

웹 소프트웨어 신뢰성

- ✓ **Instructor:** Gregg Rothermel
- ✓ **Institution:** 한국과학기술원
- ✓ **Dictated:** 전유진, 홍현석, 박금란, 구민지, 천지은

🔊 [0:00]

I don't have many slides for this one so we going to have talk one.

Um, it's not my paper so I didn't have you now three years prepared for the talk.

Now last time did I talked too fast?

Is it okay?

I will try to slow down maybe just a bit.

And it's going to be more you guys talking anyways.

Now I guess 규철 has already presented correct?

Not this.

But you presented the invariant work which is based on the top of this and it sounds like you presented quite a bit of the underlined framework.

The crawl Jax's framework which sucks.

Cause I'm presenting crawl Jack's framework and I have nothing underneath this to go into.

But this is, so I selected this because in the previous talk, I talked about automatic testing just in general.

But a lot of my experience is from sort of a critical ebriotics perspective and their technical challenges.

But there aren't big show stopping technical challenges where we say..

Ok we could do this in theory, but in practice it'll take a lot of engineering of a lot of building tools.



In web testing, there's a lot of building tools.

A lot of very strange things exist like you're talking, between two machines, which doesn't normally exist when you think about a program, you've got an interpreter for a language and you've got multiple interpreters, and each interpreter operate slightly differently which is not you know, even the notion of what the program should do, changes of web-testing.

So I wanted to do which find one paper sort of at the root of the problem, and so this is what I came up with.

Here, basically what they trying to do look through a web-page and looking through a program is very simple.

You can just fill it random and put look through the program, looking through a web application you can even necessarily do that.

So I thought I talk about that.

Um... So again, I have a list of questions, a list of things that I want you guys to think about during class and so we're going to touch upon those and then we ran a little long last time but I think particularly what you presented, part of what I'm going to present, and we can maybe get things done things little early.

So speak up.

Otherwise, you'll be here all day. Okay.

So the first question when you talk about moving from traditional testing a Java or see whatever, the web testing is what is it about web applications that makes it different?

🔊 **[03:03]**

And testing let's say a Java program.

I think you've thought a lot about a testing a Java program.

At lest you should cause that is your job.

That's what you pay to do.

What about web testing makes it different than java programs?

(Student speaking)



Right so one of the things that when you talk about what application is.

It's like an onion, right?

And then the outside is, I mean not everybody is like an onion, there is ... and we go under and under.

So one application has really multiple languages.

And when you talk about having a tool to go through the execution of a program, what you mean is a tool to go through the execution of several languages.

So I agree, that is one of the big ones.

Anyone else have any? There's one other thing I want somebody have to say.

Exactly!

So why is undeterministic?

(Student speaking)

In what sense?

I wouldn't say it's not undeterministic may be the implementation is but or it's unclear exactly about what's going on the specific point in time but, I think it's supposed to be the undeterministic at least.

Maybe in practice or not.

🔊 **[06:04]**

There's one other thing but I think you're right because you've got different execution environments which normally you don't think of with the Java application.

A Java application exists in a virtual machine this exists, it usually exists in the Java virtual machine or you care about.

This exists in the Microsoft java script that is different than the Firefox java script to which is et cetera.

(Student speaking)

What do you mean by that?



(Student speaking)

It's another layers.

(Student speaking)

Yeah, there are a lot of different types of inputs.

I don't know that's quite, I mean that's a problem of somewhat you make.

Because it's always guise.

But one other thing of is the most web applications are inherently current.

So when do you talk about testing, I mean you've got done work with concurrency testing.

So you've got two different things executing side by side.

So part of the state what's going on here, and what's going on here, and what's going between these things.

Now there is other works specifically focused on currency testing in this field.

I think it was you know.

It is one of the guys from the lab.

Cause there's a guy on a lab who do on concurrency test.

And he presented about concurrency testing and another student said "I've been a student long enough to know."

And you shouldn't try to do concurrency testing because it's too hard.

And so some of the bugs that occur in the web application relate to you've got something happening on your end and you press some buttons and crap happens on the other hand, and it needs to sense something back, and when you test it, you got to test it on a local machine.

And so it's going to be really crispy like that in between stuffs so fast that you can't even click fast to make enough difference.

But in reality, like so for example, when I use applications because I'm an American,



the web sites I go to are in the United States and so stuffs that used to work for me very well and U.S. looks like it garbage in Korea because it takes about a seconds just a send a signal to the United States and get it back.

So just a concept of have two things operating side by side causes problems.

🔊 [09:00]

Many makes it difficult for test because it's difficult thing to simulate.

All right, so what's Ajax?

What is Ajax stand for?

So you've got studied Ajax class then somebody's presented about it.

So what is Ajax then?

(Student speaking)

Probably wants to go home kind of data. Now continue continue...

(Student speaking)

Right, so it's basically a set up techniques that let you can modify the DOM in a real time and the other thing anybody, no?

[?? 10:19]A sink is

I mean that's the big deal you can think when [? 10:26] is called before it's kind of before my time.

But I guess you couldn't.

So before we talked about the problems unique , yeah I stole this slide from them so it don't quite fit.

Basically it talked about the old stuff and it talks about java script that you've got right?

So you've got the dynamic display for the DOM which basically means you can tinker with web page through Java Script and you don't have to refresh, right?

You can do things in the background.

Let's see, ok? In Ajax-based model application, you got this crap happening on your



end to the UI to the Dom.

We talked about that. And then you've got this crap happening on the server side and so I don't know if you've ever written web application uses Ajax, I worked for the United States geological survey, which is the parts of the government that deals with the maps.

And so we built some application to display all this mapping data and this is like.

This is totally banana, this is what exactly what you would expect you're going to have some decorders that handles your Ajax stuffs.

And then you're going to have some backing stuffs, in our case, it was the part that talked to the map server. And then just sends steps out.

Super simple.

Ok. Let's see... What's the next one here? Ok, all right! So the application logic is in two locations.

🔊 **[12:03]**

What is the state of an application like this? Like state of the job program is what?

You get to register, there's the disk, and then there is whatever you're located in the source code.

That's your state.

That's some state program counter, register, and basically the disk.

What's the state of the application built like this? Where's the state?

Dom structure? and? Server state here? You've got things happening here.

And? The wires. The stuffs in the middle.

So that's usually the hard part when you talk about concurrency application.

As now part of your state is invisible, you can't even really see it.

I'm not sure you're going to have some simulation method for doing this. All right.

And so then, we mention this before.

How does the state get modified? Where can modification is the program state occurred or how can they occur?

Right! So the obvious one, events.



Things actually happen because the user did something.

Where else can change occur? So loading another page. again, it comes from the user.

This thing can die! Right?

So Amazon they have, I don't know, thousands of websites, thousands of servers and so you think of it like, when you think of designing this, as we would think of designing an application like this, you don't even count that.

So server goes down, the application is broken.

We start the server and start over.

but if you're Amazon or Naver, or Google, that's not acceptable.

So sometimes, this will die because the disks are part of your state and the disk goes bad, and you have to recover from that.

Now, there's the 3rd way state can get modified?

I mean, there's more than that, but there's three big ones I want to think about.

The user does stuffs.

Things happen on the server side, maybe good maybe bad. or thank you!

So you can just lose packets.

So we would do this for the U.S. GS. When I first started, I was undergraduate student.

And we had this method of encoding maps, like Google maps just came out, like we're going to do that.

So we started during Google maps. What Google maps does is that's very clever.

And I'm sure that you've seen this as they have just tiles like somewhere they have whole bunch of mapping data and this mapping data gets compiled in the bitmaps.

🔊 **[15:05]**

So you have finite number of zooms, so you can zoom at some.

You just get different set of zoom in tiles.

We had really huge multitimes types megabyte bits that we would turn into images.



Huge giant images and then user can scrolled.

It worked great on our machines cause our machine would never failed, but you lose one of these images in between. So what worked so well, when we were testing our programs, we would like almost never worked.

Actually on the internet.

We were for the government, so it didn't matter.

It actually worked for the company. They will be super pissed. No one would use that service.

So the only part of our web application that say body used was the part when they download the whole map personal used.

The web interface was never really successful. Ok.

So there's three ways to get things modified.

Now when you talk about the state of an application what pops into your hear?

You thinking about, I've got a program and it can execute a number of different ways turn it on, type in some stuffs, state changes, click on something, the state changes.

What does that look like in your mind? So for me, it looks like state-flow graph.

Not necessarily this one, but you start from some root and things start happening.

And this transition you into different state, and other things can happen that move you further down, or potentially you can go back to where you started.

So think about graphically representing the web pages is basically set a stage.

You have big index file, you could go on links you look for different places.

So what is the state flow graph? State flow graph from Ajax side as defined by these authors, is going to be the root which is basically where you start.

You're going to have this set of [? 11:11] representing the states.

These are unique states in the application.

If you going to very specific encompassing what pager are on, what did internal state of your job as descript interpreter is what's going on in the wires, what's going on the server side, the weather, you know just everything you can think should be encoded as part of the state.

And then the edges are the way you move between state and state ok?

And this is basically one way of modeling your application. Ok?

Now what's the problem of with using of model of you application.

So when you think about a model of how things work so there's actually I forgot who said it I should remember cause I repeated all the time.

🔊 **[18:08]**

But there was a man he said, a model is, all models are essentially wrong.

But some models are useful.

And so when I said model of what an application is, you can encompass everything that goes into a web application, realistically you can think about the stuffs in the wires cause you could even view it, and if you think about what's happening in the server, too much data is really captured.

And so, the first thing they do in this paper when they start talking about web application, is that kind of bunch of craps.

And so, in this paper how do they do think about moving through web application?

What's the main method of thinking about?

What's the conceptual model they use to think about going through this web application?

What part of that web application do they actually think about that in this paper?

The Dom, so they're only concerned, with what's really going on in this side.

And so conceptually, it kind of as long as nothing really strange happens here that the entire application can be viewed in terms what's going on in the Dom.

Now it's going to cause problem. We'll see later on.

But key thinks to know is that this thing, this state-flow graph they use to describe their tool, is just a model of how things work.

It's imperfect.

We already identify ways of imperfect, all right, so!

So what are the problems that come in working an application,

(Student speaking)



I'm glad you way back there.

What are the problems in modeling a web application in terms of the state-flow graph?

So what do you want to do is work state flow graph starting at the index.

So what goes wrong? in their approach? in trying to any approach in this kind of model?

Pad-explosion.

So you're going to have how many states you think exist on the front page of Amazon?

How many states are reachable in the front page of Amazon? Thousands, right?

So you've got a track every single state and you've seen come up with some way recognizing somewhere that you've been before.

So what constitute states if you think about in terms of Dom?

(Student speaking)

The structure.

(Student speaking)

🔊 **[21:03]**

So any two Doms, that are different states.

So what about the part of the Dom includes the time?

Insight as a part of the Dom, yeah. Right?

So I guess I'm saying is you time stamp is part of your webpage. Somehow incorporate in your web page and every time you click somewhere, you going to get somewhere you've been before unless you're time-traveler.

So it's always going to be changing.

So how they deal with this?

I guess. More generally you can think of two completely, if you have two Doms and they different anyway you can think of being different and technically that's correct.

Technically they are different states.



Practically speaking that can be a little two regulus.

A little two... so they want to ignore some stuff.

How do they ignore some stuff in this paper.

Right. So they have a threshold.

They've got some way of different things these.

That's right.

Cause in your work you are doing that texture stuff, right?

You have a text-processing tools as part of your project, right?

And part of those tools tends to be looking at two pieces of text and trying to see if they are same or different and then measuring the difference.

And so that's kind of what they've done there.

Now later on and work, it's weird, it's all temporally backwards.

Later on and 규철's work which is already even presented, they come up with a different way of doing this.

You guys remember what that was?

It was something else actually.

So they had competitor.

Did you talk about the competitor?

(Student speaking)

Okay. So the competitor were things they use they just took the word to compare and made them into a noun.

So it's just a method of comparing two different part of the Dom.

And so they would want to do different things like or the page.

They would want a strip white space like you don't care about the time or things like that.

And so they would basically take some page and compile it to via the competitor into a simpler version that only captures what they thought was important.



Time goes some other crap goes.

🔊 **[24:03]**

Oh, is the... Okay.

We read bunch of these.

I am Sorry.

So anyways, so in later work by these authors they did this.

They had some way of stripping out of information that... so this is what you would use and then you could just take if two things were identical and then you really said that it's identical and the benefit of that over the threshold is what.

It can reduce the state.

I mean the threshold can reduce the state too.

Anybody?

So the benefit of using a competitor to compile or just to strip out useless information is that you semantically you've taking care to keep the parts of semantics you care about and ignore the parts you don't care about.

In the threshold it's, you know how I can say it, it's a Dom method or another Dom... it's not a bad idea but it's a crude method.

It's just looking at these things as if it's a stream of bite or a stream of text as oppose to a Dom document.

The benefit is that that's much simpler.

You don't have to ask a person to come in and right competitors.

So there's a difference.

So, all right.

You've got differences in states.

You've got a large number of path.

How do you find a path?

So in the java application it's always obvious when you're getting input.

You are reading from a file.



You are getting it from users.

It's coming off from network.

How do you change state?

How do you find new paths in the age X application.

Clickables.

Right.

Is it easy to find clickables?

What's wrong?

Why is it hard to find clickables?

Exactly.

So you've got... you know... there is work bite tiff them on gui testing which is kind of similar.

I think you probably even in sighted in this paper and I think about a gui testing is your whole gui is generated from Java.

So you can and it doesn't change.

I mean it can but it's not really what you expect to happen, and if it does change it does so on pretty programmatic way.

The Dom is a...

Did you watch the new 007?

I saw on Wednesday.

Anyway so there is one bit and it's like the swirling mast of thing consistently changing.

That's kind of what a Dom is.

It's... you can't really just... if you look at it once you are almost certain what that is.

When you look at it five minutes it will be changed.

All right.



Finding path is challenging.

Finding clickable is challenging.

Okay.

🔊 [27:00]

Now if you are going to walk a graph...this is not a great graph.

I don't know why they chose this.

But sometimes you go forwards and sometimes you go backwards.

You have to track what states you have been in to know where you have been before but in order to walk a graph like... alright.

So how's there was like one loop.

It's all you need.

If you have this and you start here then you go here and you go here, okay.

So if I start one, two, three, I have explored this entire thing but from this point there's no where I can go.

Right?

So I need to back up.

In order to explore graph you need to back up.

So what happens if you back up in this.

Right.

Right.

Exactly, exactly.

And so backing up... can you just hit back?

I mean it says I can hit back.

If I open the web page right now it will back.

Does that work?

Why?



Right.

Exactly.

So there's state here and then they state here but none of that state is captured by the back button.

And so alright, you can use back button.

The easy solution is gone.

There's actually a friend of mine does web testing, actually does web testing.

And the all dissertation came from accidentally buying the same pair of shoes twice.

So shoes on a web page and she bought shoes and did some weird thing and she press back and then bought her shoes again and she ended up with two pair of shoes and from there the whole this rotation float.

And I guess there two hundred dollar shoes so she really needed to graduate and get a job.

Now what compensate state here?

We've already talked about this.

Quickly somebody tell me.

This is a state in web application

There's a DOM, web page, server and the wires.

🔊 [30:16]

Thank you.

Alright so I back up.

I want to unroll all of that, right?

If I just back up in terms of this they want to, you know... I bought my shoes.

The shoes remain bought and when I go forward again I will buy more shoes.

Now what is that mean in terms of technical challenge and terms of tools.

In Java Script, and JPF what happens when you back up?



All of it?

So when you say... I mean what is the store, what part of the state?

Okay.

So like... it stores registers, right?

It stores registers and the hip.

In the stack, right?

So it's got the stack when you back up in this it rewinds the job of virtual machine.

The only thing it can't rewind is the hard disk and the networking connection.

And because it's NASA there's usually not a networking connection or hard disk.

It's on part of space shuttle.

So when you back up this, technically you need tools that track not only the DOM which is not that hard but also the server side at the same time and whatever crap was in the between.

So to build something that does that you need something to simulate all three of these to happening at once.

And so they don't do that here.

They kind of punt and terms of that.

Right.

So what could be done to make this sort of thing easier backing, backing up, moving forward.

Imagine you have a hundred people what could you have them build to make this type of task simpler.

Like what is the ideal scenario for in terms of capability.

We are talking about how when you move this thing backward, this thing does move backwards and why is that.

🔊 **[33:04]**

Right so when you say screen shots, what do you mean?

(Student speaking)



Snap shot's your word. There's a dude in a camera, Snap shot, of this?

(Student speaking)

Both. So what would that be in terms of a tool?

(Student speaking)

Right, but you, like....., I'm trying to get you to say a specific word but maybe an emulator?

So when you run an android, when you develop android applications, they have an emulator, and it simulates the whole hardware internally. Right?

You don't test on the, I mean you can test of the phone, but you test on your machine.

Or if you have, like.....When I was a kid, kid when that long ago, when I, super Nintendo emulators, and Sega Genesis emulators, something that pretends to be the entire thing.

So if you have an emulator for this conceptually, you could do that.

But why do they not use an emulator? Why aren't they done it in the way they have, which is to just wander around this part?.....

It's an easy answer.

(Student speaking)

Exactly, because they have three people on the paper.

And there's, Okay. So this is.....

You know the previous thing we have talked about before, which is me, writing basically, there was not me building anything.

But this is really where you see why there's not a whole lot of work in web testing, so when Greg told me he was doing a class in web testing, the first thing I thought was that, there's no way you can fill six months, or four, how long is it, four months of class with web testing because so few people do it.

In part because right from the very first day, you know the Java stuff starts by assuming that you don't have a network connection and you're not going to write to the hard drive and you don't have a data base, which is not a great assumption, but it holds sometimes.

Right?



There are some applications that are like that.

All applications in web testing are like this, so it's a great example of curving your..... shrinking down what you'd like to do in your research to fit within your technical capabilities.

That's not really the part of the paper, but I think it's something worth pointing out.

Alright, so we talked about that we talked about what it means to walk through a state-flow graph, some of the challenges of that.....right and so, here's, they basically done all the crap that we've suggested already.

Alright, and then there's problems you can go back.

He has one animation in this whole thing, and it's this.

Okay.

All right, anyways.

So there, that's all the stuff that I wanted to talk about with his stuff.

Okay so anyways, what is the purpose of this tool?

🔊 **[36:00]**

(Student speaking)

Okay. Why would you want that?

(Student speaking)

Okay. Okay that's correct. So what would you want a crawler, in general?

Search engines. Okay. So why am I presenting something that's used for search engines in a soft ware testing class, in a soft engineering class?

(Student speaking)

Right.

So what would you, if you were going to use this software testing, what would you use a tool like this for?

(Student speaking)

Test.

Input generation, right?



Right, so you, the benefit of something like this is that it gives you the core capability to wander an application, which is the bare minimum, in order to generate tests for web application.

So that's correct.

Alright, so you've got a tool.

It wanders through your web page for the purpose of wandering through your web page, this is not a testing tool right now.

Later on it turns into a testing tool that's where (mutuals) comes in , and the work that has compared hers and et cetra.

Okay, So how do you evaluate a tool like this? Like, what are your matrix?

What do you care about? I have a list

(Student speaking)

Right. So state coverage. Okay, what else?

(Student speaking)

It's get to go pretty fast, right, otherwise you can just have a guy in India do it.

That would be cheap.

What else?

(Student speaking)


Exactly.

Correctness, so that the design behind this tool is lossy, in some sense because we will consider two states, two different DOMs as equivalent under certain circumstances because of the, what do they call that?

(Student speaking)

The threshold, the threshold is, if it's not, you know, zero, you might consider two things same, which can introduce problems right?

You can have sort of hash collision type things.

 **[38:59]**

What else?



There's four concerns in the paper so.....

Although one of'em's kind of a duplicate I think, but, the other one that they talk about is scalability, which kind of relates to performance.

Because you are storing states, you can have something that goes through something very quickly, but if the, the representation of the state ends up being too big, you are run out of memory, if you have no good memory to store them, and retrieve them, it'll be fast for small stuff but slow for big stuff et cetra. Okay

So how do they end up evaluating this?

These are our metrics.

Like what does it mean to measure effectiveness?

What does it mean to measure correctness?

These are the, I don't care about performance and scalability, really, because it probably works okay, and maybe not that great because it's a research tool

So if you want it to measure let's say effectiveness, let's tackle that one first.

You want it to measure effectiveness, how would you do it?

Walk me through a study design here. For effectiveness.

I want something that covers the application. But I want to be able to measure it in a scale from zero to 100.

(Student speaking)

Right. So what's my baseline? So in branch coverage what is the goal?

(Student speaking)

To cover all branches right? How do I know how many branches there are?

(Student speaking)

Right, so with this I want to cover all the states.

Do I know how many states there are when I start?

You know it's finite, probably.

But huge, but less than infinity.



So I would have to come up with some guessing that I have covered everything.

Or being able to estimate the percentage without actually being sure.

So how does one go about doing that?

(Student speaking)

That would be good. Alright. Anything else?

Statistical methods maybe?

So over time, my coverage of the system is going to have what behavior?

At a 100% coverage, I'm done.

I want to know how close I get to this line.

🔊 **[42:04]**

Coverage. Time. What's this going to look like?

Give me a mathematical function. Y. X.

Y equals (Student speaking) Log of X very good.

The answer to all testing problems is that it will have logarithmic behavior.

The fault finding over time, the coverage over time, whatever is going to be logarithmic. Not like this, it's going to keep going up. Monotonic as well.

Okay. So after a while when this thing gets close to a 100%, after a while e can assume we're getting close to a 100%. And the reason for that is why? Could you do this with a branch coverage?

(Student speaking)

Really?

(Student speaking)

It depends, on what?

(Student speaking)

Right. So what happens when state, state, state, state..... bunch of stuff happens here, one state, most of the program.

So I can, what can happen is, I can go and I can get close to something for a long time. And then you get stuck, and then you break through, and then it spikes for a while.

So you get kind of the same thing here, which is why they haven't done that. I mean, if you combine this measure with the number of clickables, maybe then, you could do that, you could say, I have, you know, this number of clickables and I have found this number of clickables, maybe.

So here is a couple of ways.

What is the way of they used?

They added a state basically, extra clickables; and then they checked if they could find it manually added extra states.

What do you guys think of this?

I should use it. I should say that with the flatter voice.

In terms of the fact you want something is that related to some measure of coverage what they've done is just added a couple of things and saw that if they could catch it.

So thoughts?

🔊 [45:01]

(Student speaking)

It is in some senses. So the expected value is that this is detected and if they detected, they say it's good. But...

Yeah so for this one I think the idea of the monitoring the number of clickables probably the best you can come up with some estimates of how many things are run it for a while that and see how close you get at least it gives you someone object measurement shoot for.

As a scientist looking at another person evaluating their own work, manually doing something and then checking see that you can catch your own fault seems a little dirty.

You know if you are really honest, you know that's fine. But I've seen a lot of dirty things happened I don't necessary believe that anybody is this honest to do that.

All right, the effect of this is bit sloppy. The other one you want to know is correctness.

You'd like to know when you have two states, and you save the same they are actually the same with some high degree plausibility. When you have two states, you say they're different they're different according to some matrix.

How would you do that?

Give me an idea.

Arch my eyebrows to somebody.

(Student speaking)

We'd like to know the method they used for that in exactly, right? When two states are the same, they should be pretty close to the same. They should different minor ways. When you say that two states are not the same, they should be different.

You are going to have percentage of states.

There are basically four answers.

It's same, not same, estimate same, estimate not same. All right?

So you'd like is when something is the same, your estimate is the same.

If they are different, your estimate is not different and so 100 % of things should happen here, and zero percent should happen here.

This is the goal.

Basically as you feed this thing state this is what needs to happen.

🔊 [48:01]

So how would you do that?

What is the first thing you need to able to do this?

(Student speaking)

Exactly! That's the exactly basically what you're doing.

You want to make sure things are getting banned or they're belonged.

You don't want to categorize have two states which are not same be estimated is being the same. And you don't have state which are the same, being estimated is not the same.

What's the very first thing you need?



(Student speaking)

Training sample.

So what happens?

I guess what I'm getting at is we have a way of saying of two states of not the same.

You've got the threshold. You've got your differencing matrix so how are you going to measure of two things are actually are the same and not the same?

(Student speaking)

You need an oracle. You need to have some way objectively saying these two things are the same, not the same.

Where does that come from?

You've got a do it by hand. Okay. Who's going to do that?

So realistic lead the author? Who should do that?

(Student speaking)

Probably, you want to have some way of arguing that this is what means to be the same this is what means to be not the same.

And have that in the paper and say you know we don't care about time, we don't care about x, we don't care about x we went through and verified these pairs of state they are different according to this for these other state we verified they are different according to you know semantic things.

Step one is coming with a matrix. Step two of study relating to the correctness.

(Student speaking)

We don't have any data yet. We are step 2 we've got a method of measuring.

What do we need then?

(Student speaking)

Exactly you need in an artifact. You have to have something measure over. Okay?

And then step 3, right?

So what should you compare?

Should you run the application and have this happen?



Or so there is two ways to go by.

One is to run the application actually project over a system and then for each one you need to categorize tool categorize it and you need to categorize it.

🔊 **[51:07]**

And then the other way... the other way is to have to do the compare separately.

Take the piece apart and feed it states.

Cause there are some parts and there are some boxes in this; state, magic, bin.

Some would necessary be two but some bit of this tool that sucks in states and turns it into in this case is number.

I think they have threshold for saying of two things are different so they feed it state and says yes/no.

So anyway as I just pointed out I want to make is two ways to make this.

Ultimately this is the matrix you want.

The problem with doing it using Crawl Jax is what? How many states you're going to see when you run an application?

Minutes?

It is a reasonable number of states for any sizeable of application you're going to look at thousands, if not tens of thousands of states and because of that's going to impact of your artifact selection.

If you want to do this using Crawl Jax and have person come in and do the comparison.

So what I'm driving at here is that this is very expensive very hard to do in a study.

So you chose people sit down and compare some invariance.

One of the challenges he faces is finding a reasonable number of invariance so people can't sit there for week on a half going through this things.

지민 was part of the study we did with the invariance before you had eighty invariance right? and you fell asleep during the study.

I mean it was a lot of sitting and staring at text and it was a really hard.

I don't know you fell asleep, but somebody fell asleep.

(Student speaking)

You did completed.

You had the easy one actually.

One with the harder one people didn't finish.

What they've do.

This is hard, right?

Doing this is hard, coming up with a reasonable number of state is hard you could do this, but you have to tell apart your application.

So what they've end up doing their application here?

They put one state that they thought should be equivalent to another state and check that if their tool could find it, and the rest of it is manually checked whatever that means.

🔊 **[54:10]**

It might be what we've just described or maybe not.

Okay.

So again, this is not the most convincing evaluation to put one thing in the manually and see that you can catch using your tool.

It's bit disappointing.

And the other thing they use for performance and scalability is just measure in terms of time it just run certain amount of time.

And then they compare that against nothing.

There is nothing at all really.

Wow.

So you can definitely be sure that there tool will do it faster than nothing at all.

But on the whole it is just a nonsense of evaluation.

To put up some wrong numbers without any context what so ever, I mean it says it doesn't millia seconds, probably not that slow.



You know minutes or something might ask yourself could I sit there and go through it with hand faster?

That's the evaluation.

That's actually all I had.

I want to talk about evaluation and walk you guys through how you measure of effectiveness how you measure of correctness what makes this one unconvincing... etc other than that that's all I had.

You guys have any questions?

I know you have to do reviews. Is there anything you stuck out? In terms of why would you build something like this? How was it built? How was it evaluated?

Your passion for learning is overwhelming.

Okay if that's it then we're good for all I think.

I will be a little early this time it supposed to be late.