

Brukergenerert innhold i kontrollerte vokabularer

av Øyvind Kristiansen



Det Samfunnsvitenskapelige Fakultet,
Institutt for Informasjons- og Medievitenskap,
Universitetet i Bergen

Fullført som del av kravene til graden
«Master i informasjonsvitenskap»

Mai 2009

Oyvind.Kristiansen@student.uib.no

Innholdsfortegnelse

Innholdsfortegnelse	i
Appendikser	ii
Figurer og tabeller	iii
Takksigelser	iv
1 Innledning	1
1.1 Mål for studiet	2
1.2 Oversikt over studiet	2
2 Omgivelser for studiet	4
2.1 Digitale biblioteker	4
2.2 ABM-utvikling	5
2.3 Universitetsbiblioteket i Bergen / Billedsamlingen	6
2.4 Omgivelser for studiet: konklusjon	7
3 Tidligere studier og teoretisk rammeverk	8
3.1 Grunnleggende konsepter	8
3.2 Brukergenerert innhold	10
3.3 Bildegjenfinning	13
3.4 Kunnskapsorganiseringssystemer	16
3.5 Kontrollerte vokabularer	17
3.6 Bildetaggning	23
3.7 Vokabular-problemet	24
3.8 Tidligere studier og teoretisk rammeverk: konklusjon	25
4 Forskningsdesign	26
4.1 Definisjon av problemstilling	26
4.2 Metodevalg	27
4.3 Eksperimenter	29
4.4 Planlegging av eksperimentet	34
4.5 Design og utvikling av prototypen	41
4.6 Gjennomføring	56
4.7 Statistisk analyse	61
4.8 Forskningsdesign: konklusjon	61
5 Resultater	62
5.1 Beskrivende statistikk	62
5.2 Hypotesetesting	70
5.3 Resultater: konklusjon	71
6 Diskusjon og konklusjoner	72
6.1 Verdien av brukergenerert innhold i kontrollerte vokabularer	72
6.2 Gjennomførbarheten av brukergenerert innhold i kontrollerte vokabularer	73
6.3 Mulige fortrinn med brukergenerert innhold i kontrollerte vokabularer foran tagging	73
6.4 Forslag til UB	75
6.5 Konklusjoner	77
7 Videre forskning	78
Referanser	79

Appendikser

Appendiks A: Testbilder	81
Appendiks B: Samtykkeskjema	91
Appendiks C: Oppgavene til eksperimentgruppen.....	92
Appendiks D: Oppgavene til kontrollgruppen	94
Appendiks E: Utdrag fra kildekode.....	95
Appendiks F: Forhåndstilknyttede termer (av UB)	105
Appendiks G: Unike tagger (generert av kontrollgruppen).....	106

Figurer og tabeller

Figur 1: Precision og recall for en eksempelspørring.....	16
Figur 2: Økende strukturell kompleksitet i kontrollerte vokabularer (omarbeidet fra ANSI/NISO (2005: 17))...	20
Figur 3: Synonymring for naust	21
Figur 4: Bare-posttest-design	31
Figur 5: EER-diagram over den delen av bildedatabasen som er aktuell i dette studiet.	43
Figur 6: Søkemenyen i prototypen	50
Figur 7: Presentasjon av resultatene fra et bildesøk	51
Figur 8: Presentasjon av metadata for et bilde (eksperimentversjonen).....	52
Figur 9: Presentasjon av metadata for et bilde (kontrollversjonen).....	52
Figur 10: Legge inn synonym for termen naust (eksperimentversjonen).....	53
Figur 11: Visning av ett av de ti testbildene, med linker til de andre bildene	58
Figur 12: Total precision og recall uten brukergenerert innhold sammenlignet med data fra alle deltakerne i hver gruppe aktivert.	63
Figur 13: Precision og recall i eksperimentgruppen sammenlignet med precision og recall i kontrollgruppen når tagger som ikke er representert som termer som eksperimentgruppen kan bruke, er deaktivert.....	64
Figur 14: Frekvensfordeling for precision (i klasser).....	66
Figur 15: Frekvensfordeling for recall (i klasser).	67
Figur 16: Boksplokk for precision for de to gruppene.....	69
Figur 17: Boksplokk for recall for de to gruppene.....	69
Tabell 1: Gjennomsnittlig precision og recall med bidragene til hver deltaker aktivert, samt totale gjennomsnitt og standardavvik for precision og recall.	68
Tabell 2: Resultater fra <i>t</i> -testen for precision.....	70
Tabell 3: Resultater fra <i>t</i> -testen for recall.....	71

Takksigelser

Det å skrive denne oppgaven har vært en krevende, men utrolig lærerik prosess som jeg ikke kunne klart uten følgende personer. Først og fremst vil jeg rette en stor takk til min veileder, Dag Elgesem, og Solveig Greve ved Billedsamlingen ved Universitetsbiblioteket i Bergen som begge har vært svært positive og tillitsfulle under arbeidet med oppgaven. En stor takk også til de 20 personene som deltok i forsøkene, samt Bjørge Næss og Jan-Erik Bråthen for hjelp til pilottesting av forsøkene. De to sistnevnte, samt Bjarte K. Helland skal også ha takk for mer eller mindre faglige samtaler gjennom arbeidet med oppgaven. Sist, men ikke minst, vil jeg takke jentene mine for støtte og vennlige spark bak når motgangen har vært størst. Dere betyr alt for meg!

1 Innledning

Søking etter informasjon på verdensveven blir stadig mer aktuelt og nødvendig etter hvert som mengden og mangfoldet av tilgjengelig digitalt innhold øker. Det finnes mange måter å forbedre dataene som er knyttet til innhold, slik at det skal bli lettere for brukerne å finne det igjen. I mange tilfeller er det beste eller eneste alternativet å legge inn disse dataene manuelt, og dette kan være et møysommelig arbeid. Dette er en av grunnene til at det har blitt stadig mer vanlig å involvere brukerne i denne prosessen. De fleste metoder hvor brukerne involveres, innebærer tradisjonelt at brukerne på en eller annen måte beskriver innholdet direkte, med for eksempel stikkord om hva et bilde inneholder eller en tekstlig beskrivelse av innholdet i et lydklipp eller handlingen i et videoklipp. Dette kalles tagging, og det har blitt forsket en god del på slik brukervedvirkning, og det er i utstrakt bruk, for eksempel i Googles Image Labeler¹, Flickr², og YouTube³. Brukervedvirkning er det ene av to hovedemner i dette studiet.

Dataene som blir generert i denne prosessen kalles metadata, og er i seg selv med på å gjøre det lettere for andre brukere å finne det de søker etter. Men i mange tilfeller kan et enkelt konsept beskrives med to eller flere forskjellige ord. Forskjellige mennesker bruker dessuten forskjellige abstraksjonsnivåer for å beskrive konsepter. Å knytte hver variant til hvert digitale objekt er en svært omfattende oppgave. Her kommer det andre av studiets to hovedemner inn: kontrollerte vokabularer. Kontroll av vokabularet brukes for å forbedre effektiviteten av systemer som bruker beskrivende språk for å identifisere og lokalisere innhold. Ved hjelp av kontrollerte vokabularer kan man bedre konsistensen i beskrivelsen av dokumenter, og legge bedre til rette for gjenfinning. Kontrollerte vokabularer er, i likhet med brukervedvirkning, mye omtalt i forskningen, og det har blitt laget flere nasjonale og internasjonale standarder for opprettelse og vedlikehold av kontrollerte vokabularer. Det er imidlertid bare de siste par årene man har begynt å forske på hvordan man kan kombinere de

¹ <http://images.google.com/imagelabeler/>

² <http://www.flickr.com>

³ <http://www.youtube.com/>

to emnene brukermedvirkning og kontrollerte vokabularer, og om det er mulig å dra nytte av brukerne i selve opprettelsen av kontrollerte vokabularer, i stedet for bare i opprettelsen av tekstlige annotasjoner til dokumenter.

1.1 Mål for studiet

Teorien i studiet bygger i stor grad på feltet informasjonsgjenfinning⁴, som er et viktig forskningsområde innenfor informasjonsvitenskap. Målet for studiet er å undersøke gjennomførbarheten og verdien av å innføre brukergenerert innhold i et kontrollert vokabular for en digital bildesamling. Den spesifikke digitale bildesamlingen som omtales i dette studiet, og som brukes for å besvare problemstillingen, tilhører Universitetsbiblioteket i Bergen, og kalles Billedsamlingen. Digitale bildesamlinger inngår i feltet digitale biblioteker, som er en viktig satsning i ABM-sektoren i Norge. En viktig trend i denne sektoren, er at brukeren er stadig mer i fokus.

1.2 Oversikt over studiet

I kapittel 2 beskrives digitale biblioteker og emnets status i den norske arkiv-, bibliotek- og museumssektoren (ABM-sektoren). I tillegg omtales digitale bildesamlinger og ståstedet til den norske ABM-sektoren generelt og Universitetsbiblioteket i Bergen spesielt i forhold til spredning av digitalt innhold og medvirkning fra brukerne. I kapitlet argumenterer jeg for at fokuset i stor grad er på spredning, og at brukermedvirkning med fordel kunne vært høyere prioritert i planene for fremtiden.

Kapittel 3 tar for seg tidligere studier innenfor emnene som er aktuelle i dette studiet, og går gjennom teorien som ligger til grunn for studiet. Brukergenerert innhold og bruken av dette i bildegjenfinning gjennom bildetagging er blant emnene som blir presentert, i tillegg til kontrollerte vokabularer, som er en annen metode som blir brukt for å legge til rette for bildegjenfinning.

⁴ Eng.: Information Retrieval

Basert på kapittel 2 og 3, utvikles det i kapittel 4 et passende forskningsdesign for å undersøke om et kontrollert vokabular med mulighet for brukerbidrag kan være med på å forbedre søkeresultatene i en bildesamling sammenlignet med brukermedvirkning i form av bildetagging. En problemstilling blir fremsatt, og metode for å besvare denne blir valgt. Valg av og utforming av forskningsredskaper blir beskrevet.

I kapittel 5 blir resultatene fra datainnsamlingen presentert, og i kapittel 6 diskuteres funnene, før studiet avrundes med konklusjoner og forslag til videre forskning.

2 Omgivelser for studiet

Dette kapitlet fokuserer på hvilke omgivelser dette studiet er utført i og passer inn i. Det tas utgangspunkt i Norge og norske forhold, men mange av forholdene er likevel like i store deler av verden.

2.1 Digitale biblioteker

Det finnes neppe noen entydig definisjon på begrepet *digitalt bibliotek*. Følgende beskrivelse er hentet fra rapporten *Forskning på digitale biblioteker i Norge*:

«Digitale bibliotek yter tjenester til sine brukere. De inneholder informasjonsressurser i alle digitale former og formater, som for eksempel databaser med tekst, tall, grafikk, lyd og video. For å yte tjenester og å håndtere informasjonsobjekter, må dette biblioteket omfatte verktøy og personer som muliggjør produksjon, lokalisering, innhenting og bruk. Datasystemene er komplekse, gjerne distribuerte, informasjonssystemer. Digitale bibliotek ivaretar hele livssyklusen til data, informasjon og kunnskap: intellektuell produksjon, spredning, bruk og langtidsbevaring.» (Brygfjeld, Husby, Sølvberg, & Vognlid, 2001: 6)

Det er, ifølge Borgman (2000: 38) et generelt skille mellom forskere og bibliotekansatte i synet på hva et digitalt bibliotek er. Forskere fokuserer hovedsakelig på databaser og informasjonsgjenfinning – innsamling, organisering og tilgang til informasjonsressurser. Bibliotekansatte fokuserer hovedsakelig på digitale biblioteker som institusjoner eller tjenester. De har et bredere syn på bibliotek-konseptet, og ser på biblioteker som organisasjoner som velger ut, samler inn, organiserer, konserverer og sørger for tilgang til informasjon på vegne av et samfunn av brukere. I dette studiet er det naturlig nok organiseringen og tilgangen til informasjonsressurser fra et forskningsperspektiv som er hovedfokus, og da særlig med vekt på tilgang. Følgende definisjon er dekkende for perspektivet i dette studiet:

Definisjon 1: *Et digitalt bibliotek er en samling av informasjonsressurser i digital form og format som omfatter verktøy og personer som muliggjør produksjon, lokalisering, innhenting, organisering og bruk* (bearbeidet fra Brygfjeld, et al., 2001: 6).

For å håndtere organiseringen av informasjonsressursene i et digitalt bibliotek, brukes kunnskapsorganiseringssystemer⁵ (Hodge, 2000) Dette emnet omtales i seksjon 3.4.

I Norge arbeides det med å skape et nasjonalt digitalt bibliotek gjennom et samarbeid mellom norske biblioteker, og en del av dette arbeidet består i å vurdere hvilke datastrukturer og kunnskapsorganiseringssystemer som er hensiktsmessig å bruke i en slik sammenheng, og hvordan man kan gå frem for å opprette disse. Etaten som har ansvaret for dette arbeidet, heter ABM-utvikling.

2.2 ABM-utvikling

ABM-utvikling⁶ er en offentlig etat som ble opprettet i 2003 i samarbeid mellom Kultur- og kirkedepartementet og Kunnskapsdepartementet, og skal bidra til å utvikle de tre sektorene arkiv, bibliotek og museum. I ABM-utviklings bibliotekreform 2014 nevnes tre målområder: Bibliotekenes innhold og tjenester, deres struktur og organisering, samt kompetanse og forskning. Utvikling og formidling av digitalt innhold og nettbaserte tjenester er en av to hovedsatsninger innenfor målområdet innhold og tjenester, og er utdypet slik:

«En hovedsatsing er å utvikle og tilby digitalt innhold og nettbaserte tjenester. Målet er å skape et digitalt bibliotek som gir alle enkel tilgang til kunnskap og kultur. Gjennom samarbeid mellom bibliotekene og nasjonal koordinering og finansiering realiseres nasjonale tjenester som er fritt tilgjengelige for alle.»
(ABM-Utvikling, 2006b: 7)

Brukertilgang er altså en viktig satsning for de norske bibliotekene. I flere av ABM-utvikling sine periodiske tidsskrift (bl.a. (ABM-utvikling, 2006a) og (ABM-Utvikling, 2006b)) omtales tilrettelegging for brukertilgang som et viktig ledd i en demokratiseringsprosess for sektoren. Demokratisering ser i det hele tatt ut til å gå igjen som et viktig moment for utviklingen i

⁵ Eng.: Knowledge Organization System (KOS)

⁶ <http://www.abm-utvikling.no>

sektoren, og det refereres til grunnloven i grunnlovningen til hvorfor det blir sett på som en viktig satsning:

«[Bibliotekene] representerer viktige demokratiske verdier gjennom fri tilgang til informasjon, kultur og kunnskap. Bibliotekene må utvikle rollen som demokratisk møteplass og møte Grunnlovens pålegg i § 100 om å legge forholdene til rette for en åpen og opplyst offentlig samtale. Bibliotek og utdanningsinstitusjoner er viktige deler av infrastrukturen som skal beskytte og fremme ytrings- og informasjonsfrihet.» (ABM-Utvikling, 2006b: 8)

Det ser imidlertid ut til at hovedfokuset i mye av ABM-utviklings arbeid er på fri tilgang til informasjon, kultur og kunnskap, mens Grunnlovens pålegg om å legge forholdene til rette for en åpen og opplyst offentlig samtale i mindre grad er fremhevet. Med det mener jeg at ABM-utvikling har lite fokus på hvordan digitale medier best kan utnyttes for å tilrettelegge for å dra nytte av brukerne i dette arbeidet. Dette studiet er, gjennom undersøkning av brukermedvirkning i kontrollerte vokabularer, med på å undersøke en potensiell metode for å bedre tilgangen til en digital bildesamling, som kanskje også vil kunne brukes i andre omgivelser.

2.3 Universitetsbiblioteket i Bergen / Billedsamlingen

Billedsamlingen ved Universitetsbiblioteket i Bergen er, ifølge seg selv, en av landets største og mest anerkjente arkiv av historisk fotografi (Universitetsbiblioteket i Bergen, 2008). Selve den fysiske samlingen består av enkeltbilder og arkiver av varierende størrelse fra fotografer, samlere og private givere, til sammen omkring en halv million bilder. Noen av bildene er tatt av profesjonelle fotografer, som Knud Knudsen, mens andre er tatt av amatørphotografer (for eksempel Bekker Larsen-samlingen, Schumann Olsen-samlingen, Brosingsamlingen og Tromholtsamlingen). Det foregår et digitaliseringsarbeid, hvor bilder blir lagt inn etter prioritering av samlinger, og det er nå rundt 20 000 bilder som er digitalisert og publisert på verdensveven⁷. Publiseringen av et bilde innebærer at bildet blir digitalisert, og all tekstlig in-

⁷ <http://www.ub.uib.no/bilder>

formasjon som finnes om bildet blir knyttet til bildet og brukt i organisering og gjenfinning. I tillegg knytter spesialister i UB termer til bildene. Visse egenskaper med strukturene som blir brukt i denne organiseringen kan ha rom for forbedring, og et viktig tema i dette studiet er nettopp hvordan bruken av forskjellige organiseringsstrukturer kan brukes for å oppnå best mulig resultat, og hvordan brukerne kan aktiviseres i dette arbeidet.

Et studie utført av Jennifer Trant konkluderte blant annet med at «ikke-spesialister kan tilføre et nyttig antall nye tilgangspunkter, og forsterke de profesjonelle beskrivelsene i kunstmuseumskolleksjoner» (Trant, 2006: 102). Selv om mitt studie ikke foregår i konteksten av et kunstmuseum, har miljøet og egenskaper ved objektene som skal beskrives mye til felles med det som beskrives i Trants studie. Det er ofte snakk om gjenstander og emner som ikke er dagligdagse i dag, men som likevel kan ha interesse for et bredt publikum. Disse gjenstandene og emnene har gjerne spesielle navn, navn som det hovedsakelig bare er kjent av eksperter og spesielt interesserte. Det finnes utallige eksempler på dette i Billedsamlingen, og i bildene som er representert i dette studiet (se appendiks a) kan jeg nevne emneord som stovehus, karjol, fjordabåt, trekjørel og stola. Dette er ord som sjelden blir brukt i dagligtale, og som mange mennesker trolig ville brukt andre ord for. I neste kapittel beskrives blant annet en metode som legger til rette for at brukerne kan være med å beskrive emner med de ordene som er mest naturlig for dem å bruke, og som muligens kan bidra til å gjøre museums-samlinger mer tilgjengelig for et bredere publikum.

2.4 Omgivelser for studiet: konklusjon

Dette studiet utføres innenfor fagfeltet digitale biblioteker. Studiet foregår i en sektor som arbeider med å øke fokuset på brukerne gjennom i større grad å åpne opp for at brukerne kan være med på å skape og vedlikeholde innhold. Universitetsbiblioteket i Bergens Billedsamling brukes som utgangspunkt for å forsøke å utforske en alternativ metode for bruker-medvirkning, og om denne vil kunne skape en forbedring innenfor dette aspektet. Dette blir beskrevet nærmere i kapittel 4, hvor det, basert på dette kapitlet og det neste, blir utviklet et forskningsdesign som skal benyttes for å tilfredsstillere målet for studiet.

3 Tidligere studier og teoretisk rammeverk

I dette kapitlet skal studiet plasseres innenfor et studieområde. Tidligere forskning innenfor emnene som omtales i studiet skal presenteres, og teorien som studiet bygger på skal forklares.

Dette studiet kan hovedsakelig plasseres innenfor fagfeltet informasjonsgjenfinning⁸. Informasjonsgjenfinning er, i tillegg til informasjonsvitenskap, basert på fagfelter som matematikk, kognitiv psykologi, lingvistikk og statistikk, og kan defineres slik:

Definisjon 2: Informasjonsgjenfinning handler om representasjon, lagring, organisering av og tilgang til informasjonsheter (Baeza-Yates & Ribeiro-Neto, 1999: 1).

Representasjonen og organiseringen av informasjonshetene skal sørge for at brukeren har enkel tilgang til informasjonen hun er interessert i, i form av informasjon i databaser eller dokumenter eller metadata som beskriver informasjonen.

3.1 Grunnleggende konsepter

3.1.1 Internett og verdensveven⁹

Internett er et verdensomspennende datanettverk som er grunnlaget for mange forskjellige kommunikasjonstjenester, blant annet verdensveven, e-post, chat (for eksempel IRC), filoverføring (for eksempel FTP), streaming (for eksempel nettradio og nett-tv) og VoIP (Internett-telefoni). E-post og verdensveven er de mest brukte av disse tjenestene, og i forhold til dette studiet er det verdensveven som er den viktigste tjenesten.

Verdensveven er en tjeneste som gjør tekst, bilder, lyd og video tilgjengelig via Internett. Tjenesten behandler og presenterer informasjon for brukerne, i tillegg til å gjøre det mulig for brukere å sende inn informasjon. I Norge er det vanlig å betegne verdensveven som

⁸ Eng.: Information Retrieval

⁹ Eng.: World Wide Web

«nett» i dagligtale, og dette har ført til en viss begrepsforvirring, på grunn av navnelikheten med Internett. Det er viktig å skille mellom Internett som et globalt nettverk, og «nett», eller verdensveven, som en tjeneste på dette nettverket. Bruken av begrepet verdensveven i dette studiet er med på å tydeliggjøre dette skillet.

3.1.2 Innholdsobjekter og dokumenter

Når man bruker tekst og bilder i kommunikasjonssammenheng, kan det være i form av dokumenter. I dette studiet refererer begrepet utelukkende til digitale representasjoner av dokumenter, og er definert slik:

Definisjon 3: *Et dokument er en digital representasjon av informasjon, og kan bestå av ren eller formatert tekst, bilder, grafikk, lyd, andre former for multimedia og/eller hyperlinker til andre dokumenter (bearbeidet fra Hartvedt, 2007).*

For å skille mellom dokumenter i digital form og annet materiale som kan eksistere i fysisk form i tillegg til digitale versjoner, brukes begrepet innholdsobjekter.

Definisjon 4: *Innholdsobjekter er alle objekter som skal beskrives for å inkluderes i et informasjonsgjenfinningssystem, nettside eller andre informasjonskilder (ANSI/NISO, 2005).*

Innholdsobjekter er ofte dokumenter, slik som journalartikler, tekniske rapporter og bilder. Et innholdsobjekt kan eksistere utelukkende i digital form, men kan også være fysiske objekter som malerier, klesdrakter eller fotonegativer. I dette studiet dreier det seg hovedsakelig om bilder eller negativer.

For at det skal være mulig å finne tilbake til innholdsobjekter i en samling av en viss størrelse, er det nødvendig å organisere objektene ved å indeksere dem. I et system som tar i bruk et kontrollert vokabular (se seksjon 3.5), innebærer indeksering å velge foretrukne termer fra det kontrollerte vokabularet for å beskrive innholdsobjektet, men termene må ikke nødvendigvis være representert i et kontrollert vokabular. Indeksering av innholdsobjekter kan gjøres ved hjelp av dataprosessering eller gjennom menneskelige vurderinger.

Definisjon 5: *Indeksering er prosessen med å tilordne foretrukne termer eller overskrifter for å beskrive konseptene og annen metadata som er assosiert med et innholdsobjekt (ANSI/NISO, 2005: 12).*

Tagging er også en form for indeksering, som ofte utføres av brukere, og hvor det ikke tas utgangspunkt i et kontrollert vokabular. Indekserte bilder kan man vanligvis gjenfinne ved hjelp av nøkkelordbaserte søk, mens bilder som ikke er indeksert kan være svært vanskelige å gjenfinne i store databaser.

3.2 Brukergenerert innhold

*Brukergenerert innhold*¹⁰ er innhold som er allment tilgjengelig, og skapt av brukerne. Brukere har tradisjonelt vært konsumenter av innhold, men har i de siste årene i større grad fått en rolle som innholdsskapere. Innholdet kan være tekst, bilder, video og lyd, og ofte brukes teknologi for å legge til rette for samhandling. Verdensveven har i stor grad medvirket til å øke mengden av brukergenerert innhold, fordi tjenesten har mange egenskaper som gjør terskelen for å bidra lavere. Det har i de siste årene oppstått et begrep som heter *social media* (*sosiale medier*), som i stor grad faller sammen med brukergenerert innhold, fordi mesteparten av innholdet ofte er brukergenerert. Aktiviteten skjer hovedsakelig på verdensveven. Eksempler på sosiale medier hvor brukergenerert innhold er fremtredende er *wikier*, *blogger*¹¹, *bilde- og videodeling* og *vevsamfunn*¹².

- En wiki er et vevsted som lar brukere være med å legge til og oppdatere innhold. Wikipedia er et velkjent eksempel på en wiki, og er et nettleksikon som er skapt og holdes ved like av brukere.
- En blogg er en vevside som vanligvis blir skrevet og vedlikeholdt av en eller flere privatpersoner, og kan inneholde tekst, bilder, video og lyd. Det er ofte mulig for lesere av en blogg å komme med tilbakemeldinger til innlegg.

¹⁰ Eng.: user generated content. Også omtalt som user created content og consumer generated media.

¹¹ Eng.: weblog, blog.

¹² Eng.: social networking.

- Flickr er en av de største vevstedene for bildedeling, og YouTube tilsvarende for video. Hvem som helst kan opprette en konto og enkelt legge sine bilder eller video fritt tilgjengelig ut på verdensveven.
- Vevsamfunn har mange fellestrekk med blogger, men den store forskjellen er at vevsamfunn har mekanismer for å bygge online samfunn av mennesker. Det finnes interessegrupper man kan bli medlem av, og man kan lage en kobling mellom sin egen og andres profiler, ofte kalt «venner».

Thurman (2008) påpeker at de som tradisjonelt har konsumert nyheter, i økende grad er villig til å skape innhold, og bruker bloggsøkemotoren Technorati.com, som i 2005 hadde nesten 9 millioner blogger i sin indeks, som eksempel på dette. I juni 2008 hadde dette tallet steget til over 112 millioner. Den engelske utgaven av Wikipedia har over 2,7 millioner artikler. Ifølge Flickr¹³, har bildedelingsvevstedet over 3 milliarder bilder. Disse tallene sier noe om hvor viktig brukergenerert innhold på verdensveven er.

Et viktig aspekt med brukergenerert innhold, er det demokratiske; hvem som helst kan slippe til med innhold, og få en mulighet til å la stemmen sin bli hørt. Dette er som nevnt et viktig fokus i ABM-sektoren. Når man slipper til brukerne, må man imidlertid være forberedt på at det kan oppstå utfordringer.

3.2.1 Utfordringer med brukergenerert innhold

Neil Thurman (2008) har utført en spørreundersøkelse blant ti nyhetsvevsteder i Storbritannia, og intervjuet redaktørene for vevstedene. Basert på dette tar han blant annet for seg problematiske faktorer ved å slippe brukerne til med eget innhold. En av observasjonene Thurman gjør, er at 80 % av de tekstlige formatene hvor brukerne har skapt innhold ble redigert eller moderert på forhånd, sett bort fra meningsmålinger og spørreundersøkelser. Dette kan tyde på at nyhetsvevstedene har motforestillinger mot å gi slipp på kontroll, og noe av det Thurman undersøker, er hva som kan være grunnen til dette. Studiet til Thurman tar for

¹³ <http://blog.flickr.net/en/2008/11/03/3-billion/>

seg andre typer innhold enn tagging og kontrollerte vokabularer, men det er viktig å vurdere følgende, uavhengig av hvilket format det er på innholdet: *Hvorfor bør man åpne opp for brukergenerert innhold, og hva bør man tenke gjennom først?*

Et gjennomgående dilemma blant mange av redaktørene som ble intervjuet av Thurman, går ut på hvor mye kontroll det er ønskelig å gi til brukerne. I mange tilfeller har tilbydere av vevtjenester et redaksjonelt ansvar for materialet som finnes på vevsidene sine, noe som innebærer at man må være spesielt på vakt overfor brukergenerert innhold. Thurman (ibid.) nevner et søksmål mot The Sunday Herald i mars 2003 av Lord Robertson, som var generalsekretær i NATO på den tiden. Robertson saksøkte The Sunday Herald på grunn av et innlegg om ham i et meldingsforum som skal ha vært ærekrenkende, og saken endte med et forlik på 25 000 £. Denne saken er et eksempel på at det kan være svært viktig for tilbydere av vevtjenester å ha kontroll over brukergenerert innhold.

Det er hovedsakelig to metoder som kan brukes for å moderere brukergenerert innhold: fjerning av hele bidrag, eller redigering av bidrag. Modereringen kan skje i etterkant, ved at nye bidrag blir tatt i bruk med en gang de blir lagt til, og gått gjennom og fjernet eller redigert i etterkant (etterhåndsmoderering), eller man kan gå gjennom bidragene på forhånd, slik at ingen bidrag blir tatt i bruk før de har vært gjennom en eller annen form for godkjenning (forhåndsmoderering). Man kan bruke menneskelige og maskinelle ressurser for å utføre modereringen. Informasjonssystemer kan brukes for å luke ut bidrag som inneholder sensitive ord, og ved bruk av forhåndsmoderering kan informasjonssystemer brukes for å lage et «køsystem», hvor det første bidraget som kommer inn er det første som blir behandlet. I noen tilfeller brukes det også et tillitssystem, hvor brukere tidligere må ha generert en viss mengde innhold av akseptabel kvalitet for å få mulighet til å bidra i spesifikke deler av et vevsted. En grunn for å ønske å kontrollere brukerbidrag på denne måten, er at man vil forsikre seg om at det innholdet som når ut til brukerne har høy kvalitet. Når man opererer med brukergenerert innhold, er det vanlig at det samme innholdet oppstår gjentatte ganger. Usakligheter og krenkende utsagn, som rasisme og andre former for hets har dessuten ofte en tendens til å florere når brukere får fritt spillerom, som i eksempelet med Lord Robertsons stevning.

3.2.2 Motivasjon for å delta

Bughin (2007) utførte en spørreundersøkelse for å finne ut mer om hva som motiverer brukere til å delta med innhold på verdensveven. Studiet tok for seg brukere i fire forskjellige video-delingsnettsteder, og viste at de fleste hadde berømmelse som motivasjon for å delta, etterfulgt av opplevelse av glede, ønske om å dele erfaringer med venner, og ønske om at andre kan dra nytte av innholdet. Det var sjelden et ønske om kompensasjon i form av penger som motiverte brukerne til å delta. I studiet til Bughin anbefales det å identifisere og dra nytte av den andelen av brukerne som bidrar med innhold av høyest kvalitet. Jennifer Trant (2006) har utført et studie som dreier seg om brukergenerert innhold i museumssektoren, og ifølge Trant skiller denne sektoren seg på mange måter fra andre når det gjelder brukernes motivasjon for å delta. Museer har «eksisterende relasjoner med besøkende som definerer en sosial og kulturell kontekst for taggeaktiviteten» (Trant, 2006: 86, oversatt fra engelsk). Det er snakk om lærere, studenter og frivillige, som «uten å være motivert av personlig vinning, (...) donerer tid og kunnskap» (ibid.).

3.3 Bildegjenfinning¹⁴

Bildegjenfinning er en del av fagfeltet informasjonsgjenfinning, og har vært et svært aktivt forskningsfelt siden 70-tallet (Rui, Huang, & Chang, 1999: 39). Det finnes hovedsakelig to forskjellige måter å søke etter digitale bilder: Tekstbasert og innholdsbasert (Elgesem & Nordbotten, 2007: 24). Den vanligste måten for å gi tilgang til bildedatabaser gjennom verdensveven, er nøkkelordbaserte søk basert på annotasjoner som er knyttet til bildene. I tillegg finnes det systemer som kan generere indekser basert på form, farge og tekstur (innholdsbasert bildegjenfinning¹⁵). Bildegjenfinning generelt kan sammenfattes slik:

Definisjon 6: *Bildegjenfinning* handler om representasjon, lagring, organisering og tilgang til bilder i en digital bildesamling.

¹⁴ Eng. Image Retrieval

¹⁵ Eng.: Content Based Image Retrieval (CBIR)

3.3.1 Tekstbasert bildegjenfinning

Det finnes mange forskjellige måter å påvirke søkeresultatene fra et tekstbasert bildesøk, og jeg skal i de følgende avsnittene ta for meg noen av disse.

3.3.1.1 Leksikalsk analyse av teksten

Leksikalsk analyse av teksten går ut på å skille ut ordene fra en tekst (Baeza-Yates & Ribeiro-Neto, 1999: 165). Det innebærer grunnleggende å dele opp ord etter mellomrom, men det må også avgjøres hvordan tall, bindestreker, skilletegn og store og små bokstaver skal håndteres.

3.3.1.2 Eliminering av stoppord

Ord som forekommer for ofte i en tekst i samlingen er ikke særlig nyttige i å skille mellom gode og dårlige treff (Baeza-Yates & Ribeiro-Neto, 1999: 167). Et ord som forekommer i 80 % av dokumentene i samlingen er ubrukelige i informasjonsgjenfinningssammenheng disse ordene blir ofte kalt *stoppord*, og blir vanligvis filtrert ut som mulige indekstermer. Artikler, preposisjoner og konjunksjoner er vanlige kandidater for en liste over stoppord.

3.3.1.3 Grunnstammesøk

Det skjer ofte at en bruker skriver et ord i en spørring, hvor bare en variant av ordet finnes i den relevante teksten (Baeza-Yates & Ribeiro-Neto, 1999: 168). Flertall, gerundium¹⁶, og fortidsendelser er eksempler som forhindrer direkte likhet mellom et spørreord og en tekst. Dette problemet kan delvis overkommes ved å bytte ut ord med stammen av ordet.

En *stamme* er, ifølge Baeza-Yates og Ribeiro-Neto, den delen av et ord som gjenstår etter at affiksene¹⁷ er tatt bort. Et eksempel på en stamme er ordet *kjør*, som er stammen til *kjøre*, *kjører*, *kjørte*, *kjørt*, *kjøring*, *kjøringer* og *kjøret*. Stammer anses å være nyttige, fordi de reduserer varianter av den samme roten til et felles konsept. Dette kan man også oppnå med

¹⁶ Verbalsubstantiv på «-ing», som f.eks. hopping, kjøring, bryting

¹⁷ Tilleggsstavelse, f.eks. prefiks og suffiks.

kontrollerte vokabularer, og det ville teoretisk være mulig å bruke for eksempel en synonymring (se seksjon 3.5.4.2) for å oppnå det samme. Dette ville imidlertid ført til en omfattende jobb med å legge til termer, noe man i stor grad slipper ved å bruke grunnstammesøk.

3.3.2 Evaluering av søkeresultater

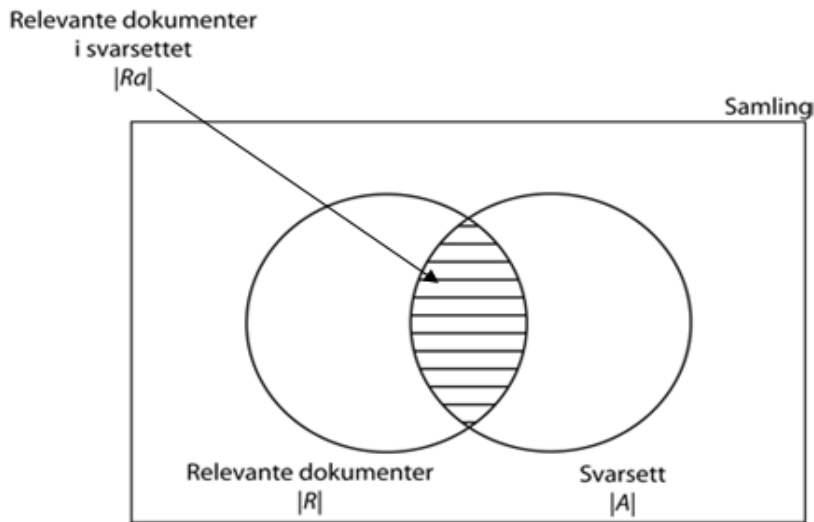
Før den endelige implementeringen av et informasjonsgjenfinningssystem, er det vanlig å utføre en evaluering av systemet (Baeza-Yates & Ribeiro-Neto, 1999: 73). Det første man gjør er å utføre en funksjonell analyse hvor de spesifiserte systemfunksjonene testes en etter en. Dette kan inkludere en feilanalyseringsfase hvor man prøver å få systemet til å feile, i stedet for å se etter funksjoner. Det er en enkel prosedyre, som kan være ganske nyttig for å fange opp programmeringsfeil.

Etter at systemet har vært gjennom denne funksjonelle analysen, bør man evaluere ytelsen til systemet. Brukerens informasjonsanmodning er vag i sin natur, og den informasjonen som blir gjenfunnet er ikke eksakte svar, og må rangeres ut fra sin relevans til forespørselen. Denne typen evaluering kalles evaluering av søkeresultater¹⁸ (Baeza-Yates & Ribeiro-Neto, 1999: 73). Precision og recall er metoder for å evaluere kvaliteten til et sett av søkeresultater ved å vurdere relevansen til dokumentene som blir gjenfunnet. Recall er den andelen av relevante dokumenter som har blitt hentet frem – hvor mye av informasjonen som ideelt sett kunne vært gjenfunnet, som faktisk ble gjenfunnet. Precision er den andelen av de fremhentede dokumentene som er relevante, og et mål for hvor godt systemet fungerer til ikke å returnere irrelevante dokumenter. Hvis hvert dokument som blir gjenfunnet er relevant, er precision 100 %. Dette kan demonstreres gjennom et eksempel. Gitt en spørring S og settet R med relevante dokumenter. La $|R|$ være antallet dokumenter i dette settet. Anta at man gjennom bestemt gjenfinningsstrategi som skal evalueres utfører spørringen S og genererer et resultatsett $|A|$. La $|A|$ være antallet dokumenter i dette settet. La så $|Ra|$ være antallet dokumenter i snittet av settene R og A . Figur 1 illustrerer disse settene. Vi kan nå definere precision og recall slik:

¹⁸ Eng.: Retrieval performance evaluation

$$Recall = \frac{|Ra|}{|R|}$$

$$Precision = \frac{|Ra|}{|A|}$$



Figur 1: Precision og recall for en eksempelspørring.

3.4 Kunnskapsorganiseringssystemer

Kunnskapsorganiseringssystemer (KOS) omfatter alle typer skjemaer for å organisere informasjon og støtte kunnskapshåndtering. Dette innbefatter blant annet klassifiserings- og kategoriseringsskjemaer som organiserer innholdsobjekter på et generelt nivå, emneoverskrifter¹⁹ (termer) brukt for å beskrive et spesifikt emne som gir mer detaljert tilgang, og autoritære filer som kontrollerer varianter av nøkkelinformasjon som for eksempel geografiske navn og personnavn. Kunnskapsorganiseringssystemer brukes for å organisere innholdsobjekter for å administrere en samling og for å legge til rette for gjenfinning. Et KOS

¹⁹ Eng.: Subject heading

fungerer som en bro mellom brukerens informasjonsbehov og innholdet i samlingen, og er dermed essensielt i implementeringen og bruken av et digitalt bibliotek.

Definisjon 7: *Et kunnskapsorganiseringsystem er et system for å håndtere organiseringen av informasjonsressursene i et digitalt bibliotek med tanke på kontroll og gjenfinning.*

Et kunnskapsorganiseringsystem er ofte implementert i form av en eller flere typer kontrollert vokabular.

3.5 Kontrollerte vokabularer

Vokabularkontroll brukes for å forbedre effektiviteten til systemer for informasjonslagring og -gjenfinning, nettnavigasjonssystemer og andre miljøer som bruker tekst for å identifisere og lokalisere innhold. Ved å bruke kontrollerte vokabularer oppnår man konsistens i beskrivelsen av innholdsobjekter, og man legger til rette for gjenfinning.

Det finnes mange definisjoner på de forskjellige konseptene innenfor feltet, men jeg har hovedsakelig valgt å holde meg til ANSI/NISO sine retningslinjer for konstruksjon, format og håndtering av enkeltspråklige kontrollerte vokabularer (ANSI/NISO, 2005). Dette er en relativt ny tekst, som bygger på tidligere forskning, og går gjennom teorien på en strukturert og entydig måte.

3.5.1 Termer

Basiselementene i et kontrollert vokabular er de individuelle termene (Foskett, 1997). Termer kan være individuelle ord, grupper av ord, eller fraser, men de fleste er enkle ord (Baeza-Yates & Ribeiro-Neto, 1999: 171). Termer er som regel substantiver, fordi substantiver er den mest konkrete delen av språket.

Definisjon 8: *En term er et eller flere ord som blir brukt for å representere et konsept (ANSI/NISO, 2005: 11).*

3.5.2 Flertydighet

Flertydighet forekommer i naturlig språk når et ord eller et uttrykk har mer enn én betydning. Dette kalles polysemi (et ord eller en frase med flere, relaterte betydninger) og homonymi (et ord eller en frase med flere forskjellige betydninger). Et eksempel på et polysem er bane, som for eksempel kan bety jernbane, men også fotballbane. Eksempel på et homonym er bolle, som for eksempel kan være et bakverk, eller en del av et servise. Ifølge ANSI/NISO (2005), må et kontrollert vokabular tilby en løsning på dette problemet ved å sørge for at hver term har en og bare en betydning.

3.5.3 Termrelasjoner

3.5.3.1 Ekvivalensrelasjoner

Settet av termer som er relatert til en bestemt term i et kontrollert vokabular er i hovedsak satt sammen av *synonymer*, *kvasisynonymer* og *leksikalske varianter* (ANSI/NISO, 2005). Dette kalles *ekvivalensrelasjoner*. Synonymer er termer hvor meningen er lik i ulike kontekster, for eksempel oksygen og surstoff, eller tidsskrift og magasin. Kvasisynonymer er termer der meningen blir sett på som forskjellig i vanlig bruk, men som behandles som synonymer for indekseringsformål, for eksempel hardhet og mykhet (Olsen, 2004: 43). Forskjellen mellom leksikalske varianter og synonymer er at synonymer er andre termer for det samme konseptet, mens leksikalske varianter er forskjellige ordformer for det samme uttrykket. Eksempler på leksikalske varianter er Ulriken / Ulrikken, eller smug / smau / smitt.

Når det samme konseptet kan uttrykkes med to eller flere termer, velges en av disse som foretrukket term. Ekvivalensrelasjoner angir relasjoner mellom foretrukne og ikke foretrukne termer på denne måten:

Foretrukket term USE FOR (UF) ikke foretrukket term.

Ikke foretrukket term USE **foretrukket term**.

Dette kan beskrives med følgende eksempel:

Eksempel 1: USE- og UF-relasjoner

smau USE smug

smug UF smau

3.5.3.2 Hierarkiske relasjoner

I tillegg kan termer ha en hierarkisk relasjon, med bredere (*broader term / BT*) eller smalere (*narrower term / NT*) termer. Dette blir beskrevet i seksjon 3.5.4.3.

3.5.3.3 Assosiative relasjoner

Det er også mulig med en tredje relasjonstype, hvor relasjonen er sidestilt og ikke-hierarkisk. Dette kalles assosiative relasjoner, og slike relasjoner blir indikert med forkortelsen RT (*related term*). Denne relasjonstypen dekker assosiasjoner mellom termer som verken er ekvivalente eller hierarkiske, men som likevel er semantisk eller konseptuelt assosiert. Relasjonstypen kan brukes til å foreslå ytterligere termer som kan brukes i indeksering eller gjenfinning.

3.5.4 Struktur

ANSI/NISO (2005) opererer med 4 forskjellige typer kontrollerte vokabularer, med forskjellig grad av kompleksitet i strukturen:

- Liste
- Synonymring
- Taksonomi
- Tesaurus

Figur 2 viser kompleksiteten til kontrollerte vokabularer i forhold til hvilke relasjonstyper som er mulig. Vi ser at en liste er den minst komplekse av de fire typene av kontrollerte vokabularer, mens tesaurus er i den andre enden av skalaen, med høy kompleksitet.

	Kompleksitet			
	Mindre			Mer
	Liste	Synonymring	Taksonomi	Tesaurus
Flertydighets-kontroll	✓		✓	✓
Synonym-kontroll		✓	✓	✓
Hierarkiske relasjoner			✓	✓
Assosiative relasjoner				✓

Figur 2: Økende strukturell kompleksitet i kontrollerte vokabularer (omarbeidet fra ANSI/NISO (2005: 17))

3.5.4.1 Lister

En liste er et begrenset sett av termer arrangert alfabetisk eller på en annen logisk måte. Lister er brukt for å beskrive aspekter ved innholdsobjekter eller enheter som har et begrenset antall muligheter. Eksempler på lister er språk (for eksempel engelsk, norsk, tysk) eller geografi (for eksempel fylke, kommune, by). Eksempel 2 viser en alfabetisk liste over 5 fylker i Norge, mens eksempel 3 viser en logisk liste over Norges 5 høyeste fjell, sortert etter fjellenes høyde i synkende rekkefølge.

Eksempel 2: Enkel alfabetisk liste

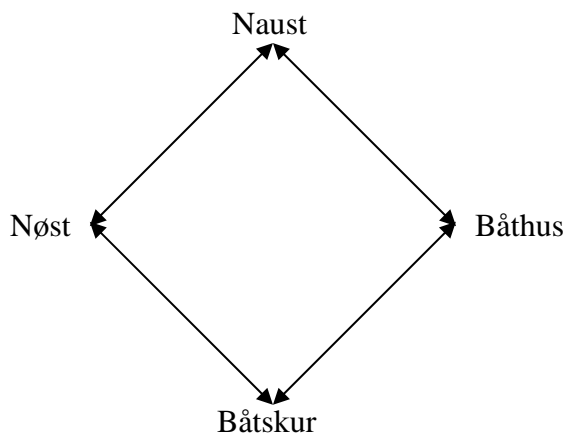
Akershus
Aust-Agder
Buskerud
Finnmark
Hedmark

Eksempel 3: Enkel logisk liste

Galdhøpiggen
Glittertinden
Store Skagastølstind
Store Styggedalstind
Skarstinden

3.5.4.2 Synonymringer

Synonymringer sørger for at et konsept som kan beskrives med flere ekvivalente termer vil bli gjenfunnet hvis hvilket som helst av termene blir brukt i et søk. Strukturen brukes ikke til indekseringsformål, men brukes bare i forbindelse med gjenfinning. En synonymring er altså et sett av termer som regnes som ekvivalente for gjenfinningsformål. Synonymringer forekommer oftest som sett av flate lister. Strukturen gir brukerne tilgang til alle innholdsobjekter som inneholder hvilken som helst av termene i synonymringen. Ifølge ANSI/NISO (2005: 18), sørger synonymringer for tilgang til innhold som er representert i naturlig, u-kontrollert språk.



Figur 3: Synonymring for naust

3.5.4.3 Taksonomier

En taksonomi er et kontrollert vokabular bestående av foretrukne termer som alle er koblet sammen i et hierarki eller polyhierarki. Hierarkiske relasjoner er basert på grader eller nivåer av overordning eller underordning, hvor den overordnede termen representerer en klasse eller helhet, og underordnede termer refererer til medlemmene eller delene til denne helheten. I

eksempel 4 kan man se et lite utsnitt av emneord-hierarkiet til Billedsamlingen. Det finnes forskjellige måter å presentere et hierarki, men i dette studiet brukes en notasjon hvor de forskjellige nivåene er representert med •••. Klær er den bredeste termen i eksempelet, og er overordnet term til skotøy, som er underordnet term til klær, osv.

Eksempel 4: Taksonomihierarki

klær

- skotøy
- barnesko
- barnesandaler

Hvis et konsept tilhører mer enn en kategori, har konseptet polyhierarkiske relasjoner. Et eksempel på dette er bunad, som kan tilhøre kategoriene mannsklær, kvinneklær og folke-
drakter. Billedsamlingens kontrollerte vokabular har per i dag ikke støtte for polyhierarkiske relasjoner.

3.5.4.4 Tesauri

Ordet *Tesaurus* kommer fra gresk og latin, og betyr «skattkammer» (Foskett, 1997: 111). Det brukes i dag for å beskrive et skattkammer av ord, eller en ordbok. En tesaurus består, i sin enkleste form, av (1) en forhåndssammensatt liste av viktige ord i et gitt kunnskapsdomene, og (2) et sett av relaterte ord for hvert ord i denne listen (Baeza-Yates & Ribeiro-Neto, 1999: 170). Relasjonstypene kan være ekvivalensrelasjoner, hierarkiske relasjoner eller assosiative relasjoner.

Tesauri er ofte domenespesifikke, men kan også være generelle. Ifølge Foskett (1997, se Baeza-Yates & Ribeiro-Neto, 1999: 170), er hovedformålene til en tesaurus (1) å sørge for et standardvokabular (eller system av referanser) for indeksering og søking; (2) å assistere brukere med å finne gode termer for spøringsformulering; og (3) å tilby klassifiserte hierarkier som gjør det mulig å utvide eller innsnevre den gjeldende spøringen ut fra brukerens ønsker.

3.6 Bildetagging

Bildetagging er en svært populær og mye utbredt metode for å indeksere innhold på verdensveven. Denne teknikken går ut på at brukerne av en tjeneste på veven knytter enkle eller sammensatte ord (tagger) til dokumenter, som bilder og videoklipp. Det har blitt stadig mer populært å bruke taggesystemer etter at et element med sosial interaksjon ble introdusert (Furnas, et al., 2006). Slike sosiale taggesystemer kobler sammen brukernes individuelle taggeaktivitet til et nettverk av tagger og ressurser. Ved å sette sammen, eller aggregere, tagger fra mange brukere på denne måten, kan man skape strukturer som kan være nyttig for å legge til rette for søk og utforskning av informasjon.

3.6.1 Folksonomier

Resultatet av å aggregere tagger kalles en *folksonomi*. Begrepet er en sammensetning av folk og taksonomi, men har lite til felles med en taksonomi, fordi en folksonomi mangler relasjoner mellom termer, slik man finner i en taksonomi. En folksonomi dannes ved at tagger som brukere har laget for å indeksere sine «egne» informasjonselementer (f.eks bokmerker i del.icio.us, eller bilder i Flickr) blir satt sammen ved hjelp av et datasystem, og brukes for å hjelpe brukerne til å finne tilbake til informasjonenheter gjennom søk og utforskning. På denne måten er det mulig for hver informasjonsenhet å bli tagget med den samme taggen flere ganger, og man får et rikere sett av metadata for hver informasjonsenhet. Hvis en term brukes flere ganger for å tagge en informasjonsenhet, er dette en indikasjon på at denne termen har en høy relevans i forhold til informasjonsenheten, og dette kan brukes for å rangere resultatene.

Folksonomier blir ofte opprettet i det miljøet de blir brukt, og på denne måten er de samme brukerne med på både opprettelsen og bruken av folksonomien, noe som er hensiktsmessig fordi det fører til at grunnlaget for å gjenfinne informasjon er skapt av de samme som skal bruke dette grunnlaget i praksis. Tagging gjenspeiler taggernes konseptuelle modell over informasjon, og gir en autentisk representasjon av språket til taggerne (Peters & Stock, 2007). Dette gjør det mulig for brukerne å dra fordel av andres erfaringer i tillegg til sine egne.

En kritikk mot folksonomier hvor alle taggene er fritt lagt til av brukere (og ikke tatt fra et kontrollert vokabular) går ut på at man mangler terminologisk kontroll (Golder & Huberman, 2006). Man mister kontroll over synonymer, homonymer polysemer og hierarkiske relasjoner mellom termer (se seksjon 3.5). Golder & Huberman (ibid.) nevner

også problemet med entall og flertall – i noen tilfeller brukes entallsendinger når brukere legger til tagger, mens det i noen tilfeller blir brukt flertallsendinger. Sistnevnte er imidlertid et problem som man enkelt kan i møtekomme, og i stor grad eliminere, ved å bruke en algoritme for å finne grunnstammen til søkeordene (se seksjon 3.3.1.3).

Tagging er en teknikk som fungerer svært godt for å øke andelen relevante dokumenter som blir gjenfunnet (recall), men som ofte går på bekostning av andelen gjenfunne dokumenter som er relevante (precision). Dette skyldes at brukerne som tagger bildene ofte ikke er eksperter på emnene som er representert i dokumentene, og derfor legger til tagger som ikke er riktige. Det kan også skyldes at brukere med dårlige hensikter med vilje legger inn feilaktige tagger (sabotasje).

3.7 Vokabular-problemet

«If everyone always agreed on what to call things, the user's word would be the designer's word would be the system's word, and what the user typed or pointed to would be mutually understood.» (Furnas, Landauer, Gomez, & Dumais, 1987: 964)

Furnas et al. beskriver en ideell verden for informasjonsgjenfinning, en verden hvor språklige forskjeller og flertydighet ikke eksisterer. I den virkelige verden er det imidlertid slik at man har slike forskjeller, og det er noe man er nødt til å ta konsekvensen av hvis man skal lage et godt informasjonsgjenfinningssystem. Basert på empiriske undersøkelser viser Furnas et al. at det er mindre enn en femtedels sjans for at to mennesker velger det samme ordet for et objekt på første forsøk, noe de kaller «the Vocabulary Problem». Furnas et al. finner, gjennom sine undersøkelser, ut at det beste resultatet man kan forvente å oppnå hvis man bare bruker én enkelt term for å beskrive et objekt, er at 15-35 % blir gjenfunnet. Dette gjelder hvis termen som er brukt av de som indekserer, er den mest populære termen for det samme objektet blant de som skal gjenfinne objektet. Furnas et al. oppsummerer dette med at «det finnes ingen enkelt god tilgangsterm for de fleste objekter» (Furnas, et al., 1987: 967).

Furnas et al. undersøker hvordan ytelsen forandrer seg hvis man bruker en *aliasliste*. En aliasliste har de samme egenskapene som en synonymring (se seksjon 3.5.4.2), og har som funksjon å knytte sammen termer som beskriver det samme konseptet. I undersøkelsene til Furnas et al. rangeres aliasene etter popularitet, og man ser at jo flere alias man legger til, desto bedre er systemet i stand til å finne det brukerne søker etter. Men man må, ifølge Furnas

et al., likevel opp i rundt 15 alias for at 60-80 % av brukernes søk skal lykkes. Forfatterens slutning er at, for at brukere som ikke kjenner til vokabularet skal kunne lykkes i å finne det de leter etter, må man tilby svært mange alternative termer. For å få til dette er det en god løsning å aktivere brukerne.

Recall kan aldri bli lavere etter hvert som man legger til flere alternative termer eller tagger, men det samme gjelder ikke for precision. Hvis en alternativ term eller tag er feilaktig, kan dette føre til at systemet gjenfinder objekter som ikke burde blitt gjenfunnet ut fra spørringen til brukeren.

3.8 Tidligere studier og teoretisk rammeverk: konklusjon

Den vanligste måten å dra nytte av brukerne på i bildegjenfinning er folksonomier bygd på brukertagging, en metode som er svært utbredt i digitale bildesamlinger og mye forsket på. Et problem med denne tilnæringsmåten kan være at det går på bekostning av presisjonen til resultatene (precision). Dette kan skje av forskjellige grunner, både tilsiktet (sabotasje) og utilsiktet (misforståelser) fra brukernes side. Alternativet til brukertagging er å la eksperter ta seg av indekseringen av bilder basert på et kontrollert vokabular, men dette kan være svært tidkrevende og kostbart. De siste årene har forskning vist at utvikling og oppdatering av kontrollerte vokabularer kan dra nytte av folksonomier (Peters & Stock, 2007). Dette er fordi tagging i tillegg til å gjenspeile brukernes konseptuelle modeller over informasjon, fører til en rask reaksjon på forandringer i kunnskapsdomenet.

De siste årene har studier som er rettet mot ABM-sektoren undersøkt om også denne sektoren kan dra nytte av å involvere brukerne for å gjøre samlingene tilgjengelig for et bredere publikum. Blant disse er Trant (2006) og Chun et al. (2006), som begge er utført innenfor prosjektet Steve.museum. I studiene ble digitale bilder av kunstverk tagget av en gruppe brukere. Ut fra en ekspertgjennomgang av de nye taggene, indikerte studiets resultater at vanlige brukere også i denne konteksten i stor grad var i stand til å bidra med valide tagger. En mangel ved denne fremgangsmåten er imidlertid at man ikke undersøker om brukerne som skal forsøke å gjenfinne bilder i samlingen har nytte av de nye taggene. Dette er en viktig forutsetning for å kunne si noe om nytteverdien av brukergenerert innhold knyttet til museumssamlinger, fordi det har lite for seg med mange relevante ord hvis det ikke er de ordene det blir søkt etter.

4 Forskningsdesign

Som nevnt i innledningen og i kapittel 2, er det stadig mer fokus på brukerens rolle både på verdensveven og i den norske ABM-sektoren og Universitetsbiblioteket, hvor studiet er gjennomført. Målet med studiet er å undersøke gjennomførbarheten og verdien av å innføre brukergenerert innhold i kontrollerte vokabularer. Med verdien menes hvilken forbedring man kan forvente å oppnå i relevans i form av precision og recall med innføringen av det brukergenererte innholdet.

Bildetagging er, som beskrevet i seksjon 3.6, en form for brukermedvirkning som man i dag finner brukt i svært mange digitale bildesamlinger for å indeksere bilder og gjøre dem søkbare. Det er ingen tvil om at bildetagging og folksonomier i stor grad er med på å gjøre det mulig å finne frem i den enorme informasjonsmengden på verdensveven. Bildetagging har en klar positiv innvirkning på andelen relevante bilder som blir gjenfunnet (recall). Andelen gjenfunne bilder som er relevante (precision) kan imidlertid ofte lide under denne formen for indeksering, fordi vanlige brukeres kjennskap til terminologien og innholdet som tagges ikke alltid er optimal. I dette studiet skal det undersøkes om man kan dra fordel av en annen form for brukergenerert innhold, hvor brukerne i stedet for å tagge bilder direkte, oppdaterer et kontrollert vokabular som er utgangspunkt for å indeksere bildene. Siden tagging kan regnes som en «de facto»-standard for indeksering av bildemateriale på verdensveven i dag, er det naturlig å sammenligne disse to formene for brukermedvirkning – tagging på den ene siden, og oppdatering av et kontrollert vokabular på den andre siden. Tagging blir derfor brukt som sammenligningsgrunnlag for å undersøke hvordan brukergenerert innhold i et kontrollert vokabular fungerer.

Selve gjennomføringen av metoden er beskrevet i seksjon 4.6. Først er det imidlertid på sin plass å presentere problemstillingen og metoden som er brukt, samt valg av og design av forskningsinstrument.

4.1 Definisjon av problemstilling

Følgende problemstilling er formulert som utgangspunkt for studiet:

Vil innføringen av nye termer lagt inn av brukere i det kontrollerte vokabularet tilknyttet UBs billedsamling kunne bidra til en forskjell i precision og recall for vanlige spørringer sammenlignet med brukergenererte tagger?

Denne problemstillingen bygger på målet for studiet, og gir et utgangspunkt for hvordan det er mulig å oppnå dette målet, ved å spesifisere hva som skal måles og hvilke verdier som skal brukes for å måle det.

4.2 Metodevalg

For å besvare problemstillingen, er det nødvendig å samle inn data, og siden brukernes handlinger og bruk av Billedsamlingen er en viktig faktor i problemstillingen, er det avgjørende å involvere brukerne i undersøkelsene. Det finnes mange metoder for å samle inn data fra brukere, og de mest populære er intervjuer, spørreskjemaer og direkte observasjon.

Før undersøkelsene ble satt i gang, ble det gjort en vurdering av hvilke metoder som var best egnet til å besvare problemstillingen. I vurderingen ble det tatt hensyn til (1) behovet for å samle inn data om hvilke spørringer brukerne bruker når de skal prøve å finne bilder i samlingen, og (2) behovet for å samle inn brukergenerert innhold som kan brukes i bildegjenfinningen for å måle forskjeller i precision og recall. Det førstnevnte behovet er nødvendig for å kunne gi et realistisk bilde av hvilken innvirkning det nye brukergenererte innholdet har på gjenfinningskvaliteten. Når det foreligger en versjon av systemet hvor det kontrollerte vokabularet er beriket med termer fra en brukergruppe, og en versjon hvor en annen brukergruppe har knyttet tagger til bildene, er det mulig å sammenligne precision og recall i de to versjonene for settet av spørringer som hentes inn.

For å finne frem til hvilken metode som var best egnet for studiet, ble det, for hvert av de to ovennevnte behovene, vurdert hvilken metode som var mest hensiktsmessig. For å samle inn data om hvilke termer som blir brukt for å søke på bildene, ville det vært mulig å bruke et spørreskjema, hvor deltakerne skriver ned de aktuelle spørringene på papir eller i et grafisk brukergrensesnitt. Intervjuer kunne også vært brukt, men ville trolig vært mindre hensiktsmessig, fordi det ville gitt deltakerne mindre oversikt over hvilke spørringer de lager og resultatene de gir, samt at det ville vært vanskelig å få hen helt presis gjengivelse av hva deltakerne sa. Et tredje alternativ som ble vurdert, var å observere deltakerne mens de utførte faktiske søk mot billed databasen, hvor de fikk se hvilke resultater spørringene gav etter hvert som de søkte. Observasjonsmetoden har klare fordeler, fordi den i størst grad gir et bilde av

den situasjonen deltakerne vil være i når de skal søke etter bildene, og deltakerne får mulighet til å tilpasse valg av spørringer ut fra hvilken respons de får fra systemet. Observasjonene kan gjøres i et feltstudie, eller som et laboratorieforsøk. For å redusere tids- og ressursbruk er det hensiktsmessig å bruke et laboratorieforsøk. Et kontrollert eksperiment hvor deltakerne brukte en prototyp av et søkesystem som logget hver spørring som ble brukt, ble derfor vurdert å være det beste alternativet for å dekke dette behovet. Eksperimenter blir beskrevet i neste seksjon.

Når det gjelder det andre behovet, hvor systemet skal endres av brukerne, ble det på samme måte vurdert om det var mulig og hensiktsmessig å bruke spørreskjemaer eller intervjuer til å samle inn forslag til nye termer i vokabularet eller tagger til bildene. Det ville trolig vært mulig å lage spørreskjemaer hvor deltakerne kunne skrive inn termer, eller å spørre om dem i et intervju, men dette ville vært en unaturlig og lite hensiktsmessig fremgangsmåte, og særlig i intervjuet ville det vært vanskelig å få en presis gjengiving av stavemåter. Igjen ble laboratorieobservasjon av deltakernes interaksjon med en applikasjon vurdert å være den mest gunstige måten å samle inn data på. Ved å føre logg over deltakernes interaksjon med et data-system, er det mulig å sammenligne forskjellige tilstander av systemet; med nye termer i det kontrollerte vokabularet, eller med nye tagger til hvert bilde. Også for å utføre denne oppgaven, ble det vurdert å være hensiktsmessig å utføre et eksperiment.

Selve utførelsen av eksperimentet blir beskrevet i detalj i seksjon 4.6.2, etter at den generelle teorien bak eksperimenter og planleggingen av dette eksperimentet er gjennomgått. Før det er det på sin plass med en kort beskrivelse av hvordan eksperimentet ble utført i dette studiet, slik at leserne kan danne seg et bilde av hvordan forsøkene foregikk.

For å sammenligne virkningen av å la brukere oppdatere et kontrollert vokabular med virkningen av å la brukerne tagge bilder direkte, ble det besluttet å bruke to separate grupper med deltakere, hvor den ene gruppen fikk i oppgave å oppdatere et kontrollert vokabular, mens den andre gruppen fikk i oppgave å tagge bilder. Bildene var de samme i de to gruppene. Oppgavene ble utført i en bildesøkingsapplikasjon som ble utviklet i forbindelse med studiet, og som fungerte som måleinstrument. Deltakerne i gruppen som skulle oppdatere det kontrollerte vokabularet fikk se bildene, og skulle oppdatere termer som var tilknyttet bildene. Deltakerne i den andre gruppen fikk se bildene, og skulle knytte egendefinerte tagger til bildene. I tillegg skulle alle deltakerne i begge gruppene bruke systemet til å søke etter bildene. Spørringene som ble brukt for å søke etter bildene ble brukt i ettertid for å regne ut relevansen til søkeresultatene i de to gruppene hver for seg.

4.3 Eksperimenter

Eksperimenter kan utføres når man trenger kontroll over situasjonen, og ønsker å manipulere atferd direkte, presist og systematisk (Wohlin, et al., 2000). Et eksperiment begynner ofte med at man har en idé om et årsak/virkningsforhold. Man har en teori, eller man kan legge frem ideen formelt i en hypotese. Et eksperiment kan brukes for å evaluere teorien eller hypotesen. De følgende definisjonene i denne seksjonen er bearbeidet fra Goodwin (2005):

Definisjon 9: *Et eksperiment er en forskningsprosedyre hvor en eller flere faktorer varieres, alt annet holdes konstant, og resultatene av den systematiske variasjonen observeres.*

Faktorene som varieres av undersøkeren kalles *uavhengige variabler*, faktorene som holdes konstant kalles *ytre variabler*, og atferden som måles er *avhengige variabler*.

Definisjon 10: *Den uavhengige variabelen er den faktoren som er interessant for undersøkeren; faktoren som undersøkes for å se om den har innvirkning på atferd.*

Uavhengige variabler må ha et minimum av to nivåer – et eksperiment involverer en sammenligning mellom to eller flere situasjoner (eller vilkår).

Undersøkeren i et eksperiment må prøve å kontrollere ytre variabler²⁰. Ytre variabler har følgende definisjon:

Definisjon 11: *Ytre variabler er enhver ukontrollert faktor som ikke er av interesse for undersøkeren, men som kan påvirke atferden som studeres.*

Så lenge de ytre variablene holdes konstant, innebærer de ikke en fare for studiet, men hvis de ikke blir kontrollert i tilstrekkelig grad, kan de påvirke atferden som måles. Dette kalles *konfundering*²¹.

²⁰ Eng.: external variables

²¹ Eng.: confound

Definisjon 12: *En konfundering er enhver ukontrollert ytre variabel som varierer sammen med den uavhengige variabelen, og som kan utgjøre en alternativ forklaring til resultatene.*

Når et eksperiment har en konfundering, kan resultatene fremstå på grunn av enten den konfunderende variabelen, den uavhengige variabelen, eller en kombinasjon av de to, og det er umulig å finne ut hvilken av de to alternativene som har forårsaket resultatet.

En viktig del av ethvert eksperiment, er å måle atferd som man antar blir påvirket av den uavhengige variabelen. Man måler effekten av å manipulere den uavhengige variabelen. De variablene vi undersøker for å se effekten av endringene, kalles avhengige variabler. Dette leder til følgende definisjon:

Definisjon 13: *Den avhengige variabelen er den variabelen som måles for å se effekten av endringene på den uavhengige variabelen.*

De forskjellige tilstandene av den uavhengige variabelen kalles *behandlinger*:

Definisjon 14: *En behandling er en tilstand av den uavhengige variabelen.*

Behandlingene blir anvendt med kombinasjonen av *objekter* og *deltakere*. Objektet i dette eksperimentet er en prototyp (se seksjon 4.5), og deltakerne er mastergradsstudenter fra Institutt for Informasjons- og Medievitenskap ved Universitetet i Bergen (se seksjon 4.4.4). Hver slik kombinasjon er et *forsøk*:

Definisjon 15: *Et forsøk er en kombinasjon av behandling, deltaker og objekt.*

4.3.1 Valg av deltakere

Valget av deltakere til et eksperiment (utvelgelsen) er viktig, fordi den har mye å si for hvorvidt det er mulig å generalisere resultatene (Wohlin, et al., 2000: 51). For at man skal kunne generalisere resultatene til en ønsket populasjon, må utvalget være representativt for *populasjonen*, det vil si den reelle gruppen av mennesker som man velger deltakere fra.

Størrelsen på utvalget av deltakere har også innvirkning på muligheten for å generalisere. Et større utvalg bedrer grunnlaget for å kunne generalisere resultatene. Variabiliteten i populasjonen er med på å avgjøre hvor stort utvalg man bør ha. Hvis det er stor variabilitet i populasjonen med hensyn til det som skal undersøkes, er det behov for et større utvalg.

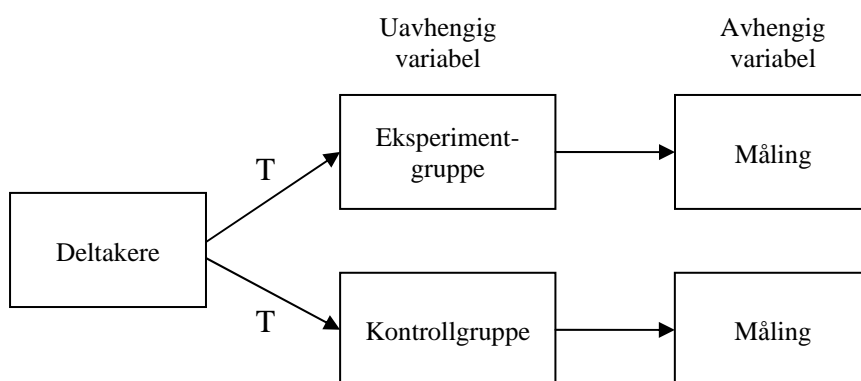
4.3.2 Eksperimentdesign

Det finnes to grunnleggende former for design for enkle eksperimenter: *bare-posttest-design*²², og *pretest-posttest-design*²³ (Cozby, 2007: 151). I begge designene har man en eksperimentgruppe og en kontrollgruppe, og målet er at man skal kunne være sikker på at den eneste forskjellen mellom de to gruppene er den manipulerede variabelen.

De to formene for design, bare-posttest-design, og pretest-posttest-design har bare en forskjell, nemlig at sistnevnte innebærer at man utfører en pretest før den eksperimentelle manipuleringen blir introdusert. På denne måten kan man få rede på om gruppene faktisk var ekvivalente ved eksperimentets start. Så lenge deltakerne blir tilfeldig plassert i grupper, er det imidlertid vanligvis ikke nødvendig med slike tiltak. Hvis man har et stort nok utvalg av deltakere, vil tilfeldig tildeling i grupper føre til grupper som er så godt som identiske på alle måter.

4.3.2.1 Bare-posttest-design

Når man skal utføre et bare-posttest-design, må man (1) skaffe to ekvivalente deltakergrupper, (2) introdusere den uavhengige variabelen, og (3) måle effekten av den uavhengige variabelen på den avhengige variabelen. Figur 4 viser en modell av designet.



Figur 4: Bare-posttest-design

²² Eng.: posttest-only design

²³ Eng.: pretest-posttest design

T-en i figuren betyr at deltakerne blir tilfeldig fordelt i grupper. I eksperimentgruppen har man et nivå av den uavhengige variabelen, mens man har et annet nivå i kontrollgruppen. Effekten av den uavhengige variabelen blir så målt på den samme måten i begge grupper, og man kan dermed bruke målingene til å sammenligne de to gruppene. Forutsetningen om at gruppene er ekvivalente og at man ikke har noen konfunderinger gjør at enhver forskjell mellom de to gruppene skyldes effekten av den uavhengige variabelen.

4.3.3 Forskningsredskaper

Det er tre typer redskaper i et eksperiment: objekter, retningslinjer og måleinstrumenter (Wohlin, et al., 2000: 62). Objekter kan for eksempel være spesifikasjoner eller dokumenter med kode. Retningslinjer brukes for å veilede deltakerne gjennom eksperimentet. Målinger i et eksperiment blir utført gjennom datainnsamling, gjerne gjennom skjemaer eller intervjuer. Man velger redskapene når man planlegger eksperimentet, og utvikler dem for det spesifikke eksperimentet. Målet med forskningsredskapene er å legge til rette for utførelse og overvåkning av eksperimentet.

4.3.4 Statistisk analyse

For å avgjøre hvorvidt resultatene av eksperimentet skyldes et årsak/virkningsforhold, eller tilfeldigheter, må man utføre en statistisk analyse. Basisen for statistisk analyse av et eksperiment er hypotesetesting. En nullhypotese, som erklærer at det ikke finnes noen underliggende trender eller mønstre, fremsettes formelt. Dataene som samles inn gjennom forsøkene brukes til å forkaste nullhypotesen, hvis mulig, og med høyest mulig signifikans.

4.3.5 Trusler mot validitet

Et grunnleggende spørsmål rundt resultatene fra et eksperiment, er validiteten av resultatene. Det er viktig å ta hensyn til validitetsspørsmålet allerede i planleggingsfasen, slik at man kan legge planer for hvordan det er mulig å oppnå tilstrekkelig validitet. Resultatene bør kunne gjelde for hele populasjonen som utvalget er hentet fra. Det kan også være ønskelig å

generalisere resultatene til en bredere populasjon. Resultatene fra et eksperiment kan sies å ha tilstrekkelig validitet hvis de er gyldige for populasjonen vi ønsker å generalisere til.

Cook og Campbell (1979, se Wohlin et al., 2000: 63) definerer fire typer trusler mot validiteten til eksperimentresultater. Disse er *trusler mot konklusjonsvaliditet*, *intern validitet*, *validiteten til tankemodellen*²⁴ og *ekstern validitet*. I de fire neste seksjonene gis en kort forklaring av hver av disse, mens validiteten til dette eksperimentet presenteres i seksjon 4.4.7.

4.3.5.1 Trusler mot konklusjonsvaliditet

Konklusjonsvaliditet handler om relasjonen mellom behandlingen og resultatet av eksperimentet. Denne kan evalueres ved å regne ut den statistiske signifikansverdien til relasjonen. Trusler mot konklusjonsvaliditeten har å gjøre med spørsmål som virker inn på evnen til å trekke riktige konklusjoner om relasjoner mellom behandlingen og resultatet av eksperimentet.

4.3.5.2 Trusler mot intern validitet

Den interne validiteten evalueres ved å undersøke om det er behandlingen som fører til resultatet, og ikke en annen faktor som undersøkeren ikke har kontroll over. Trusler mot den interne validiteten oppstår hvis en kausal relasjon indikeres uten at en slik faktisk finnes.

Modning er et eksempel på en trussel mot intern validitet. Dette er effekten av at deltagere yter annerledes etter hvert som tiden går. De kan for eksempel bli negativt påvirket fordi de synes oppgavene er kjedelige, eller de kan bli positivt påvirket fordi de merker at de lærer noe etter hvert.

4.3.5.3 Trusler mot validiteten til tankemodellen

Validiteten til tankemodellen har å gjøre med relasjonen mellom teori og observasjon. Eksperimentets behandling må reflektere tankemodellen bak årsaken, og resultatet må

²⁴ Eng.: construct validity

reflektere tankemodellen bak virkningen. Trusler mot validiteten til tankemodellen oppstår hvis eksperimentsituasjonen ikke reflekterer tankemodellen som undersøkes.

4.3.5.4 Trusler mot ekstern validitet

Den eksterne validiteten har å gjøre med generalisering, og trusler mot ekstern validitet har å gjøre med evnen til å generalisere resultatet av studiet utenfor rekkevidden til studiet.

4.4 Planlegging av eksperimentet

4.4.1 Kontekstvalg

Konteksten til eksperimentet skiller seg fra konteksten til studiet ved at eksperimentet er utført gjennom laboratorieforsøk, og ikke i studiets reelle kontekst. Oppgavene til deltakerne er ikke reelle i og med at informasjonsbehovet til studentene er konstruert og ikke et ekte informasjonsbehov. Deltakerne er studenter, og har ikke nødvendigvis den samme bakgrunnskunnskapen som de virkelige brukerne av Billedsamlingen har. I tillegg er antallet deltakere i eksperimentet begrenset. En høy grad av kontroll over eksperimentets kontekst fører til at det blir forholdsvis enkelt for andre forskere å reprodusere studiet, men det har også innvirkning på muligheten til å generalisere fra studiet. Dette diskuteres i seksjon 4.4.7.

4.4.2 Hypoteseformulering

Det er viktig å erklære formelt hva som skal evalueres i eksperimentet, og dette gjør man gjennom å formulere en hypotese. Følgende nullhypotese er uformelt fremsatt slik:

Nullhypotese: Innføringen av nye, brukergenererte termer i det kontrollerte vokabularet tilknyttet UBs digitale bildesamling vil ikke føre til forskjell i precision og recall for vanlige spørringer sammenlignet med brukergenererte tagger.

Nullhypotesen skal forsøkes å forkastes til fordel for den alternative hypotesen, som er uformelt fremsatt slik:

Alternativ hypotese: Innføringen av nye, brukergenererte termer i det kontrollerte vokabularet tilknyttet UBs digitale bildesamling vil føre til i en forskjell i precision og recall for vanlige spørringer sammenlignet med brukergenererte tagger.

Basert på denne uformelle beskrivelsen, kan nullhypotesen og den alternative hypotesen formelt formuleres slik:

1. Nullhypotese: $H_0: P_1 = P_2$

Alternativ hypotese: $H_1: P_1 \neq P_2$

2. Nullhypotese: $H_0: R_1 = R_2$

Alternativ hypotese: $H_1: R_1 \neq R_2$

P: Gjennomsnittsverdien til precision

R: Gjennomsnittsverdien til recall

Nødvendige mål: Type brukermedvirkning og precision / recall

Ut fra et statistikkperspektiv er det viktig å avgjøre om man vil utføre ensidige eller tosidige tester på dataene allerede før datainnsamlingen skjer. I dette eksperimentet er det ikke mulig å gjøre noen forhåndsantakelser om hvordan de avhengige variablene, precision og recall, kommer til å forandre seg i forhold til den uavhengige variabelen. Derfor er det en tosidig test som blir den riktige måten å teste dataene på.

4.4.3 Valg av variabler

I dette studiet er det typen brukermedvirkning som er den uavhengige variabelen. Det varieres mellom én type brukermedvirkning i eksperimentgruppen (deltakerne legger til nye termer i det kontrollerte vokabularet) og en annen i kontrollgruppen (deltakerne knytter selvvalgte tagger til bildene). For å finne ut hvilken påvirkning brukermedvirkningen kan ha på precision/recall, sammenlignes precision og recall i henholdsvis eksperimentgruppen og kontrollgruppen. Precision og recall er avhengige variabler i dette studiet, og altså de variablene det er ønskelig å måle. Dette gjøres ved å vurdere hvilke bilder som bør finnes for

hver spørring i forhold til hvilke bilder som faktisk blir funnet, for hver tilstand av den uavhengige variabelen.

4.4.4 Valg av deltakere

Utvelgelsen av deltakere kan baseres på sannsynlighet, eller ikke-sannsynlighet. Forskjellen ligger i hvorvidt sannsynligheten for å velge hver deltaker er kjent eller ikke. I mitt tilfelle er denne sannsynligheten ikke kjent, fordi det ble besluttet å ikke deltakere fra den reelle populasjonen, men i stedet bruke studenter (mer om dette i seksjon 4.4.7). Dette er altså en utvelgelsesmetode som er basert på ikke-sannsynlighet, og som kalles enkelhetsutvelgelse²⁵ (Cozby, 2007).

Det ble besluttet å bruke et utvalg bestående av 20 studenter i eksperimentet, delt inn i en eksperimentgruppe på 10 studenter, og en kontrollgruppe på 10 studenter. Eksperimentgruppen fikk i hovedoppgave å oppdatere det kontrollerte vokabularet, mens kontrollgruppens hovedoppgave var å tagge bildene. Alle eksperimentdeltakerne var mastergradsstudenter ved institutt for Informasjons- og Medievitenskap da forsøkene ble utført, og i hver gruppe var halvparten av deltakerne menn og halvparten kvinner. Grunnen til at det er brukt forskjellige grupper av personer i de forskjellige tilstandene av den uavhengige variabelen i stedet for å eksponere én og samme gruppe for begge tilstandene, er at erfaringene deltakerne gjør seg i den ene tilstanden kan påvirke resultatene i den andre tilstanden. Ved å bruke to forskjellige grupper av deltakere, unngår man dette problemet, men risikerer at forskjeller i gruppene kan påvirke resultatene. For å imøtekomme dette problemet må man gjøre det man kan for at gruppene skal være ekvivalente. Deltakere i dette studiet er tilfeldig utvalgt blant en ganske homogen gruppe; alle er studenter på samme nivå på samme institutt, og bør ha forholdsvis like forutsetninger for å kunne sette seg inn i og utføre oppgavene i forsøkene. Gjennomsnittsalderen er 26,1 år i eksperimentgruppen og 25,5 år i kontrollgruppen.

I forkant av de faktiske forsøkene, ble det utført to pilotforsøk, med deltakere som ikke skulle være med i selve eksperimentet. Før disse forsøkene ble utført, var eksperimentet nøye

²⁵ Eng.: convenience sampling

planlagt, men det ble likevel besluttet å utføre pilotforsøk for å avklare eventuelle feil eller problemer som ville kunne føre til sammenbrudd i eksperimentprosessen. Gjennom pilotforsøkene ble det avdekket småproblemer som førte til at det ble gjort mindre justeringer på retningslinjene før de faktiske forsøkene ble utført. Etter pilotforsøkene var det også mulig å anslå hvor lenge hvert forsøk ville vare, slik at deltakerne kunne opplyses om dette på forhånd.

4.4.5 Eksperimentdesign

Det ble besluttet å bruke et bare-posttest-design, selv om det er vanskelig å si om utvalget i dette eksperimentet er stort nok til å forsvare bruken denne designformen. Det ville ikke være mulig å bruke et pretest-posttest-design, på grunn av måten det ville påvirket deltakerne. Ved å utføre en pretest, ville deltakerne blitt kjent med de eksisterende metadataene i billedatabasen, noe som ville gjort det umulig å si noe om hvordan brukere som er ukjent med billedatabasen søker etter og indekserer bilder. Et bare-posttest-design ble derfor valgt.

Deltakerne ble fordelt tilfeldig mellom eksperimentgruppen og kontrollgruppen, men gruppene ble balansert med tanke på kjønn, slik at det var fem menn og fem kvinner i hver gruppe. Det ble brukt to forskjellige behandlinger: (1) brukervedvirkning basert på oppdatering av det kontrollerte vokabularet (eksperimentgruppen), og (2) brukervedvirkning basert på tagging av bilder (kontrollgruppen). Forskjellen mellom behandlingene er representert gjennom en liten forskjell i brukergrensesnittet som blir brukt av deltakerne for å legge inn data, samt en forskjell i oppgavene som ble utført av deltakerne i de forskjellige gruppene. Mens eksperimentgruppen hadde et grensesnitt for å legge til termer i det kontrollerte vokabularet, hadde kontrollgruppen et grensesnitt for å legge til tagger. Utenom dette er prototypen helt lik for de to gruppene.

4.4.6 Forskningsredskaper

Før forsøkene ble iverksatt, ble det laget retningslinjer som beskrev oppgavene som deltakerne skulle utføre. Det ble laget ett sett med retningslinjer for eksperimentgruppen, og ett for kontrollgruppen. Deltakerne fikk også en kort innføring, både muntlig og skriftlig, i henholdsvis (1) kontrollerte vokabularer og hvilke termtyper eksperimentgruppen skulle legge til

(appendiks c), og (2) tagger og hvordan kontrollgruppen skulle gå frem for å tagge bildene (appendiks d).

For å samle inn data, ble det utviklet en prototyp med funksjonalitet for å søke etter bilder. Denne prototypen er eksperimentobjekt i dette studiet. Hvert enkelt søk som ble gjort i prototypen ble også registrert. Prototypen kan dermed også regnes som måleinstrument. Prototypen, og hvordan den er utviklet og hvilken funksjonalitet den inneholder, er nærmere beskrevet i en egen seksjon (4.5).

4.4.7 Evaluering av validiteten i eksperimentet

4.4.7.1 Evaluering av konklusjonsvaliditet

Gruppestørrelsen, som i dette studiet er relativt liten, påvirker konklusjonsvaliditeten, fordi en liten gruppestørrelse reduserer holdbarheten til de statistiske testene. For å redusere ressursbruken, ble det likevel besluttet å operere med små grupper.

Konklusjonsvaliditeten er også avhengig av at alle forsøkene er mest mulig likt utført. En faktor som kan påvirke dette, er uventede hendelser under forsøkene som fører til avbrudd. Dette var heldigvis ikke et problem under utførelsen av forsøkene i dette studiet, og det forventes ikke at konklusjonsvaliditeten i nevneverdig grad har blitt påvirket av slike avbrudd. Hvert forsøk har også vært ganske likt gjennomført ved å bruke rutiner som var fastlagt på forhånd og utprøvd i pilotforsøk, og dermed like i alle forsøkene.

Validiteten til et eksperiment er avhengig av påliteligheten til målingene. Disse avhenger av mange forskjellige faktorer, som hvordan oppgavene er formulert, og hvordan forskningsredskapene er valgt ut og designet. Basisprinsippet er at når man måler et fenomen to ganger, skal resultatet bli det samme. I dette studiet blir det gjort en subjektiv menneskelig bedømmelse for å avgjøre hvilke bilder som bør regnes som relevante, en bedømmelse som blir brukt i utregningen av målingene. Dette kan potensielt føre til problemer, fordi det i noen tilfeller kan diskuteres om et bilde bør regnes som relevant eller ikke. I de aller fleste tilfellene er det imidlertid ganske klart hvilke bilder som kan anses å være relevante for hver spørring. I tillegg er vurderingen av hvilke bilder som er relevante den samme for eksperimentgruppen og kontrollgruppen.

Graden av heterogenitet blant deltakerne kan påvirke konklusjonsvaliditeten og den interne validiteten. Hvis deltakermassen er veldig heterogen, risikerer man at variasjoner på grunn av individuelle forskjeller overskygger variasjoner på grunn av behandlingen. I dette

studiet har det blitt brukt en ganske homogen deltakermasse, med studenter med relativt jevn utdanning og alder. Dette er en fordel for konklusjonsvaliditeten, fordi det reduserer individuelle forskjeller.

I dette studiet dukket det dessverre opp et uforutsett problem som har konsekvenser for konklusjonsvaliditeten. Dette problemet, og hvordan det er blitt tatt hånd om, blir nærmere beskrevet i seksjon 5.1.

4.4.7.2 Evaluering av intern validitet

I dette studiet var det en viss fare for at deltakerne i de to gruppene snakket sammen om eksperimentet mellom forsøkene i eksperimentgruppen og forsøkene i kontrollgruppen, siden alle deltakerne studerte ved samme institutt. Oppgavene til de to gruppene var likevel så forskjellige at det er liten sannsynlighet for at dette skulle kunne ha noen innvirkning på resultatene. Trusler mot intern validitet er hovedsakelig et problem når forsøkene foregår over en lengre tidsperiode enn det som var tilfelle i dette studiets kontrollerte og kortvarige laboratorieforsøk.

4.4.7.3 Evaluering av validiteten til tankemodellen

Hvis eksperimentet har en enkelt uavhengig variabel, deltaker eller behandling, er det fare for at eksperimentet underrepresenterer tankemodellen, og dermed ikke gir et komplett bilde av teorien. I dette studiet benyttes bare én uavhengig variabel, og man kan derfor si at årsaks-tankemodellen er underrepresentert.

Det er alltid fare for at deltakerne prøver å finne ut hensikten og det tilsiktede resultatet av eksperimentet, og baserer atferden sin på at de er deltakere i et eksperiment og på hva de gjetter om hypotesen. Dette kan være i positiv eller negativ forstand, avhengig av deltakernes holdning til det de gjetter om hypotesen.

En annen faktor som kan ha hatt en del innvirkning på dette studiet, er interaksjon mellom testing og behandling. Testingen i seg selv kan gjøre deltakerne mer følsomme eller mottakelige for behandlingen. Testingen blir på denne måten en del av behandlingen. Det er sannsynlig at noen av deltakerne i eksperimentet har følt at det er viktig å foreslå flest mulig nye termer, og at de gjerne vil være den av deltakerne som foreslår flest nye termer. Dette kan ha ført til at det ble lagt til et større antall termer per deltaker enn hva som er realistisk i en reell kontekst. På den andre siden kan man si at dette er oppførsel som også kunne funnet sted

i en reell situasjon, og det er mulig å stimulere brukere til slik atferd, for eksempel med system hvor brukerne får poeng for gode bidrag.

Et kritisk spørsmål i dette studiet, er om precision og recall er formålstjenlige måleenheter for det vi ønsker å undersøke. Precision og recall er mål for relevansen til dokumentene som blir gjenfunnet. I studiet ønsker vi å undersøke påvirkningen av to forskjellige typer brukermedvirkning på gjenfinningsrelevansen. Det finnes andre måter å måle relevans på, men precision og recall er mye brukt og blir ansett som gode metoder for dette formålet.

4.4.7.4 Evaluering av ekstern validitet

Den homogene gruppen av deltakere i dette studiet er, som nevnt, med på å bedre konklusjonsvaliditeten. Bruk av en homogen gruppe er imidlertid negativt for den eksterne validiteten, fordi gruppen ikke er valgt fra en generell nok populasjon. Dette er altså med på å begrense evnen til å generalisere resultatene fra eksperimentet til en reell kontekst. At deltakerne ikke er faktiske brukere av Billedsamlingen, er også negativt for den eksterne validiteten. I dette studiets eksperiment har jeg valgt å bruke mastergradsstudenter ved Institutt for Informasjons- og Medievitenskap ved Universitetet i Bergen. Deltakerne er altså ikke valgt fra den reelle populasjonen (selv om det er mulig at det var reelle brukere blant deltakerne). Dette kan være problematisk med tanke på ekstern validitet (Goodwin, 2005). Studenter på dette nivået er for eksempel intelligente og har høy kognitiv kapasitet, noe som kan føre til at den generelle populasjonen kanskje vil kunne ha større vansker med å sette seg inn i tilsvarende oppgaver. Det ideelle med tanke på ekstern validitet ville kanskje vært å bruke et tilfeldig utvalg av deltakere fra populasjonen. Ressursene i dette studiet er imidlertid begrenset, så for å unngå å bruke for mye tid på å skaffe deltakere som er mindre tilgjengelige på grunn av faktorer som geografisk spredning og tidsmangel, ble det besluttet å bruke studenter fra instituttet.

I dette studiet er objektet en prototyp med et spesifikt kunnskapsorganiseringssystem, som ikke har blitt brukt i en reell sammenheng. Strukturene som er brukt for å bygge opp kunnskapsorganiseringssystemet, er likevel basert på velutviklede standarder, og brukes i mange andre sammenhenger. Det kan derfor være mulig å generalisere resultatene fra eksperimentet til andre tilsvarende grupper som får lignende oppgaver med den samme typen applikasjon i en annen digital bildesamling.

4.5 Design og utvikling av prototypen

Et digitalt informasjonssystem er et godt verktøy for å legge til rette for brukerbidrag, hvis det er laget på en hensiktsmessig måte. Verdensveven har egenskaper som gjør at brukerne raskt og enkelt både kan få tak i og bidra med relevant informasjon. ABM-sektoren har fått øynene opp for disse mulighetene, og det foregår mye arbeid innenfor denne sektoren for å undersøke hvordan man på best mulig måte kan utnytte de mulighetene verdensveven gir for å gjøre sektoren bedre.

Det ble besluttet at en vevbasert prototyp ville være et passende forskningsredskap for eksperimentet. Med tanke på at ABM-sektoren har fokus på spredning og brukermedvirkning (se seksjon 2.1), og at vevapplikasjoner har gode egenskaper for slike oppgaver, ble denne typen applikasjon vurdert å være passende for studiet. Vevbaserte applikasjoner er dessuten svært utbredt, og noe de fleste er vant til å bruke, både i arbeid og privat.

Prototypen er bygget opp som et enkelt bildegjenfinningssystem med en versjon som har funksjonalitet for å legge til termer i et kontrollert vokabular, og en versjon som har bilde-taggingfunksjonalitet. Bildegjenfinningssystemet er forsøkt bygd opp slik at det skal gi et naturlig bilde av hva brukerne kan forvente å finne hvis de skal lete etter bilder på verdensveven. Et informasjonssystem er gunstig fordi det gjør det mulig å samle inn mye data over en relativ kort tidsperiode. I tillegg til å registrere informasjon om termer og tagger, registrerer også prototypen hva deltakerne søker på. Etter at forsøkene er ferdige, skal altså databasen inneholde alle verdiene som er nødvendig for å kunne teste hypotesene. For å spare tid på utviklingen, inneholder prototypen minst mulig funksjonalitet utover det som er strengt nødvendig for å besvare problemstillingen og evaluere hypotesene.

4.5.1 Kontekst

I dette studiet tar jeg utgangspunkt i en bildesamling av digitale bilder fra ca. 1860 til 1970 som er indeksert med forskjellige typer emneord. Bildene er produsert av profesjonelle fotografer og amatørfotografer, og har blitt digitalisert og indeksert av UB. Dette innebærer at det finnes en del metadata knyttet til hvert bilde, og det er mulig å søke i bildene basert på disse. Jeg har fått tilgang til en kopi av bildedatabasen, med alle metadata som finnes til hvert bilde. Disse metadataene er med i grunnlaget for søkene som blir gjort for å regne ut precision og recall både i eksperimentgruppen og kontrollgruppen.

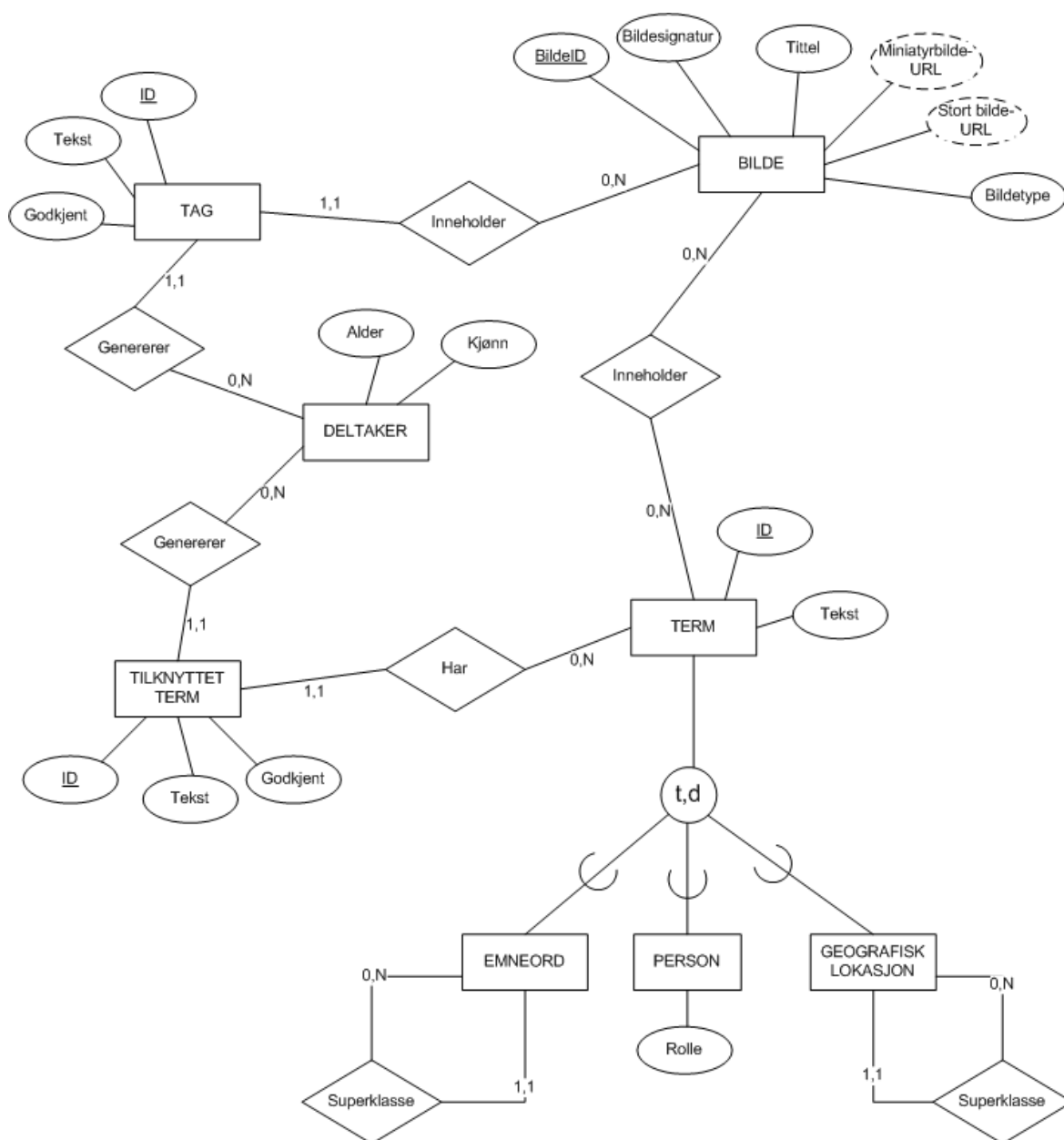
Ikke alle disse metadataene er ment å skulle være allment tilgjengelig. En del av metadatagrunnlaget er til bruk internt i UB. Jeg fokuserer bare på de metadataene som er ment å skulle brukes for å gjenfinne bilder, med et par unntak. Følgende felter er ment å være allment tilgjengelige i databasen:

- **Signatur.** Hvert bilde har en signatur som er bygget opp etter et fast system, hvor første delen av signaturen består av en kode som er basert på hvilken samling bildet tilhører.
- **Tittel.**
- **Dato.** I mange tilfeller er det usikkert når bildet ble tatt. Bildene har derfor en fra-dato og en til-dato for å plassere dem innenfor en begrenset tidsperiode. I prototypen tas det ikke hensyn til årstall.
- **Bildetekst.** Noen av bildene har en kort tekstlig beskrivelse av hva man kan se på bildet. Dette feltet er ikke tatt med i prototypen, fordi det ofte inneholder upresis informasjon som kan være misvisende.
- **Person/firma.** Forteller om hvilke personer som er tilknyttet bildet, enten som fotograf, giver av bildet til Billedsamlingen, avbildet eller andre roller.
- **Bildetype.** Et bilde kan ha en eller flere bildetyper. Bildetyper kan for eksempel være brudebilde, flyfoto, gruppebilde eller postkort.
- **Emneord.** Objekter.
- **Geografiske lokasjoner.** Lokasjonsinformasjon.

De fleste av disse feltene blir også brukt i prototypen, slik at det også er mulig å finne bilder før deltakerne har lagt til informasjon. Termene deltakerne i eksperimentgruppen legger til, blir knyttet til termer fra feltene person/firma, emneord eller geografiske lokasjoner.

4.5.2 EER-diagram

I tillegg til de eksisterende strukturene i bildedatabasen, er det lagt til en del strukturer som er nødvendige for å skape funksjonaliteten som skal testes og brukes som sammenligningsgrunnlag i dette studiet. En tabell for å registrere informanter, samt tabeller for tagger og tilknyttede termer. Følgende EER-diagram (Elmasri & Navathe, 2004) gir en oversikt over hvordan de nye strukturene fungerer sammen med de eksisterende.



Figur 5: EER-diagram over den delen av bildedatabasen som er aktuell i dette studiet.

Bilde og *term*(*emneord*, *person* og *geografisk lokasjon*) er entiteter som fantes i bildedatabasen fra før. I tillegg til en unik ID, har mange av bildene en unik signatur som er bygd opp basert på blant annet hvilken samling *bildet* kommer fra. *Bilde*-entiteten kan også ha en tittel og en bildetype, samt to forskjellige URLer som peker til bildefilenes lokasjon på serveren. En *term* er enten av typen *emneord*, *person* eller *geografisk lokasjon*, og har en tekstlig beskrivelse. *Person*-entiteten har i tillegg en beskrivelse av hvilke rolle personen har. *Emneord* og *geografisk lokasjon* er bygd opp som et hierarki hvor hver entitet har en og bare en forelder av samme entitetstype, men kan ha flere barn av samme entitetstype.

I tillegg har det blitt lagt til tre entiteter; *deltaker*, *tag* og *tilknyttet term*. *Deltaker* er et register over deltakerne i eksperimentet, hvor alder og kjønn blir registrert. En *tag* blir generert av en *deltaker* (i kontrollgruppen), og er knyttet til et *bilde*. En *tag* inneholder en tekstlig beskrivelse, og en boolsk verdi som sier om *taggen* er godkjent eller ikke. Denne boolske verdien blir brukt for å aktivere eller deaktivere tagger i forbindelse med utregning av precision og recall. En *deltaker* kan generere mange *tagger*, og et *bilde* kan ha mange *tagger*. *Deltakere* i eksperimentgruppen kan generere tilknyttede *termene*. Disse tilknyttede *termene* inneholder en tekstlig beskrivelse og en boolsk verdi som blir brukt på samme måte som den boolske verdien i *tag*-entiteten. *Tilknyttet term*-entiteten er ikke knyttet til *bilder*, men til *termene* som fantes i databasen fra før. Disse *termene* er i sin tur knyttet til *bildene*.

4.5.3 Valg av teknologi

Prototypen er utviklet som en dynamisk vevapplikasjon, med PHP/MySQL i bunnen og et brukergrensesnitt utviklet med HTML/CSS og JavaScript. Disse teknologiene ble valgt fordi de er gratis å bruke og lett tilgjengelig, og fordi jeg på forhånd hadde god kjennskap til teknologiene. Systemet skal gjøre det enkelt å søke og navigere i bildene, samt å bidra med nye vokabularertermer eller tagger.

For å kutte ned på utviklingstiden, har det ikke vært fokus på at prototypen skal være kompatibel med flest mulig forskjellige vevlesere. Prototypen er hovedsakelig testet i vevleseren Mozilla Firefox. Dette er helt uproblematisk i forhold til utviklingsmodellen som er brukt, hvor fokuset er på å skape en prototyp som skal fungere, men som ikke skal utvikles til et ferdig produkt. Utviklingsmodellen er beskrevet i seksjon 4.5.5, og i seksjon 4.5.10 tar jeg for meg andre begrensninger i prototypen.

4.5.4 Utviklingsparadigmer

Sommerville (2007: 65) nevner tre «paradigmer» for utvikling; generiske modeller som kan brukes for å forklare forskjellige tilnæringsmåter til programvareutvikling. Disse er:

1. *Fossefallmodellen*. Tar for seg de fundamentale prosessaktivitetene spesifikasjon, utvikling, validering og tilpasning og representerer dem som separate prosessfaser,

som for eksempel kravspesifikasjon, programvaredesign, implementering, testing og så videre.

2. *Evolusjonær utvikling.* Spesifikasjon, utvikling og validering blir gjort i iterasjoner. En tidlig versjon av et program utvikles fra abstrakte spesifikasjoner, og blir så forbedret gjennom gjentatt interaksjon med kunden, for å skape et system som dekker kundens behov.
3. *Komponentbasert programvareutvikling.* Er basert på at det eksisterer et stort antall gjenbrukbare komponenter. Programutviklingsprosessen fokuserer på å integrere disse komponentene til et system i stedet for å utvikle dem fra grunnen av.

For disse tre generiske prosessmodellene, er det ikke slik at den ene utelukker den andre. De brukes ofte sammen, men dette er hovedsakelig i utviklingen av store systemer. I utviklingen av prototypen i dette studiet kan det meste plasseres innenfor evolusjonær utvikling. En av fordelene med denne prosessmodellen i forhold til fossefallsmodellen er at den ofte er mer effektiv når man skal utvikle systemer som dekker de umiddelbare behovene til kunden. Det å kunne demonstrere funksjonalitet slik man kan med en prototyp, gjør det enkelt for kunde og utvikler å forstå hverandre.

Sommerville (2007) nevner to problemer med evolusjonær prototyping, sett fra et utviklings- og ledelsesperspektiv:

1. *Prosessen er ikke synlig.* Ledere trenger regelmessige leveringer for å måle fremgang. Det er tidkrevende og lite kostnadseffektivt å produsere dokumentasjon som reflekterer hver versjon av systemet.
2. *Systemer blir ofte dårlig konstruert.* Kontinuerlige endringer kan føre til at strukturen blir dårlig. Det blir stadig vanskeligere å gjøre endringer i programvaren etter hvert som systemet blir større og mer komplekst.

Når det gjelder førstnevnte problem, har det ikke vært et stort behov for å rapportere fremgang til UB, siden de ikke har store utgifter med utviklingen, og siden programmet som utvikles ikke er virksomhetskritisk ettersom UB allerede har et fungerende vevgrensesnitt for bildesøk. Det har vært tilstrekkelig med de tilbakemeldingene som har blitt gitt gjennom de mange møtene i løpet av utviklingsperioden. Når det gjelder det andre problemet, er det nok tilfelle at koden og arkitekturen til prototypen bærer preg av de mange endringene. Det er for

eksempel lagt lite vekt på å lage et system som er effektivt med tanke på eksekveringstid. Det er likevel nærliggende å påpeke at prototypen fungerer til formålet den er laget for: å besvare problemstillingen i studiet. Som vi skal se i neste seksjon, er det heller ikke meningen at prototypen skal utvikles videre til et ferdig system, så kravene til struktur og effektivitet er mindre enn i mange andre utviklingsmodeller.

4.5.5 Bruk-og-kast-prototyping²⁶

Utviklingen av prototypen har blitt gjort etter prosessmodellen bruk-og-kast-prototyping (Bell, 2005: 305). Denne modellen er en type evolusjonær utvikling hvor man lager prototyper som blir vurdert sammen med kunden, og brukt som et verktøy i kommunikasjonen mellom utvikler og kunde. Produktet av en bruk-og-kast-prototyping-prosess er en spesifisering for et ferdig system, og selve prototypen blir ikke implementert til et ferdig system. Fordelen med å bruke denne prosessmodellen er at man raskt kan utvikle den funksjonaliteten man trenger, fordi man unngår å bruke tid på å optimalisere arkitekturen og ytelsen til systemet. Prototypen er ment å skulle demonstrere de essensielle funksjonene til systemet.

Prototypen som er utviklet i forbindelse med dette studiet, er laget i samarbeid med UB, som er kunden, og med innspill fra veileder og andre systemutviklere. I begynnelsen av utviklingsprosessen forelå det et abstrakt behov fra UB sin side, som jeg delvis fikk være med på å forme ut fra mine behov om å finne en relevant problemstilling som kunne brukes i masterstudiet. Det ble holdt et møte med en representant fra UB og meg selv, i tillegg til veileder og to andre studenter som holdt på med andre prosjekter for UB. På dette møtet ble det laget en tentativ kravspesifisering hvor det som da ble ansett å være de viktigste funksjonene i systemet ble tatt med. Utviklingsarbeidet ble så igangsatt, og etter noen uker forelå den første versjonen av prototypen. Denne versjonen ble vist frem på et nytt møte med de samme deltakerne som på det første møtet. Basert på den funksjonaliteten som ble vist frem her, ble det så utvekslet kommentarer, og justeringer i kravspesifiseringen ble foretatt. Disse stegene –

²⁶ Eng.: throwaway prototyping

spesifiserings-/valideringsmøter og utviklingsperioder ble gjentatt noen ganger, før prototypen ble demonstrert for de fleste ansatte ved Billedsamlingen. På dette møtet kom det frem relativt mange forslag til endringer, og en ny utviklingsperiode ble igangsatt. Etter et siste møte med UBs representant, veileder og meg selv, ble det så avgjort at prototypen inneholdt tilstrekkelig funksjonalitet for å kunne brukes i studiet.

4.5.6 Kontrollert vokabular

Da utviklingen av prototypen begynte forelå det et kontrollert vokabular for Billedsamlingen, med blant annet en tabell for emneord (substantiver) og en tabell for geografiske lokasjoner (stednavn). Disse tabellene hadde en hierarkisk oppbygning (BT/NT-relasjoner), men ingen ekvivalensrelasjoner (UF/USE) eller assosiative relasjoner (RT) (se seksjon 3.5.3). Tabellene kan dermed ikke kalles taksonomier. Strukturen på det kontrollerte vokabularet ble forandret i forbindelse med dette studiet, slik at det også skulle være mulig å legge til ekvivalensrelasjoner mellom termer. Person-tabellen fikk også denne muligheten, slik at det ble mulig å legge til bredere termer (for eksempel tittel) og synonymer (for eksempel kallenavn) også til personer. Deltakerne fikk, for enkelhets skyld, bare ett enkelt inputfelt for å knytte ord til eksisterende termer. Dette inputfeltet skulle de bruke for å legge til ordene sine, uavhengig om de ønsket å legge til en bredere term (BT) eller et synonym. Deltakerne slapp på denne måten å ta stilling til om det ordet de skulle legge til var et synonym eller en bredere term, og de slapp også å plassere ordet sitt på riktig plass i hierarkiet. Eksisterende termer, som er lagt inn av ansatte ved UB, blir på denne måten alltid foretrukket term eller bredere term i forhold til det deltakerne i eksperimentgruppen legger inn.

I prototypen har jeg valgt å videreutvikle UBs eksisterende kontrollerte vokabular med en del nye strukturer, men med noen begrensninger i relasjonstyper. Disse begrensningene har jeg valgt å gjøre for at det skal være enklest mulig for brukere uten kjennskap til oppbygningen av kontrollerte vokabularer å kunne være med å bidra. Det er i tillegg ønskelig for UB å ha en stor grad av kontroll over innholdet, til tross for at brukerne i større grad slipper til. For å gjøre registreringen enklest mulig for deltakerne, kan de kun angi broader term-relasjoner eller ekvivalensrelasjoner til termer som allerede finnes i det kontrollerte vokabularet. Deltakerne kan ikke plassere nye termer på riktig sted i hierarkiet, men kan bare knytte nye termer til de foretrukne termene som allerede finnes. For å unngå redundans, vil et

ord som allerede finnes som stamfar til det valgte ordet heller ikke kunne legges inn som tilknyttet term av deltakerne. Slike ord vil uansett allerede være del av søkegrunnet.

4.5.7 Søkealgoritmer

Når en deltaker skriver inn en spørring i søkefeltet og sender den, går spørringen gjennom en prosess som består av flere steg, hvor enkeltord blir skilt ut, tilknyttede termer blir funnet (hvis en bruker skriver flere enn ett ord, vil eventuelle relevante bilder måtte inneholde begge eller alle ordene eller varianter av disse), og en SQL-spørring blir konstruert og kjørt mot databasen. Resultatene fra spørringen blir så prosessert, og det lages en liste over treff fra databasen.

4.5.7.1 Eliminering av stoppord

Det benyttes ingen eliminering av stoppord i prototypen. Dette kunne enkelt ha vært implementert, og ville i en del tilfeller ført til høyere recall på deltakernes spørringer, for eksempel i et tilfelle der en deltaker skrev spørringen «kvinner i bunad». Preposisjonen *i* i denne spørringen skaper problemer, fordi noen av bildene som viser kvinner i bunad ikke inneholder bokstaven *i* som en indekstern. Hvis det man søker etter består av stoppord, kan eliminering av stoppord skape et nytt problem. Et eksempel er ordet ved, som kan være både en preposisjon og et substantiv. Ved å gjøre alle preposisjoner om til stoppord, vil det bli vanskelig å søke etter objektet ved. Det ble besluttet å ikke bruke eliminering av stoppord i prototypen, for å unngå å bruke for mye tid på slike spørsmål.

4.5.7.2 Grunnstammesøk

UB har ikke vært konsekvent i bruken av entalls- og flertallsformer når de har bygget opp det kontrollerte vokabularet sitt. I noen tilfeller har de brukt flertallsendelser (roser, epler, møbler), mens de i noen tilfeller har brukt entallsendelser (salong, pelslue, due). Dette vil også brukere som skal legge til termer i det kontrollerte vokabularet eller tagge bilder gjøre. I tillegg skriver brukerne både entallsendelser og flertallsendelser når de søker etter bilder. Dette blir i stor grad tatt hånd om i prototypen ved å bruke en enkel form for grunnstammesøk (se seksjon 3.3.1.3). Løsningen fjerner blant annet entallsendelser (-et, -en, -a) og flertallsendelser (-ene, -er, -ar, -a) på bokmål og nynorsk fra søkeordene som brukerne skriver. Søker

man på et flertallsord, som for eksempel duer, vil systemet også søke etter due. Søker man etter biler, vil systemet også søke etter bil. Dette virker også motsatt vei – søker man på entallsord, får man treff selv om ordet ligger i det kontrollerte vokabularet som et flertallsord. I noen tilfeller vil denne metoden ikke fungere – hvis man for eksempel søker på ordet menn, vil ikke metoden klare å finne stammen, som er mann. Men i veldig mange tilfeller vil grunnstammesøket hjelpe brukeren med å finne flere relevante treff enn uten.

4.5.7.3 Spøringer som inneholder deler av ord fra indeksen

På det vevstedet som finnes for Billedsamlingen i dag, vil man få treff hvis spørringen man skriver er del av et av ordene det søkes i. Eksempler på dette er bil, som gir treff på bl.a. varebil og personbil, og hus, som gir treff på bl.a. bolighus, stovehus og høyhus. Dette er en enkel måte å legge til rette for at man kan gjenfinne objekter med smalere hierarkisk nivå enn det man søker etter. Medaljen har imidlertid en bakside, og det er at brukerne også risikerer å finne mange irrelevante bilder når denne metoden blir brukt. I eksemplene over vil man for eksempel også få frem bilder med termene **Gatebilde**, **Oversiktsbilde**, **Biljardbord**, **Huske** og **Husarbeid**. Dette kan i enkelte tilfeller ha en svært negativ innvirkning på precision. I prototypen er det derfor ikke lagt inn mulighet for å få treff på deler av ord. De nye strukturene vil legge forholdene til rette for at man skal kunne oppnå de samme fordelene, uten at det går ut over precision.

4.5.8 Beskrivelse av funksjonaliteten i prototypen

Prototypen er laget i to versjoner: en for eksperimentgruppen, og en for kontrollgruppen. Disse to versjonene er svært like, men med en viktig forskjell som gjenspeiler forskjellen mellom de respektive oppgavene til eksperimentgruppen og kontrollgruppen (eksperimentgruppen skal legge til termer, mens kontrollgruppen skal tagge). Hver versjon har to hovedfunksjoner, hvor den ene er lik i de to versjonene, mens den andre er forskjellig. Funksjonen som er felles for begge versjonene er bildesøkingsfunksjonalitet (begrenset til de 10 bildene i testsamlingen). I tillegg er det en funksjon som er forskjellig i de to versjonene: I eksperimentgruppeversjonen har deltakerne mulighet til å legge inn nye termer i det kontrollerte vokabularet. Termene må knyttes til en eksisterende term som på forhånd er knyttet til bildene av UB. I kontrollgruppeversjonen har deltakerne ikke denne funksjonen, men kan i stedet knytte et arbitrært antall tagger til hvert enkelt bilde. Mellom hvert forsøk ble

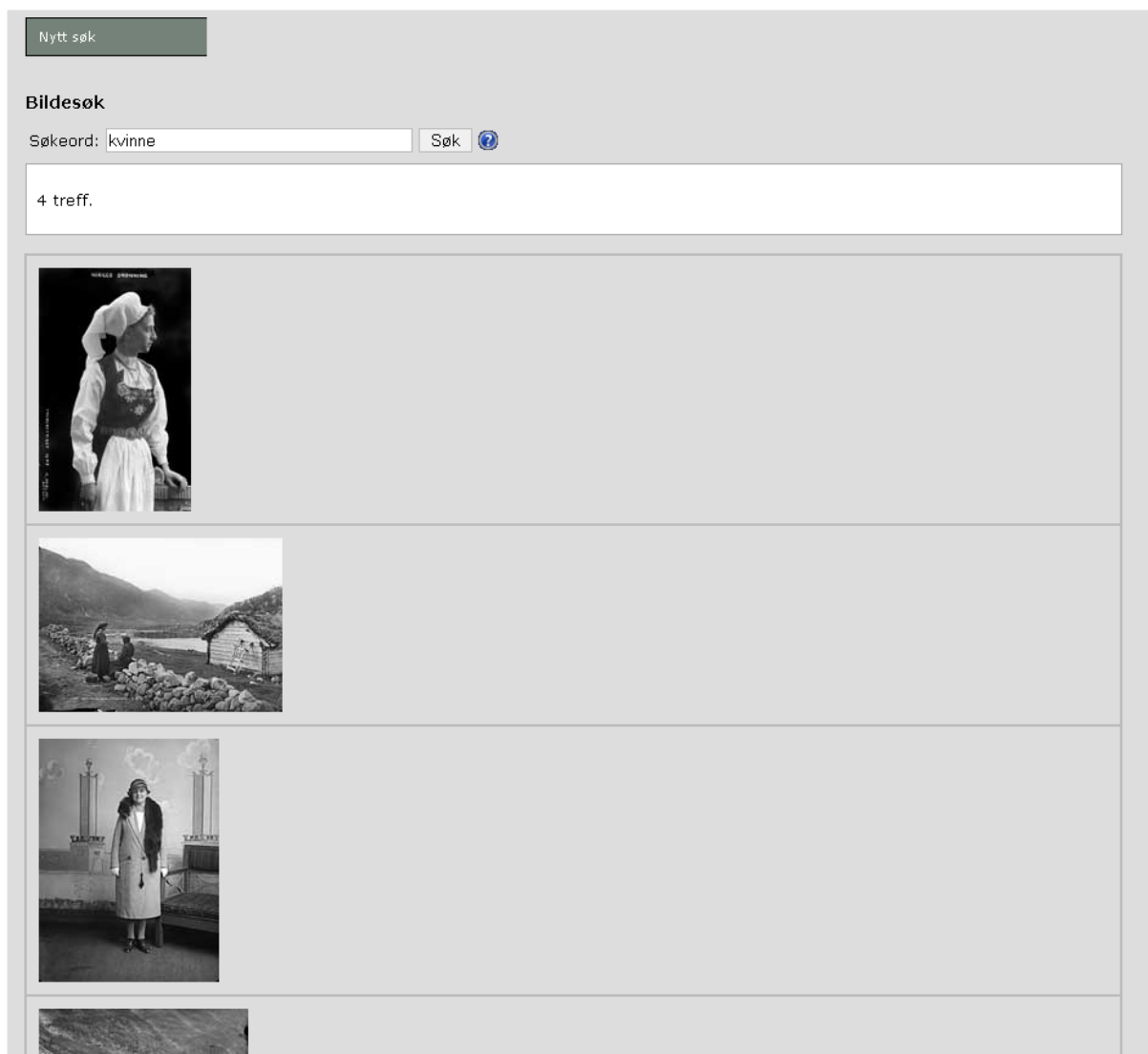
alle termer og tagger deaktivert, slik at deltakerne ikke kunne se hva tidligere deltakere hadde gjort. Tidligere aktivitet hadde heller ingen innvirkning på søkefunksjonaliteten til deltakerne – alle deltakernes søk var basert utelukkende på UBs metadata.

Søkegrensesnittet består av ett enkelt søkefelt, og en knapp for å utføre søket, samt en knapp for å starte et nytt søk. Det finnes også en knapp som bringer brukeren til en hjelpetekst som forklarer hva som blir vist på skjermen. Se figur 6.



Figur 6: Søkemenyen i prototypen

Resultatene av et bildesøk blir vist med miniatyrversjoner av hvert bilde som blir funnet (se figur 7). På denne måten får deltakerne en oversikt over hvilke bilder de finner etter hvert som de søker på nye ord. Ingen annen informasjon enn selve bildene ble vist for deltakerne, slik at deltakerne ikke skulle bli påvirket av de eksisterende metadataene når de valgte spørringer.



Figur 7: Presentasjon av resultatene fra et bildesøk

Funksjonaliteten som er beskrevet så langt, er den som ble brukt i oppgave 1 i forsøkene (se appendiks c). Resultatene fra denne delen av forsøkene, blir, som vi skal se i neste kapittel, brukt til å lage et sett av spøringer, hvor hver spørring blir vurdert i forhold til precision og recall.

Oppgave 2 var forskjellig for de to gruppene. Deltakerne i eksperimentgruppen skulle legge til termer i det kontrollerte vokabularet, mens deltakerne i kontrollgruppen skulle tagge bildene. Deltakerne i begge gruppene fikk se ett og ett bilde i bilde i prototypen, sammen med metadataene som UB har knyttet til bildet. Figur 8 viser eksperimentgruppeversjonen, og figur 9 viser kontrollgruppeversjonen.


Nytt søk


Parti fra Hausvigen, Osterøen ved Bergen
 UBB-KK-1318-0445

Fotograf: Knudsen, Knud
 01.01.1870 - 31.12.1871

Emneord:
[Stovehus](#), [Naust](#), [Fjord](#), [Dampskip](#), [Robåt](#)

Steder:
[Osterøy](#), ["Bergen og omegn"](#)





Figur 8: Presentasjon av metadata for et bilde (eksperimentversjonen)

Nytt søk

Parti fra Hausvigen, Osterøen ved Bergen
 UBB-KK-1318-0445

Fotograf: Knudsen, Knud
 01.01.1870 - 31.12.1871

Emneord:
[Stovehus](#), [Naust](#), [Fjord](#), [Dampskip](#), [Robåt](#)

Steder:
[Osterøy](#), ["Bergen og omegn"](#)

Tagger

Legg til tag 



Figur 9: Presentasjon av metadata for et bilde (kontrollversjonen)

Ved hjelp av dette grensesnittet, kunne deltakerne se om de kunne identifisere synonymer eller bredere termer til noen av de eksisterende termene. Figur 10 viser grensesnittet for å legge inn en ekvivalent eller bredere term for termen naust. For å komme inn på dette grensesnittet, har deltakeren klikket på termen naust under emneord i figur 8. Deltakeren skriver inn ordet nøst, som er synonymt med naust, og ordet naust blir lagt inn i databasen som en *til-*

knyttet term til termen naust. Dette innebærer at, hvis den tilknyttede termen er aktivert, vil et søk på nøst gi treff på alle bilder som UB har tilknyttet termen naust.

Hvis deltakeren hadde ønsket å legge til en bredere term til termen naust, ville fremgangsmåten vært nøyaktig lik. Grensesnittet i figur 10 brukes til å legge inn både ekvivalente termer og bredere termer. Deltakeren kunne for eksempel skrevet inn hus, og dermed lagt inn en bredere term til termen naust. Alle termer som blir lagt inn i dette grensesnittet, blir tilknyttede termer til hovedtermen (naust), og blir brukt på samme måte når man skal søke etter bilder: hvis en deltaker søker etter et ord som er en tilknyttet term, søker systemet også etter hovedtermen. Det vil si at et søk på nøst vil føre til at bilder som er indeksert med naust vil bli funnet, og det samme gjelder et søk på hus (forutsatt at nøst og hus er lagt inn som tilknyttede termer). Det motsatte er ikke tilfelle; søker man på naust, vil man ikke finne bilder som er indeksert med termen hus. I noen tilfeller kan det selvsagt være nyttig for brukeren å få forslag til mer generelle ting, og dette er selvsagt mulig å gjøre, men ikke fokus i dette studiet.

Denne måten å la brukerne legge til termer på har fordeler og ulemper. En av fordelene er at grensesnittet blir enklere, og terskelen for å delta blir lavere. En ulempe er at systemet ikke kan skille mellom hvilke termer som er bredere termer, og hvilke som er ekvivalente. Dette har hovedsakelig konsekvenser for indeksering, og ikke for søking. Det ble derfor besluttet å bruke denne metoden i studiet, fordi det muliggjør det som er nødvendig for å besvare problemstillingen. Det er dessuten tenkelig at det kan la seg gjennomføre å bruke en slik forenkling også i praksis, enten direkte, eller ved at ansatte går gjennom hver nye term og plasserer den på riktig plass i hierarkiet. Det kan virke som dette vil føre til en stor mengde ekstra arbeid, men man må huske på at det bare er mulig å legge til den samme nye ekvivalente termen én gang, og arbeidet vil derfor avta etter hvert som det kontrollerte vokabularet blir mer komplett.



Nytt søk

Naust (/ARKITEKTUR OG BYGGESKIKK/Bygninger/Naust)

Foreløpig ingen tilknyttede termer.

Legg til termrelasjon 

Naust er en / et eller kan også skrives som:

Figur 10: Legge inn synonym for termen naust (eksperimentversjonen)

Det kontrollerte vokabularets funksjon er å gjøre det mulig for deltakerne å finne bilder selv om ordene de bruker i spørringen sin ikke er de samme som det som er brukt for å indeksere bildene. Dette blir gjort på to måter i prototypen:

1. Ved å finne tilknyttede termer til hvert søkeord, hvis slike eksisterer. I tillegg til at søkeordene brukes slik de er skrevet, undersøker systemet om søkeordet er tilknyttet andre termer. I så fall brukes også den tilknyttede termen som utgangspunkt for søket.
2. For hver tilknyttede term i spørringen, traverseres hierarkiet i det kontrollerte vokabularet, og det søkes etter bilder som er indeksert med termen eller smalere termer til søkeordet. Dette kan illustreres med et eksempel:

Eksempel 5: Utsnitt fra emneord-hierarkiet

klær

• ...

• mannsklær

•• dress

••• ...

•• frakk

•• herrehatt

••• fes

••• flosshatt

••• sixpence

••• stråhatt

••• ...

Eksemplet ovenfor viser et utsnitt fra emneord-hierarkiet. Hvis en deltaker søker etter mannsklær, vil bilder som er indeksert med termen mannsklær bli gjenfunnet, men også bilder som er indeksert med termene herrehatt, frakk og dress, som er smalere termer til mannsklær. I tillegg vil bilder indeksert med smalere termer til hver av disse termene bli gjenfunnet, for eksempel flosshatt, stråhatt, sixpence og fes, som er smalere termer til herrehatt. Det viktige her er at traverseringen bare går én vei: nedover i hierarkiet. Søker deltakeren på herrehatt, er

det sjelden ønskelig at systemet skal gjenfinne bilder indeksert med andre typer mannsklær, for eksempel dress, og heller ikke andre typer klær. Det vil imidlertid være ønskelig å finne alle typer herrehatte, for eksempel fes, flosshatt osv.

I kontrollversjonen er det ikke mulig å knytte nye ord til termer i det kontrollerte vokabularet. I denne versjonen brukes en mer tradisjonell form for brukergenerert innhold, nemlig tagging. Tagging er i dag i bruk i svært mange bildedatabaser, og er trolig den mest brukte formen for indeksering gjennom brukermedvirkning. Det er derfor interessant å bruke denne formen for brukermedvirkning som sammenligningsgrunnlag for å se hvor nyttig det er å introdusere brukermedvirkning i kontrollerte vokabularer.

Taggingen skjer i kontrollversjonen av prototypen ved at deltakerne får se et inputfelt hvor de kan legge til en og en tag (se figur 9). Når en tag er lagt til, vises taggen i en liste under inputfeltet.

Igjen er det viktig å påpeke at deltakerne i eksperimentet aldri bruker det nye søkegrunnlaget de er med å skape. Dette blir utelukkende gjort i ettertid, i en prosess hvor de mest brukte spørringene blir valgt ut, og utført i prototypen med termene eller taggene fra én og én deltaker aktivert.

4.5.9 Kildekode

I appendiks e vises et utvalg av kildekoden til prototypen. Den komplette kildekoden kan fremskaffes ved å kontakte meg via e-postadressen på forsiden av denne avhandlingen.

4.5.10 Begrensninger i prototypen

I dette studiet er målet å undersøke forandringer i precision og recall. Prototypen som er utviklet i forbindelse med studiet, er ikke laget med fokus på å oppnå best mulig søkeeffektivitet i form av eksekveringstid, eller hvordan resultatene best mulig kan presenteres for brukerne. Funksjonaliteten i prototypen er i all hovedsak begrenset til det som er nødvendig eller gunstig for å evaluere studiets problemstilling og hypotese.

Prototypen er bare testet i Mozilla Firefox, fordi det ikke er intensjonen i dette studiet å lage en applikasjon som skal brukes i en reell sammenheng. For å besvare problemstillingen er det bare nødvendig å ha en prototyp som fungerer i én enkelt nettleser, siden jeg selv kan

velge hvilken nettleser som skal brukes i forsøkene. Det ble derfor besluttet å ikke bruke ekstra tid på å gjøre prototypen kompatibel med flest mulig nettlesere.

Prototypen utfører ingen rangering av bildene som blir funnet når en deltaker utfører et søk. Det ville vært mulig å bruke de innsamlede dataene til å rangere resultatsettet, men det er et viktig poeng for et eksperiment å kontrollere flest mulig eksterne variabler, og en slik rangering ville ført til at resultatsettene ville vært forskjellige for de forskjellige deltakerne. I stedet vil jeg, gjennom et par eksempler, kort beskrive hvordan de innsamlede dataene kunne vært brukt for å rangere resultatene.

Av de 10 testbildene var det tre bilder som inneholdt bilder av en eller flere kvinner i bunad. To av disse bildene var indeksert med termen Hardangerdrakt, mens et av bildene hadde termen Telemarksdrakt. Mens åtte av deltakerne hadde valgt å knytte ordet bunad til termen Hardangerdrakt, hadde bare to deltakere knyttet det samme ordet til termen Telemarksdrakt. Denne informasjonen kunne for eksempel vært brukt til å rangere bilder indeksert med Hardangerdrakt høyere opp enn Telemarksdrakt hvis en bruker søker etter ordet bunad. Et annet eksempel er ordet hus, som har blitt brukt 8 ganger for å beskrive termen Bolighus, 6 ganger for Høyhus, 5 ganger for Stovehus og 1 gang hver for termene Naust, Havnebod og Forretninger. På samme måte kunne denne informasjonen blitt brukt til å rangere bildene i et søk etter hus, slik at bilder med termen bolighus kom høyest opp på rangeringen. Denne måten å rangere på er imidlertid avhengig av at flere brukere legger inn de samme termene.

Taggene kunne vært brukt på en lignende måte, ved å rangere bildene ut fra hvor mange forekomster av taggen som finnes. Vi kan igjen bruke ordet bunad som eksempel. Et av de tre bildene med en eller flere bunadskledde kvinner har blitt tagget 10 ganger med ordet bunad, et annet har blitt tagget 8 ganger med dette ordet, mens det siste bare har blitt tagget 1 gang med ordet bunad. Denne informasjonen kunne også dannet grunnlag for å rangere søkeresultatet.

4.6 Gjennomføring

4.6.1 Forberedelse

For å gjøre det gjennomførbart å teste kvaliteten på søk i forhold til precision og recall, er det nødvendig å bruke et begrenset antall innholdsobjekter, fordi hvert bilde må vurderes i for-

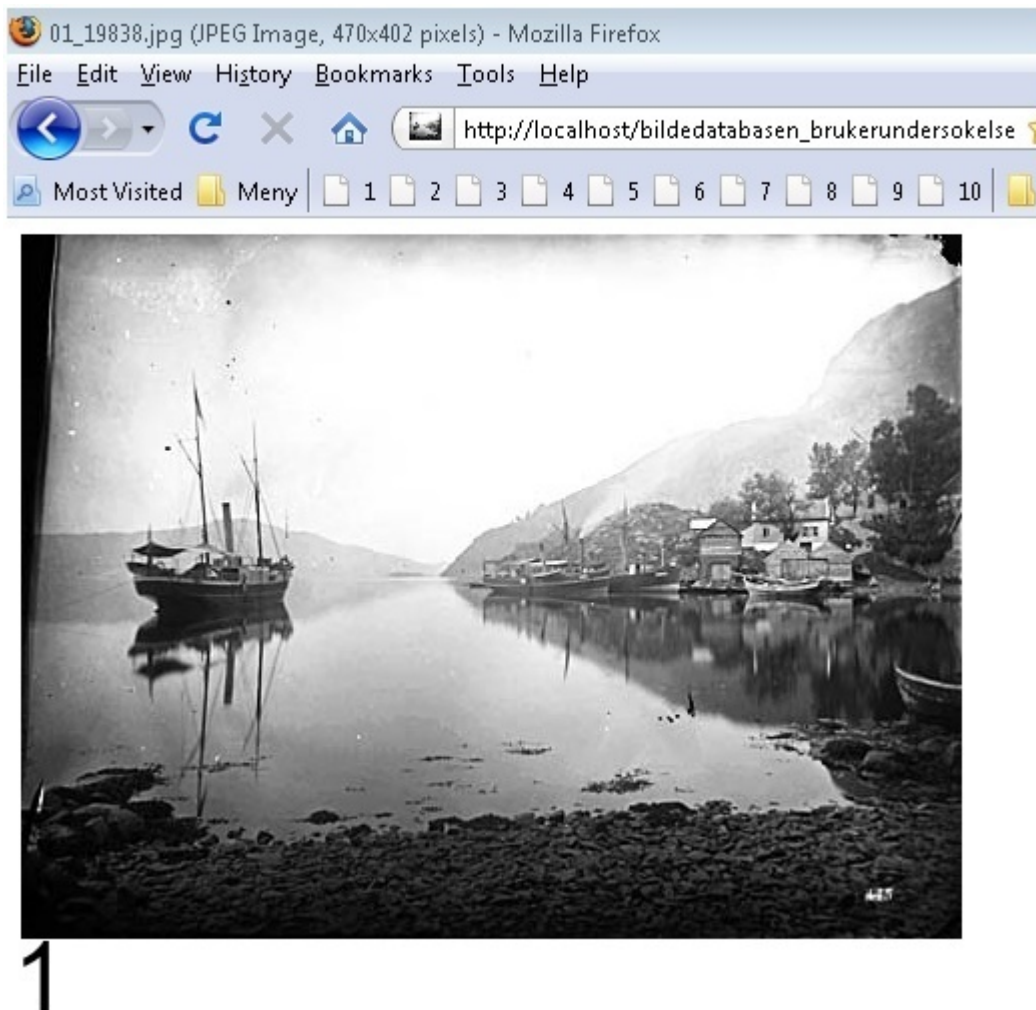
hold til relevans for hver spørring fra hver deltaker. I tillegg skulle deltakerne utføre oppgaver knyttet til hvert bilde og til termene som er knyttet til bildene, og det var ønskelig at forsøkene ikke skulle ta mer enn en time av hensyn til deltakerne. Jeg valgte derfor ut 10 tilfeldige bilder som skulle brukes i forsøkene (se appendiks a).

Deltakerne ble rekruttert gjennom direkte kontakt, hvor det ble avtalt hvor og når forsøket skulle foregå. Deltakerne fikk vite at de skulle medvirke i et studie som handlet om tekstbasert bildesøking og brukervedvirking, men fikk ikke vite den faktiske problemstillingen eller hypotesen. Det var ikke nødvendig å tilby deltakerne noen kompensasjon for å delta i studiet – alle som ble spurt sa ja. Det ser ut til at det er god kultur på instituttet for å hjelpe medstudenter med å skaffe deltakere. Mange av studentene skal selv utføre lignende studier, og ønske om å se hvordan man kan utføre et forsøk er kanskje en viktig motivasjon for å delta. En annen grunn til at alle som ble spurt var villig til å delta, er kanskje at det er vanskelig å si nei når man blir konfrontert direkte.

4.6.2 Utførelse

Forsøkene ble utført i et laboratoriemiljø, med én og én deltaker, og hadde en gjennomsnittlig varighet på ca. 45 minutter i eksperimentgruppen, og noe kortere i kontrollgruppen. Etter et muntlig samtykke fra deltakeren, ble det gjort lydopptak av forsøket. Deltakeren fikk utdelt et samtykkeskjema som måtte undertegnes (se appendiks b), og fikk så utdelt retningslinjene (se appendiks c og appendiks d). Disse bestod av to oppgaver som skulle utføres etter hverandre, samt en grunnleggende innføring i de strukturene som var nødvendig for å forstå oppgavene. Jeg var som observatør også behjelpelig med å oppklare eventuelle uklarheter, slik at deltakerne skulle ha en god forståelse av oppgavene.

Det ble brukt en bærbar PC til å kjøre prototypen. En ekstra skjerm var tilkoblet den bærbare PC-en, slik at deltakerne hadde en ekstra skjerm til venstre i tillegg til skjermen på PC-en. På den venstre skjermen var det åpnet et nettleservindu med linker til de 10 testbildene, hvor deltakerne kunne navigere frem og tilbake mellom bildene. Figur 11 viser hva deltakerne kunne se på den venstre skjermen.



Figur 11: Visning av ett av de ti testbildene, med linker til de andre bildene

På skjermen til den bærbare PC-en så deltakerne det vanlige søkegrensesnittet (se figur 6 på side 50 og figur 7 på side 51). Den første oppgaven var lik for eksperimentgruppen og kontrollgruppen, og gikk ut på å se på bildene på den venstre skjermen, og bruke søkegrensesnittet til å søke etter bilder. Spørringene som ble brukt, ble lagret i databasen. For at deltakerne ikke skulle bli påvirket av de eksisterende metadataene, ble disse ikke presentert på siden. Jeg ønsket at hver deltaker skulle identifisere flest mulig spørringer for hvert bilde uten å vite hvilke metadata som faktisk var knyttet til bildene.

Oppgave 2 var forskjellig i de to gruppene. Når deltakerne hadde begynt på denne oppgaven, hadde de ikke lenger anledning til å utføre flere søk. Deltakerne i begge gruppene ble vist en side i prototypen for hvert bilde, som viste all metadata som på forhånd var knyttet til bildene av UB (figur 8 på side 52 / figur 9 på side 52). Den andre oppgaven til eksperimentgruppen gikk ut på å identifisere synonymer eller bredere termer til termer som var knyttet til de 10 testbildene, og bruke prototypen til å knytte disse til de eksisterende

termene. Detaljene rundt hvordan deltakerne skulle gå frem for å utføre denne oppgaven ble forklart i seksjon 4.5.8.

Kontrollgruppens andre oppgave gikk ut på å bruke prototypen til å tagge de 10 testbildene. Detaljene rundt hvordan deltakerne skulle gå frem for å utføre denne oppgaven ble også forklart i seksjon 4.5.8.

4.6.3 Validering av data

Det ble ikke funnet grunnlag for å fjerne dataene fra noen av deltakerne. Mer om dette i seksjon 5.1.

4.6.4 Kommentarer til forsøkene

Det er viktig å påpeke at prototypen bare er testet i et lukket forsøk, og ikke i en virkelig situasjon. Det er mange problemer som kan oppstå i en applikasjon, særlig når den blir implementert i en reell situasjon. Et viktig eksempel på dette er misbruk, eller brukerfeil som et resultat av misforståelser. Hvis ideer fra prototypen skal tas i bruk av UB, må det i særlig stor grad tas høyde for slike problemer, og det er mulig å legge inn forskjellige mekanismer som skal bidra til å begrense feilinformasjon i databasen, for eksempel utlukning av rasistiske eller grove ord, eller begrensninger på hvor mange termer eller tagger det er mulig å legge inn i løpet av en viss tidsperiode. I eksperimentet har vi en mye mer kontrollert kontekst, men det er selvsagt fullt mulig med brukerfeil også her. Dette vil i så fall være en viktig del av observasjonene rundt gjennomførbarheten av brukermedvirkningen, da dette vil gi utslag på precision.

For å teste hvilken effekt endringene i det kontrollerte vokabularet hadde, ble det i etterkant av forsøkene utviklet et enkelt script som gjennomførte søk i testsamlingen med hver av spørringene som ble foreslått av deltakerne. Scriptet ble kjørt med data fra hver enkelt deltaker aktivert, slik at halvparten av kjøringene ble kjørt med termer fra én og én deltaker i eksperimentgruppen aktivert, og den andre halvparten ble kjørt med tagger fra én og én deltaker i kontrollgruppen aktivert. Scriptet regnet så ut precision- og recall-verdien for hver enkelt spørring, ved at det for hver spørring sammenlignet hvilke bilder som burde blitt gjenfunnet med hvilke bilder som faktisk ble gjenfunnet. Ved å finne disse verdiene, er det mulig å regne ut gjennomsnittlig precision og recall for hver spørring for hver behandling, og

sammenligne verdiene statistisk. Precision og recall er, som nevnt i seksjon 3.3.2, verdier for relevansen til en spørring. Høy recall og høy precision betyr at mange av de dokumentene som ideelt sett burde vært funnet, faktisk blir funnet, og at det ikke blir funnet mange dokumenter som ikke er relevante.

Deltakerne i studiet brukte totalt 354 unike spørringer, fordelt på til sammen 812 utførte spørringer. 280 av de unike spørringene ble bare brukt av én eller to deltakere, mens 74 av de unike spørringene ble brukt av 3 eller flere deltakere. For å finne de spørringene som er mest aktuelle, og som det er mest sannsynlig at ville blitt brukt i en reell sammenheng, ble det besluttet å bare bruke de spørringene som hadde blitt brukt av 3 eller flere deltakere, som utgangspunkt for å vurdere precision og recall. Spørringer som har blitt brukt av færre enn 3 deltakere, er ofte et resultat av skrivefeil, eller at deltakerne har «vridd hjernen» for å komme på spørringer, noe som ofte har resultert i spørringer av den mer obskure typen, som for eksempel «uklar mann i bakgrunnen som går inn en dør».

For å avgjøre hvilke bilder som var relevante for hver av de 74 spørringene, ble hver spørring vurdert i forhold til hvert bilde, og en liste over hvilke av de 10 testbildene som ble regnet som relevant ble laget for hver spørring. Følgende generelle kriterier ble brukt for å avgjøre om et bilde skulle bli vurdert som relevant for en spørring eller ikke:

- Konseptet i spørringen må komme frem tydelig visuelt med den oppløsningen deltakerne ser, eller gjennom metadata som fantes i UBs allerede eksisterende metadata.
- Spørringen må være skrevet på norsk (bokmål eller nynorsk).
- Det er likegyldig om spørringen er skrevet i entall eller flertall.
- Spørringen må ikke inneholde skrivefeil (de fleste skrivefeil blir naturlig luket ut ved å ekskludere spørringer som er brukt 2 eller færre ganger, fordi skrivefeil ofte er basert på tilfeldige feiltastinger som bare forekommer én gang).

Den mest brukte spørringen er *bunad*, og bildene som ble regnet som relevant for denne spørringen var bilde nr. 2, 4 og 7 (se appendiks a). Et annet eksempel er *fjell*, hvor bildene 1, 2, 3, 5 og 7 ble regnet som relevante. Hver spørring ble tilordnet en kommaseparert liste over bildenummer som tilsvarer numrene i appendiks a. Denne listen ble brukt av scriptet nevnt tidligere i denne seksjonen for å regne ut precision.

4.7 Statistisk analyse

I dette eksperimentet er deltakerne delt tilfeldig inn i grupper, noe som i teorien skal føre til ekvivalente grupper (Cozby, 2007). Man forutsetter at hvis gruppene er ekvivalente, vil enhver forskjell i den avhengige variabelen være forårsaket av effekten av den uavhengige variabelen. Det er likevel slik at forskjellen mellom to forskjellige grupper nesten aldri vil være null, fordi vi bare bruker et utvalg og ikke hele populasjonen. Derav følger at forskjellen mellom resultatene fra de forskjellige behandlingene reflekterer virkelige forskjeller i populasjonen pluss enhver tilfeldig feil. For å avgjøre hvorvidt forskjeller mellom de to behandlingene er signifikant eller bare et resultat av slike tilfeldige feil, er det nødvendig med en eller annen form for statistisk sannsynlighetsanalyse (Goodwin, 2005: 223). I den typen eksperiment som er utført i dette studiet kan man gjøre dette gjennom en *t-test*. En *t-test* undersøker forskjellen mellom de to gjennomsnittsverdiene av den avhengige variabelen, og avgjør om denne forskjellen er større enn det man kan forvente at ville kunne oppstå av ren tilfeldighet. Hvis forskjellen er større, og man kan utelukke potensielle konfunderinger, kan man med stor sannsynlighet konkludere med at det er en reell forskjell. Den statistiske analysen av datamaterialet blir presentert i kapittel 5, og diskutert i kapittel 6.

4.8 Forskningsdesign: konklusjon

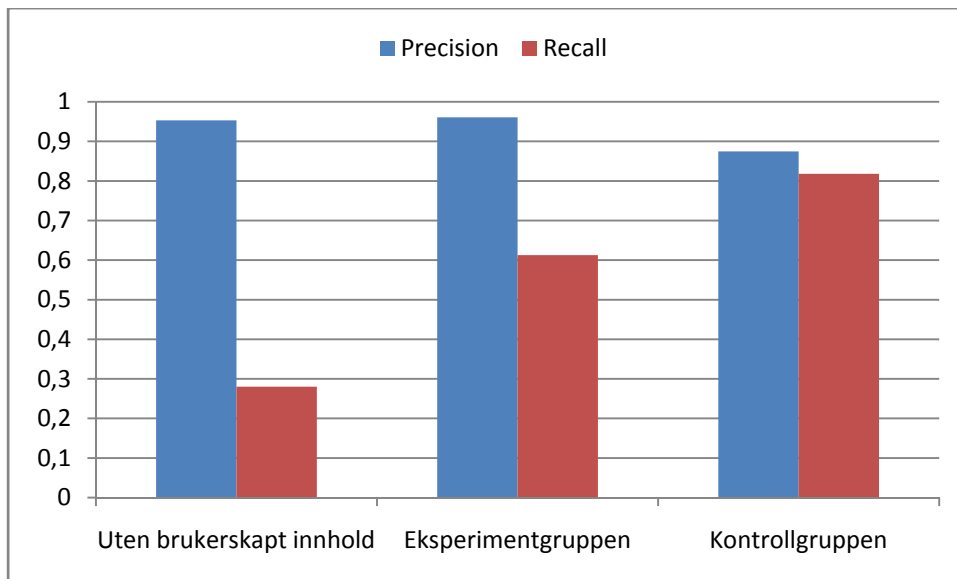
En problemstilling ble fremsatt, og et kontrollert eksperiment ble valgt som en passende metode for å besvare denne, fordi denne metoden gjør det mulig på en god måte å samle inn de dataene som er nødvendig. En prototyp som skulle brukes i datainnsamlingen ble utviklet og beskrevet. Metoden og verktøyene viste seg å være gode for formålet, og forsøkene kunne gjennomføres uten store problemer. Observasjonene som ble samlet inn, blir presentert i neste kapittel.

5 Resultater

I dette kapitlet presenteres resultatene fra eksperimentet. Som et første steg i analysen av de innsamlede dataene, brukes beskrivende statistikk for å systematisere dataene. Til slutt i kapitlet skal hypotesene testes.

5.1 Beskrivende statistikk

I figur 12 ser man hvilken innvirkning den samlede brukermedvirkningen for hver gruppe har på precision og recall. Verdiene i de første to stolpene er regnet ut ved utelukkende å bruke UBs metadata (se appendiks a for en oversikt over alle emneord og tagger knyttet til bildene før og etter forsøkene). De to stolpene i midten viser precision og recall når UBs metadata er kombinert med de nye termene som ble opprettet av deltakerne i eksperimentgruppen (uten data fra kontrollgruppen). De to stolpene til høyre viser precision og recall for UBs metadata kombinert med de nye taggene som ble opprettet av kontrollgruppen (uten data fra eksperimentgruppen). De blå stolpene viser precision, og vi kan se at precision har en liten økning når termene fra eksperimentgruppen er aktivert, mens precision går litt ned når alle taggene fra kontrollgruppen er aktivert. De røde stolpene viser recall, og basert på spørringene fra deltakerne er recall 28 % når det bare er UBs metadata som brukes. Denne verdien samsvarer med funnene til Furnas et al. som blir beskrevet i seksjon 3.7, hvor de kom frem til at det beste resultatet man kan forvente å oppnå hvis man bare bruker én enkelt term for å beskrive et objekt, er 15-35 % recall. Man kan se en tydelig økning i recall når termene fra eksperimentgruppen er aktivert (61,2 % recall). Økningen i recall er likevel klart størst når taggene fra kontrollgruppen er aktivert og termene fra eksperimenter deaktivert (81,8 % recall). Mens recall blir litt mer enn doblet med eksperimentgruppens termer, blir disse verdiene altså nesten tredoblet med kontrollgruppens tagger.

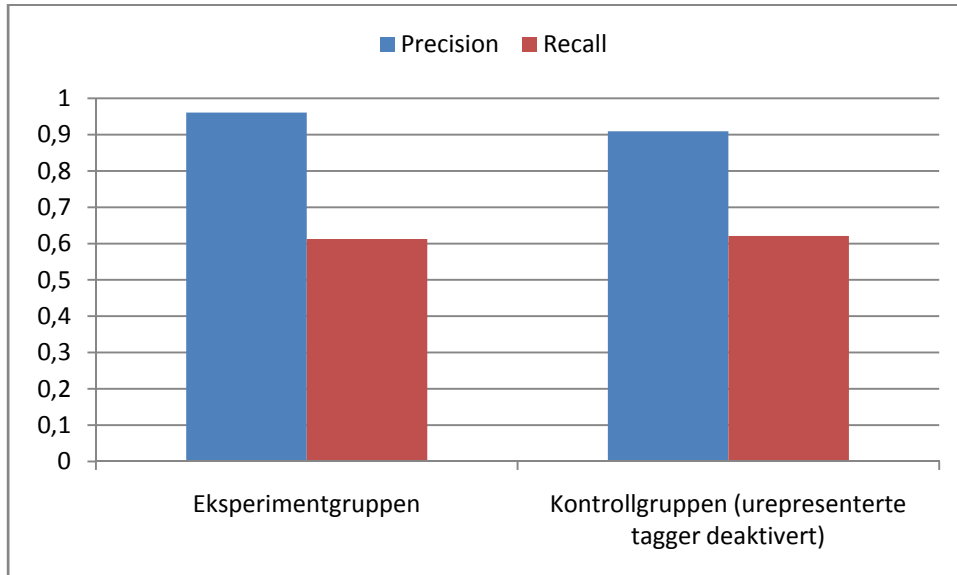


Figur 12: Total precision og recall uten brukergenerert innhold sammenlignet med data fra alle deltakerne i hver gruppe aktivert.

Disse forskjellene indikerer at direkte tagging av bilder har en større virkning på recall enn å legge til termer i det kontrollerte vokabularet. Det er imidlertid enkelt å finne en viktig grunn til dette ved å se på utgangspunktet til de to gruppene. Eksperimentgruppen kan ikke knytte nye konsepter til bilder, men må forholde seg til de 69 termene som UB allerede har knyttet til bildene (se appendiks f). Kontrollgruppen slipper å ta utgangspunkt i termer som er valgt på forhånd, og kan knytte hvilket som helst konsept til et bilde. Denne gruppen har dermed mye bedre forutsetninger for å øke recall enn eksperimentgruppen. Dette innebærer et uforutsett problem for konklusjonsvaliditeten til studiet, og evnen til å kunne trekke konklusjoner ut fra observasjonene blir redusert. Det er likevel mulig å dra visse slutninger ut fra resultatene (som vi skal se i neste kapittel).

Ved å gjøre visse manipuleringer i datamaterialet kan vi danne oss et bilde av hvordan resultatene ville blitt hvis utgangspunktet var mer likt, dvs. hvis kontrollgruppen var bundet til utelukkende å bruke tagger som allerede var representert blant UBs tilknyttede termer eller synonymer eller bredere termer til disse. Denne manipuleringen går ut på å gå gjennom alle taggene fra kontrollgruppen og deaktivere tagger som representerer konsepter som eksperimentgruppen ikke ville hatt mulighet til å legge til uten å bryte med logikken bak termene. Det er umulig å få til en helt riktig sammenligning av de to fremgangsmåtene på denne måten, blant annet fordi det eliminerer muligheten for at deltakere i eksperimentgruppen kunne knyttet feilaktige termer til de eksisterende termene, noe de kunne ha gjort hvis de hadde hatt mulighet til å knytte nye termer til bildene. Dette fører til at precision hos

eksperimentgruppen kan bli høyere enn den ville blitt hvis eksperimentgruppen hadde hatt full frihet i knytte termer til bildene. Avgjørelsen om hvorvidt de brukergenererte taggene kunne vært representert som synonymer eller bredere termer til en eller flere av de 69 termene er dessuten basert på min egen kunnskap, og siden jeg ikke er ekspert på emnene som er representert, kan eventuelle feilbedømmelser fra min side føre til feil (det finnes uansett ikke noe fasitsvar i alle tilfeller for om et bilde er relevant eller ikke). Resultatene fra denne sammenligningen brukes derfor ikke til å dra noen konklusjoner om forskjeller mellom de to gruppene. Det er likevel sannsynlig at det ved å deaktivere tagger som det ikke var mulig for eksperimentgruppen å representere gjennom oppdatering av det kontrollerte vokabularet, er mulig å danne seg et bilde av hvordan de to metodene ville ha fungert i forhold til hverandre hvis utgangspunktet var det samme – hvis deltakerne i kontrollgruppen hadde vært bundet til utelukkende å tagge konsepter som allerede var utpekt av UB, eller hvis deltakerne i eksperimentgruppen også hadde hatt mulighet til å knytte nye termer til bildene. På grunn av validitetsspørsmålet, har det ikke blitt utført en signifikanstest etter at disse taggene er deaktivert, men total precision og recall for alle de 74 spørringene er blitt regnet ut, og vises i figur 13, sammen med total precision og recall for eksperimentgruppen.



Figur 13: Precision og recall i eksperimentgruppen sammenlignet med precision og recall i kontrollgruppen når tagger som ikke er representert som termer som eksperimentgruppen kan bruke, er deaktivert

I figur 13 kan vi se at recall er ganske lik etter at taggene til kontrollgruppen har blitt modifisert (61,2 % i eksperimentgruppen mot 62,1 % i kontrollgruppen etter modifikasjon). Dette er ikke uventet, og antyder at det å gjøre premissene for de to gruppene mer lik, fører til

svært like resultater. Dette skyldes trolig at oppgavene som er forskjellige hos de to gruppene, tagging og generering av termer, egentlig er svært like – begge oppgavene går i prinsippet ut på å *beskrive konsepter på forskjellige nivåer og med forskjellige ord*. Begge gruppene øker mengden og mangfoldet av ord det er mulig å bruke for å få treff på det samme bildet. I tillegg er antallet bilder og mengden av muligheter til å finne på nye ord for konseptene som finnes på bildet ganske begrenset, og det skal ganske lite til før en stor andel av mulighetene er representert gjennom deltakernes aktivitet.

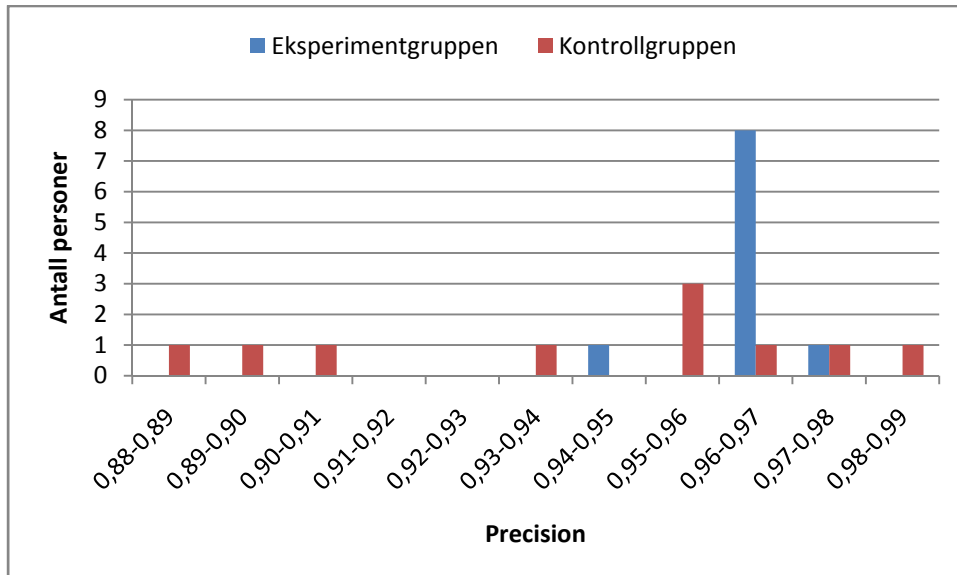
Forskjellen i precision mellom de to gruppene er større (96,1 % i eksperimentgruppen mot 90,1 % i kontrollgruppen etter modifikasjon). Dette representerer en ganske stor forskjell, og man kan spekulere i hva grunnen til dette er. Man må trolig se på hva hvert enkelt bilde er tagget med for å finne årsaken til dette. Når taggene til kontrollgruppen ble deaktivert, ble det bare vurdert om taggen kunne representeres med termer som var tilknyttet minst ett av de 10 testbildene. Hvis ikke ble den deaktivert. Det som da imidlertid ikke kommer frem, er hvis en tag også er knyttet til et eller flere andre bilder som de ikke burde vært knyttet til, noe som vil føre til lavere precision. Eksempler på tagger som feilaktig har blitt knyttet til bilder av kontrollgruppen er (tallene referer til bildene i testsamlingen, se appendiks a):

- Bergen (bilde 5)
- Hatt (bilde 6)
- fjord og naust (bilde 7)
- rev og mink (bilde 9)

Disse feiltaggingene skyldes at deltakerne har feiltolket innholdet i bildene, eller at deltakerne har tagget noe som finnes på bildet, men som ikke er vurdert å være relevant. Noen deltakere har for eksempel ikke lagt merke til at seteren på bilde 7 ligger i et fjellandskap, og ikke ved en fjord. I tilfellet med hatten i bilde 6 er det riktignok en hatt på bildet, men denne har ikke blitt vurdert å være fremtredende nok i bildet til å være relevant.

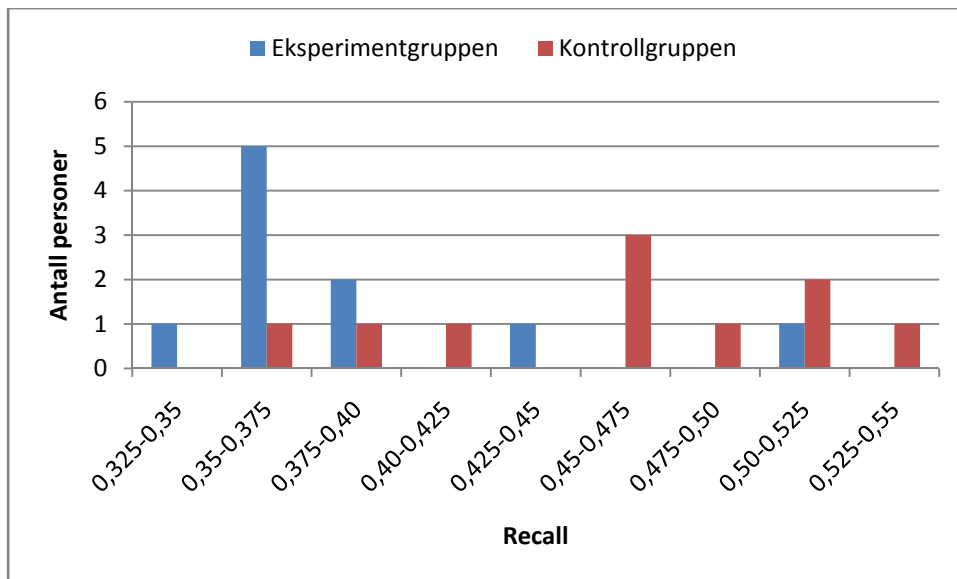
For å undersøke om observasjonene i eksperimentet er signifikante, må vi undersøke resultatene fra hver enkelt deltaker for seg selv. Figur 14 viser precision for henholdsvis eksperimentgruppen og kontrollgruppen, etter å ha delt precision inn i grupper. Den første gruppen inneholder deltakere hvis termer eller tagger førte til en precision på 88-89 %, den neste gruppen spenner fra 89-90 %, mens den 10. og siste gruppen spenner fra 98-99 % precision. Denne grafen viser at det er en relativt jevn fordeling mellom de to gruppene, men

eksperimentgruppen ser ut til å ha litt høyere precision, noe som er tydelig ut fra den høye søylen ved 96-97 %. Man ser også at variasjonen i distribusjonen er større i kontrollgruppen enn i eksperimentgruppen.



Figur 14: Frekvensdistribusjon for precision (i klasser).

I figur 15 nedenfor ser man recall for henholdsvis eksperimentgruppen og kontrollgruppen, etter å ha delt recall inn i grupper. Den første gruppen spenner fra 32,5-35 %, den neste gruppen spenner fra 35-37,5 % og så videre opp til gruppe 9 med 52,5-55 % recall. Det fremkommer av grafen at kontrollgruppen og deres tagger er den gruppen som har bidratt til høyest recall. I tillegg kan man se at variasjonen i distribusjonen er noenlunde lik i de to gruppene.



Figur 15: Frekvensdistribusjon for recall (i klasser).

Tabell 1 viser gjennomsnittsverdier for precision og recall for hver deltaker. Også her kan man se at verdiene for precision er nokså lik mellom de to gruppene, men litt høyere for eksperimentgruppen, og at det er større variasjon mellom observasjonene i kontrollgruppen. Når det gjelder recall, ser man igjen at det er kontrollgruppen som har høyest verdier, og at variasjonen mellom observasjonene er noenlunde lik i de to gruppene. Det som kommer tydeligere frem her, er at det er kontrollgruppen som faktisk har den største variasjonen mellom observasjonene både for precision og recall.

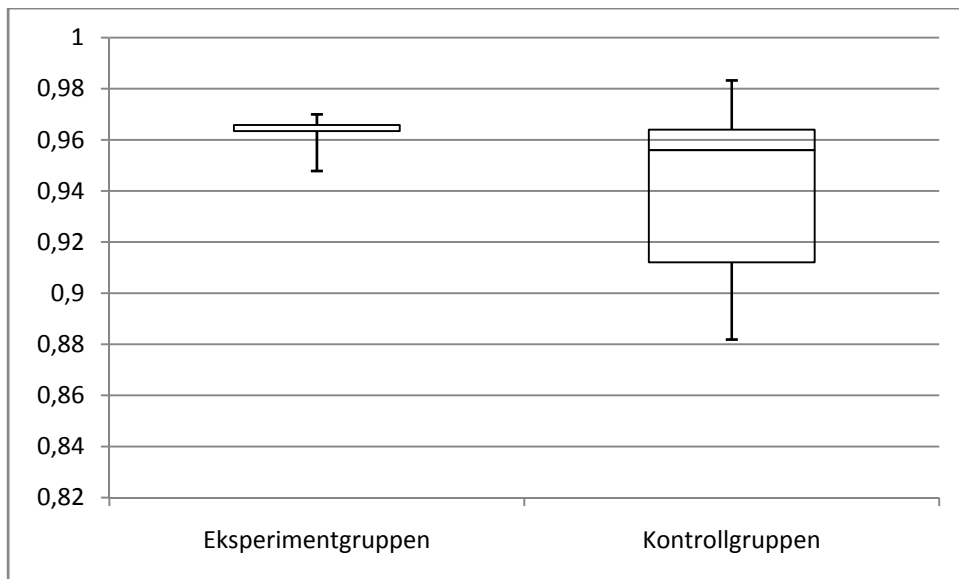
Sammenligner man recall for hver enkelt deltaker med den totale verdien av recall for hver gruppe, ser man at hver enkelt deltaker alene ikke klarer å skape en særlig stor økning i recall (total recall med dataene fra én og én gruppe aktivert var 61,2 % for eksperimentgruppen og 81,8 % i kontrollgruppen, mens gjennomsnittet i recall for én og én bruker er 38,9 % for eksperimentgruppen og 46,0 % for kontrollgruppen). Som Furnas et al. (1987) påpeker, må man ha et relativt stort mangfold av ord for å oppnå gode resultater i recall. Dette samsvarer med funnene i dette studiet. Selv om det er en del overlapp mellom ordene to forskjellige deltakere bidrar med, vil den ene deltakeren ofte foreslå et eller flere ord som den andre ikke foreslår, og vice versa. Lagt sammen vil bidrag fra mange brukere føre til et rikt mangfold som gir en høy sannsynlighet for å dekke behovene til mange brukere.

Tabell 1: Gjennomsnittlig precision og recall med bidragene til hver deltaker aktivert, samt totale gjennomsnitt og standardavvik for precision og recall.

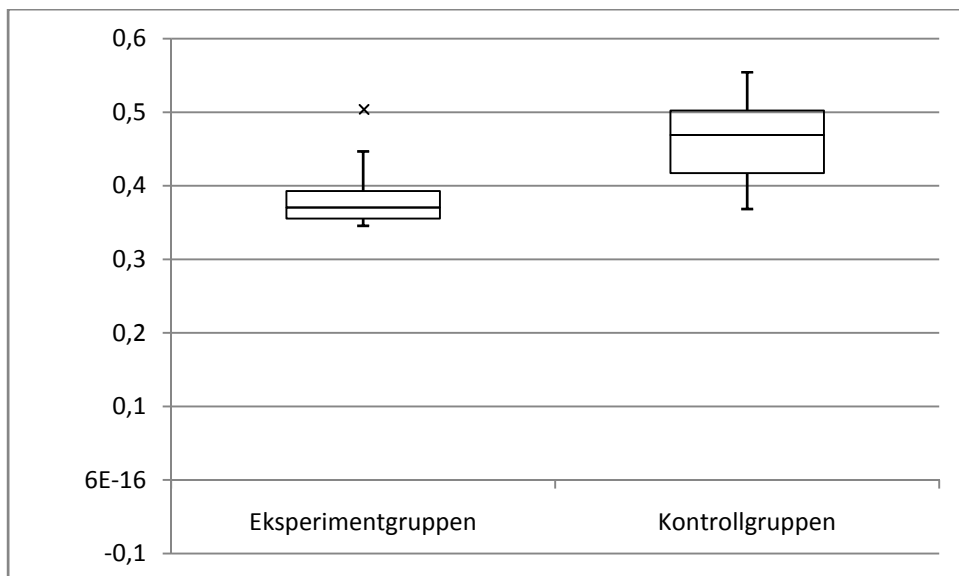
Eksperimentgruppen			Kontrollgruppen		
	Precision	Recall		Precision	Recall
Deltaker 01	0,963438	0,355909	Deltaker 11	0,881837	0,507879
Deltaker 02	0,947813	0,374394	Deltaker 12	0,98325	0,464394
Deltaker 03	0,966571	0,385758	Deltaker 13	0,959268	0,508182
Deltaker 04	0,962258	0,345606	Deltaker 14	0,965588	0,377424
Deltaker 05	0,964545	0,366364	Deltaker 15	0,898627	0,554394
Deltaker 06	0,963438	0,395303	Deltaker 16	0,958462	0,473788
Deltaker 07	0,97	0,446818	Deltaker 17	0,97	0,45
Deltaker 08	0,963438	0,352273	Deltaker 18	0,953488	0,485758
Deltaker 09	0,963438	0,355303	Deltaker 19	0,938	0,368485
Deltaker 10	0,96619	0,50803	Deltaker 20	0,903421	0,406364
Gjennomsnitt	0,963113	0,388576		0,941194	0,459667
Standardavvik	0,005833	0,051368		0,034547	0,060154

For å få en enda bedre forståelse av dataene, kan man sette dataene opp i et boksplokk. Et boksplokk gir en god oversikt over fordelingen i et datamateriale (Wohlin, et al., 2000). Ved hjelp av de fem nøkkelverdiene nedre hale, nedre kvartil, median, øvre kvartil og øvre hale er det mulig å tegne en graf som gir grafisk informasjon om verdiområde, spredning og symmetri eller skjevhet i datamaterialet. Nedre kvartil nk er 25 %-prosentilen (medianverdien til verdiene som er mindre enn medianverdien til alle observasjonene), og øvre kvartil ok er 75 %-prosentilen (medianverdien til verdiene som er større enn medianverdien til alle observasjonene). Høyden til boksen er $h = ok - nk$. Halene til boksen representerer en teoretisk grense som det er sannsynlig at alle datapunktene vil finnes innenfor hvis distribusjonen er normal. Den øvre halen oh er $ok + 1,5h$, og den nedre halen nh er $nk - 1,5h$ (Frigge, Hoaglin, & Iglewicz, 1989, se Wohlin et al., 2000). Halene blir avkortet til nærmeste faktiske verdi. Verdier som ligger utenfor den nedre og øvre halen kalles *utliggere*²⁷, og blir vist som kryss i boksplokket.

²⁷ Eng.: outlier



Figur 16: Boksploott for precision for de to gruppene.



Figur 17: Boksploott for recall for de to gruppene.

Figur 16 viser at precision er høyere i eksperimentgruppen enn i kontrollgruppen, hovedsakelig fordi det er en del lave verdier som trekker ned i kontrollgruppen. Dette kan sannsynligvis forklares med at kontrollgruppen har større mulighet til å knytte nye konsepter til bildene, og dermed flere muligheter til å gjøre feilvurderinger. Igjen ser man at variasjonen mellom observasjonene for precision er større i kontrollgruppen enn i eksperimentgruppen. Forskjellen i precision mellom de to gruppene ser dog ikke ut til å være særlig stor.

I figur 17 ser man igjen at variasjonen mellom observasjonene er ganske lik i de to gruppene. Igjen ser man også en ganske tydelig forskjell mellom recall i de to gruppene, som tyder på at det kan være mulig å bevise forskjellen i en hypotesetest.

Det er også viktig å se etter *utliggere* i datamaterialet. Utliggere er observasjoner som i stor grad skiller seg ut i forhold til de andre observasjonene, og i noen tilfeller kan det være en fordel å fjerne slike observasjoner fra datamaterialet, fordi det kan være ekstreme verdier som man ikke kan forvente å finne igjen hvis man gjentar studiet. I boksplottene i figur 16 og figur 17 viser halene en teoretisk grense hvor det er sannsynlig å finne alle datapunktene innenfor. Mens figur 16 ikke viser noen utliggere, kan man i figur 17 se at en av observasjonene for recall i eksperimentgruppen kan regnes som en utligger (50,8 % recall). Det ble besluttet å ikke fjerne denne utliggeren, fordi den ikke ble ansett som en ekstrem verdi.

5.2 Hypotesetesting

De to hypotesene, for henholdsvis precision og recall, har begge blitt evaluert gjennom en t -test.

5.2.1 Precision

Som vi husker fra seksjon 4.4.2, var den første nullhypotesen følgende:

Nullhypotese: $H_0: P_1 = P_2$

Alternativ hypotese: $H_1: P_1 \neq P_2$

Tabell 2 viser resultatene fra t -testen for precision (uavhengig, tosidig).

Tabell 2: Resultater fra t -testen for precision

Forskj. i gj.snittsverdier	Frihetsgrad	t-verdi	p-verdi
0,021918625	10	1,978308339	0,076086398

Ut fra tabell 2 kan vi konkludere med at det ikke er en signifikant forskjell mellom precision hos de to gruppene, og den første nullhypotesen kan ikke forkastes.

5.2.2 Recall

Den andre nullhypotesen i seksjon 4.4.2 var som følger:

Nullhypotese: $H_0: R_1 = R_2$

Alternativ hypotese: $H_1: R_1 \neq R_2$

Tabell 3 viser resultatene fra t -testen for recall (uavhengig, tosidig)

Tabell 3: Resultater fra t -testen for recall

Forskj. i gj.snittsverdier	Frihetsgrad	t-verdi	p-verdi
-0,071090909	18	-2,842014144	0,010815941

Ut fra tabell 3 kan vi konkludere med at H_0 kan forkastes. Recall er signifikant større i kontrollgruppen enn i eksperimentgruppen.

5.3 Resultater: konklusjon

Gjennom statistiske analyser av de innsamlede dataene, ser det ut til at selv om både oppdatering av det kontrollerte vokabularet og tagging har en positiv innvirkning på recall er tagging den mest effektive formen for brukermedvirkning, i hvert fall for å forbedre recall. Mulige årsaker til dette diskuteres i kapittel 6.

6 Diskusjon og konklusjoner

Målet med dette studiet har vært å undersøke gjennomførbarheten og verdien av å innføre brukergenerert innhold i et kontrollert vokabular for en digital bildesamling. Gjennomførbarheten er i stor grad avhengig av verdien; hvis det ikke er mulig å oppnå god verdi av det brukergenererte innholdet, er det vanskelig å konkludere med at ideen er gjennomførbar. For å måle verdien har metoden blitt sammenlignet med en allerede eksisterende metode for brukermedvirkning, som uten tvil er både gjennomførbar og verdifull, nemlig tagging. Grunnlaget for eksperimentet som er utført i forbindelse med studiet har vært at, hvis man kan oppnå tilsvarende eller bedre verdi enn tagging ved å la brukere legge til innhold i et kontrollert vokabular, kan man forvente at det vil være gjennomførbart og verdifullt å ta i bruk denne formen for brukermedvirkning.

6.1 Verdien av brukergenerert innhold i kontrollerte vokabularer

En kontrollgruppe med 10 deltakere som tagget bilder viste seg å ha en formidabel positiv innvirkning på recall, som gikk fra 28,0 % til 81,8 %. Også eksperimentgruppen klarte å gi en stor økning i recall, 61,2 %, selv uten å kunne knytte noen nye konsepter til bildene (dette tallet ville sannsynligvis vært høyere hvis eksperimentgruppen hadde hatt mulighet til å knytte nye konsepter til bilder). Det må tas med i betraktningen at det var studenter som lagde spørringene som ble brukt som utgangspunkt for å regne ut denne verdien, så man kan ikke være sikre på at disse resultatene lar seg generalisere til UBs reelle brukere. Eldre mennesker ville for eksempel brukt en del ord som studentene ikke hadde kommet på og vice versa, og det samme gjelder trolig mennesker med mer domenekunnskap. Det ville trolig ha vært en stor grad av overlapping, men i en reell situasjon vil det beste generelle resultatet kunne oppnås ved at de deltakerne som genererer innhold i form av tagger eller vokabularermer, i størst mulig grad representerer et utsnitt av den generelle populasjonen. Dette kunne kanskje gjort det mulig å oppnå tilsvarende resultater i Billedsamlingen.

Det kan tenkes at de som bruker Billedsamlingen har så god kjennskap til innholdet at de vet hvilke termer som er brukt og som bør brukes for å finne det man leter etter, men dette er ikke veldig sannsynlig. Man kan uansett ikke bare se nytteverdien i forhold til de som er brukere av Billedsamlingen i dag. Et av målene til den norske ABM-sektoren er å bli mer til-

gjengelig for publikum, og bedring av gjenfinningskvaliteten for nye brukere er helt klart et viktig virkemiddel i denne sammenhengen.

6.2 Gjennomførbarheten av brukergenerert innhold i kontrollerte vokabularer

Som nevnt, er verdien av å innføre brukergenerert innhold i kontrollerte vokabularer i stor grad en indikator på om ideen er gjennomførbar. Dette studiet har ikke lyktes i å kunne si så mye om ytelsen til brukergenererte termer i kontrollerte vokabularer sammenlignet med bilde-tagging, på grunn av en uforutsett skjevhet i utgangspunktet for de to gruppene som gjør det vanskelig å sammenligne de to gruppene direkte. Det studiet likevel har vist oss, er at det er gjennomførbart å lage et grensesnitt som gjør det mulig for brukere med enkelthet å sette seg inn i det som skal til for å legge til termer i et kontrollert vokabular og øke mangfoldet av ord som kan brukes for å finne igjen bilder, og dermed også recall, uten å knytte nye ord til bilder. Samtlige av deltakerne klarte å vise at de forstod hva oppgaven gikk ut på, og la til termer som i all hovedsak var riktige. Ved etter beste evne å ha forsøkt å gjøre forutsetningene for de to gruppene mer like i etterkant, ser det også ut til at forskjellene i precision og recall mellom oppdatering av et kontrollert vokabular og tagging er ganske små. Det er lite trolig at det vil være noen signifikant forskjell i precision og recall mellom de to metodene. Det kan likevel være interessant å se nærmere på om det kan være noen fordeler med oppdatering av kontrollerte vokabularer fremfor tagging.

6.3 Mulige fortrinn med brukergenerert innhold i kontrollerte vokabularer foran tagging

Resultatene fra eksperimentet indikerte at det var mulig å forbedre recall ganske mye ved å la brukere legge til termer i det kontrollerte vokabularet, selv om det var mange konsepter som fantes i testbildene, men som ikke var knyttet til bildene av UB, og dermed ikke kunne bygges på av eksperimentgruppen. Disse tallene ville sannsynligvis vært høyere hvis flere termer var knyttet til bildene, fordi deltakerne da ville hatt flere termer å knytte nye termer til. Man må også vurdere muligheten for at UB med vilje har utelatt enkelte konsepter, fordi de har vurdert dem som irrelevante. Hvis dette er tilfelle, vil metoden med brukergenererte termer kunne ha klare fordeler foran tagging, fordi UB ved bruk av et brukeroppdatert kontrollert vokabular

hvor de selv knyttet termer til bildene, fremdeles kunne bestemme hvilke konsepter de ønsker og ikke ønsker å knytte til hvert bilde.

Et annet poeng som kom frem gjennom forsøkene, er at precision ikke blir dårligere ved oppdatering av det kontrollerte vokabularet, men derimot nesten signifikant bedre enn ved tagging (p -verdien er nesten så lav som 0,05; se seksjon 5.2.1). En grunn til dette kan være at termene som deltakerne i eksperimentgruppen kunne knytte nye termer til representerte et objekt som deltakerne kunne se på bildet, slik at deltakerne hadde to ting å forholde seg til. Kontrollgruppen hadde derimot i mange tilfeller bare bildet å forholde seg til når de skulle legge til tagger, fordi mange av konseptene ikke var knyttet til bildene av UB. Usikkerhet om hva de egentlig så kan dermed ha vært grunnen til disse noe lavere verdiene for precision i kontrollgruppen²⁸. En av forutsetningene for dette er imidlertid at UB på forhånd har knyttet termer til bildene, noe som er tidkrevende og dyrt arbeid.

338 av 392 termer som ble lagt til av deltakerne var av typen emneord, 42 var av typen geografisk lokasjon, mens bare 12 var av typen person. Dette tyder på at det i de fleste tilfeller er termer av typen emneord som kan beskrives med mange forskjellige ord. I tillegg er emneord den typen termer som klart blir mest brukt. Mens hver geografiske lokasjon blir brukt i gjennomsnitt ca. 9 ganger, og hver person blir brukt ca. 12 ganger, blir hvert emneord brukt gjennomsnittlig ca. 44 ganger. Dette vil si at sjansen for at en ny tilknyttet term vil ha innvirkning på mange bilder, er klart størst hvis termen er knyttet til et emneord. I emneordtabellen er det 2590 ord totalt. Man vil trolig kunne øke recall ganske mye ved at brukere av Billedsamlingen går gjennom denne tabellen og knytter nye ord til termene. Dette er en stor jobb, men sett i forhold til økningen i recall det ser ut til å kunne føre til, er det trolig relativt lite som skal til for å få omtrent en dobling av recall. Dessuten vil mangfoldet i det kontrollerte vokabularet være gunstig også for bilder som blir lagt til i ettertid, og man trenger bare å gjøre denne jobben én gang for hver term i det kontrollerte vokabularet. Dette er ikke tilfelle hvis man baserer seg på tagging; da må hvert nye bilde som blir lagt til tagges av et

²⁸ Et eksempel på dette finner vi i bilde 7, hvor noen av deltakerne trodde det var en fjord de så på bildet. Hvis konseptet elv hadde vært representert blant UBs tilknyttede termer, ville det vært lettere for deltakerne å se at det ikke var en fjord, men en elv, som var avbildet.

visst antall personer for at recall skal bli god. I så fall er det over 20 000 bilder som må gjennomgås, og stadig nye blir lagt til.

6.4 Forslag til UB

Det kontrollerte vokabularet til Billedsamlingen brukes i dag for indekseringsformål og som søkegrunnlag for brukerne. Vokabularet er imidlertid best egnet for indeksering slik det nå er bygget opp, og fungerer dårlig for søkeformål. Det er basert på fagtermer, og det finnes ikke alternative ord til termene i vokabularet. Resultatene fra dette studiet indikerer at bildene i Billedsamlingen er ganske dårlig tilgjengelig for brukergruppen som har deltatt i studiet, og dette vil trolig gjelde også for andre brukere som ikke har kjennskap til hvilke termer som finnes i det kontrollerte vokabularet. Tilgjengelighet er et viktig satsningsområde for ABM-sektoren, og det bør derfor være mulig for brukere å finne bilder ved å bruke spørringer som virker opplagt uten at man kjenner til et snevert vokabular som er ment for eksperter.

Resultatene fra dette studiet indikerer at det lar seg gjøre å lage et grensesnitt som gjør det mulig for studenter å legge til termer i det kontrollerte vokabularet til Billedsamlingen uten å behøve mye opplæring på forhånd. Termene til deltakerne førte til en stor forbedring i antallet bilder som ble gjenfunnet av deltakerne fra den samme typen brukergruppe. En viktig forutsetning for å aktivere brukerne er at brukerne har en motivasjon for å delta. Det er vanskelig å si om Billedsamlingens brukere ville hatt interesse av og motivasjon for å bidra med innhold, og det er ikke sikkert at deltakelsen hadde vært stor nok til å gjøre noe særlig forskjell bare ved å gjøre et tilsvarende system som ble brukt i dette studiet tilgjengelig på verdensveven. Et problem i forhold til motivasjon for å delta når det gjelder Billedsamlingen er at innholdet ikke er brukernes eget innhold. En viktig grunn til at brukere vanligvis tagger bilder, er for å organisere sitt innhold for eget bruk, eller for å gjøre egenprodusert innhold mer tilgjengelig for andre brukere. Det at innholdet på Billedsamlingen ikke er skapt av brukerne selv, kan være et hinder for å kunne aktivere brukerne. Forskjellige brukere kan selvsagt ha forskjellig motivasjon for å delta. Noen bidrar fordi de ønsker å vise andre hva de kan. Noen bidrar fordi de ønsker at andre skal kunne dra nytte av det de har gjort. En annen motivasjonsfaktor er at brukerne synes det de holder på med er gøy. De fleste deltakerne i dette studiet sa at de syntes det var gøy å bruke systemet for å tagge bilder eller oppdatere det kontrollerte vokabularet. Man kan tilrettelegge systemet for å imøtekomme slike motivasjonsfaktorer, for eksempel ved å bruke et poengsystem, hvor brukerne kan komme inn på en topp

10-liste hvis de bidrar mye eller hvis bidragene deres fører til at flere brukere finner det de vil ha. Det er selvsagt også viktig med et system som på best mulig måte gjør det enkelt og moro for brukerne å legge til innhold.

Ifølge Trant (2006), har museer ofte miljøer av brukere som er villig til å bidra uten å være motivert av personlig vinning, og det er ikke usannsynlig at dette også kan gjelde for Billedsamlingen. Det kan derfor være en mulig løsning å forsøke å lage et system som gjør det mulig å ta imot bidrag fra hele brukermiljøet, og vurdere om motivasjonen i miljøet er stor nok til å basere seg på denne løsningen. Et annet alternativ er å betale en gruppe brukere for å gå systematisk gjennom hele det kontrollerte vokabularet, med hovedfokus på emneordene, og legge til flest mulig alternative termer. Dette vil øke mangfoldet av ord brukerne kan bruke for å finne bilder både med de termene som allerede er knyttet til bilder av UB og de som vil bli knyttet til bildene i fremtiden. Hvis det i tillegg blir lagt ned et arbeid i å knytte flere termer til bildene som allerede er indeksert, vil man kunne øke tilgjengeligheten til bildene betraktelig. Dette kan også gjøres av brukere, men det vil gå på bekostning av presisjonen til søkeresultatene. Alternativet er at UB fortsetter å gjøre denne jobben selv, noe som vil føre til mest presise resultater. På denne måten vil kunne dra stor nytte av det økte mangfoldet av ord som fører til treff, samtidig som de selv beholder en stor grad av kontroll. Det vil trolig være lettere å kontrollere brukergenererte termer i et kontrollert vokabular enn brukergenererte tagger, fordi antallet vil være mye lavere.

Et problem som alltid vil oppstå i større eller mindre grad når man benytter seg av brukergenerert innhold, er uønsket data. Dette kan oppstå på grunn av brukere som med vilje prøver å sabotere systemet, eller det kan oppstå ved at brukere med gode hensikter misforstår og gjør utilsiktede feil. Det er til en viss grad mulig å begrense slike problemer, ved for eksempel å lage datasystemer som fanger opp forsøk på sabotasje, eller ansatte som går gjennom innholdet. Det er imidlertid vanskelig å lage et perfekt slikt system, og det er kostbart å bruke ressurser på å gå gjennom innhold. Ved å bruke den sistnevnte modellen i forrige avsnitt, vil mengden av uønskede data trolig holdes på et lavt nivå. For UB betyr det at de i stor grad kunne fortsatt å gjøre jobben sin slik de har gjort til nå, og oppnådd en mye mer tilgjengelig tjeneste enn de har i dag. Dette er tross alt en viktig satsning i ABM-sektoren og i arbeidet med et nasjonalt digitalt bibliotek.

6.5 Konklusjoner

Resultatene fra studiet peker mot at de to formene for brukervedvirkning som er sammenlignet i dette studiet (oppdatering av kontrollert vokabular og tagging) har en viktig felles egenskap: I begge formene oppnår man et økt mangfold av ord man kan bruke for å beskrive hvert enkelt konsept. Dette er avgjørende for at vanlige brukere skal kunne finne flest mulig bilder som er relevante for de ordene de søker etter, noe som kommer tydelig frem av økningen i recall etter at eksperimentgruppen har økt mangfoldet i det kontrollerte vokabularet. Selv om antallet konsepter som ble beskrevet forble det samme, førte økningen i mangfold i det kontrollerte vokabularet alene til at recall ble mer enn doblet etter at 10 deltakere hadde lagt til sine termer.

Målet for dette studiet har vært å undersøke om det var gjennomførbart og verdifullt å innføre brukergenerert innhold i kontrollerte vokabularer. Det har ikke lyktes å vise direkte fordeler med denne fremgangsmåten fremfor tagging, men resultatene indikerer at det er mulig å øke recall drastisk *uten at brukerne behøver å knytte noen nye konsepter til bilder*. Dette kan åpne opp for at man kan dra nytte av brukere for å øke mangfoldet av ord som er mulig å bruke for å finne frem til bilder, samtidig som man beholder kontrollen og presisjonen man får ved å la eksperter avgjøre hvilke konsepter som er representert i bildene. I stedet for å operere med en folksonomi uten relasjoner og andre strukturer, vil man altså kunne kombinere mangfoldet man får i en folksonomi med strukturene i et kontrollert vokabular. Hvis man har et rikholdig og godt oppbygd kontrollert vokabular som utgangspunkt, er det mulig at eksperter bare trenger å knytte én term fra det kontrollerte vokabularet til hvert konsept i hvert bilde, og så får man alle bredere termer og synonymer som brukerne har skapt med «på kjøpet».

7 Videre forskning

Resultatene fra dette studiet indikerer at brukergenerert innhold er en god måte å forbedre bildesøk, og da særlig recall. Den tradisjonelle taggemetoden kommer best ut basert på observasjonene som er gjort, men dette studiet har gitt resultater som antyder at også brukergenererte termer i et kontrollert vokabular har potensial til å være nyttig i visse sammenhenger. Videre forskning innenfor de samme emnene ville kunne gitt et bedre svar på om det er mulig å oppnå høyere presisjon uten at recall blir lavere, ved å benytte seg av en modell hvor eksperter knytter termer til bilder som er oppdatert av brukerne enn en modell hvor man utelukkende baserer seg på brukergenererte tagger. Ved å sørge for at flest mulig konsepter er representert blant termene til ekspertene, vil det trolig være mulig å kunne sammenligne disse modellene på en god måte, noe som vil åpne for å kunne si noe mer om potensielle nytteverdier av å kombinere folksonomier med kontrollerte vokabularer.

Referanser

- ABM-utvikling (2006a). *ABM-skrift#32: Kulturarven til alle*. Oslo: ABM-utvikling.
- ABM-Utvikling (2006b). *Bibliotekreform 2014 - Del I*. Oslo: ABM-Utvikling.
- ANSI/NISO (2005). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, Maryland, U.S.A.: the National Information Standards Organization.
- Baeza-Yates, R., A., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*: Addison-Wesley Longman Publishing Co., Inc.
- Bell, D. (2005). *Software engineering for students: a programming approach*: Addison-Wesley.
- Borgman, C. L. (2000). *From Gutenberg to the global information infrastructure: access to information in the networked world*. MIT Press.
- Brygfjeld, S. A., Husby, O., Sølvsberg, I. T., & Vognlid, L. K. (2001). *Forskning på digitale bibliotek i Norge*. Oslo: Nasjonalbiblioteket.
- Bughin, J. R. (2007). *How companies can make the most of user-generated content*. Brüssel: McKinsey & Company.
- Chun, S., Cherry, R., Hiwiller, D., Trant, J., & Wyman, B. (2006). Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums. <http://www.archimuse.com/mw2006/papers/wyman/wyman.html>. Sist sett 22.05.2009.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & Analysis Issues for Field Settings*: Houghton-Mifflin Company.
- Cozby, P. C. (2007). *Methods in Behavioral Research (Ninth Edition)*. New York: McGraw-Hill.
- Elgesem, D., & Nordbotten, J. (2007). *The Role of Context in Image Interpretation*. Paper presented at the 2nd International Workshop on Context-Based Information Retrieval.
- Elmasri, R., & Navathe, S. B. (2004). *Fundamentals of database systems*. Boston, Mass.: Pearson/Addison Wesley.
- Foskett, D. J. (1997). Thesaurus. In K. Sparck Jones & P. Willett (Eds.), *Readings in Information Retrieval* (pp. 111-134). San Fransisco, CA, USA: Morgan Kaufmann Publishers, Inc.
- Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some Implementations of the Boxplot. *The American Statistician*, 43(1), 50-54.
- Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., et al. (2006). *Why do tagging systems work?* Paper presented at the CHI '06 extended abstracts on Human factors in computing systems.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Commun. ACM*, 30(11), 964-971.
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of information science*, 32(2), 198-208.
- Goodwin, C. J. (2005). *Research in psychology: methods and design*. New York: John Wiley.
- Hartvedt, C. (2007). *Utilizing context in ranking results from distributed image retrieval: the CAIRANK prototype*. UiB, Bergen.

- Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington, DC.
- Olsen, M. (2004). *Integrasjon og bruk av gazetteers og tesauri i digitale bibliotek: Søk og gjenfinning via geografisk referert informasjon*. Unpublished Master's Thesis, Norges teknisk-naturvitenskapelige universitet, Trondheim.
- Peters, I., & Stock, W. G. (2007). *Folksonomy and information retrieval*. Paper presented at the ASIS&T annual meeting.
- Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation*, 10(1), 39-62.
- Sommerville, I. (2007). *Software Engineering*.
- Thurman, N. J. (2008). *Forums for Citizen Journalists? Adoption of User Generated Content Initiatives by Online News Media* (Vol. 10): Social Science Research Network.
- Trant, J. (2006). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia*, 12(1), 83-105.
- Universitetsbiblioteket i Bergen (2008). Om Billedsamlingen. Sist sett 14.08.2008.2008.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2000). *Experimentation in software engineering: an introduction*. Boston: Kluwer.

Appendiks A: Testbilder og brukerskapte termer/tagger

Bilde 1



Originale termer:

Emneord: dampskip; fjord; naust; robåt; stovehus

Geografiske lokasjoner: "bergen og omegn"; osterøy

Personer: knudsen, knud

Bruerskapte termer:

Emneord: bebyggelse; bolig; bygning; båt; båthus; dampbåt; fremkomstmiddel; hus; konstruksjon; maritim bebyggelse; natur; nøst; sjø; sjøhus; skip; småbåt; transportmiddel; vann

Geografiske lokasjoner: bergensdistriktet; bergensområdet; distrikt; fjord-norge; hordaland; innland; norges største innlandsøy; vestlandet

(Ingen termer knyttet til personer)

Tagger: bebyggelse; bygd; båt; dampbåt; dampskip; fiske; fjell; fjord; fjordbebyggelse; fjøre; havn; himmel; hus; master; naust; rorbu; sjø; sjøfart; sjøkant; skip; skipstrafikk; speilbilde; steinstrand; stille vann; strand; trebåt; vik

Bilde 2



Originale termer:

Emneord: fangband; fjord; hardangerdrakt; hardangerskaut; mann; naust; skautekone, hardanger

Geografiske lokasjoner: "hardanger"; odda; sørfjorden, hardanger

Personer: knudsen, knud

Bruerskapte termer:

Emneord: bebyggelse; bunad; bygning; båthus; dame; drakt; festplagg; finstas; hardangerbunad; herre; hodeplagg; hus; kone; konstruksjon; kvinne; maritim bebyggelse; menneske; mennesker; nasjonaldrakt; natur; nøst; person; sjø; sjøhus; skaut; vann

Geografiske lokasjoner: fjord; hardanger; hordaland; tettsted; vestlandet; vestlandsbygd

(Ingen termer knyttet til personer)

Tagger: 1700- tallet; alvorlig; bringeduk; bryllup; budeie; bunad; damer; drakt; fest; fjord; fjordbebyggelse; folkeliv ved fjorden; gammelt hus; gård; gårdsbruk; hardangerbunad; kvinne; kvinner; mann; nasjonaldrakt; naust; oppstilt foto; portrett; sjøbod; skaut; stabbur; uklar mann i bakgrunnen som går inn en dør

Bilde 3



Originale termer:

Emneord: "mekaniske verksteder"; "motorforretninger"; fabrikk; mekanisk verksted; skøyte

Geografiske lokasjoner: solheimsviken

Personer: atelier kk; sunde og søyland

Bruerskapte termer:

Emneord: arbeidsplass; butikk; bygning; båt; forretning; forretninger; handelsforretning; industribygg; konstruksjon; skip; transportmiddel; verksted

Geografiske lokasjoner: fjord; puddefjorden; sjø; årstad bydel

Personer: firma; organisasjon; virksomhet

Tagger: arbeid; brygge; båt; fabrikk; fiske; fjell; fjord; havn; holzapfel; industri; kai; kran; sjø; stort naust; verksted

Bilde 4



Originale termer:

Emneord: hardangerdrakt;

hardangerskaut

(Ingen geografiske lokasjoner)

Personer: dronning maud; nyblin, karl anton

Bruerskapte termer:

Emneord: bunad; drakt; festplagg;

finstas; hardangerbunad; hodeplagg;

nasjonaldrakt; skaut

(Ingen termer knyttet til geografiske lokasjoner)

Personer: dame; kongelig; kvinne;

kvinnelig monark; monark

Tagger: 1905; bunad; bunadsskjorte;

dame; drakt; dronning; hodeplagg;

klær; kongefamilie; kongelig;

kongelige; kvinne; kvinne i bunad;

nasjonaldrakt; nasjonsaldrakt; norges

dronning; portrett; profil; profilportrett;

skaut; slje; stakk; sølge; sølje

Bilde 5



Originale termer:

Emneord: fjord; fjordabåt; kaianlegg; karjol; naust; skysskar, vognmann, kusk, sjåfør

Geografiske lokasjoner: "sogn"; lærdal; lærdalsfjorden

Personer: knudsen, knud

Bruerskapte termer:

Emneord: bebyggelse; brygge; bygning; båt; båthus; dampbåt; fremkomstmiddel; hestekjerre; hus; kai; kjerre; kjøretøy; konstruksjon; maritim bebyggelse; natur; nøst; sjø; sjøhus; skip; transportmiddel; vann; vogn

Geografiske lokasjoner: fjord; fylke; norske fjorder; sogn og fjordane; vestlandet

(Ingen termer knyttet til personer)

Tagger: bergen; bil; brygge; båt; dampbåt; dampskip; fjell; fjord; fremkomstmiddel; hest; hest og kjerre; hest og kjerre; hest og vogn; hestekjerre; hus; hvitt hus; kai; kjerre; landevei; naust; skip; skipstrafikk; skyss; varetransport; vogn

Bilde 6



Originale termer:

Emneord: fiskebåt; havnebod; høyhus; motorbåt

Geografiske lokasjoner: c. sundts gate; holbergsallmenningen

Personer: brosing, gustav

Bruerskapte termer:

Emneord: bebyggelse; blokk; bod; bryggehus; bryggen; bygning; båt; båthus; fangstfartøy; framkomsmiddel; fremkomstmiddel; havn; hus; høyblokk; konstruksjon; sjøbod; skip; transportmiddel

Geografiske lokasjoner: bergen sentrum; bryggen; by

(Ingen termer knyttet til personer)

Tagger: bergen; bil; blokk; blokker; bod; boligblokk; brygge; bygning; byhavn; båt; båter; fiske; fiskebåt; fisker; fiskeskøyte; gammelt høyhus; handel; hatt; havn; husblokk; høyblokk; kai; naust; nordnes; sjøbod; skip; trehus; vinduer

Bilde 7



Originale termer:

Emneord: bro; kvinne; spann; steingjerde; stovehus; telemarksdrakt; trekjørel

Geografiske lokasjoner: grungedal

Personer: knudsen, knud

Bruerskapte termer:

Emneord: bebyggelse; bolig; bru; bunad; bygning; dame; drakt; gjerde; hus; konstruksjon; menneske; person; steingard; steinmur; telemarksbunad

Geografiske lokasjoner: dal; østlandet

(Ingen termer knyttet til personer)

Tagger: bro; budeie; budeier; bunad; bøtte; dagslys; dal; elv; fjell; fjell i bakgrunnen; fjellet; fjord; fjøs; folkedrakt; gård; gårdsbruk; kvinne; kvinner; løe; mennesker; naust; redskap; seter; skaut; stabbur; stakk; steiner; steingard; steingjerde; steinmur; steinrøys; støl; telemark; torvtak

Bilde 8



Originale termer:

Emneord: bolighus; flagg; flaggstang; forretninger; varebil

Geografiske lokasjoner: kong oscars gate

Personer: brosing, gustav

Bruerskapte termer:

Emneord: bebyggelse; bil; bolig; butikk; butikker; bygning; firmaer; handelshus; hus; kjøretøy; konstruksjon;

nasjonalsymbol; nasjonalt ikon; nasjonalt symbol; stang; symbol; virksomhet

Geografiske lokasjoner: b; by; bygate; bygate i bergen

(Ingen termer knyttet til personer)

Tagger: balkong; bergen; bil; brostein; brusteinsgate; butikk; butikker; bybebyggelse; bygate; dag; domkirken;

flagg; forbigående; fotgjenger; fotgjenger; gate; halv stang; handel; hus; kirketårn; kong oscars gate; sol;

sorg; trehus; varebil; vågen

Bilde 9



Originale termer:

Emneord: 1928 landsutstillingen i bergen; dress; eldre; eldre parportrett; herrehatt; kåpe; par; skaut; slips; urlenke

Geografiske lokasjoner: landsutstillingen 1928

Personer: atelier kk

Bruerskapte termer:

Emneord: antrekk; ektepar; eldre portrett; finklær; gamle; gamle personer; gammel; hatt; herreklær; hodeplagg; kjetting; klesplagg; lenke; mann og kvinne; mellomkrigstid; portrett; sjal; verdigjenstander; ytterklær; yttertøy

Geografiske lokasjoner: utstilling

(Ingen termer knyttet til personer)

Tagger: begravelse; benk; dame; dobbel; dress; ektepar; eldre ektepar; eldre mennesker; fotograf; gammel; gammelt ektepar; hatt; hender; isse; kvinne; kåpe; mann; mørke klær; oppstilt foto; portrett; skaut; skjegg; slips; to eldre par; tvilling; tvillinger

Bilde 10



Originale termer:

Emneord: 1928
landsutstillingen i bergen;
damehansker; damehatt;
damesko; eldre; eldre
kvinneportrett; kvinne; kåpe;
stola

Geografiske lokasjoner:
landsutstillingen 1928

Personer: atelier kk

Bruerskapte termer:

Emneord: bilde; dame;
fotografi; fottøy; gamle; gamle
personer; gammel; hanske;
hansker; hatt; hodeplagg;
klesplagg; kvinnehansker;
kvinnehatt; kvinnelig
hodeplagg; kvinnevanner;
mellomkrigstid; menneske;
pels; person; portrett; sko;
vanter; vinterklær; ytterklær;
yttertøy

Geografiske lokasjoner:
utstilling

(Ingen termer knyttet til
personer)

Tagger: anheng; bakgrunn; benk; dame; eldre kvinne; fotograf; hatt; kvinne; kåpe; malt vegg; mink; monokkel;
oppstilt foto; pels; portrett; portrett av kvinne; rev; revpels; sko

Appendiks B: Samtykkeskjema

Mitt navn er Øyvind Kristiansen, og jeg holder på å skrive en masteroppgave som har arbeidstittelen «Brukerskapt innhold i et kontrollert vokabular». I oppgaven bruker jeg bilder fra Universitetsbiblioteket i Bergen sin Billedsamling, og jeg undersøker blant annet om det er mulig, ved hjelp av brukere, å forbedre deres system for gjenfinning av bilder.

Hvis du har noen spørsmål, kan jeg kontaktes på e-post hr.kristiansen@gmail.com eller telefon 99 48 27 74.

Før vi starter vil jeg understreke at

- din deltakelse er helt frivillig
- du kan nekte å svare på hvilket som helst spørsmål
- du kan trekke deg fra undersøkelsen når som helst

Undersøkelsen er konfidensiell, og vil bare være tilgjengelig for meg. Utdrag fra undersøkelsen kan bli del av oppgaven, men verken ditt navn eller andre kjennetegn som kan identifisere deg vil bli tatt med i rapporten.

Vennligst undertegn dette skjemaet for å vise at du har lest og forstått innholdet.

Dato

Underskrift

Appendiks C: Oppgavene til eksperimentgruppen

Oppgave 1

Se på bildene. Bruk programmet til å prøve å finne bildene i programmet. Det søkes i de 10 eksempelbildene, og det kan hende du finner andre bilder enn det du søker etter. Prøv med minst 3 søkeord for hvert bilde, selv om du finner bildet før du har brukt 3 forsøk (hvis du ikke kommer på flere ord, trenger du selvsagt ikke å prøve flere ganger).

Oppgave 2

Se på eksempelbildet først.

Se om du finner noen ord som du tror du kunne knyttet nye ord til – synonymer eller bredere termer. Bruk programmet til å legge til de nye termene. Du kan når som helst spørre om hjelp.

Hierarki – bredere/smalere termer – synonymer: en kort forklaring

-klær

-mannsklær

-dress

tweed-dress

-herrehatt

bowlerhatt

flosshatt

- Bredere term: Herrehatt er en bredere term i forhold til bowlerhatt – den fanger om flere objekter (bl.a. flosshatt). Det er alltid mulig å si at en bowlerhatt *er en* herrehatt. Motsatt er bowlerhatt en smalere term til herrehatt. Det er *ikke* alltid mulig å si at en herrehatt er en bowlerhatt.
- Synonymer: Ord som har den samme betydningen. For eksempel oksygen og surstoff, eller smug og smau.

Det er viktig at du, før du trykker på ”send”, tenker over om logikken i den setningen du legger til, stemmer *i alle tilfeller*. Selv om en mann på et bilde kan være feier, er ikke en mann alltid en feier. I dette eksemplet gjelder det samme motsatt vei – selv om en feier på et bilde er mann, er ikke en feier alltid en mann. Et annet eksempel er et bilde av et damplokomotiv. Selv om lokomotivet på bildet er et damplokomotiv, er ikke et lokomotiv et damplokomotiv i alle tilfeller – det finnes andre typer lokomotiver, som for eksempel elektriske lokomotiver eller dieseldrevne lokomotiver. Men et damplokomotiv er alltid et lokomotiv.

Eksempel på bredere term for termen håndkjerre:

Håndkjerre ER (alltid) EN kjerre

I programmet (se eksempelbildet på skjermen):

Legg til termrelasjon 

Håndkjerre er en / et eller kan også skrives som:

Eksempel på synonym for termen smug:

Smug KAN OGSÅ SKRIVES SOM smau

I programmet:

Legg til termrelasjon 

Smug er en / et eller kan også skrives som:

Appendiks D: Oppgavene til kontrollgruppen

Oppgave 1

Se på bildene. Bruk programmet til å prøve å finne bildene i programmet. Det søkes i de 10 eksempelbildene, og det kan hende du finner andre bilder enn det du søker etter. Prøv med minst 3 søkeord for hvert bilde, selv om du finner bildet før du har brukt 3 forsøk (hvis du ikke kommer på flere ord, trenger du selvsagt ikke å prøve flere ganger).

Oppgave 2

Se på eksempelbildet først.

Bruk programmet til å legge til *nøkkelord (tags)* til de 10 bildene. Nøkkelordene kan bestå av enkeltord eller sammensatte ord, og forklarer hva bildet inneholder. Legg til så mange tags du ønsker til hvert bilde. Du kan når som helst spørre om hjelp.

Appendiks E: Utdrag fra kildekode

```
/**
 * Generering av spørringer for søk i tagger og
 * kontrollert vokabular.
 */
require_once(Felles::getRot() . '/klasser/Bildeliste.php');
require_once('klasser/Tekst.php');
require_once 'klasser/Sokeord.php';

/*
 * Klargjør søkestrengen
 */

// Fjerner mellomrom i begynnelsen og slutten av spørringen.
$sokestreng = trim($sokestreng);
// Gjør backslash (\) om til dobbel backslash (\\)
$sokestreng = str_replace(chr(92), chr(92) . chr(92), $sokestreng);
// Gjør ett enkelt hermetegn (') om til to enkle hermetegn (')
$sokestreng = preg_replace('/\\'/, '\\\\', $sokestreng);

/*
 * Registrerer spørringen i databasen til bruk i statistikk /
 * utregning av precision og recall
 *
 * Informantid 1 blir brukt til testing, og er ikke
 * en reell informant. Derfor skal ikke søk utført av
 * informant med ID=1 bli registrert.
 */
if($informantid != 1) {
    $sporrning = "INSERT INTO sokeord_tbl(ord, bruker) VALUES ('$sokestreng'
, $informantid)";
    try {
        $billedb->query($sporrning);
    } catch(MySqlException $e) {
        if($e->getMysqlErrno() != 1062) {
            Feilmelding::leggTilLinje("Kunne ikke oppdatere sokeord_tbl" .
mysql_error());
        }
    }
}

$sokeord_orig = behandle($sokestreng);
$sokeord = array();
/*
 * Oppretter tabell med ett Søkeord-objekt for hvert ord i spørringen.
 */
for($i = 0; $i < count($sokeord_orig); $i++) {
    $sokeord[$i] = new Sokeord($sokeord_orig[$i]);
}

/*
 * Lager et Søkeord-objekt for hele spørringen.
 */
$hele_sokestrengen_sokeord = new Sokeord($sokestreng);

/*
 * Lager en kommaseparert streng med felter som skal
```

```

* tas med i spørringen (emneord, geografisk lokasjon,
* person, tagger osv.)
*/
$felterstreng = '';
for($i = 0; $i < count($feltnavn); $i++) {
    if(erChecked($feltnavn[$i]) && $feltnavn[$i] != 'signatur') {
        $felterstreng .= $feltnavn[$i] . ', ';
    }
}
/*
* Fjerner det siste tegnet i strengen, så strengen ikke
* slutter med komma.
*/
$felterstreng = Tekst::fjernSiste(Tekst::fjernSiste($felterstreng));

if(count($sokeord) != 0) {
    $motivider = finnMotivider();
    $antall = count($motivider);
    if($antall > 0) {
        $html = hentBilder($motivider);
    } else {
        Feilmelding::leggTilLinje("Fant ingen bilder.");
    }
} else {
    Feilmelding::leggTilLinje("Du må skrive et eller flere søkeord.");
}

/*
* Lager SQL-spørring ut fra brukerens spørring og
* finner motivid til relevante bilder.
*/
function finnMotivider() {
    global $billedb;
    global $sokeord;
    global $hele_sokestrengen_sokeord;

    /* MotivID til de 10 testbildene */
    $testsamling_motivider = array(19838, 34377, 44024, 49192, 21523, 54072
, 19981, 54986, 49712, 56967);

    /* Lager kommaseparert streng med de 10 testbildene. */
    $testsamling_streng = Tekst::lagKommaSeparertStreng($testsamling_motivi
der);

    /* Genererer SQL-spørring. */
    $sporrning = "select * from indeks_tbl
        where ( signatur regexp '[:<:~]' . Tekst::stjernerTilProsent($hele
_sokestrengen_sokeord->getOrd()) . "[:>:~]"

        or (";

    if(trim($hele_sokestrengen_sokeord->getOrd()) != trim($sokeord[0]-
>getOrd())) {
        $sporrning .= finnAlleAlternativer($hele_sokestrengen_sokeord);
        $sporrning .= "
        ) or (";
    }

    for($i = 0; $i < count($sokeord); $i++) {
        $sporrning .= finnAlleAlternativer($sokeord[$i]);
    }
}

```

```

        if($i + 1 < count($sokeord)) {
            $sporning .= ' and ';
        }
    }

    $sporning .= ")))";
    $sporning .= "
        and mo_motivid in ($testsamling_streng);";

    //echo nl2br($sporning); exit;

    try{
        $resultat = $bildedb->query($sporning);
    } catch(Exception $e) {
        Feilmelding::leggTilLinje($e->getMessage());
    }
    $motivider = array();
    for($i = 0; $i < count($resultat); $i++) {
        $motivider[] = $resultat[$i]->mo_motivid;
    }
    return $motivider;
}

/*
 * Finner alle alternative termer til et ord og dets
 * stammer ved å gå gjennom det brukerskapte kontrollerte
 * vokabularet.
 */
function finnAlleAlternativer($ord) {
    $ft_min_word_len = 4;
    $sporning = "(";

    if(strlen($ord->getOrd()) <= $ft_min_word_len) {
        $sporning .= regexp($ord->getOrd());
    } else {
        $sporning .= match($ord->getOrd());
    }

    if($stammer = $ord->getStammer()) {
        for($j = 0; $j < count($stammer); $j++) {
            $sporning .= ' or ';
            if(strlen($stammer[$j]) <= $ft_min_word_len) {
                $sporning .= regexp($stammer[$j]);
            } else {
                $sporning .= match($stammer[$j]);
            }
        }
    }

    if(erChecked('ordbok')) {
        if($ord->getVarianter()) {
            $varianter = $ord->getVarianter();
            for($j = 0; $j < count($varianter); $j++) {
                $sporning .= ' or ';
                if(strlen($varianter[$j]) <= $ft_min_word_len) {
                    $sporning .= regexp($varianter[$j]);
                } else {
                    $sporning .= match($varianter[$j]);
                }
            }
        }
    }
}

```

```

    }
    if($ord->getAlternativer()) {
        $alternativer = $ord->getAlternativer();
        for($j = 0; $j < count($alternativer); $j++) {
            $sporrying .= ' or ';
            if(strlen($alternativer[$j]-
>getNavn()) <= $ft_min_word_len) {
                $sporrying .= regexp($alternativer[$j]->getNavn());
            } else {
                $sporrying .= match($alternativer[$j]->getNavn());
            }
        }
    }
    if($ord->getStammealternativer()) {
        $stammeAlternativer = $ord->getStammealternativer();
        for($j = 0; $j < count($stammeAlternativer); $j++) {
            $sporrying .= ' or ';
            if(strlen($stammeAlternativer[$j]-
>getNavn()) <= $ft_min_word_len) {
                $sporrying .= regexp($stammeAlternativer[$j]-
>getNavn());
            } else {
                $sporrying .= match($stammeAlternativer[$j]-
>getNavn());
            }
        }
    }
}
$sporrying .= ") \n\n";
return $sporrying;
}

/*
 * Henter informasjon om bildene som matcher
 * brukerens spørring. Konstruert med tanke på
 * mulighet for å kunne velge til/fra-dato og for
 * søk etter mer enn 10 bilder.
 */
function hentBilder($motivider) {
    global $billedb;
    global $start;
    global $vis_antall;

    $fra = '0000-01-01';
    $til = '9999-12-31';

    $motivider_streng = implode(',', $motivider);
    $motivider_sporrying = "
        SELECT DISTINCT mo_motivid
        FROM tblmotiv
        WHERE (
            (mo_fradato >= '$fra'
            AND mo_tildato <= '$til')
            OR (mo_fradato IS NULL OR mo_tildato IS NULL)
            )AND mo_motivid in($motivider_streng) ORDER BY mo_motivsignatur;
    ";

    //echo nl2br($motivider_sporrying); exit;

    try {
        $motivider_resultat = $billedb->query($motivider_sporrying);
    }
}

```

```

    } catch(Exception $e) {
        Feilmelding::leggTilLinje($e->getMessage());
    }
    $motivider = array();

    for($i = $start * $vis_antall; $i < ($start * $vis_antall) + $vis_antall; $i++) {
        if(isset($motivider_resultat[$i]->mo_motivid)) {
            $motivider[] = $motivider_resultat[$i]->mo_motivid;
        }
    }
    $antall = count($motivider_resultat);
    $sokeinfofelt = lagSokeinfoFelt($antall);
    $html = $sokeinfofelt;

    if($antall > 0) {
        $bildeliste = new Bildeliste($motivider, 'm.mo_motivsignatur');
        $html .= $bildeliste->toHtml();
    }

    if($antall > 0) {
        $html .= $sokeinfofelt;
    }

    return $html;
}

/* For hvert søkeord: returnerer [felt] LIKE [søkeord] OR */
function _or($felt, $sokeord) {
    $streng = '';
    for($i = 0; $i < count($sokeord); $i++) {
        $streng .= "\n$felt LIKE '%" . $sokeord[$i] . "%'";
        if(($i + 1) != count($sokeord)) {
            $streng .= " OR";
        }
    }
    return $streng;
}

/* For hvert søkeord: returnerer [felt] LIKE [søkeord] AND */
function _and($felt, $sokeord) {
    $streng = '';
    for($i = 0; $i < count($sokeord); $i++) {
        $streng .= "\n$felt LIKE '%" . $sokeord[$i] . "%'";
        if(($i + 1) != count($sokeord)) {
            $streng .= " AND";
        }
    }
    return $streng;
}

/* Skiller ut enkeltord */
function behandle($stekst) {
    $sokeord = array();

    /* Finn tekst som står mellom doble hermetegn. */
    while(eregi('("[^"]*)"', $stekst, $treff)) {
        $stekst = str_replace($treff[1], ' ', $stekst);
        $sokeord[] = eregi_replace('"', '', $treff[1]);
    }
}

```



```

    }

    /* Legg til resterende ord */
    $stekst = trim($stekst);
    $stekst = htmlentities($stekst, null, 'utf-8');
    $treff2 = array();

    /* Skiller ut ord som står mellom to word boundaries eller '. */
    if($stekst != "") {
        preg_match_all('/(\x3b|\b|\x27)[\w&\x3b\x27]+(\x3b|\b|\x27)/u', $t
ekst, $treff2);
        $treff2 = $treff2[0];
    }
    for($i = 0; $i < count($treff2); $i++) {
        $treff2[$i] = html_entity_decode($treff2[$i], null, 'UTF-8');
    }

    $sokeord = @array_merge($treff2, $sokeord);
    $sokeord = Felles::fjernTomme($sokeord);
    return $sokeord;
}

/*
 * Brukes av den utvidede applikasjonen for å
 * generere årstall
 */
function leggTilNuller($stall) {
    $stallengde = strlen($stall);
    switch($stallengde) {
        case 1:
            $stall = "000" . $stall;
            break;
        case 2:
            $stall = "00" . $stall;
            break;
        case 3:
            $stall = "0" . $stall;
            break;
    }
    return $stall;
}

/* Å-safe strtolower */
function myStrtolower($str) {
    $str = preg_replace('/\//', '\\\\', $str);
    global $billedb;
    try {
        $resultat = $billedb->query("SELECT LOWER('$str') AS str;");
    } catch (MySQLException $e) {
        Feilmelding::leggTilLinje($e->getMessage());
    }
    return $resultat[0]->str;
}

/*
 * Generer felt med informasjon om antall treff.
 */
function lagSokeinfoFelt($antall) {
    global $vis_antall;
    global $start;

```

```

$html = "<div class=\"sokeinfo\"><p>";

if(ceil($antall / $vis_antall) > 1) {
    $html .= $antall . " treff fordelt p&aring; " . ceil($antall / $vis
_antall) . " sider. Er n&aring; p&aring; side " . ($start + 1) . ".<br />\n
";
    $html .= "G&aring; til side: ";
    for($i = 0; $i < ceil($antall / $vis_antall); $i++) {
        $html .= "<a href=\"javascript:\"" onclick=\"setStart($i); $('s
ok').submit();\">" . ($i + 1) . "</a>";
        if($i < (ceil($antall / $vis_antall)) - 1) {
            $html .= " | ";
        }
    }
} else {
    $html .= "$antall treff.";
}
$html .= "</p>";
$html .= "</div>\n";

return $html;
}

/*
* Sjekker om et av de ordene i tabellen
* har en streng som er lengre enn "Full text
* minimum word length". I så fall skal MySQLs
* FULLTEXT-søk brukes. Hvis ikke må MySQLs LIKE
* brukes.
*/
function harStrengSomErLengreEnnFtmwl($array) {
    global $ft_min_word_len;
    for($i = 0; $i < count($array); $i++) {
        if(strlen($array[$i]->getNavn()) > $ft_min_word_len) {
            return true;
        }
    }
    return false;
}

/* Genererer delstreng med konkatenering av strengen $str. */
function regexp($str) {
    global $felterstreng;
    return "\nconcat_ws('-',-
', $felterstreng) regexp '[:<:]]{$str}[:>:]'";
}

/* Genererer delstreng med MySQL MATCH av strengen $str. */
function match($str) {
    $str = str_replace("'", '', $str);
    global $felterstreng;
    return "\nmatch($felterstreng) against ('\"$str\"' in boolean mode)";
}

```

```

/** Klasse for en informant i eksperimentet. */
class Informant {
    private $id, $alder, $kjonn;

    public function __construct() {
        $arglist = func_get_args ();
        /**
         * Hvis __construct-funksjonen blir kalt med ett argument,
         * skal dette argumentet være ID'en til en allerede
         * eksisterende informant som skal logge inn i systemet.
         *
         * Hvis __construct-funksjonen blir kalt med to argumenter,
         * skal en ny informant opprettes. Argumentene skal da bestå
         * av alder og kjønn.
         */

        // innlogging av eksisterende informant
        if (func_num_args () == 1) {
            $id = $arglist[0];
            if(!is_numeric($id)) {
                Feilmelding::leggTilLinje('ID må være et tall.');
```

```

global $bildedb;
$sporning = "SELECT * FROM informant_tbl
WHERE id = {$this->id} LIMIT 1;";
try {
    $resultat = $bildedb->query ( $sporning );
    if(count($resultat) == 1) {
        $this->alder = $resultat[0]->alder;
        $this->kjonn = $resultat[0]->kjonn;
    } else {
        Feilmelding::leggTilLinje("Fant ikke informant.");
    }
} catch ( Exception $e ) {
    Feilmelding::leggTilLinje ( $e->getMessage ( ) );
}
}
}

```

```

/** Klasse for presentering av feilmeldinger. */
class Feilmelding {

    /* Tabell for å holde genererte feilmeldinger. */
    private static $linjer = array ( );

    /*
     * Boolsk verdi for som forteller om det har blitt
     * generert noen feil i denne kjøringen.
     */
    private static $feilFunnet = false;

    /* Tom __construct-funksjon */
    public function __construct() {
    }

    public function getFeilmeldinger() {
        return Feilmelding::$linjer;
    }

    /*
     * Legger til en ny feilmelding. Hver feilmelding / linje
     * er et felt i feilmeldingstabellen.
     */
    public static function leggTilLinje($tekst) {
        Feilmelding::$linjer [] = $tekst;
        Feilmelding::$feilFunnet = true;
    }

    public static function feilFunnet() {
        return Feilmelding::$feilFunnet;
    }

    /*
     * Genererer HTML-kode for å presentere feilmelding(er)
     * for brukeren.
     */
    public static function toHtml() {
        if (count ( Feilmelding::$linjer ) > 0) {
            $html = "<div id=\"feilmelding\">";
            $html .= "<ul>";
            foreach ( Feilmelding::$linjer as $linje ) {
                $html .= "<li>" . $linje . "</li>\n";
            }
            $html .= "</ul></div>\n";
            return $html;
        }
        return null;
    }
}

```

Appendiks F: Forhåndstilknyttede termer (av UB)

Stovehus	Høyhus
Naust	Bolighus
Fjord	Flagg
Dampskip	Varebil
Robåt	Flaggstang
Kvinne	Forretninger
Steingjerde	Damehatt
Bro	Damehansker
Trekjørel	Eldre kvinneportrett
Spann	Stola
Telemarksdrakt	Damesko
Karjol	Osterøy
Kaianlegg	"Bergen og omegn"
Skysskar, vognmann, kusk, sjåfør	Grungedal
Fjordabåt	Lærdal
Mann	"Sogn"
Hardangerdrakt	Lærdalsfjorden
Fangband	"Hardanger"
Skautekone, Hardanger	Odda
Hardangerskaut	Sørfjorden, Hardanger
Fabrikk	Solheimsviken
Mekanisk verksted	Landsutstillingen 1928
Skøyte	Holbergsallmenningen
"Mekaniske verksteder"	C. Sundts gate
"Motorforretninger"	Kong Oscars gate
Eldre	Knudsen, Knud
Urlenke	Sunde og Søyland
Herrehatt	Atelier KK
1928 Landsutstillingen i Bergen	Dronning Maud
Kåpe	Nyblin, Karl Anton
Slips	Brosing, Gustav
Dress	
Skaut	
Par	
Eldre parportrett	
Havnebod	
Fiskebåt	
Motorbåt	

Appendiks G: Unike tagger (generert av kontrollgruppen)

Tagger som ikke kan representeres som bredere termer eller synonymer til noen av de termene som på forhånd er knyttet til bildene av UB (se appendiks f) er markert med fet skrifttype.

1700- tallet

1905

Alvorlig

anheng

Arbeid

bakgrunn

balkong

Bebyggelse

begravelse

benk

bergen

bil

blokk

blokker

bod

boligblokk

bringeduk

bro

Brostein

brusteinsgate

brygge

Bryllup

budeie

budeier

bunad

bunadsskjorte

butikk

butikker

bybebyggelse

bygate

bygd

bygning

byhavn

Båt

båter

bøtte

Dag

dagslys

dal

Dame

damer

dampbåt

dampskip

dobbel

domkirken

Drakt

Dress

dronning

ejteoar

ektepar

eldre ektepar

eldre kvinne

Eldre mennesker

elv

Fabrikk

Fest

fiske

fiskebåt

fisker

fiskeskøyte

Fjell

Fjell i bakgrunnen

fjellet

fjord

fjordbebyggelse

fjøre

fjøs

flagg

folkedrakt

folkeliv ved fjorden

forbipasserende

fotgjengar

fotgjenger

fotograf

fremkomstmiddel

gammel

gammelt ektepar

gammelt hus

gammelt høyhus

gate

gård

gårdsbruk

halv stang

handel

hardangerbunad

hatt

havn

hender

hest

hest og kjerre

hest og kjærre

hest og vogn

hestekjerre

himmel

hodeplagg

holzapfel

hus

husblokk

Hvitt hus

Høyblokk

industri

isse

kai

kirketårn

Kjerre

Klær

kong oscars gate

kongefamilie

kongelig

kongelige

kran

kvinne

kvinne i bunad

kvinner

kåpe

landevei

løe

Malt vegg

mann

master

Mennesker

mink

monokkel

mørke klær

nasjonaldrakt

nasjonsaldrakt

Naust

nordnes

norges dronning

Oppstilt foto

pels

portrett

portrett av kvinne

profil

profilportrett

redskap

rev

revepels

rorbu

Seter

sjø

sjøbod

sjøfart

sjøkant

skaut

skip

skipstrafikk

skjegg

sko

Skyss

slips

slje

sol

sorg

speilbilde

stabbur

stakk

steiner

steingard

steingjerde

steinmur

steinrøys

steinstrand

stille vann

stort naust

strand

Støl

sølge

sølje

telemark

to eldre par

torvtak

trebåt

trehus

tvilling

tvillinger

Uklar mann i bakgrunnen som går inn en dør

varebil

varetransport

Verksted

Vik

vinduer

Vogn

vågen