

EPISODE 385

[INTRODUCTION]

[0:00:00.5] JM: A patent allows an individual or a company to lay creative claim to an invention. A patent can provide protection from having its idea being used without giving credit to its creators. Of course, the downside is that patterns can be filed and then not turned into products, inhibiting innovation. Patents can also be used offensively in a practice known as patent trolling. Large companies like IBM and Google have so many patents that they have trouble keeping track of them all. If your company has many different hardware and software products, how can you be sure that your patent collection protects you from a patent troll? Are you sure that you have patent coverage over all of those products?

Nicole Shanahan is the CEO of Clear Access IP, which is a product that indexes patents, looks for vulnerabilities in a corporation's patent strategy and finds opportunities in a patent collection for further value. The large text corpus of a patent collection is the perfect place to apply machine learning. We discussed the nature of patents, the intersection between law and software and product development process of ClearAccessIP.

[SPONSOR MESSAGE]

[0:01:25.2] JM: In the information age, data is the new oil. Businesses need data and there's no better data than real time data, which is why Amazon Web Services built Amazon Kinesis; a powerful new way to collect, process and analyze streaming data so that you can get timely insights and react quickly to new information. Here's the thing, websites, mobile apps, IoT sensors and the like can generate a colossal amount of streaming data, sometimes terabytes an hour. That, if processed in real-time can help you learn about what your customers, applications, and products are doing right now and respond right away.

Amazon Kinesis from AWS lets you do that easily and at a low cost. With just a few clicks, you can start sending data from hundreds of thousands of data sources simultaneously. Loading in real-time, it lets you process and analyze the data and take actions promptly. All you need to know is SQL. Kinesis also gives you the ability to build your own custom applications using

popular stream processing frameworks of your choice. With Kinesis, you only pay for the resources that you use. There are no minimums, no upfront commitments. To learn more about kinesis, just go to kinesis.aws, and let's get streaming.

[INTERVIEW]

[0:03:03.1] JM: Nicole Shanahan is the CEO of ClearAccessIP. Nicole, welcome to Software Engineering Daily.

[0:03:07.7] NS: Thank you. I'm happy to be here with you.

[0:03:10.8] JM: Today we're going to talk about some different areas of patents and patent management and how that relates to software as well as the engineering behind the software that you're building at ClearAccessIP, and I want to start by talking about just the different players who would be concerned about patents in the software space.

I want to talk about individuals, and corporations, and the lawyers who might be involved under different circumstances in software patents as well as the patent industry as a whole. Let's just start with an individual. When an individual files a patent, whether they're part of the company or not, what are they getting? What are they trying to get from a patent?

[0:03:58.6] NS: Fundamentally, they're trying to get a temporary monopoly to make, build, own rights to — In a commercial setting, the ability to implement that invention. In the software context, it's a bit fuzzy because sometimes it's hard to understand exactly how a software method is used. There are some that are more obvious. Software patents are usually visual, so the swipe to unlock technology in cell phones, that is a portfolio of patents.

If you look at where the smartphone, that's a large area of litigation between largest handset manufacturers. If you look at that space just alone you see several individuals who have actually patented and owned patents. If you look at the full scale ownership there's quite a few diverse players there ranging from individuals, to corporations, startups, and third-party patent holders also known as patent trolls.

[0:05:14.6] JM: Let's take the swipe unlock example. You said there's a bunch of different patents involved. Can you maybe give me an explanation for how the different participants in that patent would get their remuneration for that patent or how the origination of when somebody implements the swipe to unlock feature and all these different people who have skin in the game when it comes to the swipe to unlock feature, they have the skin in the game patent-wise, how are they finding out that they have skin in the game or how are they letting their claim to how much ownership they should have over that sort of patent?

[0:05:57.4] NS: That is a wonderful question and that's really the question of our age and it's one of the reasons why I created ClearAccessIP. It is incredibly hard to navigate the patent thickets that exist today. A patent thicket is defined as an environment in which multiple parties own different pieces to a product. For example, we have Apple, Samsung, Google as top players and those are the ones that have the money and the finances to really build up quite robust portfolios. Not only build them, but just understand the nature of them. Understand their competitive landscape. They hire out numerous vendors, law firms, search firms to help them conduct these landscape analyses of just similarities between their portfolios and existing portfolios that don't belong to them.

As they go through that, they identify risk effectively. This is exposure infringement risk. That's a very difficult way to create, if you want to look at it from macroeconomic standpoint, a functioning innovation environment and.

When you look at kind of the future of innovation, we're going to have very complex robotics. If you think the cell phone was complex, think again. Things are going to get significantly more complex and the thickets, IP thickets, are going to be significantly more difficult and expensive to navigate. Does this mean that smaller players are going to effectively be outcompeted just from the get-go because they don't have the finances to conduct these sophisticated landscape analyses? If nothing happens, if nothing changes to the industry, yes, we're going to have a very, very difficult environment for small companies to grow in. That just doesn't seem fair given our patent system was created in 1790 to really stimulate innovation in the useful arts.

Our patent system has really struggled to keep pace, honestly, since the bicycle was invented in 1920. Before, the first patent thicket was in 1850 over the sewing machine. Where we are right

now is I really believe is an inflection point for patents and it really stems from a lot of the difficulty between software and hardware of understanding how all the pieces fit together correctly.

[0:08:42.5] JM: Let's confront the question of whether it is even appropriate to care about patents, because I think there's a contingent of software engineers, at least, maybe free information hacker type people in the more abstract sense who will just say, "You know, if you care about patents, if you're even thinking about patents, that means that your anti-creativity. You're part of the man." What you're your thoughts on that deposition? Can you be a creative inventor type of person and just kind of ignore the patent system altogether?

[0:09:20.0] NS: Again, another great question. It's kind of a personal decision. I have clients that really do put faith in the patent system, but that usually doesn't happen until they raise a series a, because until then they're just trying to survive as a company.

It's interesting because the correlation between healthy finances and aggressive patent strategy go hand-in-hand. If a party is bootstrapping, I oftentimes see them take that former stance of, "Oh, the patent system is silly. I want no part of it. Software patents shouldn't be allowed anyways." The minute they have something to protect and has real commercial and revenue upside, I see them changing their tone a bit.

It kind of goes back to that old saying that I think our parents used to say that you don't become a Republican until you start making your own money. There's a bit of similarity here, is you don't become a believer in the patent system until you really have something to lose.

The way I tend to look at it is slightly somewhere in the middle. I think that the patent system is a very important system for attribution, and attribution is one of the largest simulators of creativity. If you just had rampant theft and a zero attribution of full community system, a full open source system that really lacked any transparency into who created what, people would likely be far less excited to invest in the hardest to make breakthroughs.

[0:11:09.8] JM: Do you really think that's true? I think there's a lot of inventors throughout history where they are not motivated by the money or the notoriety. They're really motivated by

the discovery and they don't actually care so much about the upside, the financial upside or the notoriety upside. They're really more just curious about things. I don't know. I guess you can design a patent system that would incentivize the people who are financially driven who do want their ideas protected and the people who are motivated by something other than notoriety and money. Those people, they're basically protected regardless. All they have to do is think up the idea and if they get infringed upon, they don't care regardless. It doesn't really matter what the patent system is as long as it doesn't somehow prevent their inventive activities.

[0:12:04.5] NS: Yeah. Again, that's a really interesting way of looking at it. What is the motive behind our innovators? It really is diverse. We have a system, a government system that is really a one-size-fits-all system, but the reality is that we have a very fluid invention system in the world today. How do we create a patent system that really can be fluid and remain fluid such that someone might create something and open source it one day, but the next day they may wish to patent its counterpart. Or in another scenario, can we allow someone that owns a patent to then open source for a period or open source it to a nonprofit group? This is known as pledging your patent.

Pledges really have only taken shape over the last decade. If you look at some of the pledges in effect today, Tesla, for example, pledged several of their battery related patents to stimulate innovation. However, the likelihood that Tesla can get away with pledging all of its patents is — I'm very skeptical of that in the long term, because inevitably there could be some kind of malicious actors that go after Tesla, in which case enforcing their patents would be necessary.

I think that we are a very sophisticated people and we're capable of really reordering the patent system to fit the nature and the diversity of our innovators.

[0:13:48.0] JM: Large companies, you sometimes see — There's like insiders the tried to report on what's Apple doing next or what is Google doing next and the insiders are always talking about the patents that Google filed recently or Apple filed recently, like, "Oh! Here's a patent that allows toddlers to talk to robots or something." It's just like, "Wow! Okay, weird business —" I can't fathom what product of limitations this is going to lead to. Can you help me understand the internal process of a large company and when a large company processes patents, when they

file patents, how they manage their patents and where does it take place in the product development cycle?

[0:14:34.7] NS: That a wonderful question, because there are similarities across all large corporations, but there are definitely some unique characteristics. Generally speaking, we have under the American Events Act that was implemented in 2012, a system that requires companies to be the first to file. This is a change from the pre-2012 act and that what we used to be a first to invent system, which just meant you could invent if you had proof that you invented, you had proof that it was in use. You have proof that it was somewhat known, then that would serve as prior art. You would be able to have protection if you had filed from that first to invent date.

Today, we are in a first to file regime, which is on par with effectively the rest of the world. Under first to file, you really have to expedite the filing and the recordation of that invention. Large corporations such as Google and Microsoft, IBM, Amazon, they have a very vigorous process of collecting information from their product teams, from their engineers. They get them in the format of invention disclosures. An invention disclosure often times is a few pages that describes kind of the background of the invention details of what makes this invention unique, why it's unique, and then a listing of the core steps of that invention. That invention disclosure than usually goes through a review process with either the in-house IP counsel or sometimes a head of R&D will review it, and the decision will be made whether or not to invest in the filing fees.

Generally speaking, they start with provisional filings, which give you about 10 months to really get your act together and understand fully how this invention is going to be packaged in a full patent filing. Provisionals are very, very cheap to file. When I meet startups or small companies, it's \$130 filing fee, government filing fee. You can usually get away with five or six pages of a written description and one or two drawings. Then he gives you about 12 months to figure out if this is an invention that you're really interested in investing in.

[SPONSOR MESSAGE]

[0:17:24.6] JM: As applications become more distributed and sophisticated, managing and maintaining them becomes more challenging. Running one of the world's largest PaaS platforms, Heroku understands the value of operations and the opportunity cost of developer time. Listen to the live podcast with Heroku engineering on February 28 to learn about Heroku's metrics platform architecture and how it laid the foundation for auto scaling. We'll talk about the importance of maintaining application health and best practices for monitoring application user experience.

I love Heroku both as a user and because they are a sponsor of Software Engineering Daily, but I want to emphasize that I do use Heroku in my own projects. I really am a fan and I'm looking forward to doing some continued episodes about Heroku.

[INTERVIEW CONTINUED]

[0:18:25.9] JM: Okay. I guess let's get to what ClearAccessIP does, because this is basically the challenge of — At least from the perspective — We've talked about some different roles, like you're an individual inventor. A patent system means X to you. You're a giant company, the patent system means Y to you. In any case, if you are a person who is inventing stuff, there's a likelihood that you will engage with the patent system in different ways. Patent management is something that becomes relevant. Explain what ClearAccessIP does.

[0:19:06.0] NS: Sure. We're the first integrated platform that does the collections. The invention disclosures which I spoke about earlier. We have an invention disclosure repository, and that is a touch point for the first seed of an idea. Then we have all of the filler pieces in between. We have a platform that keeps track of the provisional filing that then notifies you of the deadline to file the non-provisional. We have a computational logic AI. It's a clerical AI that then compiles and collects and stores and docket. Docket is a legal term for calendaring, calendars deadlines, and that saves ton of money and time just on administrative costs.

Finally, it's now integrated with an IP deal room that provides that flexibility. I said it would be great if we had a patent system that could provide something to everyone irregardless of what their motivation is. IP deal room has an incredible amount of flexibility to package your portfolio into either a strategic portfolio which you will use to protect yourself but you would never enforce

against anybody else. You can pledge. That's open sourcing the patents. You can license it or cross-license it with a partner or even a competitor.

Within IP deal room, we put in a machine learning analyst. It's a piece of AI software that reads the patent portfolio and draws insights into the landscape. Earlier, I said it's very expensive for parties to do this today such that only really the top 1% of companies have the budget to do that level of analysis. We're trying to even the playing field a bit and bring down the cost of that landscape work so that small and midsize companies now have access.

[0:21:16.4] JM: Now, in order to do those kinds of things where you're, for example, pledging a patent, do you need to integrate with some sort of government database that has index of pledged patents?

[0:21:31.7] NS: That has been such a hot topic. There is no comprehensive registry of pledged patents. There was one attempt at University of Washington, I believe — No. I'm sorry. It was George Contreras. He's a professor of law and he attempted to build a database of all pledges in the United States. He did it just as a private ambition, but there's no government centralized registry of these things. Keeping track of them is something that ClearAccessIP aims to do enabling parties to pledge their patents is something IP deal room is set up for. In terms of a government registry, there actually is none today.

[0:22:24.3] JM: In real estate, I know there's this thing called multiple listing service. I don't remember exactly how works, but I think it's basically at one point some large forward-thinking corporation decided, "Let's just build a big database of the real estate listings across the United States," and nobody else was doing it so they just built up this huge business because if you are a real estate agent or you're a broker, you have to use MLS. Then today you see these newer companies, like Redfin that are encroaching on the MLS business and I think they're building their own databases of stuff. It sounds like there's not even an MLS of patents. Is that an accurate analogy or do you have any historical context for the MLS business?

[0:23:13.3] NS: I do have context for the MLS business. I bought my first house last year.

[0:23:20.6] JM: Was that an accurate description of it?

[0:23:22.5] NS: Yeah. It is a formalized repository. It's a standardized format for representing a real estate opportunity, a real estate sale usually. Sometimes you can find rental listings I think on MLS. That's a great analogy to what we're missing in the patent business. There is no central repository that explains what the patent holder is doing with their patents. You can pretty much look at real estate and say, "Oh, that individual doesn't want to sell," or they'll be a signpost that indicates, "Okay. This homeowner does want to sell."

We don't have anything like that in the patent space. When you do these landscape analyses, you get a list of relevant patents and patent holders that relate to your line of business. However, you have no idea whether or not they're planning to litigate or they're aggressive or they need the licensing revenues from their IP. There's no way to really shop for IP today.

Part of the difficulty is it's cultural. There is a lack of trust between parties such that even contacting a potential buyer can come off as highly antagonistic. That's because there's been in a quite a bit of patent trolling going on over the last decade.

That being said, it would be great if we could have a central repository where all parties communicate with each other kind and kind of standard terms. One of the goals of ClearAccessIP when I started it four years ago was to really help patent holders tell their story through a standardized method of communicating.

[0:25:20.9] JM: That sounds great. It makes me wonder what the state-of-the-art is outside of Clear Access, because when I submit a patent, presumably, some government agent sees that patent goes, "Okay, great. I'm going to find out the ways to cross reference that with the patents of the past," but if there's no good way of doing that search, then what's to prevent duplication of patents and crossovers and errors in the patent filing process? If there's no centralized database, how are they even doing that?

[0:26:02.3] NS: There is a database of the published patent itself. That is actually available, actually, via a congressional mandate in the United States that are patents as they are published and their records, the prosecution history. The history by which you communicate back and forth with the United States Patent Office and argue for the novelty of your invention,

that processes all public. The patent itself is also public. What isn't known and what creates this layer of mystery and this layer of risk is nobody really understands or has the capacity to read all of the patents and derive an accurate view. That's where a lot of the skepticism of the patent system comes into play.

There's a lot of overlap between one person's patent and another. Examiners, U.S. PTO examiners, are just people. For them to be able to read hundreds of thousands of patents in a space and derive a very accurate view of whether or not a new patent truly is novel is not possible today. They say a lot of the stuff is more of an art than a science and that kind of leads us to this very precarious area where on one hand we're juggling really what is black letter law, and on the other hand human error. Enormous amount of human error, and that human error is just due to capacity. You really need a very sophisticated, I think, computer, computing capacity to properly implement the black letter law that we have for patents today.

[0:27:56.6] JM: The human error, are you talking about human error at even entering the data in their patent, like in the government's patent system? Are you talking about human error when patent, potential patent comes across a person in the government's desk and they're trying to figure out if this is a new patent or if this is old hat? What is the human error that you're referring to?

[0:28:20.7] NS: Really, the latter. It's when a new application comes across the desk of an examiner at the U.S PTO, that examiner is responsible for applying the statutory code to that patent. That individual has to do a very thorough reading of that patent. The U.S. PTO has its own search capabilities that returns potentially invalidating patents, but that whole process of arguing for what are very narrow claims in some instances. Then that back and forth from the patent attorney that says, "Actually, no. This this sentence means this. It doesn't actually mean what that other patent says."

Definitionally, we're creating very complex definitions and very nuanced and sometimes if you think about the potential for misinterpretation just in regular everyday communications and, now, compound that by the complexity of our technology.

[0:29:30.2] JM: I want to understand more what ClearAccessIP does. If I'm a giant corporation, like Intel, or Google, one of these companies that has a giant suite of patents. How am I using the software that you're developing?

[0:29:48.0] NS: Sure. You could use our software in a number of ways. We do partial and then we do full implementations, because the industry is very much in flux today, which means that some of the larger corporations, they're such large moving animals that it's hard for them to really change rapidly. We do oftentimes partial limitations for corporations where we go in and we say, "Okay. You need help with just data management." We have an automated clerk that can eat take all of your data and quickly organize it without a single stroke of the key on the keyboard.

Other corporations will use us for corporate development research. In that case, they'll be using the machine learning analyst and IP deal room to either review an outside party's IP or review their internal IP and get a sense of what they're missing. Any holes that might exist in their portfolio, our system is very good at identifying.

We like to identify the things not in terms of risk, because I really hate that word. I really hate that we have an innovation system —

[0:31:00.7] JM: I'm sorry. Sorry to interrupt you, but you said a hole in a portfolio. Is that like if I have some chip that I'm trying to patent or I'm trying to make sure I have patent understanding of it and Clear Access is good at identifying, "Oh, there's actually a hole in this," and you need to shore this up with some other patent?

[0:31:22.0] NS: Exactly. It's effectively an area that they do not have freedom to operate in and their portfolio does not give them that freedom to operate. It gives them other areas that allows them to execute on the technology, but fundamentally as a whole they're missing certain core steps.

[0:31:40.5] JM: How could you identify that?

[0:31:42.0] NS: How we identify that is our AI can read that family group of patents? The user will select their portfolio that relates to that product line. Let's use an example here. Let's say that there's a robot that does human mimicking and they're trying to put on the market a robot that has various facial features that really mimic the human face. They have a portfolio of patents that is probably combination of software and hardware. As they're looking at it and putting it into IP deal room, our system reads it all and says, "Okay. Here are all of the patents that will enable the human face mimicking." Then it will show you how your patents relate to this pool of patents and it can show you gaps that other people have filed in that you have actually not built out your portfolio.

One example could be a neural network sequence that is able to respond in a robot human interaction and say that's missing because the company focused more on, for example, mouth twitches or smiles. I mean that's just one example, but there are outside of robot and even in healthcare, quite a few others in which you can have a medical device that is very good on, for example, a delivery of a medication but not the housing of the applicator.

[0:33:24.4] JM: I know you're not an engineer, but can you describe how the data ingestion process — If feeding — Let's say I'm integrating with Clear Access and I'm one of these giant corporations and I feed my 2000, or I don't know how many patents a giant company has, 500,000 patents, a million patents into this Clear Access system, what's the process for processing those and indexing them and getting them into a state where you can learn about what products you have coverage over?

[0:34:01.4] NS: Sure. The process, all we need is the application serial number. Even at a large corporation, we're only looking at kind of a portfolio in the thousands. Potentially, the tens of thousands. 3M's portfolios like 7-1/2 thousand, and they're a massive corporation that's been around forever.

[0:34:25.9] JM: It's like 7-1/2 thousand and it's like, what? Seven pages each are 20 pages each?

[0:34:31.2] NS: They're roughly about 20, 25 pages on average. The system just needs the application serial number and an individual can key that in very quickly. Our system grabs

everything else. They then determine that they'd like to create an analysis through IP deal room, so all those numbers get bundled together and then there's a little button up top that says enable machine learning, and they just slide it into the on position.

At that point, our machines go through and read that entire portfolio and then it takes that reading and it compares it to the entire cluster of related patents. It uses a neural network process which has already been trained on the entire patent corpus and then it returns the top similarity results. It also looks at kind of how the organizations compare to one another. It can pull back all of the other companies that are doing work in this space in addition to their specific IP.

[SPONSOR MESSAGE]

[0:35:52.3] JM: Hosting this podcast is my full-time job, but I love to build software. I'm constantly writing down ideas for products; the user experience designs, the software architecture, and even the pricing models. Of course, someone needs to write the actual code for these products that I think about. For building and scaling my software products I use Toptal.

Toptal is the best place to find reasonably priced, extremely talented software engineers to build your projects from scratch or to skill your workforce. Get started today and receive a free pair of Apple AirPods after your first 20 hours of work by signing up at toptal.com/sedaily. There is none of the frustration of interviewing freelancers who aren't suitable for the job. 90% of Toptal clients hire the first expert that they're introduced to. The average time to getting matched with the top developer is less than 24 hours, so you can get your project started quickly. Go to toptal.com/sedaily to start working with an engineer who can build your project or who can add to the workforce of the company that you work at.

I've used Toptal to build three separate engineering projects and I genuinely love the product. Also, as a special offer to Software Engineering Daily listeners, Toptal will be sending out a pair of Apple AirPods to everyone who signs up through our link and engages in a minimum of 20 work hours. To check it out and start putting your ideas into action, go to toptal.com/sedaily.

If you're an engineer looking for freelance work, I also recommend Toptal. All of the developers I've worked with have been very happy with the platform. Thanks to Toptal for being a sponsor of Software Engineering Daily.

[INTERVIEW CONTINUED]

[0:37:56.3] JM: I see. Okay. Basically, this is a really useful tool for the IP counsel inside of 3M, for example, if they need to figure out, "Okay, does our new 3M —" I don't even know what 3M makes these days, "If our new chip —" What does 3M make? Anyway. Okay, fine. Our new battery. I don't know. I forgot what they make. They want to say, "Okay, do we have our new battery covered patent-wise?" They can basically do a measurement of the things that — They can take their battery and think about the different things that go into making that battery, and Clear Access will figure out what patents 3M owns and then do some that are related to those different pieces of technology in the battery and then doing some similarity measurements against the entire patent corpus so that you can basically have a larger dataset to kind of cross-reference against, and then the output of that would be, "Okay. Basically, here are the documents that pertain to the riskier portions of that," and this would allow the IP counsel to look more closely at those other documents. It's not going to spit out the answer, "Yes or no, you are patent protected," it's more going to say, "Here's the information you should look at to research this matter more closely."

[0:39:28.6] NS: That's right. Again, we try not to present this in terms of risk. We're really trying to present this terms of potential collaboration or just internal — Really getting more precise on internal, because there's a fine line between what is risk and what is strategy in the patent industry and we don't want to come off as a system that promotes the traditional risk assessments exposure. We want to, instead, have a system that creates transparency and opportunity.

One of the reasons why we present to the data the way we do, which is not, "These are your biggest competitors. Watch out for these patents." We could do that very easily, but I don't think that's conducive to what many of these people ought to be doing. I think that they should be looking at the invention landscape as, "Okay. They found that this approach helpful or they're doing something very similar in the space. Is that the route we want to go? Is that the bet we

want to make?" If it is, then, okay, well this is an opportunity to cross-license or to create a patent pool or potentially build some goodwill if we all just do a consortium of pledges. It's a slight mind shift change but it really makes all the difference, and our tool is trying to provide that piece, that critical piece of information at a critical time in the decision-making process.

[0:41:13.1] JM: I guess you probably don't want to build a tool that patent trolls would look at and say, "Ooh! What's our next target?"

[0:41:23.3] NS: Exactly. We have actually been approached by a few patent trolls and they've asked for different ways of tweaking the algorithm to help them be more effective and we just have to say, "No. That's not what our mission is."

[0:41:41.1] JM: You've said no to potential customers, because you just don't want to be associated with billing patent trolling tools.

[0:41:48.7] NS: Yeah. ClearAccessIP was created with a very specific mission in mind, and it would be antithetical to that mission to change our product in any way to make a patent troll more effective.

[0:42:03.6] JM: What do you think of these patent trolls personally?

[0:42:07.4] NS: Personally. I personally care about the integrity of our patent system, because I think that a structured way of storing human invention is valuable. I think that patent trolls are just like angry drivers on the road that think that those lanes that driving on the road was made for really one use, which is to go fast and to get to where you need to be as fast as possible irregardless of basic decorum on the road. Just like traffic trolls, patent trolls kind of just — I think they just got the wrong idea. What they're doing may not be illegal, but it certainly is disrespectful to what this system was built for.

[0:43:04.1] JM: Yeah. Talk some about the software development process at Clear Access.

[0:43:13.0] NS: Sure. We really are a software company, even though when we talk about what we're doing, it's really service oriented. We're definitely in enterprise software as a service

business. However, engineering, we're taking kind of the latest papers and latest breakthroughs in machine learning and applying them in our product and testing them to see if they even work for what we're trying to do.

For example, our machine learning analyst is based on a paper that was released out of Stanford just two and half years ago. A lot of work has been done off of this paper and we've really had that benefit of learning from researchers in and machine learning. We're not relying on kind of old methodologies of just computational logic and applying big data. We definitely benefit from these things, but really where we put our engineering focus and our engineering dollars is taking really — For example, we figured out how to apply computer vision algorithms to text, natural language text, and that's given us quite a bit of leverage in accessing at a low price analytical products and making those available in mass to our users.

[0:44:41.2] JM: These are for patents where the document has been scanned but it's not — Like optical character recognition hasn't been applied to it, so you have the capacity to process it and actually get the texts of those photos.

[0:44:57.3] NS: That's the one use scenario. The other way that we look at it is a lot of the computer vision algorithms were really good for drawing comparisons based on mathematical measurements. What we did with our patent corpus was we turned all of that text into effectively equations. How one concept is related to another concept rather than doing keyword based analysis. We do neural network level analysis, which requires us to turn that text, really, into numerical values. We measure than the numerical relationships between concepts versus saying, "Oh, these are the same keywords. Let's cluster them together." We say, "These concepts map at this waiting," and that way we can go through the Patton corpus with a greater degree of insight than just merely clustering words together and waiting the word value in a document.

[0:45:59.9] JM: How do you train that conceptual graph?

[0:46:03.9] NS: Some of the stuff we haven't actually made public yet, but at a high level, we've gone through various stages of training where we've had to, first, parse down and vectorize the entire patent corpus and work up from there. It was really a process of figuring out how to

decouple each patents file into the small concept parts and then we trained it actually based on the computer vision discovery that allows us said then pair parts of the patent filed to other parts of other patent files. That gives us that really deep understanding of why certain concepts are similar versus just the whole document being similar.

Really, what that is mimicking is the human analyst saying, “This is a part that relates to this other part, but the other parts don't match the same way and that's why these two patents are different,” rather than simply saying, “These patterns are similar.”

[0:47:19.9] JM: If you want to create a conceptual map for a particular patent, it seems like you would need to do some sort of human labeling for that to create that conceptual mapping for the patent, otherwise you would just have to go through the patent and do those some kind of textual count of different words. How are you getting the labeled data for a given patent?

[0:47:46.4] NS: Again, we're using machine learning to help us out with that. It would take simply too long to hand tag every patents in the world. There's about 90 million patents worldwide and I don't know of any clerical turk out there that would be able to do it.

[0:48:06.9] JM: Certainly. No. I understand. My question is more like how do you — For a given patent, if you're — You could just count the words and say, “Okay. This patent mentions the word swipe with some like TF IDF value, like a high TF IDF so you know this is a particularly related to swiping. This document is particularly related to swiping.”

The reason I asked is I think you mentioned you try to do something beyond just kind of the pure text counting. I was just curious if the conceptual mapping is really based on that kind of thing, like the traditional TF IDF stuff or if you're doing something different than that.

[0:48:58.5] NS: We are doing something slightly different. We turned each word into a vector, and through the vectoring process, what you do is you kind of create a semantic pairing that is much more fine-tuned than just waiting the word as its own object. Every vectorized word is in relation to all the other vectorized words that that word was seen around.

[0:49:35.1] JM: Okay. Interesting. That makes it kind of a conceptual thing compared to — Okay. Because the naïve TF IDF stuff, that's more of just how does this word — How frequently does this word occur in the document relative to how frequently it occurs in the average document within the corpus. That doesn't give you any impression for how it relates to the words around it in a given document.

[0:50:06.8] NS: Yeah. That method is used, but it's treated as a very superficial measurement, because fundamentally it is a very superficial measurement. We take that and we say, "Okay. How do we turn it into a more active interpretation?" Meaning, rather than just saying, "Okay. Now, at a high-level, it's seen it's waited this much." Why? Why is it seen and waited this much? It's seen and it's waited this much because these other enabling concepts surround and because we're able to look at this in terms of a very dense collection of concepts that all work together to achieve this certain next dense concept of concepts, and then we can then look at this in terms of, "Okay. There's a nucleus and these are the cytoplasm and all these other membrane parts." We look at in terms of human biology of how cells work together.

[0:51:14.0] JM: Okay. I know we're nearing the end of our time. I wanted to just ask you a few questions about business strategy because you're in an interesting position relative a lot of other software companies, because you're revenue positive. Also, this is a type of software that a lot of companies probably don't even know that they want. You're in a position where you could probably raise money more aggressively and scale up that sales team, maybe build more features that your current customers want so there's plenty of reasons to hire more engineers, or you can just kind of cruise along and figure out a way to scale up the team while also scaling up revenue. You've got some interesting sets of trade-offs, which is a great place to be. I'm curious about how you evaluate that from a business perspective, from a growth perspective.

[0:52:11.0] NS: Yeah. That's a great question. Our company falls into the legal tech category, and if you look at some of the investments, there haven't actually been that many into legal tech. If you look at the nature of some of the other ones, they were really based on this idea that you can you can collect a lot of legal related information, package it differently and then potentially sell to one of the large BMSs in legal tech space, such as Thomson Reuters or LexisNexis.

Our company is slightly different. It is, in some ways, very much a legal tech company, but in other ways it can actually do corporate strategy, so then it allows us to present ourselves to our users as being both a legal tech tool but also corporate development tool, and that means that we're a little confusing to investors, because they don't know how to categorize us exactly. We have revenue, but we are very much set up as a lean startup.

Where we are at this very moment is looking at our future in terms of how fast do we really want to scale our model and do we want to just focus on the large enterprise accounts or is there's something we can do for like almost a consumer product, such as Boxer, Dropbox. Can we be as available to implement for individuals and do we go that route or do we strictly stay to SaaS? These are things that we're working out right now and trying to figure out, and it's an exciting time. Our team is just, I think, one of the most inspiring groups. We're here in Palo Alto and I think that's we're not trying to hurry to raise a series A or rush to exit. We're really just enjoying the fact that we have this privilege to make and make available a very sophisticated low-cost patents AI.

[0:54:23.2] JM: Cool. That sounds like a good place to close off. It's been really interesting talking to you and I find the legal tech space quite interesting. I will be watching Clear Access closely, because the business is evolving in quite an interesting way. I've looked at some of the other legal tech businesses, I don't get seen anything exactly like Clear Access, so it will be cool to watch.

[0:54:49.4] NS: Thank you. It's been a great chatting with you, Jeff.

[0:54:52.0] JM: Okay, thanks Nicole.

[SPONSOR MESSAGE]

[0:54:56.2] JM: At Software Engineering Daily, we need to keep our metrics reliable. If a botnet started listening to all of our episodes and we had nothing to stop it, our statistics would be corrupted. We would have no way to know whether a listen came from a bot, or from a real user. That's why we use Encapsula to stop attackers and improve performance.

When a listener makes a request to play an episode of Software Engineering Daily, Encapsula checks that request before it reaches our servers and filters bot traffic preventing it from ever reaching us. Botnets and DDoS are not just a threat to podcasts. They can impact your application too. Encapsula can protect your API servers and your microservices from responding to unwanted requests.

To try Encapsula for yourself, go to encapsula.com/sedaily and get a month of Encapsula for free. Encapsula's API gives you control over the security and performance of your application. Whether you have a complex microservices architecture, or a WordPress site, like Software Engineering Daily.

Encapsula has a global network of over 30 data centers that optimize routing and cache content. The same network of data centers that is filtering your content for attackers is operating as a CDN and speeding up your application.

To try Encapsula today, go to encapsula.com/sedaily and check it out. Thanks again Encapsula.

[END]