

Formulating RQ By Alessio Ferrari : A Review

Muh. Riansyah

Faculty of Computer Science - University of Indonesia

Kampus UI, Depok 16424

muh.riansyah@ui.ac.id

Abstract—Using scientific method.

Index terms— Empirical

1. Solving Problem with Asking

"Every research endeavour starts with a question about the world":

- 1) a problem to solve,
- 2) a curiosity about some observed fact (subconsciously related to something relevant that you may not be able to always articulate, e.g., why do developers prefer to work at night? why are you asking this question? Because its interesting, but why it is so?)
- 3) a curiosity about some unknown fact (which are the most frequent defects in opensource code?)

And this will be the last for today that is related to one of core element of research that is the formulating research questions. So basically anything that you do in research in software engineering but also in any type of research **starts with a question about the reality**, ok ? A question about the world. This is normal in software engineering a problem to solve. So the question about the world is how can i solve this problem ? But it can be also **curiosity** about some of surveyed fact. This curiosity is often related to something that you that you feel is relevant. But you cannot say why for you is relevant. And for example it is interesting to know **why do developers prefer to work at night?** isn't something interesting ? but why is it interesting ? why you are asking this question ? you really don't know ! But probably if you understand why,for example, by interviewing set of developers,then you may understand that these will help you in better plan your software process, in better define,for example, your schedule for your company because if you find that developers prefer to work at night so why should you force them to work during the day why should you have a strict timing or if you discover that this is not true and some people prefer to work at night some people prefer to work during the day why don't you profit from that for that from that and you start having some people working during the day and some people working during the night so you have 24 hours people working.

2. Asking for Curiosity

And it can be also some curiosity about some unknown fact so some unknown fact this is the curiosity about something that you may have observed so you have an intuition and you are curious to know but also you want to know something that you don't know sufficient you want to explore so for example what are the most frequent defects in open source code which are their most typical the the fact that you can find in open source code this can be a vague curiosity you don't know why it's interesting but you may be interested in that.

3. Research as Guidance

The research question is the inquiry that guides your research: e.g., Which are the most frequent defects in code developed by people with less than 6 months experience? Which are the most frequent defects in code developed by people with 6 months to 3 years experience?

The research question is the inquiry that guides your research so it's a question written in natural language. When I say natural language, I say means English, Italian any language that you normally use a spoken language or written form in this case and the research question can be something like

"Which are the most frequent defects in code developed by people with less than six months of experience?"

So by novices,this is a research question also which are the most frequent defects in the code developed by people from six months to three years experience. so this is another research question. A research question is something that it's written is clear and guides your research and you should always refer to it.

You normally structure your research and reporting according to one or more research questions: they help to clarify your GOAL to the reader but also TO YOU. If you have more than one research question, it is good to establish a general research question (or research objective):

- 1) e.g., (mainly considering HOW aspects) To which extent certain defects types are related to the degree of experience of the developer?
- 2) e.g., (a more general one, may include also WHY aspects) Which is the relationship between defect types and degree of experience of the developer?

To design your research you know my structure your research and reporting according to one or more research question. Writing them down can help you to clarify your goal and also not just to you but also to the reader because you do research but you have to communicate it. of course because anytime anytime you research for something if you don't communicate it is useless okay it's just useful probably for you for your knowledge but it's not useful for the other researchers or the other software engineers. Normally if you have more than one research questions it is good to establish their general research question or research objective so for example if I want to consider how aspects so how things are in practice so more curiosity about observable fact . I can ask myself to each answers and effect types are related to the degree of experience of the developer. This is a general research question and the other two question that you saw at the beginning on slide so which are the most frequent defects in code developed by people with less than six months experience and the other question about the facts they encode developed by people that are older. This can be summarized with a more general research question is quite intuitive like to adjust a certain defects type are related to the degree of experience of the developer . I want to understand the relation between degree of experience and defect . Because my idea is that novice people tend to introduce more of a certain kind of defect while older people tend to do mistake that are more or less at architectural level. But I can also introduce a more general question also related to this like that consider also **why aspects**. For example which is the relationship between the defects types and degree of experience of the developer so this is more vague okay is not just that I'm searching for relationship between defect type and degree of experience.

I'm also and trying to understand what type of our relationship is there in a broad sense so why,for example, people that have less experience do certain type of defects and in this case probably I need to go further and not just do an experiment for example in which I have young people odd people and let them program and see the defects that they make and in order to interview them and to ask them why they committed a certain defect and to check with them for some motivation that can make me better understand the relationship between the effect type and degree of experience of the developer okay? This is a something quite difficult to grasp, but I invite you to to reason about that. Because the difference between **how things works** and **why things works in a certain way** is very relevant and you need always to resort to a different type of methods. if you want to understand both how aspects so how things are related with the sort of correlation, let say, a correlation function and why aspects so find the human reason behind certain phenomena.

Many times a clear formulation of the general research question comes AFTER the formulation of the more specific research questions So many times a clear formulation of the general research question comes after the formulation of the most specific research question so it is easier to write down which are the most frequent deffects in code developed by

people that has less than six months experience because it's more practical you know you know that you have to select people will assign some months experience but then you have to think what is the thing that . I want to understand relationship between age and type of defects and I want to understand also why there is if there is a relationship this relationship holds in some cases to distinguish you can between like detailed research question like the ones that . I listed before and the more general research question. *TIP: sometimes you can formulate the general research question as a Research Objective, e.g.: Understanding to which extent certain defect types are related to the degree of experience of a developer* you can sometimes formulate the general research question as a research objective so instead of a question just a sentence like the objective is **understanding to each stand a certain defect types are related to the degree of experience of a developer**. So this is what is my research objective how do I pack it into partition, into different questions. yes I start with the first questions which are the most frequent deffects for people with six months experience the most recurring the first with people with between six and three six months and three years and then . I can add the other other question for people who have who have more who have more experience and finally answering the general question that will tell me which are which is actually the relationship between the the type of defects and the years of experience.

4. Structure and Graph of Type of RQ

I will give you a guide that is taken from Robert felt you can check this PDF innig in Internet that is is free that is very simplified guide to how to formulate research questions it is very important it is not important really to let's say to to remember this graph this structure but it's important to refer to possible example actually of research questions but of course didactic li-like for for teaching you for teaching you the different types of research question is very useful to have this classification okay so we have a first of all research question in software engineer you have to type as . I told you at the beginning for we need to understand the reality and then we build a solution so my research question can be knowledge for Houston so understanding the reality or solution focus so the objective is transforming the reality okay let's start with the knowledge focused noise focus can be exploratory so . I don't know much about the phenomenon and the studies, I want to create some tentative series, I want to try out some hypothesis and I want to give evidence for example that a certain phenomenon that . I want to study can be made measured for example the consumes question should be general is exploratory it's a new thing that . I don't know anything about for example **to which extent do developers get tired of coding I still don't know if they get tired of coding** okay? and . I still don't know why they make tired of coding so this to which extent is a simple way of saying hey explore this theme ! explore this topic! okay? then I have a base rate question base rate question. And when you already know something

about certain phenomenon, we want to understand out the phenomenon. And the study appears which are the normal patterns and this can be asked when you already know a little bit so you know that people are able the people tend to get tired of coding my more detailed question is about frequency when the developer gets out oh god okay so . I ask myself something about related to the base rate that can be frequency or can be process we'd see more more example later on other so this in the first case in exploratory case . I don't know anything now . I know that the certain phenomena happen and can be measured. So I asking a base rate question in a different stage the research. And then I can also ask a more detailed question like for example how certain phenomenon under study relates to other phenomena so why do developers get tired of coding.

Here I will tell you how to to write down.

5. Explanatory

. I asked myself so which are the cows causes as you see there is already causality there which are the causes of for people to get tired of coding ok these are different levels that for knowledge focus question that . I normally reached when you have different degrees of knowledge of a system theme then you have solution focus questions and basically they describe better ways to solve a problem or a certain situation so which strategies app to achieve a certain certain goal how can we refine a certain tool to achieve certain goal in a better way so they are related with creating the new solution or refining an existing solution these are still research question as . I told you before because . I need to find a way . I need to create an importance for a solution and . I have to test the hypothesis against the reality like in the example before . I have to check that my proposal of forbidding the people to work at night actually reduces the number of bug in my code and doesn't create new new problems ok?

6. Sub-Types of and RQs Examples

- 1) Exploratory/Existence
 - a) "Does X exist?"
 - b) "Is Y something that software engineers really do?"
- 2) Exploratory/Descriptive
 - a) "What is X like?",
 - b) "What are its properties/attributes?",
 - c) "How can we categorize/measure X?",
 - d) "What are the components of X?"
- 3) Exploratory/Comparative
 - a) "How does X differ from Y?"
- 4) Base-rate/Frequency
 - a) "How often does X occur?",
 - b) "What is an average amount of X?"
- 5) Base-rate/Process
 - a) "How does X normally work?",
 - b) "What is the process by which X happens?",

- c) "In what sequence does the events of X occur?"
- 6) Relationship/Existence
 - a) "Are X and Y related?",
 - b) "Do occurrences of X correlate with Y?"
 - c) "What correlates with X?"
- 7) Relationship/Causality
 - a) "What causes X?",
 - b) "Does X cause Y?",
 - c) "Does X prevent Y?",
- 8) Causality/Comparative
 - a) "Does X cause more Y than Z does?",
 - b) "Is X better at preventing Y than Z is?"
- 9) Causality/Context
 - a) "Does X cause more Y under one condition than others?"

So here you have the different subtypes . I leave you this from the work of Robert felt that is linked in this light and I related to exploratory questions can be of three types like system descriptive and comparative. So existence is this **phenomenon exists**. So as I told before that do people get tired of coding or not is this thing that software engineers really do like for example is documentation something that they really do? because in a survey I may have in an interview, I may have understood that these people are really really annoyed by the fact of writing down documentation and they say that is one of the most time consuming activity of the work, but do they really document the software ? and this is often not the case so asking this this type of existence question is important also for understanding which one is the is the reality okay for certain phenomena actually existed deserves to be studied descriptive descriptive cases are more related to properties and attribution thing that . I can measure of a certain phenomenon like what are the properties attributes of a certain phenomenon for example when it is free not frequent. But for example how long is the documentation that they produce for example and in any case the things that are related to not to the existent busted a discrete general high-level description of a certain phenomenon and comparative like how can a certain phenomenon differ from another phenomenon then you have the base rate as . I told there are can be related to frequency of or to the process so how often or how does this phenomenon normally normally occur then you have the relationship that are again very similar to the exploratory they are similar to inspiratory but instead of focusing on a single element they relate to multiple elements for example do occurrence is the of X correlate with a certain other phenomenon or what are the reason for a certain phenomenon do . I know that a certain phenomenon cause causes another cause or prevent another phenomenon and more related to actual causes. So causality related question can be comparative of context related until you have a couple of examples. So these are just a list of typical example of question can guide you in better formulating the research question. This is quite important for you because it would be part also of the exam to be able to formulate research questions, okay ? So please pay

attention to this material! try to formulate research questions and try to give your own example and understand what is their classification as I said the classification is useful to give you a way of making sense of the different types of question. But in the end it is useful to look at examples ok? this list of examples if you replace the X and the y with some phenomena of interest for software engineers, they can be really useful for you.

7. Creating Research Questions

7.1. Overarching Research Topic

”Select overarching research topic (e.g., software development speed)”

”Do you want to create more and better understanding (Knowledge-based), or are you seeking for a solution to a problem (Solution-based)?”

How to create research question? you create research questions by starting from not a well-defined research objective because as I said before, it usually come afterwards with the topic normally so it can be software development speed and then you can ask yourself **do you want to create more a better understanding ? or you want to solve something related to development speed ?**.

7.2. Knowledge Based RQ

(Knowledge-based) e.g., what affects development speed? how can we measure development speed?

- 1) How much is known about the topic?
- 2) Not much (Explorative): how can we measure development speed?
- 3) We know the phenomenon, but not how or when it occurs (Base-rate): what is the development speed of agile teams?
- 4) We know the phenomenon, but not its causes (Relationship): what affects development speed?

If you want to create better understanding, then you are dealing with a knowledge-based question. And you want to ask yourself, something like

- 1) what effects development speed?
- 2) how can we measure development speed ?

So how can we measure how fast are the people in developing? what if then I have to ask myself? okay? I’ve understood that I have **a knowledge-based question**, I have to understand how much is known about this specific topic. If not much is known, I have to ask an exploitative question. So how can we measure, for example, development speed ? Because this is also problem. For example how can we measure a little in relation to what in relay in relation to the level software that even good software for example which are the measures that . I can put into place okay then you can have more information. So you know as set the phenomenon of development speed but you don’t know when certain sentence when or how it occurs.

7.3. Solution Based RQ

”(Solution-based) e.g., how can I improve development speed? what is the easiest way to improve development speed?”.

So what is the development speed? actually of a child tips so specific types specific types of teams. And i want to measure in this specific case what is their development speed? in the case in which you know the phenomenon and you have understood it quite correctly How things work in practice, so you can measure you can note. Also how things are working so you’ve been able to measure for example the develop speed of agile team you may want to understand what is affecting development speed so why certain teams are faster than others so in this case you want to understand the relationship before between certain unknown causes and the development speed in cases that you’ve working towards a solution based question the question can be

- 1) How can I improve development speed?
- 2) and What is the easiest way to improve the variable speed ?

You see that there is a relationship now first as I said before knowledge so first you understand what effects development speed and then you ask yourself how can . I improve it how can . I move the factors so that development speed increase what is the easiest way to improve development speed so which factors can . I use to somehow tune it so for example pair programming . I can use another language or . I don’t know . I can make people work at night work at the time that they want . I can remove the presence of the program manager . I don’t know ! After I’ve studied the factor, I can put in place a solution that solved my problem.

7.4. RQ and RO need refinement, same as Requirement

”One research question is not sufficient and you need a combination of them, so try to find a main research question or objective, and identify sub-questions, e.g., by checking the types of questions in the previous table and adapting them to your problem.” Normally, as I said before **one research question is not sufficient** and you normally need multiple question. So try to find after you have written down some base question, try to find the main research question or overarching research objective and then keep on refinement the your type of question okay? And check the types of question the previous table (table type of research) and adapt them to your problem too. So you start with the question and you refine it, you refine it, okay? you refine it! you first go up to more higher level research objective. And then go back and refine again. So to understand whether you’ve covered your actual research objective and just in the end very often just at the end of your research you finalize the research question. This is something like they seem to come at the beginning. They should come at the beginning. At the beginning you write some research question. But it is true that every time you do an experiment, every time you do

a case study, anytime you study a real phenomenon, you understand that your question need to be changed ! So your Research Objective (RO) and your Research Question (RQ) will because stable just at the end. Just like the requirements like the requirements for a system they become stable just when the system is finished also the research question they get transformed and transformed still even in their form intermediate form, they are needed to guide your process. So this is this is all for for today and we will we will meet again on on Thursday.