

# James Skelley

Technology Law - IP Prosecution - Transactions

MAIN

UPDATES

LAWMUX KNOWLEDGEBASE

SOFTWARE

FAQ

CONTACT

SUBSCRIBE

## Search Tool Validation

🕒 November 18, 2019 👤 SKELJ 📁 Uncategorized 💬 0

As an Examiner or anyone who's used PATFT or Google patents will tell you, prior art search certainly isn't a "solved" problem. In fact, it's been a favorite problem of mine not simply because of my IP practice, but because it gets to the heart of some the basic challenges in NLP and machine vision.

Recently, I've had some success with a process I've dubbed "Reverse Machine Learning" (RML) (surely someone's had the idea before, but because I'm ignorant, I have the pleasure of coining the term anew).

Under RML you attempt an ML solution, fail, but *with that expectation* so that you can then carefully seek to understand WHY you failed, and from that understanding, devise a more effective NON-ML solution.

Originally, I **TFIDF**'d a bajillion patents from the USPTO database, **VGG-19** feature reduced their images, and then began the search with a cosine-similarity keyword (supplemented by cognates) and image-based search. I actually thought it would do alright, but in reality the results were . . . meh.

Not too much better than what I'd get by Google or PATFT the old way. It also took forever (millions of assets in giant matrices – numpy was displeased). Which raised the question, why? On paper, I knew it would be slow, but (assuming I found all the proper cognates) I expected it would have done much better? Why the sucking?

*"To ask the right question is harder than to answer it."*

*-Georg Cantor*

### RECENT UPDATES

DMCA Study

Skelley Snapshots

CPC Based Deal Search (i.e., not a COVID post)

### NAVIGATION

Main

Updates

LawMux Knowledgebase

Software

FAQ

Contact

Subscribe

It turned out this was a wonderful question because the answer isn't the same for different types of technologies and types of claims. But there *is* a common theme across all the failure modes and that theme drew me to create (yet another) new search tool.

With the tool 80% finished, it was time to validate. There are many “bad” patents that you don't learn much by searching. Also, I don't want to tune for “easy” targets. So I needed to find “real” patents, proper “hard” patents, to validate against. Where to find them?

Two thoughts came to mind:

1. [Unified's Patroll](#)
2. Recently filed infringement cases from [PACER](#)

In theory, 1) should be “real” because the Unified folks think they're worth posting a contest and paying out for. If they were trivial to search, there's no sense soliciting the public for art. Similarly, 2) should be real, because bringing litigation is expensive, so (presumably) the plaintiff did their diligence beforehand and believed the prosecution references to be adequate (I didn't check for reissues).

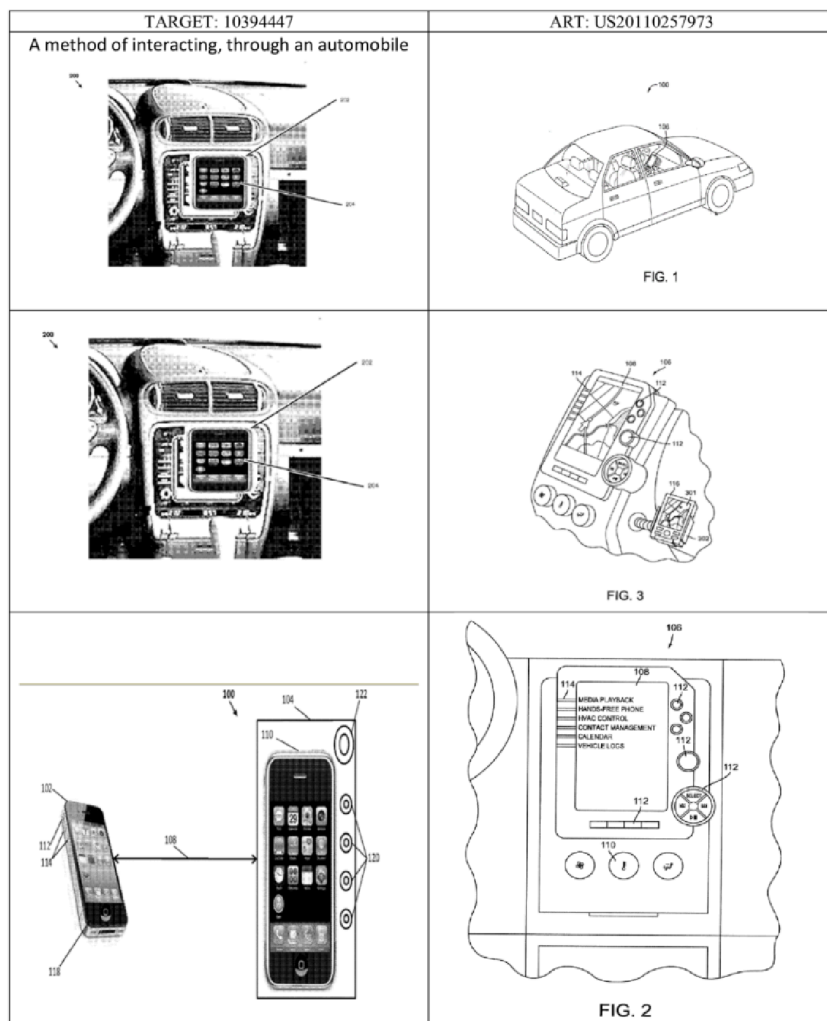
So to validate, let's select some assets from those pools and see not just WHAT we come up with, but HOW LONG it takes for us to find it.

*DISCLAIMER: If you happen to own one of these patents, please don't panic. I have deliberately NOT provided a formal claim chart below. Searching inherently involves the creation of an informal chart, which I have converted to the loose-goosey image comparisons below, but this is NOTHING like a full and proper claim chart. These searches were for diagnostic, not legal, purposes. Conversely, if you're on the receiving end of one of these, don't rely on the below. Do your own analysis. You were warned.*

## Challenge #1: Virtual Car iPhone

First up, an item from Unified's patent challenge: [US1039447](#). The priority is ostensibly April 15, 2010 (I didn't check the provisional filing to verify, so it may be later). So we need to 1) find a relevant asset before that date that 2) didn't appear in the Examiner's search

history. A handful of results emerged, including [US20110257973](#) with an (ostensible) priority date of Dec. 5, 2007.



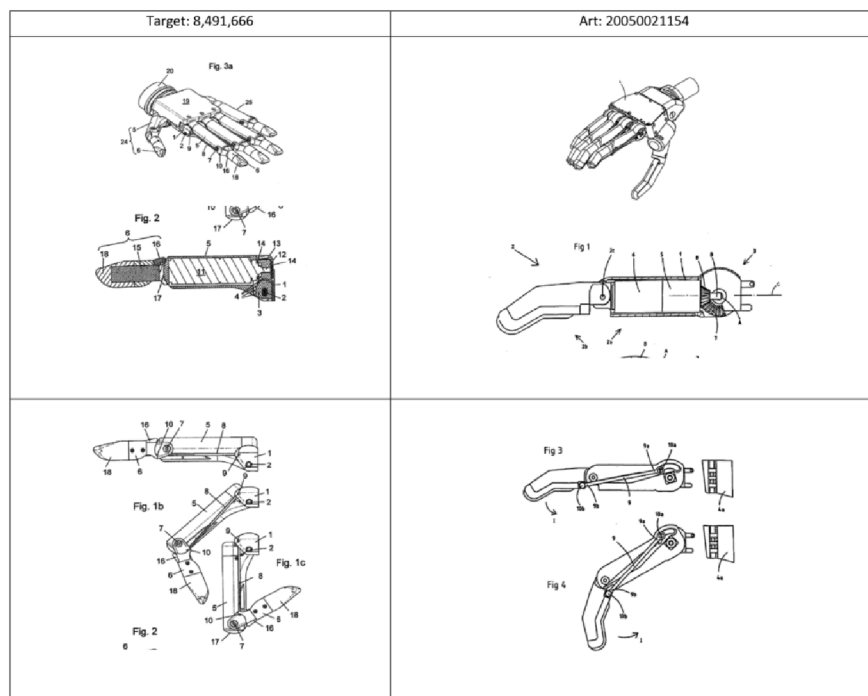
Pretty good, huh? This was actually one of the first search attempts so it actually took a bit longer than later ones, clocking in around 3 hours. That was good, though, because: 1) such a search would have taken me a day or days before; and 2) it helped teach me how to optimize the workflow. Post optimization, I expect it'd have taken around an hour. It's hard to compare the new search time with the traditional approach since, now that I have an answer, I know I'll subconsciously cheat and search to hit it if I use the old approach.

Encouraged, I decided to try another one.

## Challenge #2: Prosthetic Finger

[US8491666](#) comes to us from a recently filed infringement action in Southern California a week or so ago. Priority is again ostensibly to Nov. 8 2008 via a foreign predecessor. So we need something before

that date, which is also absent from the Examiner's search results. Crunch, crunch, crunch . . . couple hits but [US20050021154](#) has foreign priority to Aug. 27 2001.



Not bad, right? This was about ~1 hour of searching. Before that, I expect it would have taken me at least 4-10 hours, probably much longer, to have found, but again, can't say for sure.

So at this point, it was getting exciting.

Let's try another one.

## Challenge #3: Well Control

[US9334701](#) comes from a Texas action filed last week. Its priority is a little questionable, being a CIP, but let's give it the benefit of the doubt like the others and assume it goes all the way back to Oct. 20, 2011.



There actually are some proper patents out there, with goofy ideas, that no one even vaguely considered before.

I mentioned my tool to a litigation buddy and, though his response was diplomatic, I gathered he was somewhat appalled by what I was doing.

“You’re casually finding all these references – isn’t it cruel you don’t send them to counter-parties if they exist? Or at least 1.501 insert them into the files?”

There are three reasons I don’t do that:

1. As mentioned above I haven’t *\*formally\** claim charted any of these – informal search charting is just to guide the search, not to fulfill a 102/103 analysis. For all I know, there’s some esoteric comment the claim / file wrapper that makes the simplistic comparisons above specious;
2. Counter-parties should be masters of their own fate; and
3. I used to.

Regarding 3), the *universal* response was: “Who the devil are you? Go away. No we don’t want to see your silly references.” I kid you not.

However, in hindsight, that actually makes sense per 2). If I was handling a matter and some random Silicon Valley attorney offered references for “free” I’d be very suspicious as well.

There’s still more to do – it would be nice to expand to the EPO and other databases, but I’m not sure that’d be feasible. There are also a ton of features that need to be optimized.

**PREVIOUS ARTICLE**

**NEXT ARTICLE**

---

WP Template Copyright © 2016 MH Themes | Content and Supplemental Software Copyright © 2015, 2016 James Skelley  
No content on this website is intended, or should be construed, as legal advice. Rather, the content provided here merely serves to notify a wide audience of various issues' general character. If you have a legal question contact an attorney.