



## Case Tags and Outcome Analytics: Essential Functions for Litigation Analytics

Litigation analytics are most effective when you are able to find cases like yours and then quickly understand what happened in those cases. Lex Machina's case tagging and Outcome Analytics are just two examples of ways that you can determine what happened in the most relevant cases to your matter. Join Rachel Bailey, Data Relations Manager, and Marissa Skudlarek, Software Engineer at Lex Machina for this 30-minute webcast recording, where they discuss what it takes to create tagging systems and Outcome Analytics and how that benefits your practice.

### Speakers:



Rachel Bailey  
Data Relations Manager  
Lex Machina



Marissa Skudlarek  
Software Engineer  
Lex Machina

Rachel Bailey ([00:00](#)):

We are talking about case tags and outcome analytics, essential functions for litigation analytics today. Thanks for coming. A little about Lex Machina. We create legal analytics, which helps practitioners win in the business and practice of law. We will dig into some of the specifics of what that means today. And our users include some of the most well known companies and law firms in the world, including three quarters of the AM LAW 100. For this webcast today, please submit questions in the Q and A window. And we do plan to have time to answer those questions at the end.

Rachel Bailey ([00:44](#)):

So our speakers today, I'm Rachel Bailey. I am our Data Relations Manager at Lex Machina. A little bit about my background. I am coming to you from my home office in New Orleans. And before practicing in Louisiana, I practiced trademark law in the Bay Area. And so when I joined Lex Machina, I started working on trademark, trade secret, and copyright data. And so I'm one of the people who does human review, which you'll learn all about today. And I look at cases on an almost daily basis. I'm joined by one of our incredible engineers, Marissa Skudlarek. Marissa, why don't you tell us a little bit about yourself?

Marissa Skudlarek ([01:28](#)):

Hi. Yeah. So I'm coming to you from San Francisco, just a little bit north of our corporate headquarters in Menlo Park, California. As Rachel said, I'm a software engineer. I've been with Lex Machina for a little over two years, and this is actually my first engineering job. Previously to this, I worked for several years at an immigration law firm as a paralegal and in their knowledge management department, but then I was ready for a career change. I took a coding boot camp in order to make the switch into software engineering, as so many people in the Bay Area do.

Marissa Skudlarek ([02:07](#)):

And then I got hired at Lex Machina a couple months after completing the coding bootcamp. And I think partly it was that I had legal experience at the immigration law firm. And I think partly also at the bootcamp, we had to do a final project to demonstrate our skills, and mine had to do with text searching and finding word patterns. In that case, it was in public domain, literary fiction, because I'm also interested in books and theater and that whole side of life. But you will see, I'm going to be talking to you about case tags, and developing those also has a lot to do with text searching and text patterns. So I really think my background has prepared me very well for my role here in the engineering team.

Rachel Bailey ([02:57](#)):

Awesome. Well, let's get started. We'll start talking about case tags.

Marissa Skudlarek ([03:02](#)):

Great. So tags, if you've used our product, you may have seen that we offer various ways to drill down in the data and categories that we offer to see a group of cases that all have something in common. We refer to these as tags. So I'm going to be telling you a little bit about how we make sure that our tags are accurate in allowing you to sort the data in ways that are useful to you.

Marissa Skudlarek ([03:36](#)):

So the process of tagging involves two steps typically. There would be an automated query, which engineers like me will help develop in conjunction with the data team who understands the legal aspects of things better than we do often. And then after the automated query runs, there will be verification by a data team member. Rachel said that she sometimes does that. We have lots of people working on that to make sure that the automated tag is accurate.

Marissa Skudlarek ([04:09](#)):

So we offer lots of different tags. I'm going to be speaking today about federal district court. And we have what you could call general tags and module specific tags. I've given some screenshots of our software and these are the tags that you can filter with. So the general tags are things that can apply to any case from any different module, things like MDL, whether it went to trial, if it's a class action, that kind of thing. And then we have module specific tags. I've shown this, this is the list of tags that we offer for consumer protection. Now each module is actually itself a tag, like consumer protection is a tag, and then we have these sub tags, we sometimes call them, under consumer protection. And so those are specific legal concepts relating to the specific module of consumer protection.

Marissa Skudlarek ([05:09](#)):

And the reason I pulled up consumer protection is because in order to explain tagging, I thought it would be good to look at one specific tag and how we developed it. And we're going to be looking at data breach today. So data breach, as I said, is a sub tag within consumer protection. It was actually not part of our initial consumer protection offering when we first made that available to customers, but that was an oversight. A lot of our customers requested it. And so we realized we needed to develop a tag to allow people to sort and find data breach cases. We implemented it about a year ago. So it's been available for about a year. And as you might have seen on previous slide, about 3000 cases, little over 3000 cases, have the data breach tag.

Marissa Skudlarek ([05:54](#)):

This is the data breach tag definition, actually copied and pasted from the help section of our website. I don't necessarily expect all of you to be data breach experts. For me, the important thing to know is the precisely what data breach means. For you, what I want to say is we need to pin down the definition of what a tag should be looking for before I, as an engineer, can start working on it. So the people who work on consumer protection need to say, "Okay, what actually is data breach and what is not data breach?" And have a really good sense of what should and shouldn't qualify a case for that and pin down that definition. And after it's pinned down, I can start working on the automated query that will allow us to find those data breach cases.

Marissa Skudlarek ([06:49](#)):

So automated queries can be based on a whole lot of things, but especially for a lot of these sub tags in the different modules, they often rely a lot on text searching, specifically of things like complaints or pleadings, that kind of thing. And sometimes we have to, in order to have a comprehensive and accurate tag, look for a lot of different phrases, more than just a simple keyword search for data breach. It goes way beyond that. So our data breach query searches for a whole lot of different phrases and patterns. It's looking for data breach or security breach or privacy breach, and it's looking for customer information or customer data or identifiable information, et cetera, et cetera, you get the picture.

Marissa Skudlarek ([07:40](#)):

So on the next slide, this is a screenshot of a complaint for a data breach case, had to do with Uber data breach. And the phrases I have highlighted in yellow are phrases that are going to ping the query that we ultimately develop. So you see there's a lot of different things it is searching for. So important things to note, as I said, this is far more complex than just a keyword search. And we're able to do things as well with the search software that we use, like look for this phrase, but only if it's in with the first X words of the document, or look for this phrase, but only if it's within X words of this other phrase. You can get some very complicated patterns there.

Marissa Skudlarek ([08:30](#)):

The other things to note is all this is actually based on is the text of the complaints. We are not limiting it by nature of suit code. We are not limiting it to cases that were previously tagged consumer protection. So really it is searching all of the pleadings in our database for all of those data breach related phrases and patterns. It is looking at over 6 million documents and finding the few thousand that are from data breach cases.

Marissa Skudlarek ([09:00](#)):

But you might be asking, "Okay, so you have this really complex query. You're looking for all these phrases. How do you know whether it's accurate? How do you know whether you've searched for everything you might need to search for? That you're not leaving things out? And also that you're not being overly inclusive, that it's not hitting too many false positives." So what we do is we will validate the query as we're developing it. We will validate it against a sample set of cases. Few hundred. The data team will have picked out those few hundred cases, made sure it includes both true data breach cases or whatever we're developing, as well as ones that are definitely not. You need to have both true positives and true negatives in order to get an accurate read. And we have some internal tools that will allow us to sort of measure the precision and recall of that draft query against that sample set. And then we can tweak the query until the precision and recall are within what we consider an acceptable range.

Marissa Skudlarek ([10:02](#)):

And so then, once it's within an acceptable range, we can do what engineers do and put it into production, and it will run over all of our cases or as I said, 6 million documents, and apply those tags. But then especially when we're rolling out something new, a new tag, a new practice area, the data team will also review those results of the automated query and confirm that it has been applied accurately. This is a the screenshot of a case in our software. And all of those tags have been human reviewed. Because as you saw, the computer, the query, is pinging on certain phrases, but every human writes differently. There could be the possibility that some complaint has matched the query, and it is not actually a data breach case, but it requires holistic human understanding to read that complaint and make that judgment. But the query has at least brought it to our attention, and we can then do that final evaluation validation step.

Marissa Skudlarek ([11:19](#)):

So this is actually an example for data breach. This is a screenshot of one of the internal tools I use when I'm developing a query and it shows it's sort of running it, but not actually applying it. It's sort of a dry run and it shows how many it would match, just over 3,800. And also you see it's very fast. It runs in less than a minute over all those millions of documents. It's wonderful.

Marissa Skudlarek ([11:45](#)):

So that's 3,800, but this is a screenshot of our software. What is in tag data breach in Lex Machina? And you'll see that's a little bit over 3,300. So that implies that the query has a few too many false positives, but that's fine. We'd like to err in the direction of being maybe a little overly inclusive and then relying on humans to narrow it down, because we really don't want to miss anything. It's much worse to miss something. And then as I said, the human review will do that final step of making the evaluations based on reading the complaints of the edge cases. Yeah. And with that, I'm going to hand it over to Rachel to talk about outcome analytics.

Rachel Bailey ([12:35](#)):

Awesome. Thank you, Marissa. I love hearing about the ins and outs. We're very transparent at Lex Machina. And so a lot of other places wouldn't tell you that specific information, but we love to talk about our process in order for our users to have the best understanding they can of what they're looking at.

Rachel Bailey ([12:57](#)):

And I'm going to talk about outcome analytics, and we're actually going to tie this into tagging a little bit as well. So what are outcome analytics? In short, they tell you what happened in the case. In Lex Machina, our outcome analytics include findings, which are court enforceable determinations, damages, which are monetary awards, remedies. And so that includes injunctions, TROs, preliminary injunctions, permanent injunctions. And then we have some practice area specific remedies as well. So for example, in trademark law, one of our remedies is cancellation of a registration. So if someone came to court and they want their registration at the US PTO canceled, we annotate that in our system as well. Resolutions, which are the way that the case terminated. And that includes who won and at what point in the case they won. And then for state court, we have rulings, resolutions, and damages for trials as well.

Rachel Bailey ([14:06](#)):

So how does this help me win cases and win business? There are two ways to use outcome analytics. One is on a case by case basis. And then the second is analyzed in an aggregate for a group of cases. So first, we will look at an example of a case in our system, and you already know this case. You've looked it up. You had a colleague who told you look up *McGlenn vs. Driveline Retail Merchandising*. That case might have some pertinent information for you. And so you can just look it up in Lex Machina by the name or the number. And then you'll come to this case page here. And we've got some tags up here, consumer protection, data breach, as we've been talking about. As well as some basic information about the case.

Rachel Bailey ([14:57](#)):

What I really want you to look at is the case resolution. So like I said, who won. In this case, the claim defendant won. And at what point in the case, summary judgment. And then we've got findings. And the findings include no negligence, no state unfair deceptive trade practices violation, no data breach because there was no contract violation, and no data breach because there was no notification or handling violation. And so that's really quick. I just went through that very quickly to get information about exactly what happened in the case. And this happened at summary judgment and it tells you who won.

Rachel Bailey ([15:37](#)):

And then any of these blue links take you somewhere else in Lex Machina. Particularly, this takes you to the document. So if you want to know exactly what happened, you can go read that summary judgment opinion. This is important for looking up a case for a couple different reasons. One, you can find out information about the judge and see why they ruled this way and get information about the judge's experience. Two, you could get information about opposing counsel's experience. And three, this might be a unique fact pattern that's similar to your own, and you want to see what happened.

Rachel Bailey ([16:16](#)):

Now, let's look at what happened in cases like mine. So we're going to use tags and filters, search for judges, parties, or counsel, and then explore cases of interest. And you can do this in a lot of different ways in our database. I would say in millions of different ways when you use the combination of tags and filters. So we're going to look at a few examples.

Rachel Bailey ([16:40](#)):

First example is, has Judge Koh ever awarded damages in an employment case? Has a court found patent infringement in cases where Apple was a defendant? And how have opposing counsel's Security cases resolved in the past? So we're going to talk about whether Judge Koh whether has ever awarded damages in an employment case first. And again, we're looking at Lex Machina here, and I've done the filtering for us. We're looking at 265 employment cases before Judge Koh. And very quickly we can see this damages break down. We can see all different types of damages that the judge has awarded in different cases. And again, we can see 43 cases awarded damages, and we can click into those cases and see what happened using the blue numbers here. And we can get an idea of what type of monetary issues are at stake in your case. And you can show your expertise to a potential client, or it might help you determine strategy with your client.

Rachel Bailey ([17:53](#)):

Now I've done something even more specific. I've filtered down to discrimination cases, which is a subtype of tag in employment. And I do want to say that Lex Machina is the only analytics tool that has this specific information. And we can see now that there's 99 employment cases with a discrimination claim. And the reason why you want to see a discrimination claim is if your case has this, that's going to look very different than say a wage and hour claim, but it's all sort of in the bucket of employment. And so you don't want your data set to be too big and to have too much extra information. So, like I said, when we look at damages, now we're looking at the most relevant damages. We're looking at five cases instead of 43. That's a lot of work that you wouldn't have to do if you know that it's not relevant. And then also you wouldn't pass this information along to a client or someone like that where you don't want to overstate your risk, or it allows you to better understand what types of damages are available.

Rachel Bailey ([19:06](#)):

Our second example here is, has a court found patent infringement in cases where Apple was a defendant? So again, I've already done the work of narrowing this down to the 643 patent cases with Apple as a party, in which the selected party was defendant. And so we've got the patent tag, and the patent tag has already gone through the process that Marissa described and that's why the work that Marissa does is so important, is finding these cases. And then we have the filter in order to show that the party was a defendant as well.

Rachel Bailey ([19:43](#)):

And now we're looking at our findings analytics. And so we've got the finding itself, infringement or no infringement on the left. And then the number of cases where there was a finding at various events. So summary judgment, trial, any judgment event, there could be multiple findings in a case. And so these don't necessarily add up to the any judgment event. And we can quickly see 12 cases at trial with infringement findings, 12 cases with no infringement findings. And you can quickly get strategy information if you click in and explore those cases. Again, no one else shows what happened and at what stage of the case, and human review is particularly important here. So automated systems don't pick up the nuances in judicial opinions sometimes. And so we're lucky that we are able to look at these cases and tell you it was a summary judgment, and this is what happened. We're doing work for you that then you don't have to go back and do yourself.

Rachel Bailey ([20:49](#)):

Lastly, how have opposing counsel's securities cases resolved in the past? So for this one, I didn't want to pick on a particular law firm. So I anonymized it. We've got X anonymous law firm, and then we're going to look at Y anonymous law firm. And we're looking at our case resolution analytic. So on the right hand side, we've got likely settlement and procedural resolutions. We can see the likely settlement was 36%, which is kind of a nuance of this practice area. And then procedural resolutions, 51%. On the other side, we've got claimant wins and claim defendant wins 4% and 9%. And this number sticks out at me, because there's a large number of judgment on the pleadings. So that's 59 cases. And so if I'm going up against this law firm, I want to look at those 59 cases and see what their strategy was. If I am interested in what they did, I might want to copy that. And it gives you a lot of information really quickly about how cases resolved.

Rachel Bailey ([22:00](#)):

For terminated securities cases, this is for a different law firm. Now we're looking at 90% of cases that settle, or 9% with this procedural resolution. And then on the other side, fewer cases with claimant wins and claim defendant wins. Again, a large number of judgment on the pleadings, because that's a nuance of this area of the law. But if I'm going up against Y law firm, I'm going to think about strategy a bit differently than the other law firm. Additionally, if I work at X law firm, I might want to tout my record because they have so many cases and you can quickly see by the data, whether they've won cases and things like that.

Rachel Bailey ([22:48](#)):

And the last thing I want to say is that other systems don't necessarily aggregate this for you. You can kind of find it, or you can find some pieces of it. And that's what I call the DIY approach, where you have to go to a case page and find some information, but they don't put it all together. Whereas we have this all in analytic form for you.

Rachel Bailey ([23:11](#)):

So in summary, outcome analytics tell you what happened in cases like mine? We use tags and filters. So again, what Marissa and the engineering team do is so important, because that gets us to the relevant cases. And then we sort of cross search that with a judge or party or counsel in order to get a record of cases like yours and explore cases of interest quickly, find out information quickly.

Rachel Bailey ([23:40](#)):

And we already have a question. I'm very excited. It says, "Is it human analysts who are assigning the outcomes?" So in some cases for findings, it's human analysts. Whereas in other cases, we do have some machine learning that finds cases. One thing I'll tell you is that some cases in our system, nothing happens. We all have cases that settle on the fifth docket entry. So cases like that aren't necessarily reviewed quite so stringently by the human analysts. But most of the outcomes are reviewed by the human analysts. We've got another question. Could you discuss state court and how that works? Marissa, I don't know if you want to kind of jump in and talk about your experience with state court.

Marissa Skudlarek ([24:42](#)):

Sure. So state court, which we've been offering for a couple of years and expanding to more and more state courts. So from an engineering perspective, that's a little bit exciting and tricky because each state court has their own way of organizing their data, and also as well as their own state laws. So then we need to find the best way to organize and present that data to you. But in general, I would say that the tag development process is typically similar. I will say sometimes that there are state tags that aren't based on text search. So we don't actually need to have that validation set. And it doesn't take quite so long to develop the queries.

Marissa Skudlarek ([25:30](#)):

I'm thinking in particular of when we are assigning case types, basically sort of our equivalent of modules for state court. Some state courts, the case will be labeled with what case type it is. And it says "case type contracts." And we say okay, it's case type contracts. And so for that, it's a one to one correspondence. So you don't need to be doing that validation against a data set, but yeah. But for more complicated things, I'm thinking in particular of state class action is often one that needs to search for a variety of phrases, like I did with data breach. So for that, we would have that validation set and use those same tools and develop it in that way.

Rachel Bailey ([26:14](#)):

Got it. From an outcome analytics standpoint, state court is a little bit different. We like to call state court the new frontier. If you practice in federal court, then you may be familiar with Pacer, which is where we get our documents and our docket entries from, and Pacer has its own set of nuances in the way that it works, but it's evolved over time. And it's been around for a while. So we know all of those nuances. Whereas state courts, a lot of them are very new. Some don't have documents at all right now. Our offerings of state court do have documents. That's something that we look at specifically for our analytics. And so it's exciting as we get new state courts online.

Rachel Bailey ([27:07](#)):

And for the outcome analytics themselves, we have the resolution in those cases. So whether plaintiff or defendant won and at what stage against summary judgment, trial. We have rulings as well, which tell you whether it was a bench trial, jury trial, settlement. And then we have damages that tell you whether the damages were on the merits, jury damages, and attorneys fees and costs and breakdown. And so damages are some of our most high value information. And again, the human analysts specifically are looking for numbers. It's very different when a machine is out there looking for a dollar sign versus a human saying, "These are attorney's fees and costs," or, "This was compensatory damages," or, "This was settlement damages," which is something that is specific to Lex Machina, and something we're very proud of. So with that, I don't see any more questions and we're right about time.

Rachel Bailey ([28:14](#)):

I'm very excited. I want to thank everyone for joining us. If you already have Lex Machina and you have questions, please don't hesitate to reach out to either your sales rep or your customer success manager. If you do not... Oh, and the dog says it's time. And then if you don't have Lex Machina, you can visit our website and sign up with a 15 minute session with one of our experts. The demos are great because they will sometimes get information that's very specific to you. So if you want to know about your cases specifically, then please get in touch. Thank you everyone have a great day.