength. If the art corresponds to patents and/or patent applications, their known classifications are used (the primary class has higher weight than secondary classes). If the art corresponds to non-patent literature (NPL), their weights are determined by machine-learning classifiers that were estimated from the patents. Given a query patent, IPwe Analytics performs a random walk with restart on art classification node (ACN) from the node in the ACN that corresponds to the query. Upon convergence, this process determines the steady-state probability of visiting each of the nodes in an ACN starting from the query patent. The nodes in an ACN that correspond to art with a priority date that is earlier than that of the query patent and for which we have access to their full text are ranked based on their steady-state probabilities and the N highest-ranked nodes are selected. The art corresponding to this set of nodes defines the candidate prior art subset (CPAS) that is being further analyzed by IPwe Analytics in order to determine if it is relevant with respect to the query claim. By leveraging the citation network, the classification

# IPwe Platform Analytics

hierarchy and the connections between art and classifications, the CPAS contains prior art that is highly related to the query patent because there are many short paths within an ACN passing through related art and/or classifications that connect them.

IPwe believes that the methodology IPwe Analytics uses to identify prior art is superior to other methods because by using the ACN, IPwe Analytics focuses on the subset of patents or NPL that humans directly or indirectly determined to be relevant. This is a direct result of the multiple levels of random walk-based citation analysis that takes into account the state of knowledge when the patent was granted and also how the collective knowledge has improved over time. The network encodes both forward-in-time information (by traversing the links in a reverse direction) as well as the traditional backward-in-time information. For example, if at a later point in time some new patents co-cite both the patent being analyzed and also some older patents, then this forward in time information will be used to establish an association between the target patent and the older patents, even if such associations were unknown at the time that the target patent was granted. This same approach is used when evaluating motivations to combine. If patents are ranked high in the random walk seeded on the target patent, then there is strong evidence that an expert should have been aware of these different pieces of prior art and as such should have thought of combining them.

## Patent Discovery:

Using the same base algorithm used to determine Patent Validity, IPwe Analytics can perform similarity searches on any given date range in order to discover potential patents of interest based on a query. Such analysis can be used in freedom to operate queries, technology portfolio creation and technology landscaping.

If you have any additional questions, please contact your IPwe representative or email us at [analytics@ipwe.com](mailto:analytics@ipwe.com).