025 - How Much Does Google Really Understand with Dixon Jones

Dixon Jones 0:00
So when Google does a search phrase, they're then running a, they run there and saying, right what entities are involved in this search phrase, so so they're not really looking up necessarily the exact keywords in your in your pages. They're looking up at the underlying concepts. And that's a huge change for us as marketers to understand.

Podcast Intro 0:24
Welcome to the SEOLeverage, where we talk about search, marketing, and conversion.

Gert Mellak 0:33
Hello, and welcome. My name is Gert Mellak from SEOLeverage.com. This is Episode 25 of our podcast. And today, my main goal is first not to get too technical when we talk to today's guest, about how Google tries to actually understand the content of a page. It's a pleasure to welcome Dixon Jones, who is I think, not only one of the most approachable SEO speakers out there, but also the person that definitely explains Page Rank the best on YouTube. So whenever, whoever wants or doubts on page rank and how this really works, Google Dixon Jones, PageRank YouTube, and you're definitely going to get a 25 minute good overview about how this actually works, or is supposed to work. Dixon, welcome to the show. Thank you so much for taking the time.

Dixon Jones 1:18
Thanks a lot for having me Gert, that the yeah, the the PageRank video actually, if you it's on, it's on Dixonjones.com. On the homepage, I'm so proud of that one.

Gert Mellak 1:28
This is really good I use this for for my team, as well as training material once in a while I get an idea about internal linking, but start here, start how this actually was supposed to work with Google's algorithm. And then you have a good idea how PageRank interlink choosing stuff that is just transferred across the page, essentially. You're one of the most, one of the long standing SEO people out there, I would say you you almost co founded Majestic or at least where they're in the in the very early stages. The SEO industry definitely things you're the co founder.

Dixon Jones 2:00
Yeah, yeah, I mean, I was there. Right, right, right near the start. And, and but there's, there's this guy called Alex who, who, who doesn't like to get out much. And basically it's his, it is his baby. But when he was still in his bedroom, I kind of co wind of what he was doing and realized that I knew everything that all the people in the world that needed what he was building. And it

was quite incredible because he was, when he built that Yahoo Site Explorer was a system that was the only backlink system and around at the time Moz was just coming on stream with theirs. But when I realized that he wasn't using Yahoo's API, like everybody else out there was, I realized that he'd got something really fantastic. So we ended up going into, into business together. And I spent the next eight, nine, nearly 10 years working with him really, as a marketing director,

Gert Mellak 2:52
Amazing, amazing. He made it Majestic for for some of the people that may might be more on the Ahrefs side of things, there are two tools out there for for link analysis. Majestic was the first one, Ahrefs came a little bit later. I think they both tools, we use both of them in the agency for different purposes. But it is interesting for me both tools, especially Majestic is really focused on the back link side of things. Yeah, yeah. When it comes to, to the top to do type of analysis, where so what are the other topics those backlinks are in? And then I found it interesting to do now with your current company elinks.net, rather, on the on page side of things. So you transition really from a link based analysis to an on page analysis? How does this happen?

Dixon Jones 3:38
Yeah, so I mean, when I pulled back from Majestic, I kind of took a little bit of a breather, 10 years, a long time. And then another guy came along. So just like with with Alex at inLinks, Fred came along, who's based in France, and he said, like, I've got this idea for a product. And really it was, it we still do links but internally, we optimize links internally. So we use a JavaScript systems, we can, we can map all the internal links, but we also are heavily focused on the on page content now. And the reason we're doing that now. So it's again, so so basically, I'm playing number two to somebody else's number one, I, although this time, I'm the major investor so life has gotten quite good from that point of view. But my motto in life, as you know, is, is always going to business and make sure that the developer, you go into business with own small shares, and you do because then they don't leave because otherwise, businesses don't work. The fat startups don't seem to work online. You need a developer to be so heavily invested that you don't have to worry about that piece of the puzzle. But yeah, so and then then then deal with the on page part of it in a different way to all of the other tools that have come before really because Google has changed.

Gert Mellak 4:58
Absolutely. I think Google, this definitely has changed. And I think it's really interesting. As Google users, we often very often think, okay, can we just get it? Right? Can we scan the page and just get it. Our clients as well, we we educate them a lot in the initial stages instead of, it's still about not keywords necessarily anymore. But we still need to mention those topics to make sure this is a good article, this is deep enough, this is has the context enough, etc. What would you say Google really understands from a text if you put it to the web?

Dixon Jones 5:33
So Google's got a an API that people can use in a swamp. I promise not to get too technical, but it shows if you put in a bunch of text, and it will come back and say, these are the main entities

that this page is about. And, and, and the definition of entity is, is kind of interesting. So so to Google, a topic, they've in their documentation, they use the word the phrase topics. And they sometimes are talking about entities. And sometimes they're talking about just more generic topics and things. But they've, they've changed the way that they've they understand content, from trying to read a page and understand all of the words on the page and optimize the whole web based around web pages. They're now really trying to focus on concepts and ideas and topics. So for me, a topic is a Wikipedia article for that for the for the non exact people, you're not going to go a long way wrong, if you can sit there and say, right, if it hasn't got a Wikipedia article is not really a thing. It's not really a product, not really, you know, it's not, you know, it's got a Wikipedia article, then you know that Google's going to understand what the heck it is, really. So they've got their own database, Google's or their own knowledge draft of topics and ideas and concepts. And they're storing all the world's information in that way, instead of just trying to just try to have all the web pages in the in the world, it turns out that, then you've got an index, if you think in terms of databases, or think think in terms of a big Excel spreadsheet, instead of having an Excel spreadsheet with every single web page in the world, trillions of entries, trillions of URLs in the world, and then all the things that they're about, if you have a Excel spreadsheet of all the ideas in the world, that turns out to be hundreds of billions, which sounds like a lot, but it's a lot less than trillions. So you've now got an Excel spreadsheet, that's only hundreds of billions of lines long. And then all the all the all the pages on the on the internet that maybe have a reference about them. So you've got, you've got a different way of storing information, which for Google then scales much, much better. It also allows them to, to weed out rubbish, because they can sit there and say, right the concept of a president of the United States, or let's go for President of the European Union, let's not go to to American. You know, here's all the things that we know about the President of the European Union, they don't all have to be web pages, they can be, you know, things in their database that they know about, you know, when they were born when, when they when they joined, you know, their career or whatever, it will also have web pages as well, which connect and reference that kind of information. But it allows them to have a single point of truth for an idea, which is a double edged sword in its own right, but, but it allows it allows Google or any large search engine that's trying to organize the world's information to scale better. So it's not surprising that they're moving towards or have moved towards this approach. They're not killing the other approach of trying to read all the data and read all the pages. But this this, this gives them much more scale. So it then becomes a case of how do we as marketing people, integrate with and influence topics and ideas, rather than how do we get seen for a particular keyword? Because Google is not just saying, here's all the rules, here's all the web pages in the world, they've got this keyword. It's saying, right, what web pages have we already analyzed that are about the underlying topic that these these this current search phrase is looking for? So when Google does a search phrase, they're then running a run there and saying, right, what entities are involved in this search phrase, so so they're not really looking up necessarily the exact keywords in your in your pages, they're looking at the underlying concepts. And I thought a huge change for us as marketers to understand.

Gert Mellak 9:18

I think there's it isn't there many, many good points here, I think that are really worth exploring a little bit further. So Google scans it, scans an article, breaks it down into concepts into ideas into topics. Yes, and then tries to match those topics with the topics they extract from a search query. So when you type into search query Google tries tries to also understand the query, which is a part I think many people don't realize that if you type in a search query, and Google first of all tries to understand what you're trying to achieve, they don't try to match your keywords with some other keywords as they did many years ago. But they tried to understand what you were trying to achieve based on the topics you mentioned, based on the idea that's behind your query. Yeah, and yeah, hmm. Just with topics that are in the index, in certain database, certain structures, and things like

Dixon Jones 10:06
And it's something we try to emulate. So we try, I remember, you know, I had to go back a long way. And I was running an agency for 10 years before I got involved with majestic. So, you know, my SEO does go back to 1999. And, and I remember, you know, when Word Clouds came out, I used to take a Take, take two pages, and then compare the word clouds off them, and then I, but they never kind of really matched up. But it was because different pages were using different words in different places and different different concepts. And you couldn't, you know, the bridge and bridges, you know, were two different words. And so you had to go through lots of other processes to try and make get these ideas to match up. And it's kind of the end point of that, you know, there's no no real difference between a bridge and a pontoon or, or, you know, or ahm ahm, you know, I'm sure there's other words for bridges, I can't think of one right now. But there's all these synonyms that that are out there. And also, when you add the word tower to it, Tower Bridge is a very specific entity, it's not a tower, and it's not a bridge, it is Tower Bridge, you know, a tower, Eiffel Tower is not just any old tower, it's the Eiffel Tower. So I think the Word Cloud idea just broke, didn't didn't quite work. But the entity idea works much, much better for for Google and for a large search engine. So what they can then do is they can say, right, the concept of well, Eiffel and tower together ball boom, that's a one night entity, you know, no point in thinking about the word tower anymore. an Eiffel Tower is an Eiffel Tower, there's only one of them and it's there. But you can also then match up quite easily. The relationships between ideas between kings and queens is one that's, you know, a research paper that's quite quite well documented. So you can see that kings and queens are semantically related or king and queen is a semantically close topic. And then you can see that in context. So if you go for horseshoe, I don't know if, if the Swiss have got the same kind of metaphor, but horseshoe could either be in one one page of contents could be related to horses, and in another page of content could be related to good luck charms and fortune. So so they can then use this, this this map of topics and underlying context to to sit there and say right, the horseshoe in this context, I need to bring back stuff that's related to good luck. And horseshoe in this context, I need to bring back farriers and horse kind of horse shoes.

Gert Mellak 12:44
Funny, funny, my, my SEO history goes, goes back probably 20 years as well, I wasn't I was doing a lot of this time I was doing web development was slowly then growing, SEO, SEO, SEO,

Dixon Jones 12:54
You must have been six at the time.

Gert Mellak 12:57
I wish

Dixon Jones 13:00
You're younger than me.

Gert Mellak 13:01
I probably I probably started in my in my 18,20 years or something like this. And it was funny because I was really only focused on programming. I didn't want to do anything else. But it was founded in what pages I created for clients didn't rank. And so I had to figure out how this work back then every every IT guy did everything. So from from Ax2, to SEO, obviously, and everything. But SEO then has been growing over time. But it is funny because what what we did very early was handwritten note to endless models to use word clouds for this. But we did handwritten notes that I just checked what is Google ranking made handwritten note and they look they're talking about the ranking articles to talk about this ranking articles to talk about that. And it was then funny that Google at some point in transition and say, Look, we only interested in in deep articles, was it a theme at some point where they say, okay, we need articles to really catch all those phrases. And this is also something we talked to in our framework based consultant, consulting with this ERICA framework where we walk people through different stages of SEO, different things they should be taking on a regular basis. And we very often, just really ask them do some research, right? Without any tools away, just read the top 10 pages? Yeah, make some notes. And you're going to realize there are certain themes that just come up. So if you then create an article that doesn't deal with the Eiffel Tower, when we talk about the most important towers in this on this planet, you're probably missing out.

Dixon Jones 14:30
Yeah, exactly. Right. And I think this that is the key, and it doesn't really then matter too much how Google's algorithms change. If you are then writing content that is, as you know, that covers all of the important points that the the best content covers. And then hopefully, one more as well. Then, then then then you've you've written a knowledge leadership piece for that particular situation and, but wouldn't it be great if there was a piece of kit out there that could go and write find those underlying themes and topics and ideas and save you what has to have been, you know, hours of work each time just to create one article just to try and find the underlying themes that are being mentioned. And it turns out computer program is much, much better at understanding those those themes and not missing out, particularly co occurrences uhm between, you know, 10 pages, you know, is it that's, that's kind of where we come in really, with?

Gert Mellak 15:33
Would you mind explaining co occurrences for our listeners?

Dixon Jones 15:37
Yeah, so co, so co co currencies, so if you've got the top 10 pa web pages that are in the search results in, in France, for any, you know, given given phrase, then, yeah, well, let's stick with France. And so, you know, top top 10 towers in the world, and you've got, you know, Qatar bridge and Eiffel Tower and the Tower of Babel and, or whatever. So, you know, so so the thing is that, if there's going to be lists of things coming back, yes, there's going to be some that mentioned, Eiffel Tower is gonna be some that mentioned Eiffel Tower Bridge is going to be one that says Tower of Babel, or three that has, basically co occurrences is how many times that concept is, is mentioned in the in the corps of results, or seed set of results, or 10 pages that you're analyzing, I mean, you could analyze more pages, but the thing is, if you carry on going for too long, you start getting the rubbish as well. So you kind of want just the cream of the of the system, you don't want to go too deep. And you also try to avoid the looking at the the navigational parts of a web page as well, because that will set the set things off. So what you really want is the main body of the text. And you know, what the writer was doing not not all the stuff that's around it, you know, if you've got some automated bits and pieces going around a web page, that's, that's just annoying noise for trying to understand what the writer was trying to get to. But when you break that down into its underlying topics, you know, if all 10 of them are talking about the Eiffel Tower, you're pretty much not going to rank if you don't mention the Eiffel Tower. Unless Google says right, and I want something that does something completely different to all of the other ones in the results, which could happen, of course, happens as well.

Gert Mellak 17:15
Happens as well. Co occurences is definitely is also very powerful. We're seeing when we you start trying to associate a brand new brand with some established brands. So you just keep publishing content, getting content, make sure that Google connects the dots and says, Yeah, this is probably rather similar as these other solutions. If you have a comparison, and you come up with a new rank tracker, or you come up with a new text optimization tool, starting to mention this tool next to other more established brands, we feel we can't have, we don't have hard data we've been we feel this is really helpful for Google to connect the dots and just really associate this new brand, they don't know yet, they don't have an entity for yet, slowly with similar tools in the same market.

Dixon Jones 17:59
It does seem to seem to work. I mean, you know, InLinks didn't exist two years ago, and but if you, but now, it seems to be, you know, Google understands what it is, it's got its, you know, it's we've we've, you know, we've got our, you know, a little box of, you know, frequent pages and stuff, and it's got a pretty good idea of the, of the system. And it's really because I think that we are, you know, creating knowledge articles that are not competing directly head on with our competitors, we've got a trend, we have a unique bit. But, you know, and being unique and not trying to compete head on probably helps because it makes it easier for our competitors to reference us, you know, rather than us, you know, being head on, or people within within those those fields, at least, you know, accepting us as a player in that space, because our product is not at the moment competing head on with some of the big players, you know, ad for who

knows how that will develop. But it's kind of got us to, got got us this far anyway. So I've got to make the investment in it in the first place. So

Gert Mellak 19:12
Absolutely, it got to at least to a stage where we are actively working on in integrating InLinks into our application for our consulting clients. So whoever is going to use the application probably in a month or two is going to be able to get the texture text optimization, suggestions, analysis and things like those, whatever your API is going to do, probably with.

Dixon Jones 19:31
So the yeah, so this the API. So it's a it's great that you're going to put it in there. So the API actually does two things. It it runs a natural language processing algorithm, okay, the over the content over a page of content and extracts the underlying entities, but it also runs Google's natural language processing algorithm over the content and gives him good extracts those entities as well. So it's now got two lists of entities and there should be an overlap. Really, you know, if you had a perfect world, Google would say this is about these four things. And we'd say it's about these four things. But what usually happens is, Google says it's about these three things. And we say it's about these these these 30 things. And that's because we have, our job is to tell the user, that somebody, the the the SEO or the customer, all of the topics that this page is taught is is talking about. And Google's job is to try and filter out the few that it thinks is the most important things because they don't want to, you know, overload their their systems with, you know, every topic under the planet. So there's always a big gap. And we run reports every every week, that that compares the average change average thing. At the moment in the UK market, I think only about 15% of the entities that we see, are seen by Google. So so we have this concept called SEU score, which is all a bit in beta. But a bit, I think it's quite clever. So it's basically, the number of entities that Google reports in their API divided by the number of entities that are actually on a page is is what we call the SEU score. And so that then comes into into into the API as well. And it's, it's not always, you don't always want 100%, high SEU score, because then you may be just trying to say that this page is about 50 things, what you really want to do is make sure that, you know, Google definitely understands the right entities for the content reports those. So yeah, so and then you want this list of topics that you're you know, you should be talking about if you're looking at another person's page or their competitors page, then being able to see all the topics that they're talking about. If you use Google just use Google's API there, you don't have enough information on it. Because it's because they're only reporting, you know, a few of the topics, as

Gert Mellak 21:53
It seems like they're holding back quite a bit on on the information they actually would have on the page or could provide with the API?

Dixon Jones 21:58
Well, yeah, I think I found this when with all the way through Google's world. And I don't blame them that they don't want to give us all of the data. So it's nice that they give us any an API. But it's like the link data, Google kind of over the years took away the link data all the time. And, and

that's why a Yahoo had a, you know, had Site Explorer, because it was the one thing that was was undermining Google's system really. But unless technologies go out and build their own, or that the industry can use, then you're always playing to, to Google's tune. And that's not always the best way to, to try. And when you kind of need this, this extra layer of information, which somebody has got to go and build. Unfortunately, we need people like you to understand what the hell we're doing so that you could translate into usable product.

Gert Mellak 22:55
We at least try our best. What do you what do you think you're talking about entities now a lot of entities, I like to explanation everything that comes in up in Wikipedia as an entity. Again, there are also things that don't come up in Wikipedia the the entities that Google has a numeric numeric identifier for, at which point, Google takes a word like SEO leverage or agency and says this is now entity numbers such and such.

Dixon Jones 23:21
Yeah, yeah. So yeah, for sure, there are entities that are not Wikipedia, in Wikipedia. That's true, the Google understands an audit, and so they have their own independent database and, and things like Google. My business will then create an entity for, for people, most of the time, and other data sources apart from Wikipedia can help to feed Google's, Google's knowledge graph and getting to a within your, within your niche getting to a particular level of authority should trigger Google creating a an entity for your brand, for your, for your name, for your thing, for your product, whatever it may be. And I think that's, that's important that you try and get your brand as an entity if and define your brand, very, very well. But when it comes to content optimization, people, of course, are typing in generic, you know, people people are thinking, well, we still think of people as saying, you know, what is the best radio to buy, you know, or, you know, some people are people are, if they know your brand, then and you've got an entity, then your job is done, your job is done. If somebody says Nike, if Nike doesn't come up, then somebody at Nikes SEO agency was screwed up, but I assume that nike.com will come up but if somebody types in running shoes, then the game is open. And Nike has two choices, at that point. If they're a big brand, they can either try and associate their brand heavily with running shoes. But at that point, they they put some level of the distance between themselves and clothing or golf clubs or footballs or the other things that they make and sell so so it means that the concept running shoes in itself is something that any brand or any business can try and influence, because the concept of running shoes is not is not specific to a to a brand. But that becomes a. That becomes I think, more of the SEOs, you know, realm, you know, how do I rank for running shoes? Or how do I get customers that want running shoes to come to my product?

Gert Mellak 25:25
Right? Or the question? Do I really need to run rank for running shoes? Right?

Dixon Jones 25:29

Well, yeah, okay. Yeah, yeah. So yeah, yeah, I mean, is there is that and I think that, um, for a start, you know, when somebody types in running shoes, are they? Are they talking about the concept of running shoes? Or the concept of trainers? Or are they talking about the concept of running? Or they talking about the concept of shoes, there's three different entities involved in that concept, trainers running and shoes. And and it's only two words. So there is a whole myriad of questions that Google has to answer just when somebody types in a search, or what are they trying to get to? And what what is the the underlying thought process. And of course, Google has the advantage of knowing what your history was, as well. So. So going back to the horseshoe example, when if you just typed in horseshoes, Google would already know whether you have a tendency to go and have a look at fortune telling sites or whether you have a look look at, you know, farming websites, for example. So, so that history is something that we don't necessarily have as the end product. But it is, it is a good reason why we should be looking at personas, really, in our in our marketing, we why we should be trying to understand, you know, what kind of person is coming to our business or shouldn't be buying from our business. It's not just about a phrase, it's about the kind of person behind that phrase. Uhm engines is another good word. Do they mean search engines? Or do they mean combustion engines? You know? Yeah. And that's one our system messes up all the time.

Gert Mellak 26:57
Yeah, there's definitely some manual manual processes necessary. How does? How does structured data markup help machines understand content. We have all our clients create structured data markup we created for them to paste this into their website, or where we do the execution, we handle this part as well. We know InLinks helps creating this structured markup, how does this help a machine to really understand what this text is about?

Dixon Jones 27:24
It's turning. I mean, it's in the words, you know, the, it's turning unstructured data into structured or semi structured data. So it's helping a machine to file file any idea correctly, and put it in a place where it can be quickly retrieved accurately. So schemes such as an organization scheme, or and those and, for example, most businesses should be using organization scheme to help any web page on the site saying this is this is, you know, owned by or it's published by, you know, this company, or owned by this company, or, you know, is on the site that is the corporate entity for this organization. And so that kind of schema can run all the way throughout your site. And, and as long as you know, Google gets to a page that's got that schema on it understands that it's associated with this particular company behind it. And so that's great. Where I think it becomes more difficult to do schema and the bit that we concentrate on is at the page level, trying to help Google associate the content or an individual page with the underlying entities that that that page is talking about. So knowing that Google only only triggers 15% of the entities that we that we see on a page. That's okay, as long as it's the right 15%. So the schema can very much help to tell Google, what are the topics that this is mostly about. And other topics that we mentioned, are maybe secondary for they are they're still important to the page, but they're secondary to the main, the main 2, 3, one, two, or three kind of topics. And by putting in schema that sits there and says, Hey, Google, this page is about a thing called the Eiffel Tower. And in case you don't understand what that is Google, here's the Wikipedia page for the Eiffel

Tower. What we're doing is easily give giving Google giving a search engine, an easy way to map concepts to something that is already in their database, because we know that they've already got Wikipedia, well and truly, in their database, we could maybe link to a an ID, a Google a Google knowledge base ID to help them along and work that way. But they're a little bit a little bit more difficult in knowing that seeing a Wikipedia page and matching it to Wikipedia pages is great to matching your content to Wikipedia pages is great because humans can see it, we can understand it, we can understand it as humans, and the machine can understand it as well. So Google already knows everything about the Wikipedia page is Eiffel Towers Wikipedia page, it already has that mapped into its own knowledge graph, it's already got its own ID number for it. The mapping is perfect. So we're taking unstructured ideas and giving a machine a helping hand to say, okay, you could, you could read this page and try and work it out for yourself. But here's, here's some hierarchy for that, you know, here's the, and here's some pigeonholes, you know, these are the main topics, the physically Wikipedia page topics of these, these pages that are out. So yeah, it helps to helps to structure the day, literally helps to create machine learning ability, you know, cut through the long process of the machine trying to interpret it interpret what humans do when they write things.

Gert Mellak 31:09
Perfect. So so before, I'm going to ask you for some practical advice for our listeners, so they can create an optimized content, I just want to quickly recap, we have a text, the text is broken down into topics, ideas, entities, as you have explained Google doesn't understand everything. With structured data, we can help the machine understand concept by mapping them. So SEOLeverage is an ad SEO agency like this other one, or this is a microphone as the microphone that Wikipedia describes on this microphone page, or this is a coffee mug. So this is coffee.

Dixon Jones 31:45
And if you want to optimize. And if you want to optimize, then, as you as you, as you said at the start earlier on, you have to then do this not only for your own content, but you also need to check your competitors content for the target that the the target phrase that you're hoping to look at. And you need to break those down into to see that gap analysis between the topics that they're talking about. And they're already the ones that are ranking and what you're talking about. So you can see the gaps and fill those up. Yeah.

Gert Mellak 32:18
Absolutely perfect. So so let's, let's say someone wants to rank higher the ranking maybe on on page two, what should they be doing or focusing on in order to optimize their content to rank on page one? Given?

Dixon Jones 32:34
Well, three things. Actually, one, we haven't mentioned that. So I just said, so the three things are, make sure that you're talking about all the topics that that your competitors are talking about that are higher than you, particularly the top, top 10. And that can sometimes that's just one topic, there's one bit of the puzzle that you haven't got in there, and you haven't spotted it

yet. And it seems subtle to you. And it may it may not be relevant because Google might, Google might have got it wrong. So moving my Google might have suddenly thought that if you don't mention, you know, I don't know, the Queen of England, when you're talking about the concept of royalty, then you know, then you say this, you don't have to have the Queen of England in the system, but everyone else is talking about so Google's associated that as an important thing to have in that. So it may not be obvious what what it is, is missing. But But you know, it may be one or two topics and then suddenly it bumps you up there and you've got it. But the the other two things, yes, the schema. But the other thing then is cannibalization of your content. One of the problems that we see quite a lot is that particularly smaller businesses, you're focused, you've you've made the great wise leap of saying I'm going to focus on one thing, I'm not going to try and be Nike and do footballs and trainers and, you know, I'm going to concentrate on on a tailor made sports shirts are some of that. But then you find yourself writing, you know, 50 pages on different aspects of of a sports shirt. And you've talked about the concept of a sports shirt a lot. But you've got no way of telling Google where the authority in your own website is for the concept of sports shirt. And schema helps out because you sit there and say, right, this is the page about sports shirt, you then shouldn't have another pages saying this is only about sport shirt, you can have say this is a sport shirt in the context of rugby, or this is a sport shirts in the context of you know, but basically if you if you if you can have a the the authority around sports shirts in one place, then what you can do is then internally link all other mentions of the concept of sports shirts through to the page that is the authority on your website for the concept of sports shirts. So you can have many different topics that you're talking about on one page of content. You can talk about sport shoes, running shoes, talking about, you know, tennis rackets etc. You can have all of those things talked about in a big blog post. But you still want to link the tennis racket page mentioned to the tennis racket page and the sports sports shirt page of the sports shirt authority. So you need to to take those the important concepts and link them through to the to the page that is the authority on your site about that topic. Otherwise, there's going to be what we call cannibalization. And it's really basically splitting the authority for any particular topic around your website instead of concentrating it in to one particular page.

Gert Mellak 35:33
That's so important. I think one of one of the most important things that come out of our site audits we do on websites, is to prune a website, get rid of quite often, yes, duplications edit these canonicals or something like this. Obviously, structured data then as well and focus a lot of interlinking. We've seen sites come from page seven to page one, just with internal linking, just making sure Google understands what which one of these 20 I think it was 26 different pages that talk about the same topic should actually be ranking. And it seems like when you give good like a clear guidelines that you don't have to, to worry whether this URL that you relate etcetera, things in steam seemed to stabilize. So very often, when clients say, Look, I don't have a stable rating, my climb method comes up on on number three, number six, page two, page three, page one again, and we need when we take we see it's different URLs coming up, not the same URL, because Google every time this ranking should come up, they pick one of the euros and based on the degraded optimization of that page, the ranking then shifts heavily. So cannibalization is definitely one of the most important

Dixon Jones 36:37
And, and Google have also made a big noise recently about passage indexing. So the idea we try interpret as interpret as the the idea that instead of indexing pages at a time, the indexing, you know, a section of the text, which you can already see in different parts of the system. But they're they're treating a section or a paragraph of text potentially, as rather different to a page or text. So if you could, you could have a section of the page that that is authoritative on a particular subject and the rest of the page does Google doesn't care about. And I think that's kind of interesting. There was a lot of research way back in the nineties and before actually, around, breaking a page into chunks. So one of the things that Majestic does, is it breaks a web page into 40 different sections. So every 5% of the page is a different section. And it works out where the links are, whether they're internal or external, by by section on the page. And uses, you know, uses that to, to create chunks. And that's turned out to be very, very effective in understanding the difference between an influential link on and a not so influential link, and then and so so there's a lot of things around chunking, stroke, passage, indexing stroke, trying to ignore, you know, bits of the page that are our contextually irrelevant to them. But But main body content, all of these things are helpful for a search engine to pull out the important bits from a page and leave, leave the rest.

Gert Mellak 38:15
For me, it was a deposit in this game was really interesting for me, because it was like, for me, if Google said, Look, we can't work with big texts, we can't understand big texts, but we can understand chunks.

Dixon Jones 38:27
Naturally, I will we're looking at that I look at so the search engine understanding score that we've kind of got in beta is really interesting, because one of the errors that we've got is that if you've written long, long, long amount of text, then Google has a very low SEU score. So you know, maybe maybe they give up over it after a number of words, but basically carry on writing, you're going to talk about more and more and more and more and more topics. After after, after a half dozen topics, Google, you know, after you know, 20 topics, Google doesn't really want to complicate its own its own world. So writing more and more and more, doesn't help Google get better. It, it muddies the issue, unless it's extremely focused. And it just the way in which places repeating the same topics again and again, which also can be a little laborious for a machine to understand. So I'm not surprised, I think as you may be that long form content is, is going to be less important in its day in the future than it than it currently seems to be. But who knows? There's now, is now, is now and speculating.

Gert Mellak 39:36
We'll see. We'll see what happens is definitely going to be interesting. We raised this in the in the discussion before we recorded last time. Also the artificial intelligence how this is going to to have an impact on what Google does actually with content, how they recreating recognize and I'm not going to recognize if this is automated if this is all making bad because it's automated, it's, because essentially, we are describing a random mechanical process right? We are describing a process where I analyze, I have to like InLinks, for example, with machine learning

in place, that even does the job for me, I don't have to do my handwritten notes anymore. This does the job for me and does it much better. We're describing a mechanical process, whatever is in mechanical process, the sense of automation automatically goes with it. Whereas Okay, this is now mechanical. And this is like the intermediate step between me and my handwritten notes and the machine just creating the text for me based on on the analysis. So I think this is some tendencies that are really going to be interesting how this is going to evolve,

Dixon Jones 40:36
I'm really worried about how that's going to pan out, you know, I think being able to, to find out, what you're supposed to be talking about is one thing, then being able to press a button, then it generates readable text as a result of that is, is now is now possible. And and there's the GPT-3 is a is an API that's out there that can do that. And I would imagine that you map out kind of output onto GPT3's, you know input, and you end up with some pretty, some pretty impressive takes as hard for Google to ignore. And, and yet, it's just machine generated text. So I'm a little bit worried about that. I don't know whether we're going to integrate it any way because our customers demand it, or whether we're going to fight it and then let our competitors do it first. But it's, it can't be, it can't be good, that all the world's information is then getting created and generated by a machine, totally. It can't be good for society and and I don't quite know how that's going to pan out. But I guess that's going to be Google's big problem in in a few years, wait a year,

Gert Mellak 41:49
They've been worried, I think they've been worried about this evolution for at least 10 years, really, really worried because a lot of their algorithm updates a lot of their what they have put out there, etc. If you see it from a point of view that AI is coming. And at the base of a button, I'm going to be able to create 50 websites with original content created by a machine that's going to compete with other 500 with a network of 500 maybe. And this is also also so close to us GPT-3 entire websites, you tell Look, I need a website with a logo on the left and columns on the ride and and this content and these menu items, etc. And it's just created for you, we come to the playing with with AI, I just had my riders to another AI test round on on a couple of AI takes tools as well. Because I do believe with all the all the researchers there all the tools that are there like yours, we are not starting with a blank document anymore, right, we already have some information, we create content briefs based on inLinks for our clients, but other tools out there that give guidance on what we should be writing about. So we already have a lot of information in place. So having a first draft written by at least a first draft written by an AI, I think is something that might be coming. I don't think it's good either. But it's just an evolution we as we have been seeing in the last years, we see ad copy being used, AI generated ad copy, it's short text, so it's easy, and it's very has very clear tone, etc. So the easier it gets, the narrower it gets the the easier it is probably to use AI. Right now my the feedback from my writing team, we have five writers here and they said, Look, if if I have to edit everything that the AI gives us, I can also write it on my own. It's not not there yet. But what happens in a couple of years. Right? What's happen

Dixon Jones 43:35

I think, I think also if you've got genuine content writers, they don't want to just be editing, you know, there's a difference between an editor and a content writer. And if if we turn the content writers into editors, then we'll lose our content writers. So we got to be careful about that one as well.

Gert Mellak 43:53
Absolutely, interesting discussion. I want to come to an end we all were already beyond the schedule. Thank you so much, Dixon, for all the value you've put up there all the explanations you've you've been giving. It's definitely interesting. Everybody who wants to know more can probably get in touch with you and Dixonjones.com?

Dixon Jones 44:12
Yeah, well, or play with InLinks dot net if you want to, there's a there's a free version of that as well. And or use your API.

Gert Mellak 44:22
It's API integration is coming. So definitely the the contemporaries etc. We used to put together based on the InLinks are going to be available within our application. Thank you so much for coming and have a good day.

Dixon Jones 44:34
Thanks a lot Gert. Cheers. Thanks. Bye