

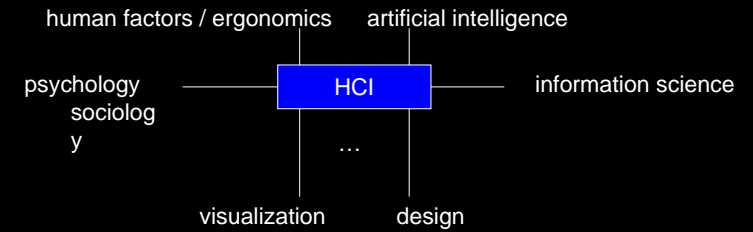
## Kapittel 14

# Evaluering av informasjonssystemer

INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## HCI and related fields



INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Design process

Four basic activities in the design process

1. Identify needs and establish requirements
2. Design potential solutions ((re)-design)
3. Choose between alternatives (evaluate)
4. Build the artefact

These are permeated with three principles

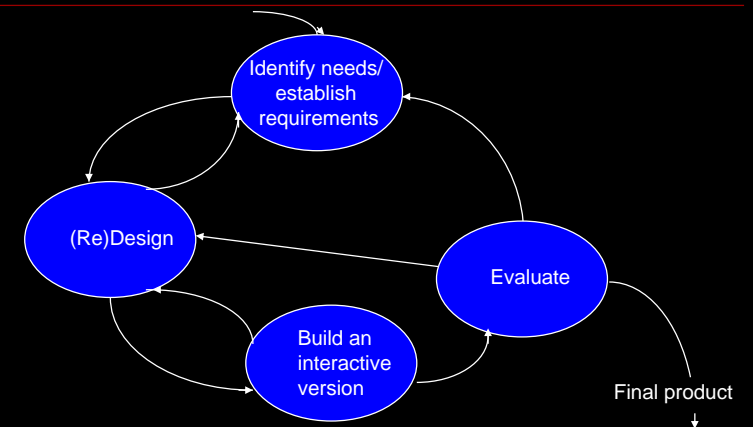
1. Involve users early in the design and evaluation of the artefact
2. Define quantifiable & measurable usability criteria
3. Iteration is inevitable

Lifecycle models show how these are related: we present just a few

INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

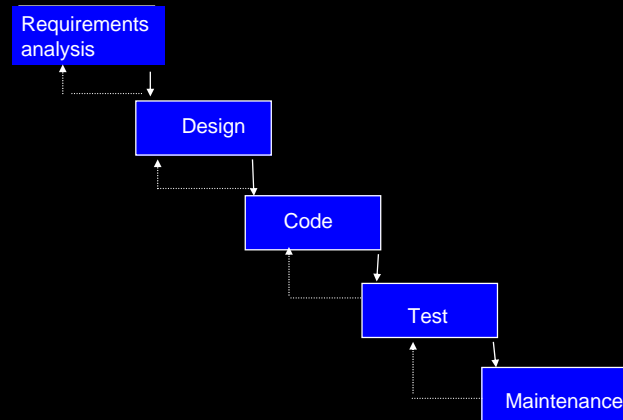
## A simple interaction design model - exemplifies a user-centered design approach



INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

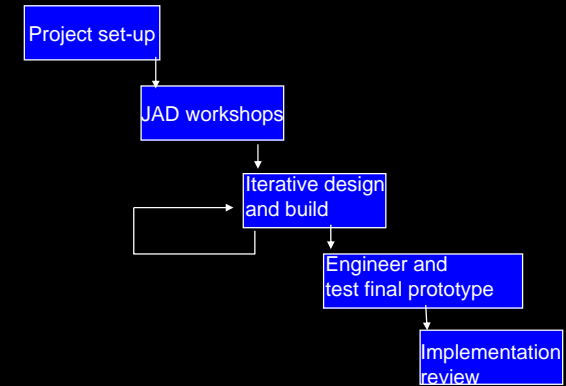
## Traditional 'waterfall' lifecycle



INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## A Lifecycle for RAD (Rapid Applications Development)

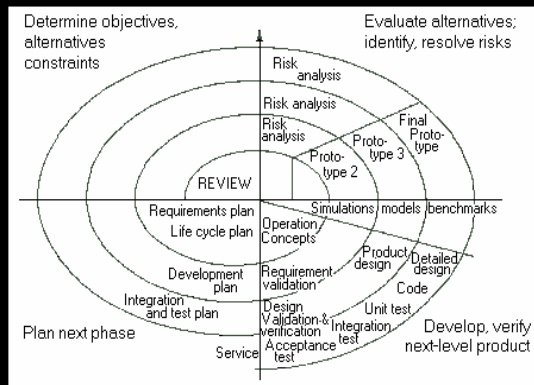


INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Spiral Lifecycle model

From [ctr.umkc.edu/~kennethjuwng/spiral.htm](http://ctr.umkc.edu/~kennethjuwng/spiral.htm)



INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## The Star lifecycle model

- Suggested by Hartson and Hix (1989)

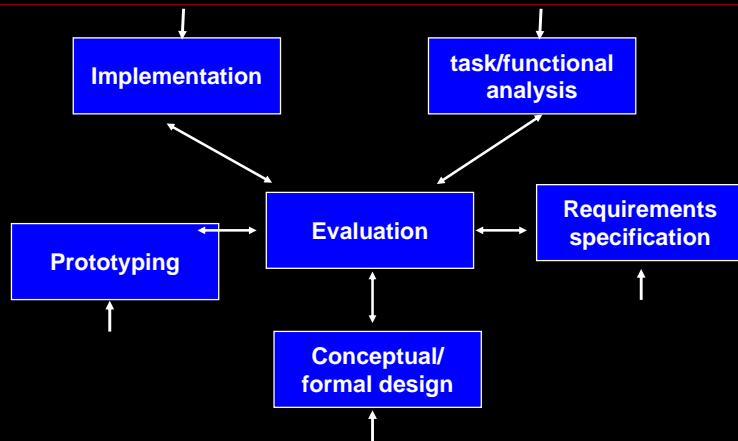
- Important features:

- Evaluation at the center of activities
- No particular ordering of activities. Development may start in any one
- Derived from empirical studies of interface designers

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## The Star Model



INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Hvordan fungerer dette for informasjonssystemer

1. Hvordan kan vi vite om systemet er riktig?
2. Hvordan kan vi vite om systemet er "godt"?
3. Hva vil det si at systemet er "godt"?
4. Hvordan kan vi vite noe om verden?

Hvorfor og når skal systemet evalueres?

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

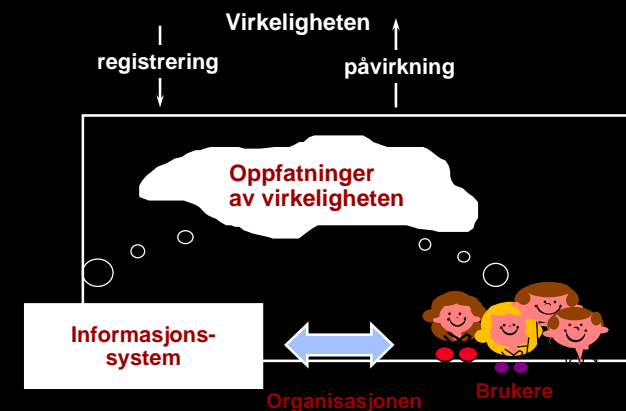
## Hvorfor (og når)?

1. Ved overlevering
  - Er kontrakten oppfylt?
  - Kan prosjektet erklæres avsluttet?
2. Underveis
  - Fossefallsmodell
    - Kan en fase erklæres fullført?
  - Iterativ/inkrementell modell
    - Er prototypen/versjonen/leveransen OK?
    - Hva kan vi lære av hvordan denne versjonen virker i praksis med tanke utvikling av neste versjon?
3. Ved hver milepæl
  - Hva kan vi lære mer generelt (med tanke på senere virksomhet av "alle" slag)?

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Informasjonssystemet gjenspeiler «virkeligheten»



INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Hypotetisk-deduktiv metode

- Framgangsmåte for å vite noe om verden
- Hver del-aktivitet må følge vitenskapelige kriterier

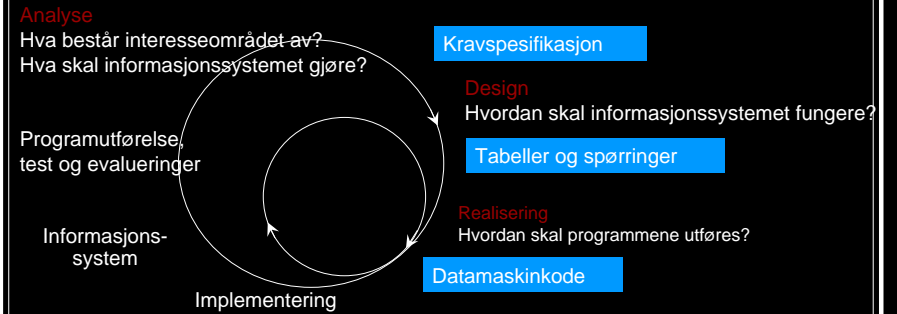


INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Evalueringer i forhold til kravspesifikasjonen

- for å vise om vi har laget **systemet riktig**
- Å evaluere i hht. kravspesifikasjon: å verifisere systemet



INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Syntaktisk test

- Hypotese: Programmet er syntaktisk korrekt
- Undersøkelse: Kompilering
- Konklusjon av vellykket resultat:  
→ Vet at vi har et syntaktisk korrekt program

INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Funksjonell, empirisk test

- Hypotese: Programmet behandler data korrekt i henhold til kravspesifikasjon
  - Use case
- Undersøkelse:
  - Velger data som er representative i forhold til de mulige data
  - Kjører programmet med utvalgte testdata og korrigerer det til det håndterer disse i henhold til kravspesifikasjonen (inf3120)
- Konklusjon av vellykket resultat:  
→ Vet at programmet håndterer disse data

INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Funksjonell, teoretisk test

---

- Hypotese: Programmet behandler data korrekt i forhold til kravspesifikasjonen
- Undersøkelse:
  - Tar utgangspunkt i utsagn om programmet, gjerne en invariant
  - Beviser matematisk at programmet opprettholder invarianten (inf3230)
- Konklusjon av vellykket resultat:
- Vet at programmet er korrekt i forhold til invarianten

## Empirisk inspeksjon

---

- Hypotese: Programmet oppfylder kravspesifikasjonen
- Undersøkelse:
  - Andre programutviklere leser programmet og stiller systematisk spørsmål ved alle setninger om hva setningen gjør og hvorfor den er programmert slik (inf3120)
- Konklusjon av positivt resultat
- Vet at programmet tilfredsstillers andres kritiske vurdering

## Key points

---

- There are many issues to consider before conducting an evaluation study.
- These include the goals of the study, the approaches and methods to use, practical issues, ethical issues, and how the data will be collected, analyzed and presented.
- The DECIDE framework provides a useful checklist for planning an evaluation study.

## Inspections

---

- Several kinds.
- Experts use their knowledge of users & technology to review software usability.
- Expert critiques (crits) can be formal or informal reports.
- Heuristic evaluation is a review guided by a set of heuristics.
- Walkthroughs involve stepping through a pre-planned scenario noting potential problems.

## Heuristic evaluation

- Developed Jacob Nielsen in the early 1990s.
- Based on heuristics distilled from an empirical analysis of 249 usability problems.
- These heuristics have been revised for current technology.
- Heuristics being developed for mobile devices, wearables, virtual worlds, etc.
- Design guidelines form a basis for developing heuristics.

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Nielsen's heuristics

- Visibility of system status.
- Match between system and real world.
- User control and freedom.
- Consistency and standards.
- Error prevention.
- Recognition rather than recall.
- Flexibility and efficiency of use.
- Aesthetic and minimalist design.
- Help users recognize, diagnose, recover from errors.
- Help and documentation.

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

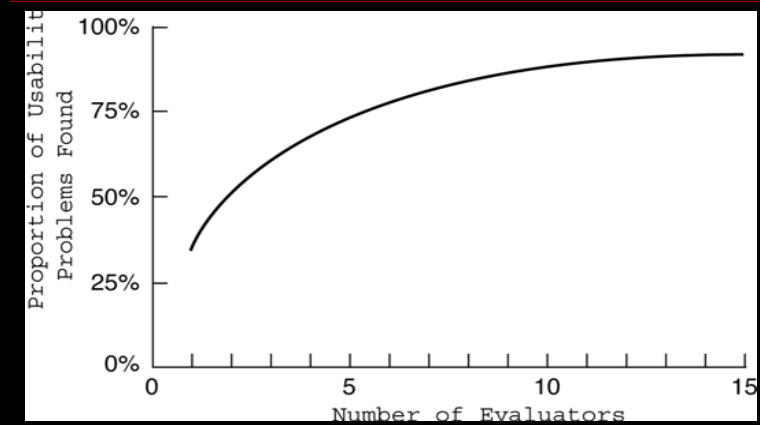
## Discount evaluation

- Heuristic evaluation is referred to as discount evaluation when 5 evaluators are used.
- Empirical evidence suggests that on average 5 evaluators identify 75-80% of usability problems.

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## No. of evaluators & problems



INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

### 3 stages for doing heuristic evaluation

---

- Briefing session to tell experts what to do.
- Evaluation period of 1-2 hours in which:
  - Each expert works separately;
  - Take one pass to get a feel for the product;
  - Take a second pass to focus on specific features.
- Debriefing session in which experts work together to prioritize problems.

### Advantages and problems

---

- Few ethical & practical issues to consider because users not involved.
- Can be difficult & expensive to find experts.
- Best experts have knowledge of application domain & users.
- Biggest problems:
  - Important problems may get missed;
  - Many trivial problems are often identified;
  - Experts have biases.

### Cognitive walkthroughs

---

- Focus on ease of learning.
- Designer presents an aspect of the design & usage scenarios.
- Expert is told the assumptions about user population, context of use, task details.
- One or more experts walk through the design prototype with the scenario.
- Experts are guided by 3 questions.

### The 3 questions

---

- Will the correct action be sufficiently evident to the user?
- Will the user notice that the correct action is available?
- Will the user associate and interpret the response from the action correctly?

As the experts work through the scenario they note problems.

## Pluralistic walkthrough

---

- Variation on the cognitive walkthrough theme.
- Performed by a carefully managed team.
- The panel of experts begins by working separately.
- Then there is managed discussion that leads to agreed decisions.
- The approach lends itself well to participatory design.

## HCI perspektive

---

Verifikasjon av systemet mot kravspesifikasjoner er ikke garanti at systemet skal ha de forventede ytre virkninger. Man trenger å få tak i andre metoder og undersøkelser som validerer systemet

HCI metoder og teknikker som for eksempel DECIDE, empiriske brukertester, evaluering av brukbarhet, tenke-høyt-test, intervju etc.

DECIDE rammeverk er introdusert med 6-7 foiler på engelsk

## DECIDE: a framework to guide evaluation

---

- Determine the *goals*.
- Explore the *questions*.
- Choose the evaluation *approach* and *methods*.
- Identify the *practical issues*.
- Decide how to deal with the *ethical issues*.
- Evaluate, analyze, interpret and present the *data*.

## Determine the goals

---

- What are the high-level goals of the evaluation?
- Who wants it and why?
- The goals influence the approach used for the study.
- Some examples of goals:
  - Identify the best metaphor on which to base the design.
  - Check to ensure that the final interface is consistent.
  - Investigate how technology affects working practices.
  - Improve the usability of an existing product .



## Explore the questions

---

- All evaluations need goals & questions to guide them.
- E.g., the goal of finding out why many customers prefer to purchase paper airline tickets rather than e-tickets can be broken down into sub-questions:
  - What are customers' attitudes to these new tickets?
  - Are they concerned about security?
  - Is the interface for obtaining them poor?
- What questions might you ask about the design of a cell phone?

## Choose the evaluation approach & methods

---

- The evaluation *approach* influences the *methods* used, and in turn, how data is collected, analyzed and presented.
- E.g. field studies typically:
  - Involve observation and interviews.
  - Do not involve controlled tests in a laboratory.
  - Produce qualitative data.

## Identify practical issues

---

For example, how to:

- Select users
- Stay on budget
- Stay on schedule
- Find evaluators
- Select equipment

## Decide about ethical issues

---

- Develop an informed consent form
- Participants have a right to:
  - Know the goals of the study;
  - Know what will happen to the findings;
  - Privacy of personal information;
  - Leave when they wish;
  - Be treated politely.

## Evaluate, interpret & present data

- The approach and methods used influence how data is evaluated, interpreted and presented.
- The following need to be considered:
  - Reliability: can the study be replicated?
  - Validity: is it measuring what you expected?
  - Biases: is the process creating biases?
  - Scope: can the findings be generalized?
  - Ecological validity: is the environment influencing the findings? - i.e. Hawthorn effect.

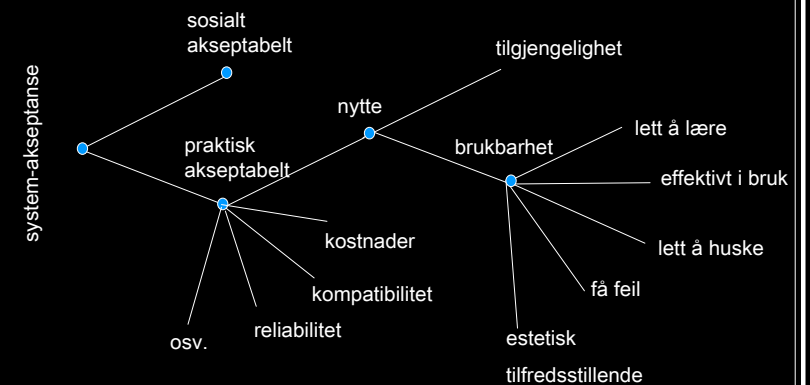
## Andre valideringsmetoder

- Empiriske brukertester
- Observasjoner
- intervju

## Heuristisk evaluering av brukbarhet

- 2 - 3 brukbarhetsspesialister
- Gjennomgår hver detalj i prototypen
- Vurderer i forhold til kjente retningslinjer for hensiktsmessig design
  - Retningslinjer for brukergrensesnittdesign
- For hver retningslinje som brytes noteres et mulig bruksproblem
- Enkel, billig første-evaluering

## Brukarbarhet er sammensatt



etter Jakob Nielsen: *Usability Engineering* (1993), s 25

## 12 krav til gode brukergrensesnitt (Gerhard Skagestein kap. 9.3)

1. Brukergrensesnittet må følge etablerte standarder
2. Brukergrensesnittene må være mest mulig innbyrdes konsistente
3. Brukeren skal hele tiden ha kontrollen
4. Systemet må gi (informative) tilbakemeldinger
5. Systemet må oppfattes som trygt
6. Brukergrensesnittet må vise hvilke handlinger som står åpne for brukeren
7. Brukergrensesnittet skal ikke være basert på at brukeren må huske noe
8. Brukergrensesnittet må ikke gi brukeren for mye å velge mellom på en gang
9. Unngå en total utveksling av hva brukeren ser på skjermen
10. Brukergrensesnittet må tilby snarveier
11. Brukergrensesnittet må gi et intuitivt bilde av systemet
12. Brukergrensesnittet må være "vakkert"

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## 8 gyllne regler for grensesnittdesign (Ben Shneiderman)

1. søk konsistens (terminologi, prompts, menyer, hjelp, farger, form, fonter)
2. lag snarveier for hyppige bruker (forkortelser, spesielle taster, gjemte kommandoer, makroer)
3. gi informativ tilbakemelding
4. design dialoger som lukkes (gruppør handlingssekvenser, sekvens: begynnelse, midt, avslutning)
5. forebygg feil og tilby enkel feilhåndtering
6. tillat enkel omgjøring av en operasjon
7. støtt brukerens grunnlag for kontroll
8. reduser belastning av korttids-hukommelsen

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Heuristisk evaluering (forts.) Gjennomføring og resultater

- Prototyper, skisser og papiretterlikninger (mock-ups)
- Evaluatørene må være noen andre enn dem som har medvirket i designet
- Evaluatørene bør ha erfaring i menneske-maskin interaksjon og i grensesnittdesign
- Én evaluator finner 1/3 av brukbarhetsproblemene
- Tre evaluatører finner 2/3

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Gjennomganger (walkthroughs)

- Utgangspunkt i vanlige oppgaver systemet skal utføre
- Brukbarhets-spesialister går detaljert gjennom designet
- Noterer mulige feil
- Vurderer måloppnåelse
- Tester målbare størrelser
  - Antall sider som må blas gjennom for å finne det man ønsker
  - Antall tastetrykk på som trengs for å finne det man ønsker

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Tenke-høyt test, evt. med intervju \*

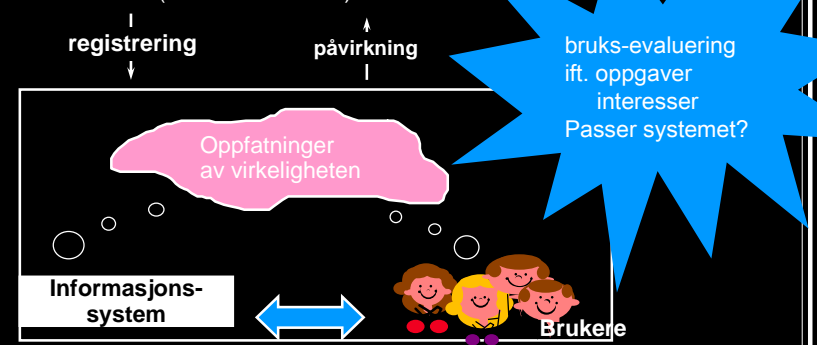
- Et lite antall test-personer, stoppe når intet nytt
- Riktig utvalg? (målgruppe)
- Utforme oppgaver de skal løse
- Be dem si høyt alt de tenker
  - Minn forsøkspersonene om å snakke når de blir tause
- Feilkilder
  - De sier det de tror vi vil høre
  - Det vi sier får oss til å tenke
- Video-opptak, tidtaking og notater
- Eventuelt intervju før og etter sesjonen
- Analysere brukernes forståelse, misforståelser og feil
- Mer tidkrevende enn heuristisk evaluering
- For systemer som skal benyttes mye og eksternt

INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Bruksevaluering

- for å få vite om vi har laget **det riktige systemet**
- Å evaluere i hht. behov og forventninger: validere systemet  
Virkeligheten  
(interesseområdet)



INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Empiriske brukertester

Hypotese: Vi har laget det riktige programmet

Undersøkelse

La framtidige brukere anvende programmet til sine oppgaver

1. Observasjon

Observér hvilke problemer de har og hvilke feil de gjør

Observerbare problemer

2. Måling

Mål hvor lang tid de trenger for å lære programmet eller løse en oppgave

Tell antall feil, antall tastetrykk, ...

Tall

3. Intervju

Spør om det var dette de trengte eller ville ha

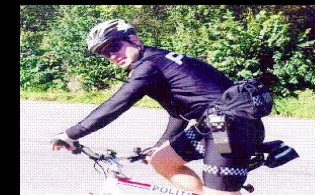
Brukernes subjektive oppfatninger

INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Observasjon

- Der datasystemet blir brukt
  - Arbeidsplass
    - Samspill med andre datasystemer
  - Hjemme
    - Gamle browsere
    - Langsomme linjer
  - Offentlig sted
    - Lys
    - Skriftstørrelse
  - Håndholdt hvorsomhelst
    - Hendene ledige?
- Notater
  - Video-opptak



INF1050, Evaluering av informasjonssystemer  
Alma Culén

11. April 2007

## Typiske problemer som kan oppdages

- Navigasjonsproblemer
  - Veksling mellom skjermbilder
  - Hvor er jeg?
- Feil inndata
- Musbevegelser eller snarveier
  - Tabulator mellom felter
- Fyller inn noe annet enn tiltenkte data i felter

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Intervju

- Spørre brukere om
  - hva de bruker systemet til
  - hvor mye de bruker det
  - deres personlige oppfatning eller opplevelse av systemet
- Ikke nødvendigvis noen sammenheng med hvilke resultater som oppnås ved hjelp av systemet
  - Brukere synes endringer er brysomme
  - Brukere vil ha eksisterende funksjonalitet pluss litt til

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

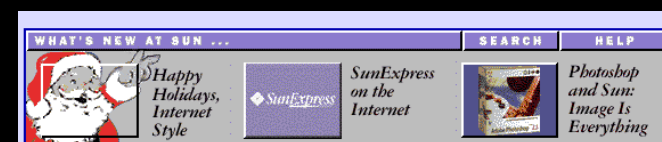
## Logging

- Hele eller deler av populasjonen
- Billig for web- og databasesystemer
- Vanskelig å tolke
  - Hva betyr det at en side er aksessert 68 592 ganger i forhold til formål som?
    - Profilere butikken
    - Holde på kundene
    - Selge mer
- Identifisere hver bruker
  - Når de oppgir identitet ved for eksempel betaling
- Logg med oppfølgende intervju
- Kan gi svar på hvorfor brukerne navigerte som de gjorde

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Uttforming inkonsistent med vanlig oppfatning av knapper



Before: The top row doesn't look much like buttons, does it?



After: No doubt that the top row contains buttons that you can click on

416 % økning i bruken av knappene over 2 måneder, 48 % økning i bruk av web-tjenesten

INF1050, Evaluering av infolasjonssystemer

11. April 2007

## Målsetting – formål og midler

- En målemetode
- En verdi som skal oppnås med metoden (target value)
- Basert på personers oppfatning
  - Mål som skal oppnås: 95% av publikum skal synes at det nye systemet er bedre enn det gamle.
  - Målemetode: Å svare på et spørreskjema der hvert av systemene rangeres på en skala fra god til dårlig.
- Basert på observerbar bruk.  
Mål som skal oppnås:
  - 2/3 av kundene skal komme igjen. Telling av hvor mange kunder som kommer mer enn 1 gang.
  - Wap-systemet skal gi trafikkinformasjon i løpet av 30 sekunder for øvede brukere. Tidtaking i reell bruk.
  - Nytt system skal gi en statistisk signifikant forkortelse av tiden for å finne trafikkinformasjon. Tidtaking i lab.forsøk.
  - Gjennomsnittlig antall brukerfeil  $\leq 2$ . Feiltelling under reell bruk.

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Oppgaver i samsvar med systemets formål eller midler

- Enkel tilgang til informasjon om produktene
  - med formål å selge mer
  - "Besøk web-butikken, og bestill og betal følgende varer: ..."
  - Enkelte av produktene gitt i oppgaven tilbys ikke av butikken
  - Måten å oppgi kvanta i oppgaven (for eksempel 10 stk.) stemmer ikke alltid overens med butikkens (for eksempel 6-pakning eller kg).
- Å profilere butikken
  - Intervju etter høyttenkningssesjonen
  - "Hvilket bilde av butikken synes du web-tjenesten gir?"
  - "Hva skiller denne butikken fra andre?"
  - "Hvordan vil du karakterisere dens vareutvalg?"
  - "Hvordan var servicen?"
  - "Hvilken stil har den?"

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Laboratorietest av måloppnåelse

- Evaluering av et middel som vi tror leder til et mål
- Sett et mål (target) for middelet
  - 80% av testpersonene kan gjengi et hovedaspekt ved butikkens image
  - Gjennomsnittlig tid for å finne ønsket trafikkinformasjon er 30 sekunder
- Lag et eksperimentoppsett
  - Gi et antall forsøkspersoner oppgaver
  - Telling / måling av hva de gjør

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Sammenliknende laboratorietest

- Lag to design og utfør samme test på begge
- Del inn forsøkspersonene i grupper
- Halvparten av forsøkspersonene tester det ene designet først
- Den andre halvparten tester det andre først

## Valg av evalueringsmetoder

- Tidlige evalueringer
  - Sparer store kostnader knyttet til senere endringer
  - Tvilsom validitet
    - Kan evaluere noen midler som vi ikke vet om er relevante for tjenestens formål
    - Under betingelser som ikke forekommer i virkeligheten
- Feltevalueringer
  - Større mulighet for å evaluere om formålene oppfylles
  - Mer data å håndtere
  - Ukontrollerbare sammenlikninger
  - Kostbare endringer

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Avsluttende kommentarer

- Evaluering gjøres altfor lite
- For sterk tro på rasjonelle metoder i stedet for å lære av konkrete prosjekter
- Evaluering er ofte politisk
  - Skal leverandør få betalt nå eller gjøre mer for egen regning?
  - Hvem har skylda? (..bør ikke få forfremmelse, lønnsøkning, nytt oppdrag, ...)
  - "Etterpåklokskap!"
  - "Har ikke tide"
  - "Det gjelder å se framover! Vi må konsentrere oss om dagens utfordringer. Verden er helt annerledes nå"

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007

## Informatikkens vitenskapelige metoder

- Alle de nevnte evalueringsmetodene inngår
- Informatikk bruker dermed både matematiske, naturvitenskapelige, samfunnsvitenskapelige og humanistiske metoder
- Informatikk skiller seg fra disse fagene ved at vi også konstruerer det vi skaffer oss viten om, vi
  - analyserer
  - designer
  - koder
- Informatikk er dermed mer et teknologifag enn matematikk, naturvitenskap, samfunnsfag eller humaniora

59

## Noen linker om brukbarhetstesting & tips

<http://sigchi.org/>  
<http://sigchi.org/chi2004/>  
<http://www.uie.com/>  
<http://usableweb.com/>  
<http://usability.gov/guidelines/>  
<http://www.usabilityfirst.com/>  
<http://www.dialogdesign.dk/>  
<http://www.nngroup.com/>  
<http://www.useit.com/jakob/index.html> el. <http://www.useit.com/>  
<http://www.jnd.org/>  
<http://www.asktog.com/tog.html>  
<http://www.upassoc.org/>  
<http://jthom.best.vwh.net/usability/> og mange flere ....

INF1050, Evaluering av infolasjonssystemer  
Alma Culén

11. April 2007