웹 소프트웨어 신뢰성

- ✓ **Instructor: Gregg Rothermel**
- ✓ **Institution: 한국과학기술원**
- ✓ **Dictated: OES 봉사단**

Or some.. some way that you can do something better than current solutions.

And that's what the project has to be involved with.

I've talked about four different types of projects you might do here.

Oh.. I should not space back and forth because you can't follow me actually stand still a little bit.

So I keep forgetting that.

And four of them I talked about here.

Research fairly, simply is formulating and offering a solution to a problem and by itself.

That's what it is.

But it also usually involves components, some of the other components that of a project about to talked about.

You might choose, if you are looking into a new area.

I'd want to talk about this carefully with you but if you're looking into a new area, one possibility is to do a survey of the literature in that area.

If it's an area where has been a lot of literature that's what you could do for this but it's not just enough just to read papers and say what they have in them.

You have to try and say something new about the area.

You have to use the search to trying find out what problems have been addressed, what has not seen addressed and to trying get to something out of that.

Still sometimes for some people that's the best choice because maybe there's just trying to think about the area.

Implementations involved taking some existing algorithm and implementing it.

Just an implementation you know I feel almost never enough for a paper and so there’s get to be some good reasons what you want to focus on is implementations I thinks there’s get to be some good reasons here too, hopefully would be something.. hopefully, there might be some small research portion to it or some empirical portion to it.

But that’s a.. it is something we can talk about.

Experimentation involves taking existing tools or algorithms and doing some study of them and various types of studies are possible.

Wewill look at papers that include examples of studies.

So we can.. I mean you can imagine if there are papers that have proposed certain algorithm you could do some study of those of implementations of those to try and see whether they perform efficiently enough for, effectively enough for.

You can compare techniques or algorithms.

So the type I hope most of you will end up doing is our research or experimentation and if you do research or research one that has probably some.. you will do some reading of literature because you have to know what the background is.

The.. you won’t be do full survey.

You might be doing some implementations if you are coming up with technique and trying to study it empirically.

If you do experimentation you can focus on experimentation without creating new algorithm.

You might (have to) do some implementation to run the experiment.

But there are wide range of things to do and you will have to come up with an idea or something to do but we will have to talk about it or you will have to do proposal about it.

So that I can make sure it’s adequate and also that is not too much to do in relatively short period that a class has to offer.

So there are some examples from the none web realm and let’s look at those examples for a minute.

So what I classified is a.. these are all projects that someone has done in the past in prior version of the course that were not focused on web dependability.

So prioritization for interaction testing.

I will read from this description interaction tests exercise combinations of system components configuration testing involves selecting some of these combinations.

So in many actually this would apply to a GUIs and web based applications too when you have lots of choices of things to test the different combinations of perimeters may interact different ways.

But in practices are often too many combinations let you test all of them.

And so interaction testing tries to interaction testing through a combinatorial testing tries to choose some combinations to test.

But this particular.. I just say there's been a lot of research on combinatorial for testing of interactions.

But this particular project was to look at prioritizing or ordering configurations if you have a large number of configurations to test and you have a finite period time you'd like to test somehow the most important ones first or the ones that will make the most difference, ones that are most likely to have faults in their various reasons for it.

And so in this project the person who did it was trying to find a way to, first of, to define kind of the importance of configuration and then to find a way to prioritize them so that you could do that in testing.

And they did this project and ended up being a paper.

They ended up doing their whole ( ) on this topic.

So as far as for the course, I don't remember exactly what they did but I am pretty sure they.. the paper focused on methodology and then I don't think they had much in the way implementation because it was relatively complicated to do.

So it was primarily a methodology paper with what we called extended example.

So not a tool implemented that would run but unextended example that you show how would it work and for a short period of time that's what we might have to do here for some of the project you come up with.

If it's not feasible to implement and get any data on but that's where we get to look at your project and then decide how much it can propose.

Example that implementation, a project this is from while ago too recently just have been reached on a topic of testing spread sheet actually we started that 14 years ago and automatically generating test cases for spread sheets.

In that case, we had developed an approach and we had looked at actually two approaches.

One called you might not know what that is.. It's okay. It's dynamic test case generation and then one is randomly generating inputs.

There had been some other techniques developed not in the spread sheet realm for using constraint solving and other algorithms for spread sheet.

And so this project was involved implementing one of those other already-defined techniques in their spread sheets.

So there's a little bit research there figuring out, you know, how do I apply techniques in the spread sheet. Okay?

But it was mostly implementation and then there was, then there were able to do some.. together some data on the implementation afterward in kind of a case study comparing at to the existing approaches.

So it's really matter of what's the majority of the world probably implementation.

But there is a better research and some study in there.

A survey that someone did involve looking at test design and processes.

We will be talking about this today and next class but there're many different ways in which organization is do testing.

And many different kinds of organizational structures and many different processes by which people release software and it seems like under different processes, different testing approaches might be more or less appropriate.

And so at the time that I was.. that some a student was interested in this I had not really studied test design and process issues.

But it was clear that there were some papers in the area, you know, before they proposed it they made sure they knew the papers in the area and there were interested in working in this area and so it seemed like there was enough material to do a survey on and they ended up doing that and stealing out of that or getting out of that an idea of what has been done in the area and how can you classify the work that has been done and what are the open problems that's what you really like to get in something of that sort.

And the last one, experimentation, and this was an interaction testing processes well.

This was not prioritization but it was simply experimenting with interaction testing and as the second paragraph says that they.. you'd have set up.. they have set up an existing system that we have.. had to make it possible to do to make interaction test suite for it and then they were able to creates some suites study them and see which one we are more affective or more efficient.

So the primary contribution here allude some implementation and getting things set up primarily an empirical study.

Now I should say that in two or three weeks, we will have a couple of classes that are about empirical studies how do you do them, what are components, how you do them right.

So in a little while you have more of idea what goes into those.

We will also look at papers that contained them.

It turns out that there really is a whole procedure, process for properly doing empirical studies.

In the same way that there's processes for proving algorithm but in computer science there often are not classes in that topic.

Well classes in algorithms where you learn how prove algorithms but often there are not classes that teach (here's) how to do empirical study and soft engineering the study turned out pretty important because most of the techniques we develop our heuristics they offer partial solutions heuristic always wondering how effective can be more effective, how efficient can be more efficient and studies are needed for that.

So those are the classes of things and some examples but what does that mean for you.

There's two aspects of that.

Your project has to involve, I think you don't need to.. I will just turn the light up and right be back.

The focus of this course is web applications.

Now that's a very broad term on purpose because I wanted it to include anything that is involves programming for the web.

And so what are the examples of web applications..mash-ups and you usually? use those like Yahoo pipes and usually is(end users and?) professional things okay?

Mash-ups just web macros.

Are people write web macros? There are web macros authoring tools.

I mentioned these two because I get to has students do some research on them and their end user approaches.

But of course there are E commerce web app.

There are service compositions.

There’s whole bunches of different classes of web applications.

There’s Ajax.

I don’t know where they put that in there but I know someone in here who’s interested in that potentially.

So you are going to have to choose some domain of web applications.

That’s of interests you.

And as long as it really involves some form of programming for the web and something that’s important enough that you can argue then need to look at it.

That’s going to be okay.

So your first.. one of your jobs is to figure out what domain you want to look at.

And that’s often what we do in researches even you want to start to focus on something.

You don’t want to be too narrow or too broad but here in this class you got a limited period time so you got to be careful what you pick.

The second thing this course is about is dependability.

Now dependability means fitness for intended use.

Fitness for intended use.

So there’s lots of different aspects of that: correctness, safety, reliability, performance, usability and lots of things that go into whether something is sufficiently dependable or suitable for intended use.

Now we’ve started to look at some techniques.

When we start to read papers about the kind of bring these two things together, you will see that there happen people working bring these things together.

There are papers for instance on..

Well, I have got papers on helping find errors in mash ups and on helping find areas in web macros and there's going to be papers on testing some of the types of applications that happen in this realm.

And you will find some of those and by looking at those and that's where.. that's one way that you might find some problems that haven't been solved.

And that you can work on for this class.

We've been looking at more generally some of the underlying.. Let's say algorithms, (or) technologies that are used in techniques for measuring dependability or assessing or increasing dependability through..

We've been looking at initially here at program analysis techniques.

And these are techniques that in some of the work on dependability applied here have been? used.

And that's when we start reading papers, you will see that.

So for instance, last time the very in the class I mentioned something called program slicing they can be used to follow dependencies and the same concept occurs in some of these things and so you will see some people have looked it that possibility but there probably will be some open reach problems and solving that.

But that's why I was looking at program analysis techniques and we are going to look at testing techniques and generally speaking first and it some point here we will look at empirical studies.

That's all kind of general not specific but when we look at papers you will see there are things they have been applied.

So these are kinds of basics you need to have.

Somewhere in here and here will be reading papers that are specifically about research people have been done dependability for web applications and you will start to see what's happening with that.

That might be right after Chu-Seok.

The problem is you would really got to be thinking about the project now. Okay?

So there's nothing stopping you from going out and finding some papers involving I am not trying to bring in a bibliography of something papers next week.

That are about dependability applied to web applications but you could obviously go out and start googling things like testing service compositions and start seeing what's there and you might have started looking at papers because what you have to do is again, find some domain, find some quality of programs that you want to help with and then do some sort of project around that.

Now research project might be creating a new technique for building one of these properties for some class systems.

An empirical project might be taking some existing techniques that it had been proposed implementing them and comparing them to one another empirically.

A survey might say if you find that there’s been a lot of work done in some area here and you want to find out what all that is.

That would be that.

And implementation well there would have to be some proposed algorithms that really haven’t been strongly implemented that you want to check on somehow.

So your job is to come up with a problem.

Now and that I am happy to interact with you on that and obviously you will see as I said large enough and not too large something in by that I mean something that would fit into the class periods the dates for the.. the dates on which due things are do are showing on the call for papers and to do things little more formally I will have you each write a little proposal and it’s really only needs to be a page or two but the prop.. I will hand some information out on this but just… so you can start thinking about it the proposals going to have to say paragraph 1 here’s a problem. It usually starts out with I will use mash ups.

Mash ups are increasingly being used by users to program the web.

However they contain errors.

Prior work on this has provided some techniques but we still don’t know how to do something and that’s your problem.

That could be one paragraph, there could be a second paragraph on prior work.

The basic thing is here’s the problem, here’s why it is still a problem, why it matters.

That’s your intro and then here’s what you proposed to do toward that.

That’s really the outline

Proposal about a research project work could be an experimentation project and then here’s what you propose to do, we've proposed to create a technique and then some sort of outline of what’s your milestones are because in a research project you need time to think and you need time to read papers so it does not work to wait to the last week.

We will not get it done.

You got to get started and keep going.

If you've done research you know that already.

So I want some milestones and I want to check on you about as far as how you are keeping those.

Now, one bit of possibility good news is if you have current research topics that you are looking at and I know what several of you are looking at if you can fit that in the here that's great.

I'd like you to benefit from this not just do busy work. Okay? So, you have mentioned performance once, right? Had you not? Someone had mentioned performance a big interested in performance.

Did you mentioned that once?

Maybe.

Okay, well, you don't have to be interested in it but if you were okay and there are obviously some of these classes of systems there are performance issues through put response time they are waste to measure them.

I don't know the literature there at all.

But obviously you can look and see what had people done for ways to let's say assess performance, measure performance improve I am not sure whether the problems are okay?

But if that's something you are interested in, you can focus on them.

And obviously you have got to see whether there is something you can do for this course if you don't think so you will have to pick something else. Okay?

Regression testing of Ajax program.

Okay if you are interested in that well you can see what's been done and if you are just working on that see what's been done and see if you can do something read some of the existing papers.

On that trying come up with technique so whatever your research is if you can find a way to benefit your research by this project we will find a way that to benefit each other that's great okay?

So you might think of it from that point other ways I will say it is if there’s particular type of system you are studying but you don’t know much about dependability well at least you know all about that type of system like you have some domain knowledge so maybe that makes it possible for you to do something with that type of system with that other people haven’t done.

Let’s see was there something else out there.

So I have you a prepare proposals I haven’t set the date for that and I will have you present them here informally say what they are so everyone knows what everyone is doing that’s not could be a long presentation but because I think it's useful if everyone hear what everyone is doing.

Also I’d mentioned everyone would do a paper presentation that paper should be something very closely related to your project to paper you have to read anyway.

You will possibly again to pick on performance if you do a performance of something here performance of e-commerce systems probably just been some work on that topic or there’s be a paper on performance issues generally and that’s a paper present to us cause you will going to have to know that paper that way we learn something about the area projects in and then later on when you prepare the paper and give it a talk in the class we have a little bit a background to understand that work.

Now, questions. Yeah.
[Question]

Um Okay um testing is let me see it this way testing is let me start differently there are lots of properties that we want systems to have correctness, safety, reliability, performance usability others too.

Let me put in other one down there.

Maintain ability.

How well is a code clear enough that we can maintain it overtime without it breaking down quickly.

Other’s reason is add in that one right now.

For any of these, there are different types of techniques that might be used to improve, or measure them.

So, testing is a class of techniques that is used to measure and to improve certain factors about systems I was certainly used a form of testing to consider correctness to consider reliability.

Safety not so much because safety is an absolute in testing can get that.

Maintain ability, I don’t know how to test maintain ability.

So testing is if you want we have there are many attributes in dependability and I guess the third thing I guess I didn't put down here really was I could put techniques and under techniques there are various things um testing proofs usability studies and some of these apply these and some of these some of them buy find others so you are not respect to testing you can find, in some sense, you can find a domain, find an attribute and given those too there might be different ways to approach those.

Now, usability I don’t know how to test for that either and essentially how showing you inputs I’d really need to do use ability studies with actual people.

So um at your comments and are right there is something testing doesn’t apply to and you are not require to respect your attention to test there are many other things.

So another thing I will say is look prior look at the paper we did on mash ups in that paper now we talk about the web macro.

Web macro, you know what macro is you in-code something now you run again and again and again.

A web macro tends to be done by capturing some actions in a capture replay tools such as you can play it again and again and again and there are tools for making web macros.

So they are for often done operations but a problem with web macros is overtime let me that may stop working and if you record in macro in a certain setting and it depends on some website out there that changes.

It may stop working and so and sometimes it’s obvious sometimes it’s not sometimes your macros still returns to result.

It’s actually incorrect result and you don’t know it.

So we didn’t do testing but we put in into the macros, we put in checks actually assertions that were checking whether things remains the same or things were still working and something broke and they could operate behind the scenes and if something broke that they could detect and they could tell you so it’s not a testing thing and it’s more of a checking thing or detecting areas when they happen rather testing will see is taking system throwing lots of inputs at it to see how good it is.

Checking just says well I have got some code and anytime you run it, I can check some aspect of it.

So that help so long, ask more if you need to.
[Question]
Well, that’s why that’s why we will discuss it and there’s a proposal.

So it does need to be dependability issue. But, really that's quite broad there's many aspects of dependability.

Do you have an idea in your mind that you want to mention pacifically or yet or? No?

Okay.

Other thoughts or questions? I know some of you are already working on coming up with things but if you those of you because have some good idea we can talk about it.

If you get some ideas and want to run them buy them let's do that.

One way to start is to use these little examples.

This is a proposal examples on the first thing you might trying to do is if you got an idea could you write these paragraph for two that describes the project that's what you kind of can got to come to me with and say you know and that expends into a one page proposal later but that's one place you going to start.

Ok,

If no more question, I will start with material.

That's? I guess the Back one. Is that ok?

Yeah, I just thought something else, I could say I mean dependability is very broad and it is not all there's whole there is another side to dependability.

I tend to be someone who I am interested in testing and verifying things and so usually I assume some programmers have created that the program of some sort and now I am trying to measure it, assess it.

The other side to dependability is a there is a saying and testing you cannot test quality in.

You have to build it in if you got a system that's got lots of faults in it testing it.

You can remove some faults but it doesn't raise the quality bar as much as if you built it better to begin with.

Ok, part of dependability also is how you create more dependable applications so there are people interested in programming language environments for creating system that are more dependable in those environments can you help users not create errors so there is.

If you are interested in and the system the programming side of things system creation side there are dependability issues there are too so I don't know if that goes toward in anything but keep that in mind it's not all about here it is how do I look at it afterward so also but how do I build it right to begin with.

Alright, but testing is something that we do after it's built pretty much and that's what I want to go through some fundamentals of right now and then next class look at some specific testing strategies and techniques so what is testing is the process of it says as dynamically exercising a program that means running it executing it dynamic means execution to determine whether it meets specific quality needs such as matching it specifications, conformance to specifications, performance, reliability, or robustness.

Not maintainability it's hard to test for that alright but certain requirements, now in this slide conformance to specifications or you are doing what it's supposed to do is the thing is look that most often.

I will focus on that but there are other things you can do testing for so what are we doing when we try to tests? or what are some test what are some of our objectives?

Well one is to uncover faults you've created this program as experience programmers you know that programs often contain faults cause it's hard to write them and we test them to try to find fault.

So we can correct them before we will use it so in that context in some sense good tests how are ones that might find faults it's interesting to think about that because when people begin they sometimes think well hey the test didn't reveal it fault that's good ok?

But in from a, from maybe from the point of you or someone who develop the system that's good right?

I mean you want a, you think you are smart, you want to prove you are smart, you want your test to show that the system works as intended but someone who's paid to test once could break it.

Good tests have high probabilities of finding faults and that's what you used to trying keep in mind because quite frankly however good a programmer you are when the programs get complicated there will be faults in them and you want to find them so successful test cause failures find faults.

You haven't discovered but the big thing about testing is it takes time, person power, man power woman power, ours engineer's salaries so they are trade-offs you obviously you have to stop testing eventually and so how long should you test, how much testing should you do, what type sort of testing that's always the trade-off which mean the cost of testing and what benefit it yields and that's what we find most of what we do is about.

So how we do testing? Well there are actually is I think that is a company that has some protocol monkey testing I wrote this before that product exist but. Okay. Let's suppose this you work for an engineering company and I come to you and I say I will give you a room full of computers and trained monkeys and you put your software on the computers and the monkeys are going to hammer away at it and when things break messages will be sent to a console or somewhere and the monkeys are rewarded with banana chips or something so they keep going and the things get tested.

So <Kind of random testing?>

It's a kind of random testing approach isn't it?

Although how random will the monkeys be.

Smart monkeys know how to not spend too much energy right?

So, he might not be random but it certainly sounds like it could be if we train them right? Okay?

Maybe we will reward them if sufficient randomness in what they do.

Other good things here about this when as supposed to something less random?

Possibly cheap I am hoping monkeys cost less to maintain the engineers and custodians to clean up after them but um okay. Are there any other things?

[Question]

Yeah, yeah, yeah.

So you have got monkeys maybe more like the users of the software than software engineers are.

It depends on the software obviously right? Okay?

They can they might test well for something we call robustness the ability the system to respond to unexpected inputs.

Well the testers particularly beginning testers tend to test things you know the kind of the mainline of functionality and not robustness.

So there's a couple of possibilities here and even tales of the monkeys and just do some random test generation and gain some of those possibilities too potentially.

And so there have been people of research random testing and there are some evidences in some cases it can be as strong as other forms.

But, it does have disadvantages too.

Systems are used in particular ways by users.

And users spend far more time on certain functionality and in some sense I'd like to balance my testing so I spend more time testing the stuff they're going to use more often.

Whereas if you are doing you know if you are doing 1% of your time on the functionality use 50% of your time maybe that's not enough. So it may not distribute the effort well enough.

Also, well, random testing it may miss lots of important things in the software.

There maybe things in specification that say if such and such an event occurs to software will I don't know do some self-healing or something and maybe the randomness will never exercise that.

So you may miss behaviors.

So there is pluses and minuses. Let's try something else.

/3 번 양원목님/
I will give you the program.

Now you test it! I will just give you. I don't have one here.

I will just give my some drive.

Put this on your computer in test. Does that work? Why(?) not? You don't time?!

I'm giving you time. You got time. Here is the program test.

What will going to do? Ye Ye Ye.. Is it program of that?

You don't know what is supposed to do.

Okay that's the problem. You don't know what suppose to do.

I tell you it's "sqrt". Can You test it now? Why not? That's right.

You're probably all thing in the square root. It's squart.

It runs hydronic conjectures on some machine.

I tell you it computed square roots, it square root computation program.

Can you test it? Why not? It does not take negative numbers or not. It supposes to do or not.

You can type in see one what happens but is that it right or wrong along.

What's it's..wait a minute...

I will tell you more, I will tell you more.

It takes a real number X. If X is not nonnegative computed square root.

And if negative principle X prints must be non-negative. Okay? How about now?

That's better? But, so, so, so, you give it 4.1 and an output something. Is that correct?

I don't know the square root of 4.1.

But, I will make something up may meet not be closed to 2.01214.

Is that correct? Assume it look at up the table. Okay. What bothers you? Yes.

That's true. That's right. That's other inputs have to try. That's right. That's also another even with this one.

I can tell you what precision what I want to be.

So, a couple of issues still Okay?

The point is we need specification.

Some idea what program is supposed to do.

And I can take many forms, if we were being good software engineers we want you to written up a precise specification for this system.

And what's next is that's what I test and test against and practice you see a wide range of specifications or model s or description s of system.

UML user case all to the things one way or another you need some specification.

Some terminology, a failure is what we call externally visible incorrect behavior.

Something incorrect comes out of the program the string segmentation fault.

Externally, incorrect behavior that you can see. A fault is the part of the code that causes failure.

That's not the best definition you could be missing code.

So I just say incorrect portion of code, that's the fault.

The fault, the incorrect portion of code, leads a failure an observable output.

That’s wrong.

An error that’s more something better programmer does a mental mistake.

As a programmer, I made an error by using less than sit of less than equal to the fault itself was the wrong competitor there the failure might be if I choose the value of make the fault appear.

Bug is an informal term.

That’s often used and one of problem with sometimes not clear whether means fault or failure.

And the sometimes we use it anyway but, debugging is what happens when you given in failure you are trying to find the fault causes.

Debugs take failures find causes fault. Testers find failures.

Now, often open times in organizations, the same person is doing involve they are finding failures and debugging but they are two separate things.

They are organizations with testers are in charge of finding the failures specifying full what it tested reproduce them and handing them on to the other people who do the debugging.

And an Oracle is term we used for a devise or procedure for checking the correct this output and I see devise is a procedure the oracle might be the specification and an engineer comparing output to the specification.

That’s a manual oracle. Or it might be assertions in J-unit tests.

Or a number of things they will be checking for correctness. There is a term we will use. Little more

A test case is set of input data and an expected output.

Obviously, when we want to test program not just drawing input at the program we need to check the output and again it’s checking for whether output is correct.

There are many ways we can do that.

We can check precisely for output and we can check lets precisely.

All depends on program and how much we know about specification.

Talking about random testing.

Temporally for a minute.

One other things you can do it random testing we mentioned about monkeys if you put a lot of inputs in this system.

And whereas an engineer designing input specifically does not have much time and can't find many inputs.

A problem with that door though.

If you are throwing a million of inputs at the system.

Someone still has to judged with output correct.

So, even if cheap throw inputs at it, it can be expensive to check the output.

Now, sometimes, if you can automatically oracle somehow, you can do it.

That's hard to do.

So Sometimes, people make what's call it's really born(?) with primary side.

There are many types oracle humans want they might air of checks automate checks.

It's partial oracles.

An example of a partial oracle, it's one that simply look of system as crashed.

You are on test has system crashed.

Well fine.

No crash fault that maybe other fault judge of part of correctness.

Now is random testing you can use that's a partial oracle and the people who do that.

Other things you can check partial oracle to but test case is input data and expected output.

And the expected output is really oracle comes in and that it could be different degrees of a measurement of en expected output.

And the Test with this correction a set of tested cases.

Test plan is description of testing process.

So often overall approach of specific test so if you under in a particular organization particularly if the organization that building software where recently US for some safety for this purpose.

You might be required have a test plan.

I will treat company those making software for the pharmaceutical industry.

And the US food and drug ministration oversee is that and they don't tell you how you have developed software.

But they do say you we have to have a process that you follow you have document of that process.

And so we have had the test plan and keep track of tests for that software.

Let’s talk about faults and a failure for a minute.

Consider the code. Squarez.

What's wrong with it? Is it correct? why not? it's two time Z. it's not Squaring Z it is time Z. OK? T

The fault it's a incorrect piece of code.

What's would be time of z. What's the failure?

Can you depend on the input.

What if I want it is input two.

Input to there is no failure.

There is a fault, but no failure.

What if I want learn it with input zero. No failure.

Any other input it will produce an incorrect output.

So there is a fault. An most inputs you see a use it be failure.

The error here is the fact I wrote it load it this way is the mental mistake I may want to made writing riding in.

With a When zero in is two we have something called coincidental coincidence correctness which is when a with program happens to be correct output even though with the fault in it. So it is coincidentally correct.

And that's the problem for of testers.

Because, it says that you can pick a set of outputs that wouldn't rebuilt fault the fault, even if the fault there.

The problem is to find right inputs the rebuild if it is there.

Here is not that hard if you pick more than two more inputs, you going to get fault.

Just the probablitic speaking. Or pick the right input.

Now how about this one? It's just a fragment.

But, you do you see a of potential problem of there. It could be divided by zero.

This test at least certainly doesn’t prevent why cling and two Maybe code up there does, but pretend it doesn't.

The fault again is .. it is hard to pin point the fault is this this's incorrect line of code? not necessarily.

It is perfectly normal operation. Is this maybe regarding against equal to or maybe incorrect some appear excepting input where I am suppose to not allow to.

Someone in here is code or missing code.

That allows this to be reached to be why equal to.

And If that happens, you haven't get failure.

That failure only happens for one value.

Why? Any other value wants to expose.

So, much harder is tester to get the failure exposing value and that's again where the difficulty the tester comes in.

We will be the difficulty come in.

So, I set early that a primary reason tested find the faults and requires making test cases but keep in mind a lot more going on in testing.

It's not just finding inputs and applying of program.

There is a whole oracle issue.

Judging with output is correct, there is hard we set up systems run able test may sound treble but if you are testing a box one of boxes is on the seven, seven, seven. you've got a whole bunch of stuff round it.

To put it the test inputs monitored outputs.

That's one example of testing a Medi-system, but setting up many systems requires some effort to run the tests anyone want to be able to run them efficially.

And that's going to very with systems too. long, long ago, I knew someone who worked trade ATNT, and they had a have physical lab with telephones where we are some point in testing software.

They can resolve that lab and test involved doing things with phones dialing numbers.

Not particularly efficient compare to automate ones and other test too.

But the test can have different to costs.

And the efficiency comes into it.

We need to be able to run about tests because you just sit there to cost one time program you troughing away.

Most program they are going to evolve as thing change.

And then, we got the whole process building systems how to we fit testing into the process we use.

We look at some of those.

Anyone know roll of water fall model is? Water fall outlet obviously lot of difference process building software.

Most primitive but just let me illustrate something. Water fall outlet you do requirement analysis design, implementation, some testing and maintain and testing fit here indifferent ways this always says during implementation will do something called unit testing I say but, down more and minute system.

But, if you are creating your tasks from specification you can start creating up of hear specification available.

If you are creating tasks related to sequence I don't know diams make they design face, so their testing activities can go throughout this life cycle.

To keep in mind and different time which is testing.

So, some of face testing unit testing involve testing individual program unit are different language have different unit individual program unit portrang procesures sehas function job and method classes any other those unit so, when you test those individually I'm saying each those of unit.

Let's say procesure or procesure language them.

That's unit testing.

There is some complexity to get the minute if one thing you can do often I should say often all languages we treat the class develop task for the class because doesn't make much sense to test new operation or something like that.

If you test new combination mathmethically in the class.

In the integration testing, what happen to when we are integrate or put together unit.

So perhaps, doesn't necessarily check some other interaction can happen so we start to put thing together and integration test of collection of things.

And the system testing is the way describe testing entire system usually the hardware you can do simulate another thing well and the regression testing is what happen after changes. You got the version you tested and if you save your test, if your version changes you can reexercutive applicable test.

That's regression test.

How do you things involving say integrating testing if you do unit testing now you put together. strictly speaking there is different ways to do it, strictly speaking, I could top down bottom up I could say Okay, I start joining together I take main routine and test actual code of A. now going to code add actual code B, now going to code add actual code D. you can build up that way, or you can start to test D. now calling test B calling D A calling B calling D. in any of these door, you need to have thing we called stuff drivers, stuff or drivers.

/4 번 김형태님/

If I have just C, how do I just test it. I need something to something that calls it. That’s what we called the driver.
Something that we call it.

And that driver, if I don’t have A.

That’s going to be some mark up, some prototype that just make calls to C giving a values so I can test it. Same thing goes if I’m integrating A and C. I need a driver for main. So integration testing requires us to build some stuff in order to test the things we are integrating. The other way around, if I’m testing just main, what would I do when it calls A and B? If I don’t have the code. I have to put in stubs. It might just something that A returns some value. But they fill in for the complete procedures. And that takes work writing all those.

So, there are different strategies used in integration testing to trying lessen that work.

And sometimes you might choose not to integrate one by one by one but integrates sets.

Let’s get main with A and B and build two stubs.

You don’t have to with cost effectiveness.

But just be aware that when you are thinking of integrating you got a think how to do this.

So just as an example here’s a driver If we do have a square root program and I want a test run some different values and the square root going to be embedded and whole calculator later I don’t have the calculator code.

But this is a driver that call square root on ten different… eleven different values… 10 different values and I can look at the results.

Here is some stubs.

This main routine calls routine called usage and it calls function F and goes on and does some other things that I haven’t shown here.

We’ve got to provide something that does things so I will just have usage return and then a have F return a value. Now to test different things I may need to substitute different values in there now and then.

If I want to test different operations of this.

That would become part of adjust in the inputs.. (Q: What is main difference between test driver and test stub? I’m not sure what is the difference…) The driver fills in for the calling procedure.

A stub fills in for the called procedure. So, a driver… a driver… so if I want to test A and I have just A. I don’t have any other code. I need to create something that would call A. That is a driver. I need to create something that will play the role of C.

That’s the stub. A caller.. A driver substitutes for a caller. A stub substitutes for a called routine. OK? (Q: drivers and stubs are available in unit test?) You need to write them for the unit testing too. Yes. (Q: How about in the field?) You know, I don’t know. There maybe.. Probably are some tools.. I’m just guessing. There probably are some tools.

Where if you got something they will create… automatically creates some stubs for you.

Or at least templates for the stubs, so you can fill in some values.

I seem like something someone do pretty easily.

Course you doing this with.. You are creating a new piece of software each time so… so in that sense there's not a library of drivers and stubs..

Not that I know of that tries to do this but it should be fairly easy to create something at least make the template for fill them in and say well fill in return value or something so there should be help for that and of course once you made this program I mentioned that we like to save our test.

We'd like to save our drivers and stubs too.

So when it evolves, I have the drivers and stubs. I may need to edit those If this is my program and I make some changes to B and I want to unit test to B and I have its prior drivers and stubs maybe I can use them, maybe I have to edit them.

Depends on what changes I made.

So this means that testing involves more than just finding inputs, involves maintaining all of this stuff we have that help us the stubs and drivers and inputs. So they still work. Yeah now I want to a go back and google automatic stubs creating or something like that.

Automatic test stub creation and see.. sounds like something that someone should have done. Other questions before I go on with this? In the early days of research on testing, people did try be more theoretically about this and say what is it really trying to do.

Because people were writing program and they are running inputs on them and people were trying to suggest methods for testing but the question is what are you trying to achieve? It's always good to have some more precise idea of what you trying to achieve so you can judge even whether it's achievable.

And so, one of the early testing theory papers pose this as what they called the test case problem. What they mean by that is as testers what we are trying to solve that's the test case problem. And they said well, maybe it's this problem. Given program P where P has an input domain, D. Program P. Input domain D.

Even if the program… Even if the square root takes one real number it's got an input domain that's affectively infinite. T, a member of D. A test case. And F, a hypothetical correct version of P. so we have P. P has this domain F, it's hypothetical correct version. What do we mean, hypothetical correct version. (specifi..) specification.

You know a description of what we think as we suppose to do. Choose test suite T subset D. If the D is affectively infinite, you are not going to choose D you want to be a subset. Such that for each fault in the program. There's a member in here that detects the fault.

Now let me draw that in another way. I will say like this. Here's what you want to do in your testing, you got an input domain. There's some faults in the program. Some of these inputs fail. In fact.. if I have infinite amount of time I could enumerate the one's it fail. Will say this one fails, this one fails, this one fails, this one fails. I want to say what you want is to choose T such that it include all these. In fact if you could choose exactly these, you solve this statement of problem. You'd like an algorithm. That would say, here pick these. If it picks more, Ok maybe picks more it's imprecise. But it includes all those, you've gotten what you want. If you had these algorithm, and you could run it on the program. You would be able to tell whether program is correct or not. Because accentually, it's that thing. Hey if there's a fault in there, we get a test it to expose this. Therefore, I can expose all the faults Ok? And… that sounds great. I mean yeah I wish I could have that. If I could have that, that's exactly what I want.

But the problem is it's theoretically impossible to do this. And that takes a proof that I won't go through here, but it's the proof that basically shows that if there were an algorithm that given any program ok? With any faults cause that's what we are interested in.

Something worth all the time any program, any faults. If there were an algorithm, that could always give you this let's say ideal test suite. It would solve the hunting problem the hurting problem. The hurting problem cannot be solved therefore, there is no algorithm. What that comes down to is it all sounds good.

But we are not going to be able to come up with the technique does it.

And so whatever it is we are trying to do is testing it is not to ensure correctness we can't get that ok? And.. that's kind of disappointment in some ways but it does mean, we have lots of research we can do and our techniques aren't going to ensure correctness but they are going to try and do the best that we can at detecting faults so.. the bottom line there is for any program domain the perfect test suite exist.

I mean it's the test suite that contains all these inputs that fail. It's just that there's no algorithm for computing that. So we want to do what the best we can but we are not going to get it exactly.

Now if we can't get perfection or correctness, what can we do? Out of this came one idea that's been advanced in testing called test adequacy criteria. I should say this testing can show correctness unless we try all inputs under all possible circumstances and the maybe on the occasional program, Like Hello world, we can do that but in general we can't.

So then as testers our problem almost become which inputs do we use and when do we stop. And adequacy criteria helps us with that. What we need is a strategy for choosing inputs we know it won't show correctness but we hope it provides very qualitative terms sufficient confidence and help us to decide whether we can stop. And so there are so many adequacy criteria.

Adequacy criteria is accentually a rule that tells us whether our testing is sufficient. I can't remember how many more slides are in here? Let me look ahead. Ah few more we might not get through them but. We pick up what we left of soon.. but I've got a couple more minutes.

Give you some examples over the years people are developed many different types of adequacy criteria actually I should give you one simple example that you probably all know. It's called Statement coverage. You've got a piece of code written and see statement coverage says find tests that ensure that each statement is executed at least once that's an adequacy criteria if you do that your testing is considered adequate in terms of that criteria. So that's adequacy criteria stated in terms of code. Now there are two broad classes of criteria we can consider there are other but too broad classes.

Black box testing, sometimes called specification based, sometimes functional, uses the requirements functional requirements things in your specification or designed documents to test the system and your adequacy is judged relative to kind of coverage of requirements and you don't look in the software.

So the software is black box.

You don't look inside of it. A white box criteria, sometimes called code based criteria or structural, look at the code and they focus on driving the test from the code and the structure of the code and evaluate them in terms of coverage.

So statement coverage is a white box criteria and there are others, many others in between. So one of the other is better?

Let's suppose we have specification what we have here? We have a specification for this system that says even or odd returns 0 if number is even and -1 if it's odd. And so.

Black box on this if we looking at the specification not the code we are going to let it choose even number or odd number ok? And what we get as a result? What is this return 0 if the number's even.. negative what oh. Of course it doesn't return -1. If the number is odd it's going to return one.

So specification based testing you are trying even number, trying odd number sees that the value doesn't match the specification.

So you detect the error. Statement coverage says execute each statement once.

You can do that by throwing just an even number on that and you won't detect the error. This is very very simple and simplistic example.

But it's showing a case where specification based helps you get something that the code base doesn't.

So, this is just saying black box detects the fault whit box with 100% coverage does not. Pretty print prints and integer input to it. If the spec just says pretty print prints and integer, I can't get away from just one integer or I can try several integers but this programmer implicated something a lot different than that statement the programmer basically if the number is less than or equal to a thousand they will print the integer. If it's greater, they will print the number in them thousand.

So it's silly example but basic programmer sometimes do their own thing and don't precisely implement the specification.

And if we test for the specification since we can't try every possible input we might not catch this.

But code coverage will ensure that we cover the statement and will see that we aren't doing what the spec says.

So the two criteria have different strength and weaknesses.

They detect different things.

So for instance features that are totally missing from the code I forgot to implement this feature covering all the code isn't going to show you that looking at the requirements. Basic test requirements should.

So missing features tend to be detected better or things that are specified product tend to be better by black box testing. Additional features or things that programmer added that weren't in the spec.

Programmer do this called “gold planting”.

They think that they can improve it add gold to it somehow.

Specification based testing won't get those and so the code based looks at that.

And there's intersection too.

So what we tend to recommend to practitioners is that they do both black box and white box.

And there's different ways to do that in the process.

We are almost at the end here. So we will look next class at some adequacy criteria on the black box side we are going to look at requirements coverage and we will look at several white box criteria just to show you some basics.

So we will look at each of those.

But few things to keep in mind is aphorism.

An aphorism? It just a saying it just a saying.. that's been for a while I mentioned this one earlier testing detect the absence of faults only.. no I mentioned this one earlier, testing can't detecting absence of the faults only there presents.

That goes back to that testing problem.

With testing I can't ensure that all faults are absence. I can find some faults fro you but I can't ensure the absence.

So all I can do is show you some faults, not guarantee that there aren't any.

Then there's you can't test quality in, you have to build it in.

That's where we get back to good design, good requirements are the foundation of everything.

Testing detects only certain classes of faults … look at that later.

And it's cheaper to find and fix faults early engineering process if you do all the coding don't do any testing at all trying integrated trying working.

Now do system testing as you find fault.

It's going to cost a lot more to the fix them.

Then if you do unit testing and catch some faults there.

So… Testing throughout the process is better that gets us to the end for today.