CLEARSTONEIP

WHY SEMANTIC SEARCHING FAILS FOR FREEDOM-TO-OPERATE (FTO)

White Paper

This three-part paper explains why conventional techniques, particularly "semantics-based" searching, fall short for freedom-to-operate (FTO) searching and analysis. It then puts forth a solution for avoiding these problems. Part I is an introduction to the differences between the searches. Part II identifies the deficiencies of semantic searching in relation to FTO analysis. Part III explains how these deficiencies can be overcome.

I. The Differences Between Patentability and FTO Searching

Not all patent searches are the same.

This seems an obvious point. But how well understood are the conceptual distinctions between the various types of patent searches? We are quite familiar with a "patentability search," which attempts to answer the question:

Is this concept novel and non-obvious?

We are also familiar with an "invalidity search," which attempts to answer the question:

Should this patented invention have been considered novel and non-obvious?

These types of searches are conceptually similar, and may be collectively referred to as "patentability searches." Now consider, in contrast, the question posed in an FTO search:

Is this product likely to infringe an active patent?

Based on these different underlying questions, three critical distinctions between FTO and patentability emerge.

A. In FTO, relevance of patent results is determined by *claim scope*, not *description*.

Patents necessarily include a technical description and legal claims. While the technical description must enable the claimed inventions, the actual scope of what is claimed may vary significantly from what is described in the technical description. For example, practitioners generally aim for detailed technical descriptions yet broad all-encompassing claims.

In the vast majority of FTO cases, the claims of patents that describe features of a product undergoing FTO do not actually cover those features. Usually these claims are significantly narrower in scope. Other times, the subject matter of the claims is simply directed to other disclosed aspects.

The reverse scenario is also a significant concern. Patents that do not describe features of a product undergoing FTO could certainly have claims that cover one or more of its features. For example, consider a cup having a handle. A patent may never describe a handle, but may claim a cup with circumferentially asymmetric mass distribution. A handle could likely fall within those bounds. These scenarios are quite common

B. Products tell a thousand stories.

Questions of *patentability* are often limited to a concept or a fixed set of concepts. The question hinges on a specific claim that is, by definition, a single textual sentence. On the other hand, *FTO analysis* centers on actual products. A product, by its physical presence, could be described in thousands of ways. For example, even a simple device implicates all of its structural components, its mass characteristics, its geometric characteristics, processes underlying its manufacture, and processes involving its use.

Anticipating all the ways in which a product can be described is serious guesswork. Opting to focus on some ways and not others is an arbitrary exercise.

C. Missing patents in an FTO search could be dire.

A final distinction between these types of searches lies in the consequences of missing key patent references. Missing key patent references in a patentability search is certainly not desirable. However, I would venture a guess that, if they were forced to choose, most companies would prefer to have a potentially invalid patent issue than a potentially infringing product launch.

Also, finding *some* relevant patents in a patentability search is helpful. In fact, perhaps, in an invalidity investigation, a few good references is all it takes; no need to lose sleep over the prospect of other patents lurking about. In other words, there are *pro rata* rewards to locating relevant patents in patentability searches; the more relevant patents we find, the better we understand the landscape of a feature.

Not true for an FTO search. Finding relevant patents “along the way” does not bring an analyst any closer to the finish line or provide any greater satisfaction that their work is complete. Finding some patents of concern is little indication of whether other patents exist that may also be of concern. That one missed patent could spell complete disaster for a product line or, worst case, a business.

These observations may not be news to experienced patent analysts, who have long understood the unique difficulties associated with FTO patent searching and analysis. What is notable, however, is that conventional analytics tools have not evolved to recognize these distinctions. They apply virtually the same processes to both *patentability* and *FTO*, despite their compelling distinctions.

II. Why Semantic Searching Fails for FTO

A. How Semantic Search Platforms Work

There are countless patent searching software platforms available. Each has unique features, but broad commonalities exist. Available platforms tend to offer some combination of natural language, Boolean, classification and semantic searching. Semantic searching is the primary focus of this discussion, as it is the most evolved.

Semantic patent searching generally refers to automatically enhancing a text-based query to better represent its underlying meaning, thereby better identifying conceptually related references. This process generally includes: (1) supplementing terms of a text-based query with their synonyms; and (2) assessing the proximity of resulting patents to the determined underlying meaning of the text-based query. Semantic platforms are often touted as critical add-ons to natural language searching. They are said to account for discrepancies in word form and lexicography between the text of queries and patent disclosures.

Based on this, it would seem that semantic searching is powerful and effective. Well, it is... for some types of searches (*e.g.*, patentability or invalidity searches). However, it is surprisingly ineffective for FTO. And this has everything to do with the distinctiveness of FTO as discussed in Part 1.

B. The Effect of Semantic Platforms on FTO

Semantic platforms, by their nature, assume a certain paradigm. They purport to interpolate the underlying meaning of a text-based query. This is great in cases where an analyst knows which technical concepts are relevant. For example, in a patentability or invalidity search, the analyst has a specific claim under review with specifically-recited elements. FTO searches do not fit this paradigm.

Consider the distinctions discussed in Part 1 of this series:

(1) ***In FTO, relevance of patent results is determined by claim scope, not description.*** *The technical aspects described by a patent's disclosure are distinct from its claims.*

In a patentability search, the semantic platform will return precisely what the searcher desires – patents describing the subject concept of the query.

For FTO, the platform will not. *Some* patents describing a product feature under review may contain claims covering such feature. However, the vast majority will not. The claims will instead be drawn narrower by requiring additional aspects and specificity. Accordingly, semantic engines necessarily output a high proportion of non-relevant patents (*i.e.*, they are "noisy").

The reverse scenario is also problematic. Many patents will exist that do not describe a specific product feature, yet will have claims sufficiently broad to cover the feature. Semantic engines will rarely identify these types of patents. Even if identified, they are likely to be assigned a low relevancy rank given their much broader scope. This makes sense in a patentability search, but not in an FTO context.

For this reason, semantic platforms suffer two deficiencies at opposite ends of the spectrum: (1) they are under-inclusive as they are prone to missing relevant broad patents; and (2) they are over-inclusive due to their noisiness with respect to patents with narrow or otherwise non-relevant claims.

(2) ***Products tell a thousand stories****. Products, due to their physical existence, can be described in thousands of ways. Each way could be a basis for infringement. Patentability searching, instead, is more discrete.*

Semantic search tools force analysts to play an arbitrary game of "guess the element." They require that analysts examine features of a product and pick out just the right ones worthy of review. Even for experienced analysts, this exercise is more conjury than skill. It is simply impossible to accurately predict which aspects of a product are likely to be the basis of infringement in an FTO analysis.

In practical terms, semantic platforms unduly force analysts to pit accuracy against timeliness. If an analyst is selective, many relevant references will inevitably be missed. If, on the other hand, the analyst is cautious and queries many product features, the results will be unworkably noisy.

(3) ***Missing patents in an FTO search could be dire***. *Finding relevant patents in an FTO search is no indication whether additional relevant patents exist. An entire technology space must be cleared. In patentability searching, producing a few close results is more acceptable.*

Because of points (1) and (2) above, semantic-based results are likely to contain a large number of patents, perhaps ranked by purported relevance. In a patentability search, an analyst may be comfortable reviewing only the first tier of patent references (*e.g.*, the top one-hundred or so). However, the purpose of FTO is to assess and minimize liability risk. Reviewing only the first arbitrary tier of references would undermine this mission. FTO is not concerned with which patents *most predictably* cover a product; FTO means ensuring that *no* patents cover the product.

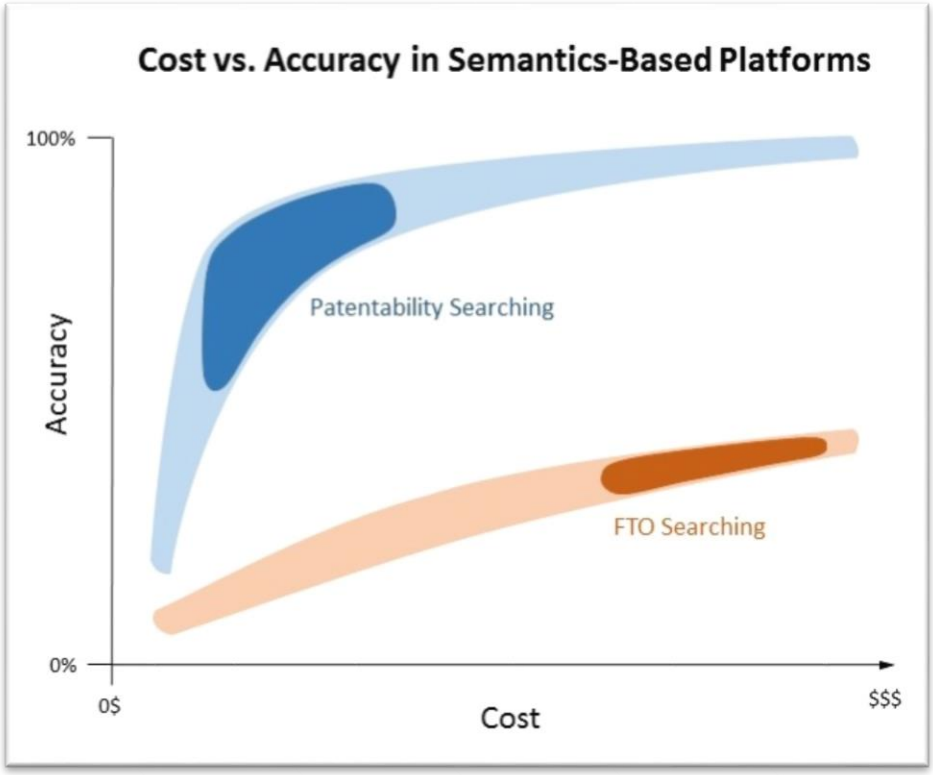
C. Summing Up Semantics

Conducting FTO searches using semantic platforms produces noisy results that are also prone to significant omission of relevant patents. This presents the analyst with a dilemma. The analyst must choose between: (1) reviewing a compact set of references that is likely incomplete; or (2) reviewing a comprehensive set of references that likely contains a significant amount of noise.

If interested in whether these findings relate to you, perform a simple test. Dig up your last comprehensive FTO search. Review the patent references that you ultimately deemed relevant. Do they generally fall within the same patent classes (as opposed to being scattered over the classification map)? Do they all pertain to a predictable technical feature (as opposed to relating to the product in unexpected ways)? Do you believe they could have all been retrieved using just a few keywords? If your responses are generally "no," then your experience is quite typical. If your responses are generally "yes," you've experienced a surprising amount of luck. I suggest buying a lottery ticket.

The illustration on the following page shows how semantic search platforms handle patentability and FTO searches differently in terms of accuracy and cost ("cost" essentially being a proxy measure for work time). A high proportion of missed references results in an inaccurate search. A high proportion of noise results in a costly search. The darker shaded regions represent where industry cases typically fall.

The point here is that semantic platforms can deliver effective results for patentability searches at a reasonable cost but, when it comes to FTO searching, the effectiveness of the platforms is limited even at great cost.



This all leads to the question of whether FTO searches are innately high-cost/low-accuracy processes or if we are just not handling them correctly. Many in the patent industry seem resigned to the belief that improving FTO is a futile endeavor. This point-of-view is understandable but incorrect. FTO *can* be made accurate and low-cost. It just takes a fresh approach.

III. What You Should Be Doing Instead

Part I and Part II of this series explained how semantic and similar keyword-based platforms are ill-suited for freedom-to-operate analysis. We saw how, for various reasons, these platforms show little for their cost.

We can overcome the shortfalls of conventional search tools by building a new FTO solution from the ground up.

Let's take a look at some necessary characteristics of such a solution:

A. The solution must recognize claim scope, not just patent disclosure.

Easier said than done, right? Patent claims are notoriously complex and are often intentionally vague or broad. Despite its complexity, claim scope could be effectively navigated with the right platform.

First, the solution ultimately must leverage human analysis in some form. We should dispel the notion that artificial intelligence, such as semantic-based algorithms, can properly interpret claims. Sure, they may be fine at retrieving patents that *disclose* pertinent subject matter, but there is a fatal disconnect when it comes to claim coverage. Semantic algorithms simply cannot read and process delineations of scope. They are essentially language-similarity detectors and can't differentiate between *claimed* concepts and those that are *merely disclosed*.

Second, the solution needs to handle claim concepts in an *eliminatory* or *deductive* framework. This is a significant departure from the status quo. Conventional platforms *amass* or *aggregate* sets of potentially relevant patents to create a large set for deeper review. However, in FTO, it is far more efficient to arrive at a review set by first *eliminating* irrelevant patents from a large initial set based on a claim scope determination. Several reasons for this were discussed in a previous blog post.

The correct framework places the most relevant question at the forefront, not the back end. For FTO, the question is whether a particular product embodies each claim element of a patent. It is not whether a patent discloses similar subject matter.

B. The ideal FTO solution accounts for the infinite ways of describing a product.

The ideal FTO solution should not require an analyst to identify keywords or specific terms ahead of time because, as discussed in Part II, there is never a single "right" way to do so. And if an analyst were to try to capture all of the ways, semantic platforms would retrieve an impossible amount of results.

The solution to this problem must remove this guesswork from the equation. Building on the eliminatory framework described above, the solution should present to the analyst an organized menu of claim concepts. Instead of considering what to bring into a search, an analyst only needs to consider which of the displayed claim concepts do not correspond to the product.

The menu of concepts should be displayed in an organized manner, for example an index-based system that an analyst can navigate. The index will present a list or taxonomy of technical concepts that each represent patent claim elements. In this way, the analyst can simply make a determination on an element-by-element basis as to whether it relates to the product at issue.

A rough semblance of a concept-based index exists in the form of official patent classification systems used by patent offices around the world, such as the former U.S. Patent Classification system (USPC) and the newly adopted Cooperative Patent Classification system (CPC). But these systems are still extremely cumbersome for FTO for several reasons: (i) they are not keyed to specific claim elements but, rather, general inventive concepts; (ii) they have no capacity to distinguish among different independent claims of a single patent; (iii) they are not nearly specific enough to be effective; and (iv) while they are updated from time to time, they are effectively static indices that are difficult to modify and adapt.

The ideal FTO solution includes a dynamic, easily modifiable taxonomical index of elements that are programmatically connected to specific patent claims. The index has high granularity but allows the analyst to operate as broadly or as specifically as desiredwithout reducing efficiency.

C. The ideal FTO solution does not sacrifice completeness for relevance.

In Part II, we discussed the problems that arise when a search platform provides results in a "ranked" order. We saw that these ranking algorithms could be arbitrary since they are based primarily on similarity of language or terminology. Highly relevant results from an FTO perspective can be placed far down the result list. Missing pertinent patents in an FTO analysis is far more consequential than missing a potentially relevant reference in a patentability search.

The ideal FTO solution should be equipped to capture all potentially relevant patents, readily bring them to the surface, and do so efficiently. It will avoid burying highly relevant patents and make them easy to locate.

* * *

Clearstone Elements™: The Ideal Solution

Our Clearstone Elements application is the ideal FTO solution. It is an interactive platform that, as a core capability, provides a comprehensive taxonomical index of technical elements drawn directly from human analysis of patent claims. As an analyst navigates the hierarchy, he or she selects elements that are *not present* in a product under review. The software will automatically eliminate from the initial set the patents that require the selected element for infringement. After just a short period, typically less than an hour, 90-95% of the initial patent set is usually eliminated, leaving the most critical and relevant patents for closer review.

To see this more clearly, take a look at this real-time video, in which more than 10% of the initial patent set is eliminated in less than 45 seconds.

These kinds of results and efficiency are simply not achievable with any other system. This is how the "noise" is removed from search results.

An interesting phenomenon occurs in FTO. A large proportion of patents tend to be dismissible from an initial patent set based on only a few, general requirements. This is due in part to the noisiness of conventional search tools, but also to the peculiarities of claim-drafting.

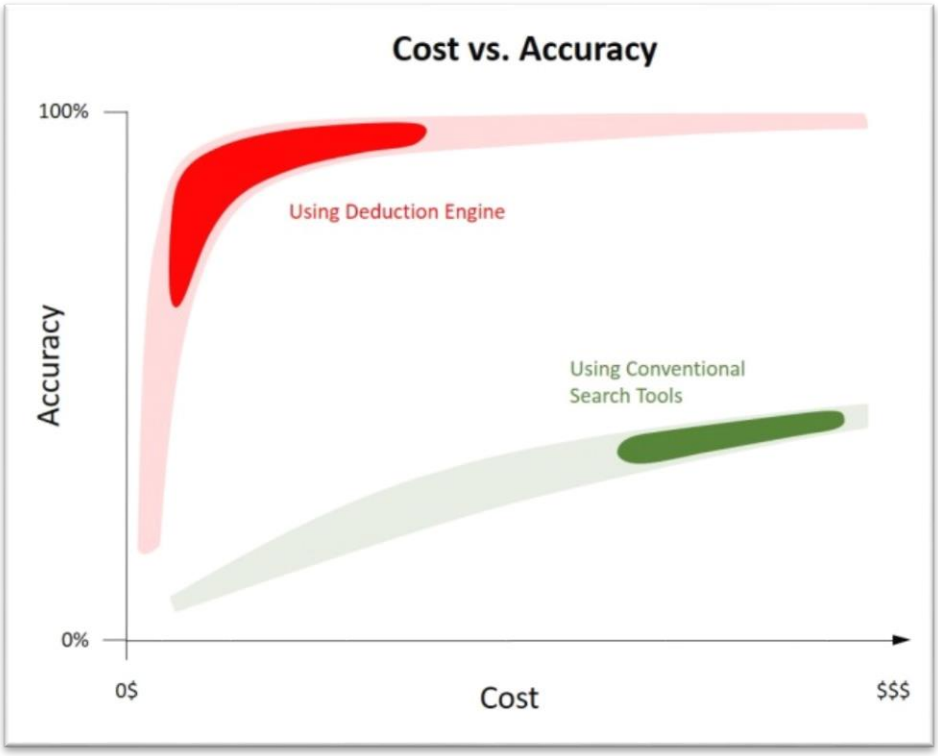
As an example, consider the golf club field. There are about 4,500 active patents in this field, a 100-plus year old industry. Obviously, these active patents are directed to nuanced, highly incremental improvements. Yet, of this universe of highly specific patents, the claims of about 62% require a golf club head to be an "iron-type," "putter-type," or "wood-type." What this means is that, if one only applies those three broad technical concepts in Clearstone Elements, they could eliminate 30-40% of patents from any particular search. Imagine what is possible by applying a few more concepts.

Determining if this phenomenon occurs in your industry is simple enough. Review a random swath of patents from the initial patent set of your last comprehensive FTO investigation. For patents that you excluded, what were the reasons? Were they excluded for requiring broad, sweeping technical concepts (or were you compelled to dig deep to understand the fine points of novelty)? Did these reasons frequently recur

(or were they unique)? I suspect that most patents were dismissible based on broader concepts that frequently recur throughout the patent set. Clearstone Elements leverages this phenomenon and more to achieve incredible results.

* * *

The following graph illustrates conceptually how the addition of a deduction-based platform such as Clearstone Elements can shift the cost-accuracy curve for FTO.



Because of the ease of objectively eliminating large portions of patent references with little work, high accuracy could be achieved at little cost. Conventional tools are unable to achieve this efficiency.

Another important aspect of Clearstone Elements is that the analyst does not have to know beforehand which aspects of the product may present infringement issues. The patented concepts are presented on the screen in the taxonomical index. The analyst only needs to decide whether the product embodies the concept or not. This is how the system ensures that critical patents are not missed – they are only removed from the initial set upon a deliberate decision by the analyst based on displayed concepts.

Taking this a step further, analysts can create a "product record" upon completing their review of the index. This product record is essentially a fingerprint of the product as it relates to the indexed elements, and, in turn, how the product relates to the initial set of patents. The product record can be opened and modified later on to quickly reflect any changes that are made to the product during development to achieve an incredible result: ***The Elements interface will instantly display a list of patents that become of issue solely due to the product changes***. This capability is truly unprecedented and is

key in streamlining product development through enhanced communication between product designers and the legal department. See the blog post, *Bridging the Divide Between Patent and Engineers*, for more.

Many more interesting and powerful results are being achieved with Clearstone Elements as a foundation, which will be discussed in future articles. The methods discussed here will pave the way for a new industry standard for all varieties of patent claim analysis since they represent the correct analytical approach (not to mention how enjoyable it is to interact with the application and watch the patent counter drop!). We hope you join us on this exciting journey.