**Dr. Pennington Interview**

Dr. Pennington: No. We have our own algorithms now that we’ve developed to transform the data into certain ways that we can work with it.

Dr. Gates: I’m going to ask Dr. Pennington to give you a scenario that might help you because it’s going to be coming up. Actually coming up in a little bit … you know I’m jumping .... on platforms. On platforms, we talked about web based and we talked about mobile right? And so earlier she gave me a scenario on how that might work-- how someone might use the system. Wo we’re going to go back to how you use the system and I think it might help with that question that you just answered. So we we’re talking about imagining…

Dr. Pennington: Yea well so one scenario that I can imagine, is that I’m out on the field and I get a text on my cell phone that says that there’s a problem or an anomaly with one of the sensors and I would want to be able to walk over to that sensor and try to look at, on my phone, what is the error that is being recorded, maybe look at the property that’s generating the error. Look at the data that’s being generated by the sensor, and maybe I might want to look at those three different kinds of data together, in different ways, um, so I might want to look at the data itself but have a plot…, like let’s say I did a property that said temperature should be between, you know, x and y, during this time period. So I would want to see it in a graph of temperature, threshold, that the property defines, and then I would want the anomaly to be highlighted for me. And then I might want to drill down on the property, I might want to switch to the property, I could look more closely at the property. Or I might want to look at the anomaly more closely. So there’s different things I might want to look at but like I said that there’s those three different kinds of data that I’m interested in. So I want to go over to the sensor, I want to be able to look at those three different kinds of data, maybe in different ways, and I want to be able to fix the problem, if it is a problem, right there or may be something of interest. So I need to be able to interact with all those three different kinds of data in real time, at site, on my mobile device.

Dr. Gates: So this is a use case scenario. So I’m expecting to see this much depth. Any questions?

Student: Just to clarify, you said that the three types of data that you would like to see work on right there right off hand would be the property, the anomaly, and what else I’m sorry?

Dr. Pennington: …and the data itself that’s being generated by the sensor.

Dr. Gates: Now if you were at your desk, what would you like to do?

Dr. Pennington: Same thing. It’s always those three things, now I might want to see them in different ways. I might want to see like a graph of the data where like a certain of that stuff overlaying on it. I might want to see a map view of all of the sensors and being able to maybe hover over a particular location, see something about the data that is being collected at that location. Or I might want to see the raw data. That’s a possibility.

Dr. Gates: So one of the questions that came up is: do you want it to be a stand-alone application or web-based?

Dr. Pennington: I want it to be web-based because I could be in the Arctic and get notified that there’s something wrong with my information coming in the dessert. And I might not have… maybe my cell phone has died and I need to go over to somebody’s camp and use their computer so it needs to be on the web. These things happen, cell phones die.

Student: You keep talking about the graph and displaying stuff do you want any kind of predictions of future data at all or do you just want what we have getting in what we typed, as far as displaying wise?

Dr. Pennington: I think future predictions would be great if we can do that, if we could analyze the data, the patterns, maybe get developed and be able to predict, that would be wonderful. I think that’s an actually, a good use of data properties is you know, we don’t know sometimes what those properties are gonna be, in fact, most cases we’re making that up as we go. What are the properties ‘cause we don’t really know very well. So building the capacity to predict, you know, given the properties and the way that I have it designed, this is what the property would predict, next hour, in two hours, three days. And being able to see, think about, does that really make sense? Does that really um… Because one of the problems that we’re gonna have, we’ve already had it, is that as the scientists start designing their properties, you know those properties are being specified in um, sort of, data base and um and thinking about them from an information scientist point of view. But the scientists don’t always have the right point of view. I guess what I’m trying to say is that if you had a scientist sit down and you asked him to develop some sort of logic, they don’t necessarily have the training to anticipate the implications of what they’ve just specified. So if you have a scientist to sit down and design a property, they don’t necessarily have the background, to anticipate, the implications of what they’ve just specified. So being able to do some sort of predictive thing and say “Ok, you’ve designed this property this way, based on what you said, here’s what we might expect”, would help them in being able to troubleshoot their properties. That was a good question, I hadn’t thought about that before but that would be an outstanding…

Student: Expanding on his question, you mentioned you guys have some expected values and that’s why you’re trying to pick the anomalies. So, adding to that let’s say you have a graph, normally it would just go from one start or beginning to where you had the last reading, so you would put the expected result and highlight them in some way maybe to show this is what’s expected and then around it have it make the comparison between the actual results and the anomalies? Is that what you’re getting at?

Dr. Pennington: Yea. ‘cause sometimes, ok so one thing that you might want to do is compare across sensors, so I don’t want to just talk about well temperature or maybe … so for instance precipitation, there are certain characteristics, patterns and habits of temperature and humidity and barometric pressure, that we know about, and that we could specify in a property, to get at, is it raining or is it not raining? But there may be other things, comparisons that we try to do… that weren’t so well known. So as we start to develop more complex properties, yea it’d be nice to be able to say now that…. And It’s also just and error-checking mechanism when you develop a property. Here’s the way that you developed your property, Now here’s the way, here’s a simulation of what we might expect. Did that answer your question?

Student: Just an idea I had so, when you expect, obviously, what you’re real data is going end up being, so let’s say uh that the real data from what you expect is constantly being more and more off each time so you have that so and so recording so you know what these expectors have been this off and have a notification for you to see so you could maybe readjust whatever algorithm that you have used..

Dr. Pennington: And I can somehow envision that you start reading these anomalies, report it, you find a pattern in the anomalies that indicates that your specification is wrong, it’s not that there’s any error or that there’s anything interesting happening. It’s that your specification is wrong.

Student: I actually had a two part question. Um, you keep talking about the properties but the way that you’re kind of explaining them is making me wonder, are these ever expanding properties that people can define for the database?

Dr. Pennington: Yes.

Student: And then the second question about that, you were saying that we have to be able to go back and actually have them specify who was the author at that time and who was the current author or whoever is the care taker for that data. So, my question is on that part would they have to have the database widely available in a database format that everyone can change and in real time view the changes?

Dr. Pennington: Maybe. I think for now it’s more local. That I could envision that this all worked out well it became widely adopted yeah there might be some sort of community registry, you know, I assume it would be a database, you know, that’s for you guys to decide not for me to specify. I don’t know how you would do it, but I could imagine there would be some sort of community registry where everybody in the community is sharing and looking at these properties.

Dr. Gates: So at this point, you always want to get the “what” and not the “how.” But, having said that, what Dr. Pennington is saying is that we have to store them locally so like within this group-- within her research group--they want to keep the properties stored. Ok. So they know they need to be stored in a repository, whether that’s Excel or an SQL Database or whatever. That’s part of your analysis which you’ll do later. What she’s saying is that you’re collecting these properties and this metadata. And then there may be a point in time when we start saying ok we want to now be able to move those into a more national type of database, which the community would be building, right? That says ok-- now I’m going to take these [properties] and move them up for a broader view. So if we’re thinking about designing for change, which is going be one of the principles you learn about, we’re going to scope it, but you’re always thinking about [change]. And I said this on the first day of class, if you remember, that you start thinking about in the future we may want to do that. We want to be able to think about that because it impacts your design. So it’s a good question.

Student: Well let me also redefine something as well, for the updates of the information, do those have to be real time updates?

Dr. Pennington: Ok, which information are you talking about?

Student: The properties, yes.

Dr. Pennington: I um…..

Dr. Gates: Can I help with that maybe?

Dr. Pennington: Sure

Dr. Gates: If I’m thinking that the properties are stored someplace-- I’m a researcher and I’ve set up a sensor. So here’s a scenario: I could be grabbing them [properties] and putting them in there [i.e., in the system for monitoring sensors]. If I’m grabbing and reusing somebody else’s [property], you know you want to know that this [property] is being monitored and it’s reused. If I change it [the property] and change the parameters, then I think that’s what Dr. Pennington was talking about. We want to then store that information. I reused it, but I change the parameters.

Dr. Pennington: Yea, and you don’t want to overwrite, because it may have changed, you know, this property the

**Dr. Salamah**

dependency we are talking about. Of course we didn’t give you any of these examples today, because we gave you the not dependent today, but time dependent doesn’t mean that there is no time associated with the scope because the scope has to be built with some sort of a time stamp over reading. So don’t let that confuse you. “Not time dependent” doesn’t mean that the property will not have any time in the textual description of it. Is that clear? Okay. All right.

Dr. Gates: We will open it up for questions. First questions from what Dr. Salamah has said, and then we will go back and you can ask questions off of your interview questions. So you had a question.

Student: From these terms that we are using just today, are we able to assume that the scientist or the users of the program are gonna know these terms or at what level of vocabulary?

Dr. Salamah: That’s a very good question. I think that is what Dr. Gates was talking about. That is something that you’ll probably need to investigate and see. The most important thing is for you to understand what this means, right? And try to interpret it the most with the language of the scientist and to what is more appropriate for the scientist.

Dr. Gates: So the scientist, as Dr. Pennington said last time, did not want to use these words. So when you study, you’ll notice and see the different levels of users. You should try to develop this where someone doesn’t have a background and then maybe other people, as they gain knowledge or expertise, then maybe they can very easily see “oh this is absence”. You won’t know that until the scientist reviews and you start prototypes. That is a really good question.

Student: For the response relationship, is that going to be over 2 or is it just going to be a two sensors relation between each other?

Dr. Salamah: At this time it is two.

Dr. Gates: Design for change, remember that it is one important principle. We are just sticking with two, but you should be assuming that it could be multiple.

Dr. Salamah: I am going to add for that point of design for change. If you notice this list of data that we have which is not final by any mean, this is different from what you got last two weeks or so. So after you develop the system this set of data is expanded, so design for change.

Dr. Gates: When we have two different sensors-- when we talk about two scopes

Dr. Salamah: That will be later when we talk about relations

Dr. Gates: Two different sensors-- we assume that they are in sync

Student: so they are parallel scopes

Dr. Salamah: Right so the Boolean statement might be in one single statement, so one part may apply to one scope and second might be applicable to the second scope

Dr. Gates: That’s a good question.

Dr. Salamah: typo for the existence. The formal definition should read “should hold” at some reading with the scope

Dr. Gates: We’ll correct it and post it on Piazza.

Dr. Gates: Team 1 questions

Student: question about the time stamps, is it just one gigantic reading from one period to the next, or is it some discrete intervals of time stamps that are read

Dr. Gates: That’s set up by the scientists. They can say were going to collect it by the hour or were going to collect it continually, or every 5 seconds, or every minute

Dr. Salamah: However it is set up, the scope is being built by distinct time stamps. If we’re reading every hour for example then we have 24 readings. Now the scope could be any one of those, and subset of those 24 hours

Student: Question for the formats. We talk about the interface having to display the information, but did we need to have us implement the need to download the information on a different format, like a text file document file.

Do you we need to provide ability to download reports on pdf or.

Dr. Gates: It depends on what you’re talking about with respect with information. If you’re talking about the data itself, you’re not concerned with downloading the data in any format--that’s the job of some other system. What you’re concerned more about is capturing the properties and then possibly graphing the data to show the relationship to the properties--doing a visualization of that data.

There should be a lot of questions on how do you visualize the data, and how you set that up. Dr. Pennington said that I could be in front of a machine, or I could be in front of this visualization wall. I may be showing temperature data. I might be showing a lot of data on this wall; if I’m doing near real time I may be viewing a continuous feed of data and how its mapping to my properties. I’m going to stop there I’m trying to prompt you to ask more questions. We’re more concerned about the properties, displaying the properties and getting a report about what the anomalies were.

Student: So instead of logging back into the system and having to request the same properties being displayed we could have a document that’s already premade, they can open that up and it already has that information

Dr. Gates: So if I hear you right, I’m a scientist in the field and I’m working at the Jornado [site]. I should already have a lot of properties. I should be able to reuse those properties, and should be able to look at those [properties], and change them-- select from that and create a subset [of properties].

Student: What I’m getting from him is that it’s ok to implement a caching system, he wants to download something and re-use it instead of submitting another request, so should we have a way to cache certain information

Dr. Gates: Caching is only so large and it [what is cached] goes away. If you keep on adding new information, that goes into your cache. The reuse for me is more of a repository. I don’t think it’s a cache. You could be monitoring over months. I mean you set it up, and you’re monitoring for months. Unless you going from site to site to site, what you set up is what you set up for that period of time. You’re not setting it up for 24 hours typically. You’re setting it up for weeks, so these files are huge. That’s a big problem for the monitoring side of it and