

R: Dan heb ik nu de opname gestart en dan ga ik zo meteen ook mijn scherm met jou delen. Als het goed is kan je nu mijn scherm zien en ook een bestand.

P: Ja.

R: Oké, mijn onderzoek richt zich op het kwantificeren van software product kwaliteit.

Dus we kijken specifiek naar de kwaliteit van producten en niet naar code kwaliteit. Ik denk dat dat een belangrijk verschil is om te onderscheiden. Daarvoor heb ik eerst verschillende modellen uit de wetenschap met elkaar vergeleken en daar kwam één ISO-standaard uit die het meest geaccepteerd is in de wetenschap. Dat is ISO 25010 en die zegt software product kwaliteit kan je benaderen aan de hand van acht kwaliteitskarakteristieken. Wat ik vervolgens heb gedaan, ik heb vervolgens discussiegroepen gehouden, zowel binnen <organisatie naam> als buiten <organisatie naam> waarin ik aan de deelnemers heb gevraagd hoe zou je deze acht karakteristieken nou concreet kunnen meten.

Dus wat voor data punten kunnen we uit een github halen, kunnen we uit een project management tool zoals Jira halen om dit karakteristiek te benaderen. Dat is dus de eerste stap geweest. Daar is dus een lijst van kwantitatieve measures uitgekomen. Sommigen zijn vaker genoemd, vandaar dat de lijst ook niet heel lang is. Dus het zijn er tussen de 10 en 20 per karakteristiek, want sommigen zijn vaker genoemd. Dus er is meer evidence ook voor bepaalde. En eigenlijk de volgende stap waar we binnen het onderzoek geïnteresseerd naar zijn twee dingen. Allereerst willen we weten waar kan je deze informatie uithalen.

Dus ik noemde bijvoorbeeld eerder al, ga je kijken naar een repo, ga je kijken naar een project management tool. Moet je ergens anders de informatie vandaan halen. Dus wat we zo meteen gaan doen, we gaan langs die, ik geloof dat er voor deze sessie in totaal 55 zijn, als ze ze niet allemaal kunnen behandelen is dat ook niet erg.

Ik heb gewoon een hele rits aan interviews die ik ga houden, dus dat is echt niet erg als we niet aan toekomen en vervolgens gaan we dus per measure kijken wat is naar de source, waar halen we vandaan en daarnaast gaan we kijken naar de feasibility van het meten. En je kan je namelijk voorstellen dat sommige van deze measures vrij lastig zijn om te meten. Soms is de data er niet of misschien beschikt een team over het algemeen niet over technische expertise om de appropriate infrastructuur te bouwen om het überhaupt te kunnen meten. Oké, aan de hand daarvan heb ik dus twee dimensies bedacht.

De eerste is de availability of data en die gaan we benaderen aan de hand van drie categorieën: low, moderate en high. Je ziet daarboven de beschrijvingen, dus wanneer classificeren iets als low, dat is wanneer de data available is, maar minimale effort required is om hem available te maken. Bij moderate is het mogelijk available, maar is effort required in het preprocessen en cleanen en we hebben high wanneer de data eigenlijk unavailable is en het dus moeilijk is om aan de data te komen. Misschien is de data een beetje niet intuïtief, maar het gaat om de feasibility en vervolgens hebben we ook nog de dimensie technical expertise, dus low, er zijn bijna geen technische skills nodig om deze bepaalde measure te kunnen meten. Bij moderate kan je verwachten dat ze wel te meten zijn als je de appropriate infrastructuur ervoor bouwt en dat meeste developers dit wel zouden kunnen. En de measures die onder high zullen vallen, daar zijn echt hele specifieke technische skills voor nodig, waarbij je dus echt extensive knowledge van een bepaalde technologie nodig

hebt om dit te kunnen doen.

P: Oké, dus hier gaat het niet over feasibility?

R: Ook over feasibility.

P: Ja, maar als het high feasibility is en je hebt advanced technical skills nodig, dan werk je nu in twee omgekeerde dimensies.

Misschien kunnen ze het best even omdraaien.

Of je moet zeggen, het gaat daadwerkelijk over availability of data, dan is het high availability, moderate availability.

R: Ja, het is een beetje de feasibility om het te meten gegeven dat je data hebt, maar ik denk dat het het best is om ze even om te draaien.

Is dat makkelijker voor het benaderen?

P: Kijk, het maakt mij niet uit. Ik kan het wel bepalen, maar je gebruikt nu twee verschillende indices voor je twee verschillende dimensies en dat is wel verwarrend.

R: Dat snap ik inderdaad. Ik denk dat het het best is om ze even om te draaien. Ik denk dat deze skills het makkelijkst zijn om te benaderen. Dus de availability van data is laag wanneer die mostly unavailable is en je technische expertise is laag wanneer je weinig skills nodig hebt. Ik denk dat die dan het makkelijkst te lezen is, right?

P: Ja, dus even kijken hoor. Dus nu is het zo dat... Ik weet niet of ze om te draaien, want kijk, nu is het zo dat... Bij availability of data is high is goed en bij technical expertise is high is slecht.

R: Right, right, exact.

P: Dus je... Ik denk dat je volgorde wel goed is, maar je titelt niet. Ja, zeker, ja.

Dus ik zou wat je net deed ongedaan maken. Het zo houden, maar dan zou ik zeggen... Het gaat niet om de availability, maar het gaat naar de difficulty of obtaining data.

R: Misschien is dit hem dan, hè? Hoog is moeilijk. En dan technical expertise required. En dan is hoog is moeilijk. Dus laag is goed, hoog is slecht. Ik denk dat we hem zo duidelijk hebben inderdaad, om het te benaderen.

P: Oké.

R: Alright. Goed, het is ook belangrijk om te weten dat je kan natuurlijk eens of niet eens zijn met een gegeven measure. Dat is in dit geval niet waar we naar kijken. Dat is nog een andere stap die we hierover heen gaan slaan. Dus we gaan puur kijken naar waar kunnen we deze mogelijk uithalen. Hoe moeilijk is het om de data te verkrijgen en wat door technische expertise is required. Het is dus best wel een gestructureerde vorm.

Dus in feite zou je ook kunnen zeggen, dit kan ook een middel van een questionnaire.

Maar het belangrijke wijze neem ook vooral mee in je gedachtengang. Waarom zou het lastig zijn om data te optelen? Waarom is hier technische expertise voor nodig?

Zodat we net iets meer context informatie krijgen. En ook zodat ik iets extra kan uitleggen per measure. De eerste categorie waar we dus naar gaan kijken is performance efficiency. En de standaard die definieert dat als de performance of a product relative to the amount of resources used under certain conditions. Oké, en de eerste measure die men heeft bedacht is memory utilization. Dus gegeven een software product. Hoe zou je memory utilization kunnen meten? Is dan de vraag.

P: Ja, goed. Een klein beetje context. Want die gaan we nog vaker tegenaan lopen.

Dit heeft te maken ook weer met het observatie. Wat je gebruikt van het systeem die je gebruikt. Als jij echt een systeem gebruikt, als je voor mijn part een mainframe gebruikt, of een fysieke server. Ja, dan is dit makkelijk te meten. En weinig technische expertise voor nodig. Maar als je kijkt naar <organisation name>, wij gebruiken vrij hoge abstraction.

In de vorm van serverless of containers of wat dan ook. En hoe meer je, hoe abstracter het wordt, hoe minder relevant deze data wordt en hoe moeilijker te meten wordt.

Bijvoorbeeld bij Lambda. Ik heb over het algemeen geen idee. En bij containers en bij stepfunctions al helemaal niet. Dus de waarde van de data die we hebben, die we kunnen

meten. Is het de context van <organisatie? Is het de context van de teams waarmee ik heb gewerkt? Of is het de context van computersystemen in het algemeen?

R: Ja, we proberen denk ik te kijken op het hoogste abstractieniveau. Dus computersystemen in het algemeen.

P: Het laagste abstractieniveau?

R: Het laagste abstractieniveau, ja.

P: Oké, als dat het is. Oké, als dat de context is, dan zou ik zeggen...

Low difficulty and low technical expertise. Maar dus wel met die kanttekening.

Ja, exact ja.

R: Ik denk dat dat overal de kanttekening is die we hierbij gaan maken.

P: Waarom ik het zeg is, als je 20 mensen dit interview gaat afnemen, dan moeten ze wel allemaal onder diezelfde context werken. Dus ik krijg heel uitheenlopende...

R: Exact ja, ja.

P: CPU is hetzelfde, low en low. Ja, ik denk dat die allemaal hetzelfde zijn.

R: En wat zou dan de source zijn? Moet je hier logging voor opzetten? Hoe zou je dit kunnen meten? Voor memory utilization,

P: je hebt gewoon een monitoringsysteem met die dit doen.

Dus ik zou zeggen, het inrichten van standaard monitoring. Voor zowel memory als CPU.

Ja, ik zou wel zeggen, ja. Standard monitoring tools, zou ik zeggen. Je moet er wel iets voor inzetten, maar het is wel gewoon table stacks.

R: Exact ja, de meeste tech stacks zullen dit wel supporten.

P: Ja oke. Zelfde. Ja, deze vallen eigenlijk onder hetzelfde kopje. Wanneer gaan we kijken naar de latency per user? Wat bedoel je daar precies mee?

R: Dus per gebruiker op jouw platform, hoe lang duurt het voordat zij. Hoeveel milliseconden duurt het voordat zij antwoord krijgen op een request? Dat is hoe die beschreven stond in onze focus group.

P: Maar wat betekent per user?

R: Dus stel, je hebt je hele userbase. Die vuren allemaal requests af op jouw service.

En daar pak je de gemiddelde latency van.

P: Ja, is dat niet average latency?

R: Ja, inderdaad, dit is hoe ze beschreven zijn in onze focus groups. Maar het is dus ook waardevol om te weten als dat onvoldoende informatie nog heeft.

P: Nou, wat je nodig hebt is application monitoring. Dus het is niet standard monitoring, niet OS monitoring, maar echt op applicatieniveau. Kan je het zo doen dat we die definities daarnaast kunnen houden? Misschien de tweede scherm even of zo?

R: Ik heb op dit moment geen tweede scherm hier bij me.

P: Nee, maar als je gewoon even dit scherm iets kleiner maakt en even een tweede venster ernaast opent of zo.

R: Dan kan ik het best misschien zo hiernaast zo slepen. Dan moet ik hem even bijpakken.

Oké, laten we iets inzoomen, want dit is natuurlijk niet leesbaar. We waren al in het begin van het programma. Dan willen we deze definities hiernaast hebben. En als het goed is kan ik hier pinnen, right? Weet je hoe ik hem kan pinnen?

P: Nee, maar je kan toch twee vensters, eentje links, eentje rechts doen.

R: Dit does the job.

P: En oké, aan de link kan je misschien even dat menu invouwen. Dan hebben we iets meer schermoppervlakte. Op dat pijltje drukken.

R: Oké. Waar waren we? Dus Latency per user, Feasibility of Availability. Oké, dus de difficulty. Waarom is deze moderate ten opzichte van de andere die bijvoorbeeld onder low zijn geclassificeerd?

P: Omdat deze data niet standaard beschikbaar is. Dat is afhankelijk van de tool die je gebruikt of de applicatie die je gebruikt en of dat tool die je gebruikt of de applicatie die je gebruikt en of dat je daar iets voor in moet richten om dat te meten. Dus daar moet je iets van effort voor doen.

R: Oké, en die effort, valt dat onder basic technical skills of heb je hier meer technische skills voor nodig?

P: Dat ligt ook wel weer aan je definitie, maar ik zou zeggen moderate, je hebt iets van kennis nodig. Misschien is het low, maar het ligt heel erg aan je definitie.

Kijk, als je een totale leek dit laat doen dan is het moeilijk en als je iemand laat doen die het al twintig jaar doet dan is het heel makkelijk. Ik zou het maar even zo laten.

R: Oké, de volgende gaat hand in hand met de vorige, dat is de latency per increasing user. En wat de deelnemers hier eigenlijk proberen te zeggen is, stel je hebt situatie, een soort A-B test, stel je hebt situatie A, 10.000 users op je platform, stel dat je het number of users increased naar 20.000, wat doet dit met de gemiddelde latency? Dus hier heb je twee metingen nodig ten opzichte van één. De source zou waarschijnlijk hetzelfde blijven dan?

P: Ja, ik zal moderate en moderate zetten.

R: Oké, het format waarin we nu doen, is het duidelijk voor jou? Werkt dit?

P: Ja, al vind ik zowel de vragen als de mogelijke antwoorden niet super scherp.

R: Dat snap ik. Hoe zou ik die scherper kunnen maken? Natuurlijk ook met oog op vervolgende interviews?

P: Ik zou het meer over harde waarden dan fuzzy waarden doen.

Wat is low, moderate en high? Wat zijn basic technical skills? Dat is net wie je het vraagt en hoe die zich op die dag voelt. Dat is niet bepaald wetenschappelijk.

R: Wat voor soort numerieke waarden zouden we hier dan kunnen koppelen?

P: Je zou bijvoorbeeld aan jaren ervaring kunnen koppelen met bepaalde platformen of bepaalde systemen. En als je het hebt over data obtinen, doe je hoeveelheid werk in uren die je nodig hebt om zoiets in te richten?

R: Misschien moet ik dan eerst die scale beter uitwerken. We kunnen nu natuurlijk wel doorgaan met deze hele scale, alleen als je denkt dat er weinig waarde in zit.

P: Ja, ik vind het wel lastig zo.

R: Dat snap ik. Dan denk ik dat dit misschien niet de way forward hiervoor is. We kunnen nu nog wel 40 minuten hier dan mee doorgaan. Alleen als de uitkomst voor jou te ver weg staat van wat de realiteit is, dan heb je helemaal niks aan de informatie.

P: Ja, en waar we over begonnen met de context van wat voor systeem gaat het, zou ik ook duidelijker en eenduidiger vastleggen. En nu dat we begonnen met het veranderen van die definities, nu kijk ik naar die vragen, dan staat er overall weer visibility. De visibility is dus weer tegenovergestelde van wat we in het begin zeiden. Want nu zeggen we bij sommige dingen low visibility, wel hetzelfde bedoelen dat het high visibility heeft. Dus ik zou al die dingen even scherp bekijken en de interviewvorm strakker zetten voordat we verder gaan met het vragen beantwoorden.

R: Dat denk ik ook. Wat zou je zeggen over het abstractieniveau? Anders stop ik gewoon even de recording, want dit zijn natuurlijk niet de echt onderzoeks dingen die we willen doen.