The Value of Hard Work in the Age of Big Data and AI for IP Analytics

by Philip Totaro, CEO of Totaro & Associates www.totaro-associates.com

I had a high school English teacher who encouraged his class to think critically about a topic by 'rubbing some brain cells together.' In an age of Big Data analytics and artificial intelligence (AI) algorithms, it is possible to lose something by shifting the burden of research analysis to a computing infrastructure, no matter how robust, mature or modern.

The phrase 'garbage-in, garbage-out' is often used to describe analysis tools which apply a perfectly sound analysis methodology to a data set which was not comprehensively cultivated or qualitatively assessed in the first place. There is value to human intelligence that can compile a dataset which is worth analyzing in the first place, but many seem unwilling to put effort into doing even that. Some hope instead that AI will be able to do the heavy lifting for them, but if you task a computer to find something for you, how would you know if you received an incomplete set of results?

Is close enough really good enough, or is that just the excuse of the lazy?

Modern computing technology was developed as a tool to help facilitate the calculation of complex mathematical operations, and humanity has wonderfully capitalized on this capability to develop tools and software capable of a significant amount of analytics and data post-processing. The one thing that has not been achieved yet is to develop a computer that can replicate human intelligence, no matter how many Jeopardy questions it can answer correctly.

There is, of course, value in leveraging this computing infrastructure to unlock the mysteries of the universe, but it seems as though many believe we can replace human ingenuity with computers that can automate portions of our lives. Computers can do amazing things, but it's all in how you program them. If you don't want 'garbage-out' then put some effort into feeding 'quality-in.'

I myself am an innovation analyst. Most people would look at the exercise of reading and cataloguing 75,000 patents as a boring activity. After doing just that over the past 4 years, I would tend to agree, but I love learning about innovation, and if reading patents is one way to help cultivate the understanding of how things work, then so be it.

Other innovation analysts, who are entrenched in the conventional methods of analysis in the modern age, have told me 'that's seems like a waste of time and resources. Why can't we automate that process? Why not use a smart algorithms or heuristics to find the results you're looking for? Can't we leverage off AI to facilitate that kind of work?'

My response to them is simple, 'How do you find a needle in a haystack?' If you intelligently search and catalog the entire haystack, you never need to search it again. Even if the haystack grows, all you need to capture is the delta. With that in mind, my company used the results of that 4 year cataloguing effort to develop an innovation analysis toolkit to post-process the results.

While the initial results set used to develop our tools was focused on one industry, wind energy, the methodology and tools can now be leveraged in any industry. All it takes is technical subject
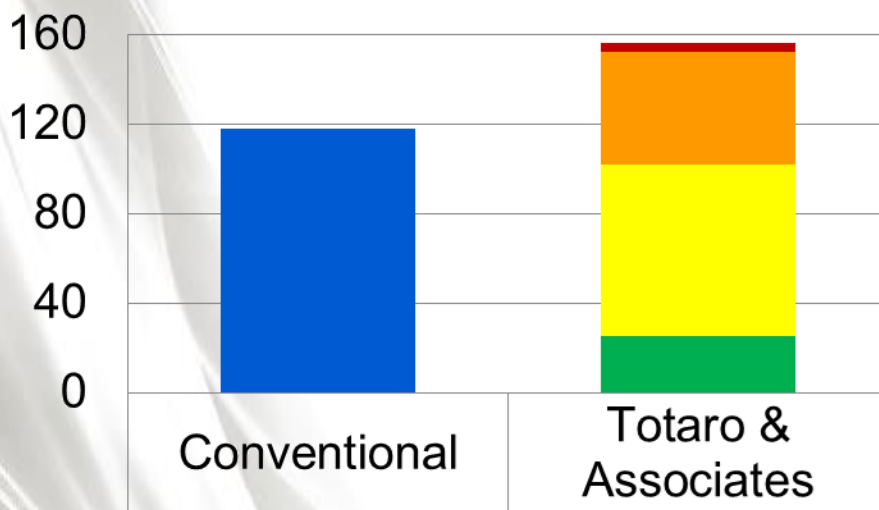
matter expertise, industry expertise, and a willingness to put effort into analyzing a patent landscape in order to get a meaningful result out of the tools that were developed to post-process the 'quality-in' data set.

It doesn't have to take years of analysis work anymore to get there, because if a patent landscape already exists, and has been qualitatively assessed, it only needs to be dropped in, and the insights will flow.

We developed the qualitative analysis method and tools to address the most pressing issues regarding intellectual property landscape analysis:

- Can conventional tools compare the claim breadth of patents to industry usage of a product comprising that patent protected innovation? No, because reliance on IPC codes or other existing classification methods won't identify what patents are relevant to the rest of the industry. People with both technical and industry subject matter expertise can.

- How do conventional tools identify prior art which is un-cited, but should have been? They don't because citation analysis doesn't link what's not there. Subject matter experts can identify this, and using the right tools, an answer to this question is available at the push of a button.

- Can conventional tools tell me, 'If my product contains a particular technology, in what jurisdictions of the world will I have to make a design change in order to avoid infringing on a competitor patent?' Sadly, no. Certainly not at the push of a button. But a data set that has been qualitatively analyzed could.

- Can conventional tools tell me what patents are for sale that I could buy that would neutralize an NPE / PAE / patent troll? Maybe, but the analysis is not likely based on a qualitative assessment of the patents and their claims. You also have to know which patents are for sale, which is not always public.

We decided to benchmark our results against competitor tools based on 'intelligent search algorithms' to see if our qualitative approach would be better. We ran a keyword search in several conventional tools for "wind AND (turbine OR energy OR power)" AND "power factor control."

1400 McKinney St, Suite 2207
Houston, TX 77010-4058 USA
www.totaro-associates.com

Phone: +1 832 915 0010
eFax: +1 206 202 0593
info@totaro-associates.com

	Conventional	Totaro & Associates
■ No Classification	118	
■ High		4
■ Medium/High		50
■ Medium		77
■ Low		25

Keyword-based Prior Art Search	Search String	Totaro & Associates Patent Landscape
118 Results	"Wind Turbine" AND "Power Factor Control"	156 Results with Risk Classification (L, M, M/H, H)

Conventional results produced 118 patents, although upon inspection, at least 3 were false positives (i.e. they contained the keywords we searched for, but upon inspection, the patent was not relevant to the scope of technology). Comparatively, our results showed there were actually 156 patents of relevance, and our qualitative analysis approach identified those patents which comprised claimed technology that was widely used within the industry by companies other than the patent holder.

The qualitative assessment was available at the push of a button; there was no need to weed through 118 patents to find the needle in the haystack. So why was our approach so much better?

Some people believe patent examiners are not always subject matter experts, and they can't be trusted to come up with a complete list of citations or perform a thorough examination of a new patent filing. While that may be true in certain instances, most people do not realize that it is often a sub-contracted IP search company with almost no technical and/or domain expertise who are entering the data in the PTO database and performing a prior art search on behalf of a patent examiner.

1400 McKinney St, Suite 2207
Houston, TX 77010-4058 USA
www.totaro-associates.com

Phone: +1 832 915 0010
eFax: +1 206 202 0593
info@totaro-associates.com

Conventional search and IP analytics tools which are based on International Classification codes, citation link analysis and keyword-based searches contained in the patent spec will provide an inaccurate or incomplete picture because the initial data entry is fraught with inconsistency, and even typos.

The problem is that most people who have not gone through a comprehensive patent landscape would not know that they are receiving results that are incomplete and inaccurate. Intelligently compiled search algorithms are incapable of providing a more accurate result because it's not 'quality-in' to start off with.

One response I've heard from those who question our methodology is that, 'There are just too many patents to search and individually fact check the way you do.' Our qualitative patent landscape analysis philosophy provides the added benefit of a data integrity check to the PTO data, so our analysis ensures 'quality-out.' Other tools can't guarantee that due to that lethargy that exists to tackle a complex matter, or to use algorithms to shortcut doing the work.

Seems as though Albert Einstein had plenty of inspiration prior to his annus mirabilis papers on topics such as special relativity, the photoelectric effect and other groundbreaking theories from taking his time at the Swiss patent office to read the patent filings of others and understand how to innovate above and beyond.

With the multitude of people around the world interested in innovation analysis and pursuing the development of new tools, I would love to see those resources directed at cultivating knowledge and exploring the millions of other patents out there. It might be a big number, but it is finite.

Let's not be afraid to get our hands dirty doing the hard work that can expand human knowledge. Let's intelligently leverage the computing infrastructure that has been developed while cultivating human intelligence at the same time. Let's not let laziness or a seemingly complex problem get in the way of achieving great things in innovation. Let's rub some brain cells together… you'll be amazed at what happens.

1400 McKinney St, Suite 2207
Houston, TX 77010-4058 USA
www.totaro-associates.com

Phone: +1 832 915 0010
eFax: +1 206 202 0593
info@totaro-associates.com