

Rasjonelle beslutninger ved valg av metoder og verktøy

Magne Jørgensen
magnej@simula.no

Eksperimenter (ikke del av foilsett)

- Hvor raskt bestemmer man seg
 - Speed estimation
 - Gruppe A -Task 1 (5 minutter) + Task 2 (1 minutt)
 - Gruppe B - Task 2 (5 minutter) + Task 1 (1 minutt)
 - Gruppe C Task 1 (1 minutt for å bestemme seg, 4 minutter revisjon) + Task 2 (1 minutt)
 - Hypotese 1: Selv når man gjør estimering bottom-up, har man et bilde av hva dette bør ende opp med – dvs bestemmer seg kjapt.
 - Hypotese 2: Mer tid fører ikke til vesentlige endringer, kun til større sikkerhet.
- Vekt på nærmeste analogi (lignende til Saab-Volvo eksperiment)
 - Lag situasjon der det normative vil være å estimere “aggregert”, men mange velger en “closest analogy”-strategi.

Materiale (del ut?)

- Tore Dybå, Barbara Kitchenham and Magne Jørgensen, Evidence-based Software Engineering for Practitioners, IEEE Software, Vol. 22, No. 1, Jan-Feb 2005.
- Xxx
- Ide: Oppsummering av eksperiment fra forrige gang (spesifikk vs generell informasjon)

[**simula** . research laboratory]

Et eksempel fra virkeligheten

- En utviklingsavdeling ønsket å bytte ut dagens utviklingsverktøy med målsetningen å få "mer effektive og moderne utviklingsmetoder".
- Etter flere besøk hos og demonstrasjoner av mulige leverandører, samt konferering med "referanse kunder" og egen analyse ble et verktøy valgt. Dette var en investering i minst 10-20 millionersklassen + store opplæring/endringskostnader.
- Utviklingsavdelingen gjorde i etterkant målinger på effektiviteten med det nye verktøyet (noe som i seg selv er svært uvanlig).
- Effektiviteten viste seg å være lavere enn med de tidligere verktøyene. Noe kunne bortforklares mhp type system som ble utviklet, innkjøringsproblemer m.m., men trolig langt fra alt.

Selv med mye ressurser, ble det gjort et dårlig valg.

[**simula** . research laboratory]

Hva gikk galt?

- Valgprosessen hadde i hovedsak evaluert etter prinsippet “jo mer funksjonalitet, jo bedre”. De viktigste egenskapene var ikke-funksjonelle, f eks produktivitet, og vanskelig å vurdere.
- Demonstrasjonene hadde (som naturlig er) fokusert på hva verktøyene var gode til og ikke svakhetene. Problemer ble ikke oppdaget før det var for sent.
- Referansekundene var i) valgt av leverandøren, og ii) hadde selv investert mye penger i dette verktøyet. Langt fra balansert informasjon.
- Problemstillingen var uklar og “all” relevant valginformasjon ble derfor hentet inn og evaluert. Oppsummering var derfor uklarl og åpnet for stor grad av “magefølelse”.
- Ingen av de som deltok i valgprosessen hadde trening i systematisk innhenting og vurdering av informasjon. Trolig en stor grad av naivitet mhp evne til å motstå villedende påvirkninger – som middager med leverandører.

[**simula** . research laboratory]

En bedre prosess: Kunnskapsbasert systemutvikling

- Tore Dybå, Barbara Kitchenham and Magne Jørgensen, Evidence-based Software Engineering for Practitioners, IEEE Software, Vol. 22, No. 1, Jan-Feb 2005.
- Kunnskapsbasert systemutvikling (derunder kunnskapsbasert valg av metode og verktøy) vektlegger:
 - Å formulere spørsmålsstillinger som er avgjørbare
 - “Er metode X bedre enn metode Y?” er et dårlig spørsmål
 - Å innhente kunnskap (fra forskning og praksis) som er relevant for egen kontekst, representativ og så nøytral som mulig
 - Demonstrasjoner og evalueringer av referansekunder er det stikk motsatte.
 - Å være trent i å gjøre bedre vurderinger av holdbarhet til kunnskapen
 - Svært få har noen som helst skolering i argumentasjonsanalyse eller evaluering av empiriske studier
 - Å sammenstille informasjon av ulike typer til en konklusjon relatert til problemstilling
 - Kunsten er ofte å evaluere effekten av manglende og/eller usikker informasjon på konklusjonen, og å ikke påvirkes av uviktig informasjon.

[**simula** . research laboratory]

Illustrasjon på Kunnskapsbasert Systemutvikling: Windows eller Linux?

- Kontekst: En organisasjon skal utvikle et større IT-system og må bestemme seg for om de skal basere dette på Windows eller Linux-plattform.
 - **NB:** Dette er et område jeg selv har liten kunnskap, selv om jeg har brukt begge operativsystem. Området er i hovedsak valgt fordi å illustrerer stegene i kunnskapsbasert systemutvikling.

[simula . research laboratory]

Steg 1 – Formulering av problemstilling

- Formuler en avgjørbar problemstilling: Hva er kriteriene for valg? Vil man i praksis la seg influere av analysen? Etc.
- Eksempel på mulig delproblemstilling: **Vil “Total Cost of Ownership” (TCO) være lavere ved bruk av Linux eller Windows som plattform for dette IT-systemet.**
 - Her bør en definisjon av hva som skal inngå i TCO beskrives. Dersom det er kostnader som er vesentlige, trolig ulike for Windows og Linux, og lite målbare, bør man vurdere om problemstillingen er tilstrekkelig avgjørbar til å fortsette analysen.
- Åpenbart vil andre forhold være viktige ved dette valget, som f eks riskovurderinger, kultur, eksisterende kompetanse, synergieffekter med andre systemer og sikkerhetsproblematikk. Hvert delproblem man ønsker å undersøke bør være basert på avgjorbare problemstillinger! Dårlige spørsmål gir dårlige svar!

[simula . research laboratory]

Steg 1 – Formulering av problemstilling

- De fleste analysene om Windows vs Linux jeg har funnet hopper over steget med å formulere en avgjørbar problemstilling. Dette medfører at man lett tolker resultatene som mye sterkere enn det de er.
 - Dersom man f eks ikke er i stand til å beskrive hva som skal til for at man skal mene at A er bedre enn B på et område, så vil en mengde målinger lett medføre at man fokuserer på det som er målt og glemmer det som ikke måles.
 - Eksempel: Anta at vi har kalkulert at Linux har 40% mindre investeringskostnader enn Windows og at vi vet lite om andre forskjeller i kostnader. Vi vil da lett legge for mye vekt på det målte, og for lite på det vi ikke vet noe om.
 - Stor usikkerhet bør tilsi at man predikerer “ingen forskjell”!
- Mange analyser burde vært stoppet pga manglende/uavgjørbar problemstilling. Hvis vi ikke vet hvordan vi skal avgjøre om noe er bedre, gir det oftest lite mening å starte analysen.
- Dersom avgjørelsen er så å si gitt på forhånd, bør man vurdere om det er verdt å starte en analyse. Når man vet hva utfallet bør bli, vil analysen bli dårlig.

[simula . research laboratory]

Steg 2 – Innhenting av informasjon

Eksempler på søketjenester:

- IEEE Xplore (<http://ieeexplore.ieee.org>) provides access to IEEE publications published since 1988.
- The IEEE Computer Society Digital Library (www.computer.org/publications/dlib) provides access to 22 IEEE Computer Society magazines and journals and more than 1,200 conference proceedings.
- The ACM Digital Library (www.acm.org/dl) provides access to ACM publications and related citations.
- The ISI Web of Science (www.isinet.com/products/citation/wos) consists of databases containing information from approximately 8,700 journals in different research areas.
- EBSCOhost Electronic Journals Service (<http://ejournals.ebsco.com>) provides access to over 8,000 e-journals.
- CiteSeer (<http://citeseer.nj.nec.com>), sponsored by the US National Science Foundation and Microsoft Research, indexes PostScript and PDF files of scientific research articles on the Web. Access is free.
- Google Scholar (<http://scholar.google.com>) indexes scholarly literature from all research areas, including abstracts, books, peer-reviewed papers, preprints, technical reports, and theses.

NB: Vektlegg artikler fra journaler av høy kvalitet og “reviews” dersom det er mye å velge mellom.

[!

Steg 2 – Innhenting av informasjon

- Dersom det finnes lite dokumentert erfaring/kunnskap:
 - Identifiser personer, organisasjoner og miljøer (mest mulig uavhengig av leverandør) som kan ha relevant erfaring. Sørg for at denne identifiseringen er systematisk og leder til et representativt utvalg av informasjonskilder.
 - Etterspør informasjon fra disse. (Lettere enn man ofte tror!)
 - Evaluer informasjon om relevans, hva erfaring bygger på (evidens), sannsynlig grad av objektivitet, samt sammenheng mellom erfaring og konklusjon.
- IKKE baser evalueringen på kunnskap fra:
 - Tilfeldige søk på web og lesing av de 4-5 første treffene.
 - Referanseklunder valgt av leverandører.
 - Studier der informasjonskilder (f eks forsker) eller oppdragsgiver har sterke interesser i utfallet.

[simula . research laboratory]

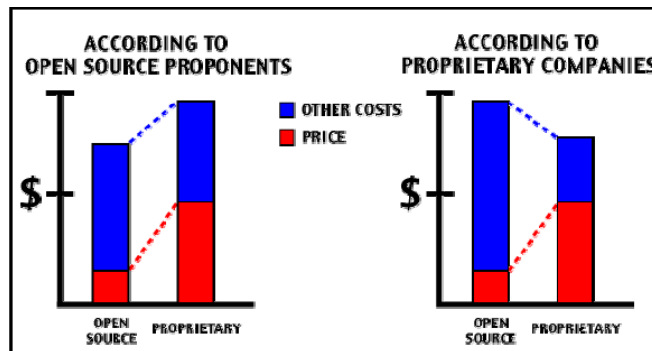
Steg 2 – Innhenting av informasjon

- Mitt søk på "Google scholar":
 - *Windows AND Linux AND "total cost of ownership" [AND review].*
- Flere hundre treff
- Siling bør inneholde vurdering av både relevans for egen organisasjon, kvalitet og sannsynlig minstemål av objektivitet
- Min strategi i dette tilfellet:
 - Alle analyser utført av organisasjon med sterk egeninteresse i utfallet utelukkes.
 - Kun analyser der det var sannsynlig at den som gjennomførte analysen hadde kompetanse i empiriske studier (universitetet, analysebyrå, ...) ble inkludert.

[simula . research laboratory]

Steg 2 – Innhenting av informasjon

- Utfall (min vurdering):
 - Lite trolig at analysene inneholder akkurat en spesifikk organisasjonen er ute etter, men det fantes flere studier som var bedre enn “ikke noe” (ukvalifisert synsing).
 - Stort sprik i resultatene (noe som i seg selv er et resultat), f eks a la oppsummeringen fra http://www.netc.org/openoptions/pros_cons/tco.html.



[simula . research laboratory]

Steg 2 – Innhenting av informasjon

- Dersom det er lite som er relevant og/eller har høy nok kvalitet, så bør man vurdere:
 - Egeninnhenting av informasjon fra brukere som ligner en selv starte (**NB:** Uten å falle i referansekunde-fella!)
 - Gjennomføring av egne empiriske studier. (Er man stor nok kunde, så stiller mange leverandører opp på selvdefinerte benchmarking-tester!)
- Eksempel på hvorfor referansekunder ikke kan brukes:
 - “Microsoft Norge ønsker å knytte til seg flere referansekunder. Fortell oss hvordan dine forretningsmuligheter har blitt **styrket** ved hjelp av løsninger og produkter fra Microsoft, og vi forteller det videre. Som referansekunde får du ikke bare muligheten til å bli **profilert** som et selskap som tar ny og kostnadseffektiv teknologi i bruk - hvis du er raskt ute med å registrere din løsning kan du også bli med i trekningen av 10 **gavekort**.”
 - <http://www.microsoft.com/norge/news/archive.aspx?year=2002>
 - Ikke representative. Egeninteresse i å fremstille noe som vellykket. Tette bånd til med leverandør. Fordeler ved å være referansekunde (Gaver virker utrolig effektivt!)

[simula . research laboratory]

Steg 3 – Vurdering av informasjonen

Sjekkliste for vurdering av rapport/artikkel:

- Hva er hovedpåstandene eller konklusjonen?
- Hvem er det som fremsetter argumentasjonen?
 - Uavhengig vurdering sannsynlig?
 - Egeninteresse?
 - Hva slags erfaringsbakgrunn og perspektiv kan ventes?
- Hvilke argumenter støtter opp under hovedpåstandene eller konklusjonen?
- Hvor god er sammenhengen mellom argumenter og påstand/konklusjon.
- Hvor gode er argumentene?
 - Basert på synsing?
 - Basert på egen erfaring? Og hvor relevant og uavhengig er den?
 - Eksempelbasert? Selektivt valgte eksempler?
 - Relevant og kvalitetsmessig gode undersøkelser gjennomført?

[**simula** . research laboratory]

Steg 3 – Vurdering av informasjonen

- Brukes språklige og/eller andre virkemidler til usaklig påvirkning?
 - Tekst som ikke deltar i hovedargumentasjonen (det kan være mesteparten) vil ofte ha som funksjon å påvirke følelser og skape sympati med konklusjonen.
 - Ord er "ladet" og kan brukes til å skape sympati eller antipati, uten at det er en del av den rasjonelle argumentasjonen.
 - Sjekk alltid hvor stor del av "argumentasjonen" som er relevant for konklusjonen. Vurder hvilket formål resten av argumentasjonen har.
- Er argumentasjonen ensidig eller gir et feilaktig bilde?
- Finnes en tilforlatelig dokumentasjon av svakheter/mangler ved argumentasjonen.
 - Hvis ikke. Er det fordi at argumentasjonen er perfekt, eller fordi det er noe å sjule.

[**simula** . research laboratory]

Steg 3 – Vurdering av informasjonen

- Er det ting som IKKE er med, som burde vært med.
 - Det er en typisk menneskelig svakhet å kun vurdere det som står, og ikke det som IKKE står beskrevet.
- Hvor relevante er resultatene for DITT FORMÅL
- Vær ekstra nøye med vurdering av argumentasjon dersom du er intuitiv enig i (eller har fordel av) konklusjonen. Vår forsvarsevne mot "teoriladet observasjon" er lav, og må trenes. En fin trening er å forsøke å finne feil med argumentasjoner du sympatiserer med.
- Vær skeptisk, men ikke forkast argumentasjon som verdiløs, fordi om den har mangler.

[**simula** . research laboratory]

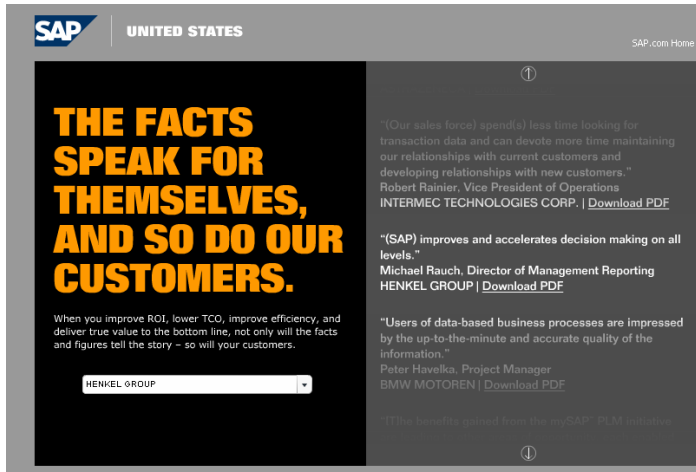
Steg 3 – Vurdering av informasjonen

- Noen ganger er det relativt enkelt å finne svakheter i informasjonen:
 - <http://www.microsoft.com/casestudies/casestudy.aspx?casestudyid=200945>
 - *"Benchmark tests showed that SQL Server 2005 running on Windows was the most viable solution. One of the key factors influencing the technical team's decision to choose Microsoft was the dependability of Microsoft software. The team wanted a solution that performed consistently and provided timely, reliable service."*

[**simula** . research laboratory]

Steg 3 – Vurdering av informasjonen

Andre ganger er det ingen argumentasjon, kun påstander. Da bør man vurdere i hvilken grad man kan anta at en påstand er bygget på “evidens”.



[**simula** . research laboratory]

Steg 3 – Vurdering av informasjonen

- Andre ganger kan det være svært vanskelig å avsløre:
 - Anta at du hadde lest IDC-rapporten som viser at Windows har lavere Total Cost of Ownership (<http://www.microsoft.com/windows2000/docs/TCO.pdf>).
 - Resultatene virker overbevisende, IDC er en seriøs organisasjon og vant med slike studier.
 - Det mangler imidlertid detaljert informasjon om hvordan scenarier er valgt og kalkulasjonene gjennomført. Hvordan skal dette influere vurderingen?
 - At sammenligningen ikke nødvendigvis er “fair” viser følgende sitat fra BusinessWeek:
 - *“IDC analyst Dan Kusnetzky says the company selected scenarios that would inevitably be more costly using Linux. Also, he believes Windows should be cheaper to operate, since it has been around longer, giving Microsoft more time to develop software to manage the operating system. “Microsoft has had a lot more time to work on this. I wonder why the win wasn’t bigger,” Kusnetzky says. Microsoft insists that it didn’t rig the contest and chose the most popular uses for the software.”*

• www.businessweek.com/magazine/content/03_09/b3822610_tc102.htm

[**simula** . research laboratory]

Steg 3 – Vurdering av informasjon
BANAN-testen:
Har noen av disse bananene noe gult i seg?



A



B

[**simula** . research laboratory]

Steg 3 – Vurdering av informasjonen

- Dersom man ikke er i stand til å vurdere analysen – særlig når problemstillingen er dårlig definert og analysen komplisert – bør man ikke stole på rapporter eller der oppdragsgiver er analysefirma har egeninteresse i utfallet.
- Dette gjelder dessverre selv om analysen er gjennomført av respekterte firmaer eller universitetsansatte forskere.
- Våre sanser (og enda mer våre vurderinger av resultater) påvirkes av hva vi forventer å se. Har du selv sterke egeninteresser i utfallet, bør du ekskludere deg selv fra videre analyse!

[**simula** . research laboratory]

Steg 3 – Vurdering av informasjon

- Vi påvirkes trolig for mye av spesifikk og for lite av aggregert informasjon!
- Resultater fra en av oppgavene på forrige JavaZone illustrerer dette:
 - Vektlegging av informasjonskilde om nytten av et kurs om testing for deltakerne.
 - 15% ville vektlegge en positiv kursevaluering av en venn som hadde deltatt mer enn en uavhengig studie fra et respektert universitet.
 - 80% ville vektlegge egen erfaring mer enn den uavhengie studien.
- Eksperiment på personers valg av Volvo eller Saab:
 - *“Imagine a hypothetical choice between buying a Volvo or a Saab, assuming Volvos have a better long-term repair record. Now imagine you hear from a friend who drives a Volvo that their car just died on the highway-now which brand would you be likely to pick?”*
 - De fleste velger å vektlegge informasjonen fra en venn mer enn aggregert informasjon om gjennomsnittlig pålitelighet, selv om det er mange grunner til å vektlegge aggregert informasjon mer.

[simula . research laboratory]

Steg 4 - Sammenstilling

- Ta kun med det viktigste i sammenstillingen. Mindre viktige data har en tendens til å vektas for mye og gjøre vurderingen dårligere.
- Unngå rasjonaliseringer av hva man føler er riktig (dersom det ikke er det egentlige målet med evalueringen)
 - Dersom magesfølelse og analyse tilsier to ulike ting, følg analysen (gitt at ens egen fornøydhet med valget ikke er av svært stor betydning).
- Min egen vurdering av de rapportene jeg leste er:
 - Det er få STORE systematiske forskjeller i TCO mellom Linux og Windows, men muligens en tendens til at Windows har hatt lavere TCO (noe som lett kan ha endret seg til i dag).
 - Manglen på rapporter som ikke er bestillingsverk eller på annen måte partiske er slående. Noen få har likevel en tilfredstillende metodikk.
 - Usikkerheten/variasjonen i resultatene er såpass høy at små forskjeller ikke bør tillegges stor vekt, dvs TCO bør dermed trolig tillegges liten vekt ved valg mellom Linux og Windows (gitt at rapportene beskriver situasjoner av høy relevans for organisasjonen selv).
 - **NB:** Jeg har ikke brukt mer enn ca. 4-5 timer på informasjonsinnhenting og evaluering, så ikke legg alt for mye vekt på denne oppsummeringen.

[simula . research laboratory]

Metoder som brukes til å villede

Hentet fra: changingminds.org/index.htm :

- Alignment: When everything lines up, there are no contradictions to cause disagreement.
- Amplification: Make the important bits bigger and other bits smaller.
- Appeal: If asked nicely, we will follow the rules we have made for ourselves.
- Arousal: When I am aroused I am full engaged and hence more likely to pay attention.
- Association: Our thoughts are connected. Think one thing and the next is automatic.
- Assumption: Acting as if something is true often makes it true.
- Authority: Use your authority and others will obey.
- Bonding: I will usually do what my friends ask of me, without negotiation.
- Closure: Close the door of thinking and the deal is done.
- Completion: We need to complete that which is started.
- Confidence: If I am confident, then you can be confident.
- Confusion: A drowning person will clutch at a straw. So will a confused one.
- Consistency: We like to maintain consistency between what we think, say and do.
- Contrast: We notice and decide by difference between two things, not absolute measures.
- Daring: If you dare me to do something, I daren't not do it.
- Deception: Convincing by trickery.

[**simula** . research laboratory]

Metoder som brukes til å villede

- Dependence: If you are dependent on me, I can use this as a lever to persuade you.
- Distraction: If I distract your attention, I can then slip around your guard.
- Evidence: I cannot deny what I see with my own eyes.
- Exchange: if I do something for you, then you are obliged to do something for me.
- Experience: I cannot deny what I experience for myself.
- Fragmentation: Break up the problem into agreeable parts.
- Framing: Meaning depends on context. So control the context.
- Harmony: Go with the flow to build trust and create subtle shifts.
- Hurt and Rescue: Make them uncomfortable then throw them a rope.
- Investment: If I have invested in something, I do not want to waste that investment.
- Logic: What makes sense must be true.
- Objectivity: Standing back decreases emotion and increases logic.
- Passion: Enthusiasm is catching.
- Perception: Perception is reality. So manage it.
- Pull: Create attraction that pulls people in.
- Push: I give you no option but to obey.

[**simula** . research laboratory]

Metoder som brukes til å villede

- Repetition: If something happens often enough, I will eventually be persuaded.
- Scarcity: I want now what I may not be able to get in the future.
- Similarity: We trust people who are like us or who are similar to people we like.
- Specificity: People fill in the gaps in vague statements.
- Substitution: Put them into the story.
- Surprise: When what happens is not what I expect, I must rethink my understanding.
- Tension: I will act to reduce the tension gaps I feel.
- Threat: If my deep needs are threatened, I will act to protect them.
- Trust: If I trust you, I will accept your truth and expose my vulnerabilities.
- Understanding: If I understand you, then I can interact more accurately with you.

[**simula** . research laboratory]