



Opgave 8

Studie af videnskabelig artikel

Gruppe: 3D

Mikkel Enevoldsen

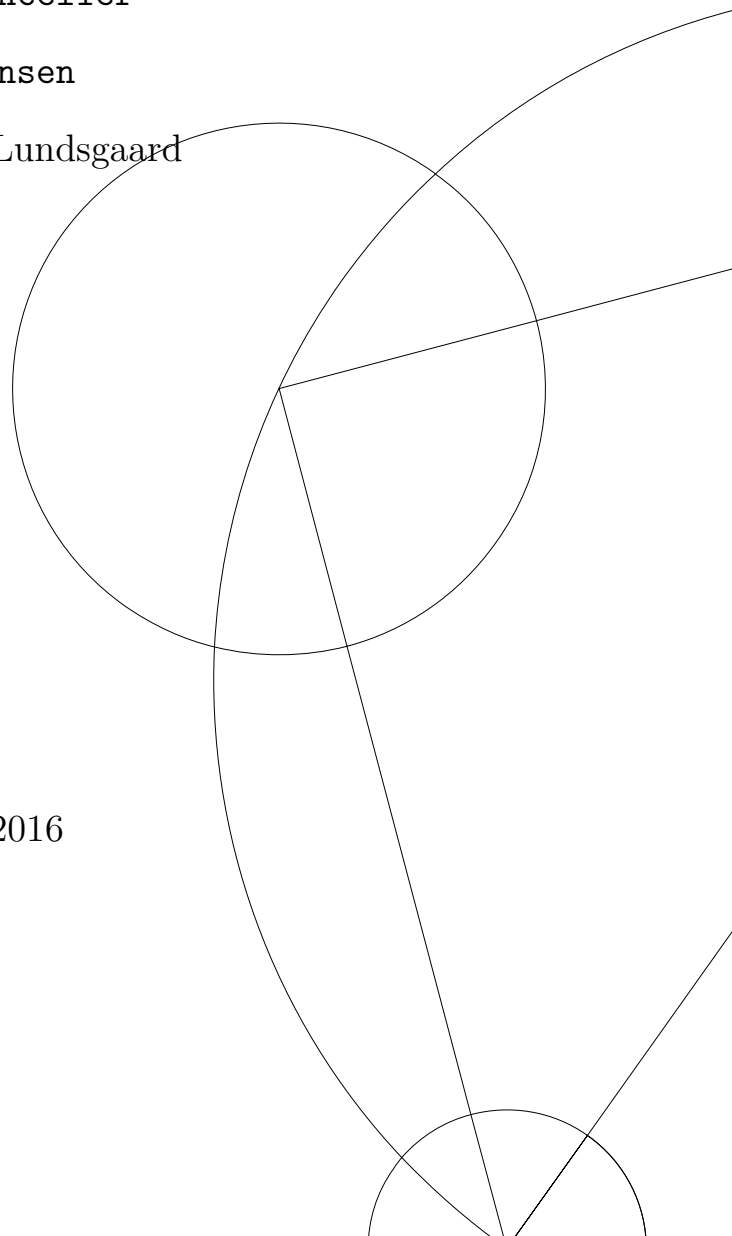
Kristian Høi

Dominique Chancelier

Carsten Jensen

Instruktor: Jesper Lundsgaard

March 18, 2016



Resumé

Da man gennem tidligere undersøgelser har vist, at en endeløs mængde af testdeltagere ikke vil give en endeløs mængde af unikke resultater, undersøger artiklen hvorvidt det samme gælder for evaluators. Til dette valgte man fire testdeltagere, der alle var vant med computeren men ikke med programmet der testedes. Yderligere havde man fire evaluators, som hver analyserede testdeltagernes tænkehøjttast. Evaluatorsne rapporterede problemerne og alles rapporter sammenlignedes, hvorefter duplikater blev fratrasket, indtil man kun havde unikke problemstillinger tilbage. Evaluatorsne lavede en top 10-liste, over de problemer de fandt mest kritiske. Det viste sig, at de ikke alle kom frem til samme resultater. Mens én evaluator kun fandt 46% af problemerne, så blev kun 20% af de samlede problemer fundet af alle fire evaluators. Hvad angår top 10-listerne, var der blandt andet ikke ét problem der optrådte på alles liste. Evaluatoreffekten er derfor en faktor der skal medregnes, når vi kigger på usabilitytest.

Et meget centralt begreb i artiklen er UPT - unique problem token. Alle UPT'er beskriver hver især et unikt problem fundet af evaluatorsne ved en usabilitytest. I artiklen skrives det eksempelvis at der i deres test i alt blev fundet 276 problemer ud fra deres ni opstillede kriterier. Nogle af disse viste sig dog at være duplikater, og da vi kun er interesserede i unikke problemer kortedes dette ned til 93. Fra vores egen usabilitytest kan vi eksempelvis fremhæve problemer vedrørende det røde flammeikon, der hos GroupRoom opstår når der er sket noget nyt, hvilket mange af vores testdeltagere ikke forstod. Derudover gør GroupRoom ikke brugeren tilstrækkeligt opmærksom på at vedkommende skal validere sin konto over den oplyste mail. En opgave det viste sig at alle vores testdeltagere enten havde problemer med.

Table 2 og figure 2 stemmer de nogenlunde overens. Rent visuelt kan vi også se, at figure 2 tilnærmelsesvis viser at 20% af UPT'erne var fundet af alle evaluators, som table 2 dikterer skulle ske. I samme stil passer resten også i deres kategori - 13% af alle UPT'er blev fundet af præcis 3 evaluators, mens 20% også passer fint til dem der kun blev fundet af 2 evaluators. Slutteligt skulle den sidste strimmel med kun enlige observationer denne ca. 46% af alle UPT'erne - hvilket også må siges at være tilfældet.

Vi fokuserer på formlen på artiklens anden side, og bruger denne til at udregne antallet af UPT'er. I en test med fire deltagere og fire evaluators vil der således være $19.35 \cdot 4^{0.505} \cdot 4^{0.661} = 97,427$ UPT'er. Da vi finder det svært at have en halv UPT, må der herfor skulle rundes op til 98.

Artiklen kan fint relateres til vores egne erfaringer i usabilitytesten. Vi er ad flere gange støt på situationer, hvor vi internt i gruppen ikke kunne blive enige om graden af Jo flere evaluators, des flere UPT'er finder man, men effekten synes, som de også nævner i artiklen, at være stavrende efterhånden som antallet af evaluators går op. Derfor er det i enhver usabilitytests interesse at finde en gylden middelvej for antallet af evaluators, da for mange synes at være spild af ressourcer - mens ved kun én evaluator, ifølge undersøgelsen berørt i artiklen, opdages kun 46% af det samlede antal UPT'er.

Da flere evaluatore er også har en tendens til at finde forskellige problemer, og de er oftest uenige i deres prioritering af eksempelvis top 10-liste over største problemer, som det ses i forsøget i artiklen. Derudover skete det i kun 20% af gangene, at alle fire evaluatore fandt det samme problem. Det vil sige, at 80% af alle fundne problemer, var der mindst en evaluator der anså det for ikke at være et. Dette er naturligvis problematisk, og vi ønsker denne evaluatoreffekt mindsket. Dette kan vi tilgå på flere måder, for eksempel kunne en klarere definition på graden af problemer ville kunne give en mere objektiv analytisk fase af usabilitytests. Et andet bud, kunne være hvis evaluatorerne ikke har set produktet de tjekker før. Ved dette undgår vi, at evaluatore har dannet sig et indtryk og dermed holdning af et produkt før de analyserer testdeltagerne. Ultimativt ville det give alle evaluatore mere enslydende forudsætninger for at analysere en test og man undgår således, at de underbevidst har dannet problemrapporter på forhånd. Det ville dog kræve en testleder der på forhånd havde gennemgået systemet, til at styre testdeltagerne igennem usabilitytesten.

Det der primært overraskede os mest, var det faktum, at fire mennesker med markant viden, ikke kunne finde frem til en prioriteringsliste der var nogenlunde enslydende. At måderne at tilgå og analyserer en usabilitytest kan resultere i så forskellige resultater og så stor spredning af UPT'er.