

**Pia Christine Bruun Syversen**

---

# **Anbefalingsystemer for litteratur i den digitale hverdagen**

## **Recommendation systems for literature in a digital world**

**Masteroppgave i bibliotek- og informasjonsvitenskap**

**Høgskolen i Oslo, Avdeling for journalistikk, bibliotek- og informasjonsfag**

Denne masteroppgaven tar for seg anbefaling av litteratur. Den går grundig igjennom de anbefalingsmetodene som finnes på området. Innholdsbaserte, kollaborative, demografiske, kunnskapsbaserte, kontekstuelle, sosiale og hybride metoder. Oppgaven kartlegger litteratur som anbefalingsdomene og bøker som gjentander som skal anbefales. Deretter drøftes metodene innholdsbaserte og kollaborative anbefalingsmetoder med reelle data fra det norske nettstedet Bokelskere.no. Formålet med denne diskusjonen er å få oversikt over hvilke konsekvenser de forskjellige metodene får for anbefalingssystemet. Dette er et sosialt nettsted for litteraturinteresserte og brukes som et case i forbindelse med oppgaven. Arbeidet med oppgaven kan sees på som et mulighetstudium for å kartlegge hvilke valg som må tas ved utvikling av anbefalingssystemer for litteratur, med fokus på den tekniske siden.

This master thesis addresses the recommendation of literature. It thoroughly reviews the recommendation methods that are available in this area. Content-based, Collaborative-based, Demographic, Knowledge-based, Contextual, Social, and Hybrid methods. The master thesis provide an summary of the literature domain and how books can be recommended. Content-based and Collaborative recommendation methods are discussed thoroughly using real data from the Norwegian site Bokelskere.no. This is a social site for literature enthusiasts and is used as a case study in this thesis. The thesis can be viewed as a feasibility study to identify the choices that must be taken in the development of recommendation systems literature, focusing on the technical possibilities.

**Høgskolen i Oslo, Avdeling for journalistikk, bibliotek- og informasjonsfag**

**Oslo 2011**

## Takk

Da jeg begynte som fersk student for fem år siden var jeg klar og bestemt. Jeg skulle få meg en bachelor og jeg skulle få meg en jobb, med bøker. Ferdig, punktum, finale. Men så skjedde det noe underveis, vet ikke helt om jeg kan sette fingere på det eller om jeg burde i denne sammenhengen. Uansett satt jeg der etter endte tre år med energi til å søke mer lærdom. Det har jeg ikke angret på et sekund, kanskje et par sekunder sånn rett før eksamen men de teller vel ikke nå. På masternivå har jeg møtt engasjerte lærere med viten og viljen til å lære bort aktuelt stoff. Takk.

Da jeg bestemte meg for tema i oppgaven og ble møtt med Ragnar Nordlie som veileder begynte et spennende år. Ragnar tar veilederrollen rolig, men med stor kompetanse og åpen dør. Du har en stil som gjør at jeg ønsker å yte mer og det har vært en positiv opplevelse å samarbeide med deg. Takk.

Min samboer Per Hansen Løfshus har ikke kjent meg som noe annet enn student, men vi har vært flinke til å bruke den tiden vi har innimellom jobb, skole og andre plikter. I tillegg til moralsk støtte har han pløyet seg igjennom oppgaven flere ganger og kommet med språklige kommentarer. Jeg setter veldig pris på alt du har gjort for å tilrettelegge for meg og gleder meg til en hverdag, uten skole, sammen med deg og Herman. Takk.

Mer språklig støtte og oppbakking har jeg fått av min venninne Ann-Christin Foss. Uten et fnugg av kunnskap om anbefalingssystemer har hun kastet seg over den ene siden etter den andre og kommet med forslag til språklige krumspring, eller innhentning av for store krumspring. For ei som kan skrive til krampa tar seg er dette absolutt et stort fortrinn. Det har hjulpet meg å ha en støttespiller som deg. Takk.

Hvor stopper man en takk som denne? Mine foreldre Gunn og Frank har ikke hatt en eneste finger med i spillet når det gjelder denne oppgaven, ikke fått lest en eneste liten setning. Det betyr ikke at de ikke har en del av æren. Uten deres støtte hadde jeg aldri kommet meg hit. Klisjé? Kanskje. Sant? Ja. Jeg er glad for at dere alltid stiller opp og ønsker det beste for meg. Takk.

Høgskolen i Oslo, Avdeling for journalistikk, bibliotek- og informasjonsfag

Oslo 2011

1 Innledning.....	8
1.1 Bakgrunn .....	12
1.2 Problemstilling .....	14
1.3 Språklige utfordringer .....	14
1.4 Masteroppgavens hovedtrekk.....	15
2 Teori om anbefalingssystemer .....	17
2.1 Hva er en anbefaling og hvorfor anbefale? .....	18
2.1.1 Data i et anbefalingssystem.....	21
2.2 Forskjellige anbefalingsmetoder .....	23
2.2.1 Innholdsbaserte metoder.....	23
2.2.2 Kollaborative filtreringsmetoder .....	27
2.2.3 Demografiske metoder .....	30
2.2.4 Kunnskapsbaserte metoder.....	30
2.2.5 Kontekstuelle metoder.....	31
2.2.6 Sosiale metoder .....	33
2.2.7 Hybride metoder.....	36
2.3 Egenskaper i et anbefalingssystem.....	37
2.3.1 Brukerpreferanse .....	37
2.3.2 Prediksjonsnøyaktighet .....	38
2.3.3 Dekning .....	41
2.3.4 Tillit.....	42
2.3.5 Selvtillit .....	43
2.3.6 Nyhet .....	44
2.3.7 Serendipitet.....	44
2.3.8 Mangfold .....	45
2.3.9 Nytteverdi.....	46
2.3.10 Risiko .....	47
2.3.11 Robusthet.....	47
2.3.12 Personvern .....	48
2.3.13 Tilpassningsdyktighet.....	49
3 Anbefalingsdomenet.....	50

3.1	Anbefaling av litteratur .....	50
3.1.1	En boks egenskaper .....	50
3.1.2	Hva er en god bok?.....	53
3.1.3	Lesevaner .....	56
3.1.4	Utvalg av bøker og utgavekopier .....	58
3.1.5	Fag- og skjønnlitteratur .....	59
3.2	Anbefalingssystemer for litteratur.....	62
3.2.1	LIBRA .....	62
3.2.2	Amazon.com .....	63
3.2.3	LibraryThing .....	64
3.2.4	Andre litteraturanbefalingssystemer .....	65
4	Metodologi .....	67
4.1	Et mulighetsstudie .....	68
4.2	Case .....	70
4.2.1	Det sosiale nettstedet Bokelskere.no .....	71
4.2.2	Likhet mellom brukere på Bokelskere.no .....	74
4.2.3	Anbefalinger på Bokelskere.no .....	75
4.2.4	Data hentet fra Bokelskere.no .....	79
5	Analyse med drøfting av hovedfaktorer.....	82
5.1	Trinn en: Litteraturanbefaling som domene.....	83
5.1.1	Ressursrealterte valg .....	83
5.1.2	Brukerne i en ressurs .....	85
5.1.3	Datagrunnlaget .....	88
5.2	Trinn to: Egenskaper i et anbefalingssystem.....	89
5.2.1	Drøfting av potensielle egenskaper .....	90
5.2.2	Sammendrag av utvalgte egenskaper for litteraturanbefalinger.....	98
5.3	Trinn tre: Potensielle anbefalingsmetoder.....	101
5.3.1	Drøfting av litteraturanbefalinger med innholdsbaserte metoder .....	102
5.3.1.1	Gjenstandsrepresentasjon .....	103
5.3.1.2	Bok kontra boksamling .....	110
5.3.1.3	Brukerprofil .....	111

5.3.1.4 Sammenligning og rangering .....	112
5.3.2 Drøfting av litteraturanbefalinger med kollaborative metoder.....	115
5.3.2.1 Likhetsmål, hva brukerne mener .....	117
5.3.2.2 Sammenligning av naboer .....	127
5.3.2.3 Resultat og rangering .....	129
6 Oppsummering .....	131
6.1 Oppsummering av underspørsmål.....	131
6.2 Anbefalingsmetoder for litteratur .....	133
6.2.1 Innholdsbaserte litteraturanbefalinger .....	134
6.2.2 Kollaborative litteraturanbefalinger .....	135
6.2.3 En hybrid løsning .....	137
6.3 Refleksjoner om oppgaven.....	137
6.3.1 Den teoretiske fremstillingens begrensninger .....	138
6.3.2 Hva mister den anbefalte verden? .....	138
7 Videre forskning.....	141
7.1 Videre forskning på Bokelskere.no .....	141
7.1.1 Refleksjoner for Bokelskere.no.....	141
7.2 Videre forskning på litteraturanbefalinger .....	143
7.3 Videre forskning tilknyttet bibliotek .....	144
8 Litteraturliste .....	146
8.1 Kilder.....	146
8.2 Personlig kommunikasjon .....	153
8.3 Tabeller.....	154
8.4 Bilder .....	155
9 Vedlegg .....	156
Vedlegg 1: Ønskebok.no .....	156
Vedlegg 2: Antall representerte bøker .....	158
Vedlegg 3: Hovedsiden for en brukers bok.....	160
Vedlegg 4: Tall om den bestemte bokelskeren .....	163
Vedlegg 5: Å følge en annen bruker .....	164
Vedlegg 6: De du følger leser nå.....	165

Vedlegg 7: Bokanbefalinger.....	166
Vedlegg 8: Datamodellen til Bokelskere.no .....	168
Vedlegg 9: Utgavekopihåndtering av <i>Stolthet og fordom</i> .....	169
Vedlegg 10: Omtale .....	171
Vedlegg 11: Bokhyllenavn.....	172
Vedlegg 12: Brukere med flest like bøker .....	173

# 1 Innledning

I dagens vestlige samfunn lever de fleste mennesker et liv som inkluderer både den fysiske og den digitale verden. For mange begynner dagen med å vekkes av mobiltelefonen, som nå er en liten datamaskin i seg selv. De sjekker e-post og de leser digitale utgaver av avisene på nettet. For studenter og akademikere finnes store deler av pensum og artikler digitalt i databaser. Handel gjennom nettbutikker øker, lesebrett blir mer utbredt og mennesker tilhører ofte ett eller flere sosiale nettsteder. Bibliotekene blir stadig mer fokusert på sine digitale tjenester, og brukerne er ofte inne og sjekker bibliotek katalogen på nettet slik at de blant annet kan reservere bøker selv. Formatutvalget vokser, lydbøker på mp3 er etablert og lesebrettordninger i norske biblioteker er i oppstartsfasen. Antallet nettkunder er i ferd med å sprengte alle grenser. I Norge har ni av ti husstander internett, Statistisk sentralbyrå<sup>1</sup> forteller at i 2008 hadde informasjonssektoren her til lands 100 000 ansatte.

Som følge av denne enorme mengden med digital informasjon er begrepet information overload blitt en realitet man må forholde seg til. Pijpers (2010) forklarer begrepet grundig og har samlet flere definisjoner. Den enkleste definisjonen er at det er så mye informasjon tilgjengelig at det ikke lenger er mulig å bruke den effektivt. På et mer omfattende nivå definerer han information overload som et fundamentalt problem for mennesket. Dette konkretiserer han gjennom tre punkter. Punkt en, frustrasjonen over ikke å ha tilgang til relevant informasjon som man vet eksisterer. Punkt to, redsel for å ha oversett relevant informasjon. Punkt tre, manglende evne til å lese, bearbeide og bruke informasjonen grunnet for stor mengde og tilgang til den. Pijpers sier også at forskere har slått fast at menneskets hjerne ikke er konstruert for å behandle så mye informasjon som nå finnes tilgjengelig og at information overload kan føre til stress, nedsatt tilfredsstillhet og nedsatt fysisk helse.

Informasjonsmengden på nettet er enorm, og brukere trenger av den grunn hjelp til å finne frem til riktig nettside, dokument, ressurs eller produkt. Med de enorme mengdene informasjon og litteratur som finnes på verdensbasis, nå kun et tastetrykk unna, er behovet for hjelp gjeldende også her. En mulig måte å hjelpe en bruker på er å tilby personalisering og innenfor denne teknologien har man anbefalingssystemer. Det finnes mange definisjoner av personalisering, den formuleringen dekker begrepet godt:

---

<sup>1</sup> [www.ssb.no/ikt/](http://www.ssb.no/ikt/)



Personalization is the use of technology and customer information to tailor electronic commerce interaction between a business and each individual customer. Using information either previously obtained or provided in real-time about the customer, the exchange between the parties is altered to fit the customer's stated needs [...]  
(Personalization Consortium, gjengitt etter Vesanen 2007).

Definisjonen er i utgangspunktet utarbeidet for netthandel, da denne bransjen er sterkt representert innenfor personaliseringsteknologisk forskning. Bransjen er opptatt av å tilfredsstillere kundene sine, i tillegg til å selge produkter. Selge produkter gjør de ved at kunder gis mulighet til å finne produkter som passer spesielt for dem personlig, gjennom personalisering. Jeg mener den definisjonen fungerer godt også når man snakker om andre nettbaserte ressurser. Selv om de nettbaserte ressursene ikke har fokus på å selge noe er utviklere generelt sett interessert i å tilfredsstille brukerne sine. Tilfredsstilte brukere er ofte lojale brukere, noe som gjør at brukerne velger å bli værende som aktive brukere av den aktuelle ressursen.

Personalisering på nettet er i følge Smeaton & Callan (2005) å tilpasse presentasjonen, innholdet og selve tjenesten i forhold til brukerens kontekst, i denne sammenhengen brukerens: oppgave, bakgrunn, historie, informasjonsbehov, tilgang og oppholdssted. Personalisering blir da en måte å møte brukerens behov for at brukeren skal føle seg ønsket og vil derfor være det overordnede begrepet for flere teknikker deriblant anbefalingssystemer. Et anbefalingssystem har i følge Ricci, Rokach & Shapira (2011) som hovedoppgave å foreslå en eller flere gjenstander som trolig vil være nyttig for en bruker. Et anbefalingssystem i forbindelse med netthandel er et viktig redskap for å selge produkter plukket ut spesielt for å passe en bruker. Anbefalingssystemet på en nettbasert ressurs er trolig en ekstra funksjon og det brukes for å anbefale gjenstander en bruker kan være interessert i å vite om, tilpasset hva slags ressurs det er snakk om.

Etter hvert som bibliotekcatalogens omfang øker, grunnet både samlingsutvidelse og samarbeid med andre bibliotek, er det sannsynlig at brukerne der vil trenge hjelp. Å anbefale litteratur er allerede en stor og viktig del av et biblioteks tjenester. En bibliotekar vil kjenne seg igjen i spørsmålet: Nå har jeg lest alt av Jo Nesbø, Gert Nygårdshaug, og Karl Ove Knausgård hvilken bok vil du anbefale? Når stadig flere av bibliotek tjenestene blir digitaliserte er det logisk

at også litteraturanbefaling blir tilgjengelig på denne måten. Dermed er det på tide å se på hvordan litteratur blir anbefalt på andre arenaer.

For å tilby en bruker anbefalinger kreves det observasjon av brukeren eller at brukeren selv oppgir informasjon om seg selv. Personvern er et aktuelt politisk tema som det arbeides med i forhold til bibliotek på verdensbasis, Norge intet unntak. Ser man på hvordan prosessen med Datalagringsdirektivet blir mottatt både her i landet og internasjonalt, kommer det tydelig frem at personvern setter følelser i sving. Datalagringsdirektivet er et EU-direktiv som omhandler lagring av abonnements-, lokaliserings- og trafikkdata og er i utgangspunktet myntet på å bekjempe terrorisme og kriminalitet. Argumenter om brudd på personvern, mangel på klare regler for avgrensning og omfang, utgifter ved lagringen og fare for overtredelser settes opp mot fordelene ved å ha denne typen data tilgjengelig ved mistanke om ulovlig virksomhet. Direktivet har møtt voldsom motstand her i landet og debatten om å verne om privatlivet har rast i flere år (Datatilsynet 2011). Å ta fatt på det politiske aspektet ved å benytte seg av brukerdata i bibliotek er en masteroppgave i seg selv. Spennende, men ikke det jeg ønsker å arbeide med i denne oppgaven.

Jeg har lenge vært oppmerksom på personalisering og anbefalingssystemer på nett, men min interesse for anbefalingssystemer ble virkelig vekket da jeg ble informert om at nettstedet Bokelskere.no<sup>2</sup> ønsket å utarbeide et anbefalingssystem for sine brukere. Bokelskere.no er et sosialt nettsted for bokinteresserte, og er Norges svar på LibraryThing<sup>3</sup>. Her kan brukere blant annet registrere egne bøker, gi emneord, skrive anmeldelser og gi terningkast. Nettstedet har også en stor mengde diskusjoner om litteratur, alt fra innholdet i en bok, hva brukeren foretrekker å lese og hvor favorittstedet for lesing foregår. Med dette i tankene, og en bibliotekars sans for litteratur har interessen for å finne måter å generere gode litterære anbefalinger økt.

Jeg ønsker i denne masteroppgaven å finne frem til gode anbefalingsmetoder for litteratur. Da dette viser seg å være et meget omfattende arbeid har jeg gått inn for å systematisere hvilke anbefalingsmetoder som finnes, hvilke kriterier som er viktig å vektlegge i et anbefalingssystem for litteratur og hva som skiller litteraturanbefaling fra annen type anbefaling. Dette vil legges frem som en teoretisk drøfting som kan benyttes når anbefalingsmetode velges

---

<sup>2</sup> <http://bokelskere.no/>

<sup>3</sup> <http://www.librarything.com/>

og et anbefalingssystem skal utvikles. På denne måten kan masteroppgaven sees på som en mulighetsstudie innenfor nettbaserte ressurser.

Det er viktig å snakke om hva slags type anbefalinger jeg vektlegger i oppgaven siden litteraturanbefalinger kan genereres på bakgrunn av teori fra to fagområder: informasjonsvitenskap og litteraturvitenskap. Innen informasjonsvitenskap finnes grunnlaget for en algoritmisk analyse for likhet mellom brukere og/eller dokumenter. Innen litteraturvitenskap vil analyse av verket og brukernes bestemte ønske i forhold til hva leseropplevelsen gir dem være sentrale. Jeg presiserer at jeg i denne oppgaven fokuserer på teorier innenfor informasjonsvitenskap og algoritmiske prosesser for å finne likhet. Jeg har ikke vektlagt spesifikke algoritmer men hvilke konsekvenser forskjellige anbefalingsmetoder får, og hvilke valg som må tas innenfor de forskjellige metodene. Det er derfor snakk om anbefalingsmetoder innenfor ressurser som har mange brukere og hvordan brukere kan tilfredstilles på et personlig plan. Siden det er snakk om anbefaling av litteratur, vil det være behov for å ta for seg hva som skiller det å anbefale en bok fra å anbefale andre gjenstander. Jeg akter derimot ikke å gå dypt inn på litteraturanalyser og resepsjonsteori, som er opptatt av hvordan litteratur blir verdsatt og hva den får en bestemt bruker til å føle.

For å gjennomføre oppgaven vil jeg først gjennomgå aktuell forskning på området. Det er et stort område å kartlegge og det finnes mange aktuelle anbefalingsmetoder, samtidig finnes det mange ressurser som benytter seg av anbefalingssystemer derfor har kapittel 2 og 3 blitt omfattende. Deretter vil jeg gjennom teoretisk diskusjon legge frem en eller flere anbefalingsmetoder som jeg gjennom kunnskap fra tidligere forskning innenfor området anser som gode kandidater innenfor litteraturanbefaling. Jeg benytter meg av Bokelskere.no som et case for å konkretisere denne diskusjonen med reelle datakilder fra et sosialt nettsted. Jeg får igjennom dette et overblikk over hvilke praktiske og tekniske utfordringer som ligger i utvikling av et anbefalingssystem. Jeg håper at dette arbeidet kan være et springbrett for anbefalinger innenfor norske ressurser som ønsker å tilby litteraturanbefalinger, deriblant bibliotek, når biblioteksektoren har kommet så langt personvernsmessig at dette er en reell mulighet.

## 1.1 Bakgrunn

I dagens digitale samfunn har de fleste mennesker på et eller annet vis stiftet bekjentskap med et anbefalingssystem. Kjente nettbutikker som Amazon.com<sup>4</sup>, iTunes<sup>5</sup>, CDON.com<sup>6</sup> og Bokkilden.no<sup>7</sup> benytter seg av denne teknologien i mer eller mindre skreddersydde løsninger. Brukere blir også møtt med en type anbefalinger i biblioteker som bruker Bibliofil<sup>8</sup> som biblioteksystem. Anbefalinger omgir oss i vår digitale hverdag, derfor er det viktig å vite hva det er og hvordan det fungerer.

Anbefalingssystemer var i første omgang myntet på mennesker som ikke hadde tilstrekkelig kunnskap eller erfaring til å ta et valg når de sto ovenfor et stort utvalg av gjenstander. Senere har spesielt aktører innenfor netthandel bidratt til å utvikle anbefalingssystemer, da med økt profitt i sikte. Anbefalingssystemer er i økende grad med på å hjelpe mennesker overkomme information overload-problemet, men er også mye brukt på forskjellige nettressurser, som sosiale nettsider og i netthandel.

I hovedsak snakket man om to eller tre hovedmåter å gi anbefalinger: Innholdsbaserte metoder som bygger på teknikker hentet fra informasjonsgjenfinningsfeltet og genererer anbefalinger basert på likhet mellom innholdet i gjenstander. Kollaborativ filtreringsmetoder som bygger på teknikker som høster informasjon om brukerne og generer anbefalinger basert på likhet mellom brukere. Til sist hybride metoder som bygger på sammenslåing av teknikker innenfor de forskjellige metodene. Burke (2007) skisserer derimot hele seks klasser for anbefalingsmetoder i sin klassifisering av anbefalingssystemer basert på informasjon utvikler har til rådighet og anbefalingssystemets funksjon:

- Innholdsbaserte metoder; i denne fremgangsmåten er en innholdsanalyse av gjenstandene som skal anbefales og genererer anbefalinger basert på gjenstandenes attributter og vurderinger brukerne har gitt gjenstandene.
- Kollaborative metoder; denne fremgangsmåten gir anbefalinger basert på brukernes profil. Systemet finner likheter mellom brukere, og anbefaler på bakgrunn av denne likheten gjenstander lignende brukere har gitt uttrykk for at de liker.

---

<sup>4</sup> <http://www.amazon.com>

<sup>5</sup> <http://www.apple.com/no/itunes>

<sup>6</sup> <http://cdon.no>

<sup>7</sup> <http://www.bokkilden.no>

<sup>8</sup> <http://www.bibsyst.no>

- Demografiske metoder; gir anbefalinger basert på en demografisk profil av brukeren. Informasjon om kjønn, alder, bosted og arbeidssituasjon er viktige datakilder.
- Kunnskapsbaserte metoder; gir anbefalinger basert på slutninger om brukerens behov og preferanser. Metoden bruker blant annet kunnskap om gjenstanders attributter og brukers forhåndsbestemte behov i forhold til utvalget av gjenstander.
- Kontekstuelle metoder; er foreløpig lite utprøvd, men blir mer og mer populære å forske på, spesielt for mobilapplikasjoner som benytter seg av konteksten brukeren befinner seg i. Et eksempel på dette er å få værreport basert på mobilens GSM-signal.
- Hybride metoder; går ut på å slå sammen andre anbefalingsmetoder, så under denne samlebetegnelsen fins det mange muligheter. Man kan slå sammen forskjellige metoder, eller forskjellige tilnæringsmåter til metoden og det er stadig mer populært å benytte seg av hybride metoder for å få bukt på kjente problemer innenfor anbefalingssystemer.

I tillegg til de seks overnevnte anbefalingsmetodene som Burke har skissert trekker Ricci et al. (2011) frem en metode til. Siden sosiale medier og web 2.0 interaksjon har blitt en del av den digitale hverdagen har dette ført til at sosial data blir tilgjengelig nett.

- Sosiale metoder; baserer seg på preferanser som brukerens venner har og muligheter som genereres av web 2.0 applikasjoner. Tanken bak denne fremgangsmåten er at brukere stoler mer på folk de kjenner. I tillegg vil systemer som tillater tagging og presisering av begreper gjennom den semantiske webben falle innenfor denne metoden.

Det kan diskuteres om denne siste metoden skal stå for seg selv eller være underlagt kollaborative metoder, siden den benytter seg av lignende fremgangsmåter. Da den åpner for flere brukerskapte datakilder enn kollaborative metoder har jeg valgt å omtale den som en selvstendig metode. Siden de syv forskjellige metodene kan brukes innenfor flere forskjellige domener har jeg gått igjennom alle metodene i kapittel 2.

## 1.2 Problemstilling

I denne masteroppgaven arbeider jeg med en problemstilling som lyder:

*Hvilke anbefalingsmetoder kan vise seg mest hensiktsmessig å benytte i et nettbasert anbefalingssystem for litteratur, der brukeren står i sentrum?*

Jeg mener at arbeid med dette kan hjelpe utviklingen av anbefalingssystemer av litteratur i Norge. Da dette er et nærliggende felt for bibliotek og utforsker en mulig anbefalingsteknikk som kan benyttes der, mener jeg dette feltet er spesielt interessant. Mengden informasjon og digitale ressurser øker og det er derfor viktig og utforske metoder utarbeidet for å hjelpe brukere håndtere dette. Det er en trend i samfunnet å oppholde seg mer og mer i den digitale verden, det handler om å møte brukerne der de er. Samtidig øker mengden bøker en bruker kan velge å bruke tid på, en tenker seg at litteraturtype og sjangerinndeling kan hjelpe de et stykke på vei. Selv etter tradisjonelle inndelinger er det enorme mengder litteratur brukere kan fordype seg i, derfor kan et anbefalingssystem hjelpe de å finne aktuelle bøker.

For å kunne svare på den overordnede problemstillingen har jeg utarbeidet et sett underspørsmål som jeg bruker underveis:

- Hvilke forskjellige anbefalingsmetoder finnes?
- Hvilke anbefalingsmetoder kan en utvikler benytte seg av i forhold til tilgjengelig data fra et sosialt nettsted?
- Hvordan skiller det å anbefale en bok seg fra anbefaling av andre gjenstander?
- Hvilke egenskaper er viktig å legge vekt på når en skal utvikle anbefalingsmetoder for litteratur og hvilke konsekvenser får dette?

## 1.3 Språklige utfordringer

Forskning på anbefalingssystemer er internasjonalt og strekker seg over flere fagområder. Denne oppgaven er basert på artikler og forskningsrapporter skrevet på de språkene jeg behersker norsk, svensk, dansk og engelsk. Engelsk er det dominerende språket i egenskap av å være et internasjonalt språk. Begrep som brukes i engelskspråklig litteratur har ikke alltid en norsk oversettelse. Jeg har valgt å beholde det engelske begrepet der det ikke eksisterer et dekkende norsk begrep. Der det er mulig å fornorske et begrep uten å miste viktig informasjon har jeg valgt dette. Det er en trend innenfor det digitale fagmiljøet å låne, men ikke fornorske engelske

begreper, og det kan dermed være med å forvirre en leser. Enkelte steder har jeg valgt å presentere det engelske begrepet før jeg oversetter det til norsk og på den måten forsøker jeg å unngå språklige forvirringer.

Begrepet internett og nettet er velbrukte begreper i Norge, men ikke alltid benyttet korrekt. Internett et verdensomspennende nettverk for datakommunikasjon og informasjonsspredning. Populære tjenester er e-post, Chat, filoverføring og World Wide Web (nettet). Problemer oppstår på norsk siden vi har forkortet World Wide Web til nettet. Jeg bruker betegnelsen nett eller nettet om World Wide Web. Dessuten vil jeg bruke den norske måten å ordlegge meg og vil derfor bruke den semantiske webben i stedet for The Semantic Web. Fotnoter er brukt både til begrepsavklaring og url-er.

## **1.4 Masteroppgavens hovedtrekk**

Denne masteroppgaven består av utvelgelse av anbefalingsmetode for å skape litteraturanbefalinger basert på algoritmisk likhet. Utvelgelse og utprøvelse av metoder og algoritmer vil være et avgjørende steg når man skal designe et anbefalingssystem. Siden det er mange mulige anbefalingsmetoder å velge mellom ønsker jeg å gå systematisk igjennom de potensielle metodene for å generere anbefalinger. Deretter går jeg gjennom anbefalingsdomenet. Selve analysekapittelet går igjennom tre hovedkriterier. For det første anbefalingsdomenet og egenskaper tilknyttet gjenstanden som skal anbefales. Det andre er anbefalingssystemer og egenskaper tilknyttet dem. Det tredje er anbefalingsmetoder basert på tilgjengelige datakilder fra den aktuelle ressursen anbefalingssystemet skal betjene. På den måten vil utviklingen av litteraturanbefalinger presenteres med teoretiske diskusjon tilpasset en konkret ressurs.

Jeg har med andre ord valgt å gå grundig igjennom utviklingsprosessen og presenterer en dyp drøfting av de valgene en utvikler står ovenfor. Jeg ønsker på den måten å sitte igjen med de aktuelle metodene for litteraturanbefalinger som jeg mener er best i forhold til det aktuelle domenet og tilgjengelige datakilder. Oppgaven vil begrense seg mot en type ressurs, sosiale nettsteder. Derfor vil nettstedet Bokelskere.no presenteres som et case. Dette gjøres fordi jeg på den måten kan gi konkrete eksempler med reelle data. Det er ikke prioritert å bruke tid på å presentere økonomiske valg eller algoritmer som kan implementere anbefalingssystemet på det aktuelle nettstedet.

I kapittel 2 gjør jeg rede for teori om anbefalingssystemer og forskningslitteraturen på området. Først tar jeg for meg hva en anbefaling er og hva slags data som trengs for å anbefale gjenstander. Jeg går gjennom forskjellige anbefalingsmetoder, gjør rede for problemområder, sterke sider og utfordringer forbundet med de forskjellige metodene. Jeg tar også for meg hvilke egenskaper anbefalingssystemer har og hvordan disse kan påvirke anbefalingene.

I kapittel 3 tar jeg for meg litteratur som anbefalingsdomene. Jeg går grundig gjennom boken som anbefalingsgjenstand og legger frem spesielle aspekter ved denne. Jeg går også inn på spesielle utfordringer som knytter seg til bøker og lesere. Deretter legger jeg frem to aktive ressurser som benytter seg av litteraturanbefalinger og et tidligere forskningsobjekt.

Kapittel 4 er viet metodologiske valg, og jeg forklarer hvordan min oppgave er bygget opp som et mulighetsstudium og hvilke begrensninger jeg har gjort. Jeg legger i det kapittelet frem hovedkriteriene for analysedelen. Deretter presenterer jeg det sosiale nettstedet Bokelskere.no som et case og går gjennom tilgjengelig data derfra. Jeg vil presentere min valgte analysemetode; teoretisk drøfting.

Kapittel 5 er analysekapittelet som inneholder drøftingen som resulter i en teoretisk modell for litteraturanbefalinger. På bakgrunn av trinn forbundet med anbefalingsprosessen vil jeg presentere analysens tre hovedtrinn. Trinn en: Anbefalingsdomenet. Trinn to: Utvelgelse av egenskaper i et anbefalingssystem. Trinn tre: Drøfting av anbefalingsmetoder. Avslutningsvis legger jeg frem den teoretiske modellen for litteraturanbefaling som bygger på de tre nevnte trinnene.

I kapittel 6 vil jeg oppsummere masteroppgaven og arbeidet med litteraturanbefalinger. Jeg vil presentere funn for oppgaven og anbefalingsmetoden jeg foreslår for Bokelskere.no Jeg vil til slutt evaluere egne erfaringer underveis.

Kapittel 7 er viet videre forskning. Jeg vil her presentere videre forskning innenfor Bokelskere.no, litteraturdomenet og innenfor bibliotek.

I kapittel 8 finnes en detaljert litteraturoversikt over kilder brukt i forbindelse med oppgaven og oversikt over personlig kommunikasjon.

Vedlegg utarbeidet til oppgaven og oversikt over tabeller og bilder finnes i kapittel 9.



## 2 Teori om anbefalingssystemer

Det finnes forskjellige måter å generere anbefalinger for en bruker, derfor trengs en grundig gjennomgang av hele fagområdet. Det brukes forskjellige anbefalingsmetoder på forskjellige anbefalingsdomener og forskjellige egenskaper ved anbefalingssystemer må vektlegges. Det finnes forskjellige grunner for utviklere å tilby brukere anbefalinger og det finnes kjente problemområder innenfor de forskjellige metodene.

Litteraturanbefalinger kan genereres på bakgrunn av en algoritmisk analyse for likhet mellom brukere og/eller gjenstander, eller på bakgrunn av krav en bruker setter til anbefalingssystemet for domenet som skal betjenes. Felles for begge fremgangsmåtene er at de ønsker å tilfredsstille brukeren og dens behov. Innenfor domenet for litteraturanbefalinger snakker vi om gjenstander i form av bøker, dette vil jeg gå dypere inn på i kapittel 3. Mens anbefalinger generert på bakgrunn av algoritmiske analyser for likhet mellom brukere og/eller bøker benytter metoder fra informasjonsvitenskap, vil anbefalinger generert på bakgrunn av brukers leseropplevelse trekke på fagområdet for litteraturvitenskap. Forskning på det sistnevnte området vil i stor grad bygge på forståelse av leser som person og dermed resepsjonsteori. Ridderstrøm (2011) oppsummerer resepsjonsteorien slik; resepsjonsteorien er en moderne litteraturteoretisk retning der det er fokus på hvordan det litterære verket blir mottatt av leserne. Hovedinteressen blir dermed flyttet fra forfatter og verk til leserens medvirkning når det gjelder tekstens mening. Hvordan leseren opplever verket avhenger av leserens ytre påvirkninger; alder, kjønn, erfaringer, lesenivå, religion, kultur, sosiale skillelinjer og mål med lesing.

Ønskebok.no<sup>9</sup> er et eksempel på et anbefalingssystem som foreslår bøker basert på egne forutsetninger brukeren ønsker boken skal oppfylle, konkretisert i kriterier (jmf. Vedlegg 1). Brukeren legger inn tre kriteriesett for et boksøk, velger i hvor stor grad kriteriene er gjeldende og søker etter relevante treff. Brukeren kan velge mellom 12 kriteriesett, som er bygget opp av motsetninger. Eksempler på disse er; glad/trist, forventet/uventet, mild/brutal, uten sex/mye sex og tradisjonell/uvanlig. Deretter får brukeren presentert forslag på bøker, hvis det ikke finnes samsvar mellom søket og tilgjengelige bøker må brukeren endre noen av kriteriene. ABM-Utvikling, i samarbeid med flere folkebibliotek står bak Ønskebok.no. Hvis en bruker finner en bok han ønsker å lese vil en lenke guide ham videre slik at han til sist kan søke i sitt foretrukne bibliotek.

---

<sup>9</sup> <http://www.onskebok.no>

Ønskebok.no var et spennende forsøk på et anbefalingssystem som skal hjelpe brukere som ikke helt vet hva de leter etter, og ta hensyn til brukerens ønskede til leseropplevelsen. Det er nå i ferd med å avvikles og skal evalueres. Det var en spennende vinkling på anbefalinger, men var avhengig av at brukeren aktivt oppsøker tjenesten og frivillig la inn kriterier. Jeg er mer interessert i den algoritmiske analysen av likhet mellom brukere og/eller dokumenter som dominerer anbefalingslitteraturen, det er denne typen anbefalinger folk flest møter i sin digitale hverdag. Dermed seg jeg en nytteverdi i å undersøke hvilke av de forskjellige metodene som finnes som passer best til litteraturanbefalinger. Av den grunn har jeg valgt å legge fokus på fremgangsmåter som faller innunder den algoritmiske analysen for likhet. Det vil selv med denne typen fremgangsmåte være viktig å gjøre rede for boken som gjenstand som skal anbefales, og lesere som brukere. Dette er viktig fordi forskjellige kriterier og egenskaper, samt hvordan boken skal brukes og hva slags type leser brukeren er kan påvirke resultatet og dette går jeg inn på i kapittel 3. Først må det klargjøres hva en anbefaling er.

## 2.1 Hva er en anbefaling og hvorfor anbefale?

Jeg har tidligere poengtert at anbefalingssystemer hjelper brukere med å overkomme informasjon overload-problemet. Tanken er at det finnes så mange mulige valg der ute så brukeren trenger litt hjelp. Samtidig må det skilles mellom brukers behov og utviklers motiv. Før jeg går inn på disse forskjellene ønsker jeg å skape et klarere bilde av hva en anbefaling er. Det finnes personlige og ikke-personlige anbefalinger. Et godt eksempel på ikke-personlig anbefaling er VG-lista topp 20<sup>10</sup>. En slik anbefaling er et sammendrag av hva som er populært i samfunnet der og da. Det blir hentet inn informasjon om hva som selger mest for øyeblikket, og en liste hvor de mest populære sangene rangeres fra mest til minst populær blir presentert.

Jeg nevnte innledningsvis at Bibliofil-bibliotek tilbyr en type litteraturanbefaling. På søkesidene til Bibliofil-bibliotek vil man få frem en karusell som automatisk viser frem aktuelle bøker. På nettsidene til Moss bibliotek<sup>11</sup> ligger karusellen nederst på søkesiden. De aktuelle bøkene blir i denne sammenhengen anbefalinger, utvelgelsen av disse blir gjort på bakgrunn av data som skapes i tidligere søkesesjoner eller utlånssituasjoner. Det eneste en aktiv bruker kan gjøre for å påvirke den inneværende søkesesjon er å velge hvilken kategori karusellen skal vise.

---

<sup>10</sup> <http://lista.vg.no> (Norges offisielle hitliste er basert på siste ukes salg hos ca. 100 musikkforhandlere over hele landet.)

<sup>11</sup> [http://www.moss.folkebibl.no/?option=com\\_websok&Itemid=1&lokalsok=/cgi-bin/websok-portal](http://www.moss.folkebibl.no/?option=com_websok&Itemid=1&lokalsok=/cgi-bin/websok-portal)

Kategorier kan være blant annet; mest utlånt, mest reservert, mest vist i det siste, mest utlånt for barn og nyeste titler. Dette er dermed som tidligere nevnt bibliotekets forsøk på å anbefale litteratur for brukeren på en ikke-personlig måte siden systemet ikke har noe informasjon om den aktive brukeren. Biblioteket antar at bøker som blant annet søkes opp og lånes ut ofte er populære for folk flest og de blir derfor anbefalt. Forskjellen fra VG lista eksempelet er at brukeren kan velge kategori og kan dermed påvirke utvalget noe. Disse to anbefalingssituasjonene faller utenfor hva man vanligvis forsker på innenfor anbefalingssystemer, men reflekterer bruk av kollektive preferanser. Det som mangler er den personlige vinklingen på anbefalingene.

Når det gjelder personlige anbefalingssystemer så er det to ting som er viktige for utvikler å ha klart for seg før han velger å tilby brukeren av en ressurs anbefalinger. Hvilke gjenstander som skal anbefales og hva er formålet for anbefalingene innenfor anbefalingsdomenet? Dette er avgjørende for valg av riktig fremgangsmåte og anbefalingsmetode. En brukers behov vil variere alt etter hva slags gjenstander som blir anbefalt, men det finnes noen generelle funksjoner et anbefalingssystem tilfører en ressurs. Herlocker, Konstan & Riedl (2000) trekker frem hvordan et anbefalingssystem kan være nyttig for brukeren. De sier at et anbefalingssystem kan hjelpe brukeren å nå målet for å bruke ressursen. I tillegg vil et anbefalingssystem hjelpe brukeren med å: Finne noen, eller alle gode gjenstander og finne gjenstander som inngår i en serie eller på annen måte passer sammen. Schafer, Frankowski, Herlocker & Sen (2007) utdyper dette med seks punkter:

1. Hjelp brukeren å finne nye gjenstander han kanskje liker. For å hjelpe brukeren overkomme information overload-problemet.
2. Gi råd om en bestemt gjenstand. Brukeren har funnet en bestemt gjenstand men er usikker på om han vil like den.
3. Hjelp brukeren å finne en bruker, eller flere brukere han kanskje liker. Noen ganger har brukere utbytte av å se hva andre brukere gjør.
4. Hjelp en gruppe brukere å finne nye gjenstander de kanskje liker. Noen anbefalingssystemer er laget for å hjelpe en større gruppe finne gjenstander sammen.
5. Hjelp en bruker å finne en god miks av nye og gamle gjenstander. Spesielt i forhold til restauranter og dagligvarer er det ikke noen grunn til å luke vekk gamle og likte gjenstander.
6. En bruker kan trenge hjelp som er spesielt tilpasset domenet. I en database med forskningspapirer kan det være et behov for å finne alle potensielle dokumenter.

Ricci et al. (2011) redegjør for fem hovedgrunner for hvorfor utviklere bør interessere seg for anbefalingsteknologi. Denne er i utgangspunktet fokusert på netthandel:

1. Øke antall solgte gjenstander. For en kommersiell ressurs er dette hovedgrunnen til å implementere et anbefalingssystem. For en ikke-kommersiell ressurs vil det også være ønskelig at brukeren velger anbefalte produkter. Et godt anbefalingssystem vil anbefale gjenstander som brukeren liker. Brukeren blir dermed fornøyd og fortsetter å bruke ressursen.
2. Selge flere forskjellige gjenstander. Tenk deg en nettbutikk som Amazon.com som gjerne ønsker å selge fra de millioner av bøker som finnes på lager. Samtidig ønsker de ikke å annonsere mindre populære, eller eldre bøker. Ved bruk av anbefalingssystem, hvor en kilde til informasjon er tidligere kjøp, vil bøker som ellers ikke blir annonsert vises til brukere som trolig er interessert i dem.
3. Tilfredsstille brukeren. Et godt anbefalingssystem vil gi brukeren nyttige, spennende og relevante anbefalinger og på den måte øke brukerens tanker om ressursen. Brukeren vil føle seg ivaretatt og ønske å komme tilbake. I tillegg er det mer trolig at en fornøyd bruker vil benytte seg av anbefalingene systemet foreslår.
4. Øke brukerens lojalitetsfølelse. En bruker er trolig mer lojal mot en ressurs som gjenkjenner brukeren og hvor brukeren blir behandlet som en velkjent og anerkjent bruker. Ressurser som husker hva brukeren har sett på, kjøpt og evaluert kan tilby skreddersydde anbefalinger. Jo oftere brukeren benytter seg av ressursen jo mer lærer ressursen brukeren å kjenne. En lojal kunde fortsetter å bruke den samme ressursen.
5. Økt forståelse for brukerens behov. Et anbefalingssystem kan lagre en profil som forteller hva brukeren foretrekker. Denne informasjonen kan hjelpe skaperne av ressursen til å få en høyere grad av forståelse for brukere av ressursen. Denne kunnskapen kan brukes til blant annet design og layout av ressursen eller eventuelt å fortelle hvilke gjenstander en nettbutikk eller et bibliotek bør kjøpe inn og tilby.

En nettbutikk, en søkemotor og et bibliotek vil ha forskjellig motivasjon for anbefaling. Foreløpig er det foretatt mest forskning på anbefalingssystemer i forbindelse med netthandel, dette er trolig fordi det handler om penger. Jeg velger likevel å bruke listen ovenfor da flere av punktene kan overføres til andre nettbaserte ressurser siden det er et enormt antall nettsider som ber om oppmerksomhet jobber de hardt for å tilfredsstille brukerne sine. Denne

serviceinnstillingen har sine røtter i handel, derfor er det nødvendig å holde seg oppdatert med hva som skjer på den fronten. Shcafer et al. (2007) forenkler utviklers motiver med å vise til tre punkter:

1. Anbefal gjenstander. Gi brukeren en rangert liste med anbefalinger slik at han kan finne gode gjenstander.
2. Beregn anbefalingsgrad for en gjenstand. Hvilken vurdering er det trolig at en bruker vil gi en bestemt gjenstand?
3. Anbefal gjenstander innenfor grenser brukeren setter. En bruker kan ha behov for å avgrense på alder.

Uansett ressurs er anbefalingssystemets funksjon viktig å ha tenkt godt igjennom slik at man velger en fremgangsmåte som passer ressursen og målet. Det er utvikler som velger hvorfor ressursen skal tilby en anbefaling og vil alltid ha en grunn for dette. Om denne grunnen kommer av et ønske om å hjelpe sine brukere av et sosialt nettsted eller om den kommer av en netthandels behov for å øke salg.

### **2.1.1 Data i et anbefalingssystem**

For å kunne betjene et anbefalingssystem er man avhengig av innsamlet informasjon om ressursen som benytter seg av den. Denne typen informasjon samles inn som tilgjengelig data fra ressursen anbefalingssystemet skal tjene. Alt etter hva slags metode man benytter seg av og hvilke muligheter selve ressursen tilbyr vil datakilder og datainnsamlingsmetoder variere. Ordet data blir i denne oppgaven brukt slik:

[...] a collection of objects and their attributes, where an attribute is defined as a property or characteristic of an object. Other names for object include record, item, point, sample, observation, or instance. An attribute might be also be referred to as a variable, field, characteristic, or feature (Amatriain, Jaimes, Oliver & Pujol 2011 s. 40).

I et anbefalingssystem finnes det to klare kilder til data; gjenstandene som skal anbefales, og brukerne som benytter seg av anbefalingssystemet. I tillegg til disse har man også mulighet til å lagre data om transaksjoner mellom bruker og gjenstander. Ricci et al. (2011) oppsummerer datakildene slik:

## **Gjenstander**

Gjenstander som blir anbefalt kan karakteriseres gjennom deres verdi. En verdi kan være positiv hvis den er til nytte for brukeren, eller negativ hvis den ikke er til nytte for brukeren. Når det gjelder representasjonen av gjenstanden vil den ofte bli representert med et sett attributter. En bok vil bli representert med attributter som tittel, forfatter, sjanger, ISBN, målgruppe og sideantall. En gjenstand vil også ha en kompleksitet forbundet med seg angående beskaffenhet. Dette går direkte på den kostnaden brukeren utsetter seg for ved å følge anbefalinger. Det kan være økonomiske kostnader eller kostnader forbundet med tidsbruk. Gjenstander med lav kompleksitet er nyheter, nettsider, bøker, Cd-er og filmer. Gjenstander med høy kompleksitet er digitale kamera, mobiltelefoner og Pc-er.

## **Brukere**

Brukere er en betegnelse på alle som benytter seg av anbefalingssystemet, de kan være forskjellige og de kan ha forskjellige grunner for å benytte seg av anbefalingssystemet. Det er viktig å utvikle en brukerprofil som representerer den bestemte brukeren på en korrekt måte, derfor må informasjon om brukeren samles. I anbefalingsmetoder som kollaborativ filtrering, lager man en liste med brukerens vurderinger, men man kan også bygge en demografisk brukerprofil bestående av informasjon som kjønn, alder, yrke, utdanning og bosted. En brukerprofil kan også bestå av data samlet inn om brukeradferd. Hvilke nettsteder brukeren ser på kan for eksempel være relevant i et anbefalingssystem for nettressurser. En brukerprofil vil ofte bestå av flere typer data. Jeg vil komme tilbake til selve innsamlingen av data.

## **Transaksjoner**

Transaksjoner er forholdet mellom gjenstander og brukere, samt andre forhold. Det en bruker gjør med gjenstander eller andre handlinger brukeren foretar seg. Transaksjoner kan ofte oppstå ved menneske-maskin interaksjon. Transaksjoner kan brukes om tilfeller der en bruker vurderer en gjenstand eller handlinger som å se nærmere på en bestemt gjenstand i en nettleserøkt. Eksempler på transaksjoner er vurderinger, anmeldelser, data om hva en bruker har klikket seg inn på, søkt på og på andre måter interagert med.

I et anbefalingssystem snakker man primært om to forskjellige innsamlingsmåter for data. Alag (2009), Ricci et al. (2011) og Lops, Gemmis & Semeraro (2011) legger frem begrepene eksplisitt og implisitt data. Eksplisitte data er ”frivillig” data som brukeren bevisst eller aktivt gir fra seg. Anmeldelser, omtaler, tagging og vurderinger faller innunder denne kategorien. Brukeren har selv satt seg ned og gjort beviste handlinger for å registrere denne dataen. Implisitte data er indirekte angitt data som systemet selv fanger opp og lagrer om brukeren, her er det ofte snakk om handlinger brukeren gjør. I en ressurs kan dette være spesifikke gjenstander som brukeren har sett eller søkt på. En del nettsteder overvåker brukere også utenfor systemet og registrerer data derfra. Blant annet kan blogger brukeren skriver, eller andre nettsteder brukeren har sett på registreres.

## **2.2 Forskjellige anbefalingsmetoder**

Uavhengig av hvilken anbefalingsmetode som velges er hovedfunksjonen å forutsi hvilke gjenstander som er verdt å anbefale for en bestemt bruker, gjenstanden skal være nyttig for brukeren. Målet er å få brukeren til å bli så interessert i en gjenstand at han velger å benytte seg av den, Amazon.com vil selge produkter mens Bokelskere.no vil vise brukeren bøker han ønsker å lese. Innledningsvis har jeg nevnt syv anbefalingsmetoder, som bruker forskjellige fremgangsmåter for å generere anbefalinger. Anbefalingsmetodene er som tidligere nevnt inndelt i klasser basert på datakilder de benytter seg av, målet med anbefalingen og grunnlaget for analysen, for å utarbeide anbefalingene. Felles for de syv metodene er at de operer med et likhetsmål, men har forskjellig måter å finne dette på. Jeg vil i dette avsnittet gå gjennom de syv anbefalingsmetodene og illustrere dem med eksempler.

### **2.2.1 Innholdsbaserte metoder**

I innholdsbaserte metoder er det ”innholdet” i gjenstanden som bestemmer om en gjenstand blir anbefalt eller ikke. Metoder for denne typen anbefaling stammer hovedsakelig fra informasjonsgjenfinningsfeltet. Et innholdsbasert anbefalingssystem prøver å anbefale gjenstander som ligner de gjenstandene den aktuelle brukeren tidligere har likt. Likhet mellom gjenstander er kjernen i denne metoden. Systemer som bruker innholdsbaserte anbefalingsmetoder analyserer gjenstandene som en bruker har evaluert, brukt eller kjøpt, og prøver gjennom dette å lage en profil basert på brukerens interesse. Deretter foretas det en relevanseevaluering i forhold til mulige gjenstander og gjenstander knyttet til brukerens profil

(Anand & Mobasher 2005; Lops et al. 2011). Det er lett å forklare denne typen anbefalinger for den aktive brukeren ved å si: Du har tidligere kjøpt produkt X det er derfor trolig at du vil like produkt Z, siden det ligner på produkt X.

Ved innholdsbaserte metoder brukes forskjellige algoritmer for å måle likhet mellom gjenstander og ved vurderingen av hvor relevant gjenstanden er for en bruker. Lops et al. (2011) forklarer hvordan denne utvelgelsen skjer i systemet fordelt på tre hovedkomponenter; Content Analyzer, Profile Learner og Filtering Component.

Content Analyzer foretar en innholdsanalyse av datakilden, gjenstandene. Innholdet karakteriseres ofte som gjenstandens attributter. Hvis gjenstanden er en bok vil attributter være tittel, forfatter, emne og/eller tagger, ISBN og andre bibliografiske data. Disse skal representeres i et format som er lesbart og sammenlignbart. Ved ustrukturerte data, for eksempel anmeldelser, vil en tekstanalyse være nødvendig før selve innholdsanalysen. Ser en nærmere på dette trinnet er det tydelig at en Content Analyzer ofte bruker teknikker hentet fra informasjonsgjenfinningsområdet. Representasjoner av gjenstandene blir hentet inn og bearbeides til blant annet nøkkelord eller konsepter. Jannach, Zanker, Felfering & Fredrich (2011) poengterer at gjenstandenes attributter ikke er det egentlige innholdet i et dokument, men er en metadata representasjon av innholdet. Det egentlige innholdet i et dokument kommer bare frem ved fulltekst representasjoner. I forbindelse med denne typen gjenstander vi har så må gjenstandens attributter og da spesielt emneord, forfatter, tittel og nøkkelord representasjon av for eksempel omtale være nok til å bygge en representasjon av innholdet. Jeg bruker dermed begrepet innhold om disse datakildene.

Profile Learner samler data om brukeren, og prøver å generalisere dataene for å opprette en profil for brukeren. Som oftest vil generaliseringen foregå gjennom maskinlæringsteknikker som kan bygge opp en brukerprofil. For å samle data om brukeren er, i likhet med overnevnt trinn, metoder hentet fra informasjonsgjenfinningsfeltet. Det finnes i følge Herlocker et al. (2000) relevance feedback som gir en forståelse for hva en bruker liker eller eventuelt ikke liker. Et eksempel på dette er et nettsted hvor brukeren har gitt terningkast på bøker eller filmer. Dette er et eksempel på eksplisitt feedback, fordi brukeren aktivt legger inn informasjonen. Det finnes i følge Alag (2009) og Lopes et al. (2011) tre hovedmåter å samle eksplisitt relevance feedback. Liker/like ikke, skalert vurdering og tekstlige kommentarer. Liker/like ikke-måten gir brukeren mulighet til å signalisere om en gjenstand er relevant eller ikke relevant. Skalerte vurderinger



som terningkast eller stjerner gir brukeren en mulighet til å gradere sine tilbakemeldinger. Tekstlige kommentarer som anbefalinger og anmeldelser gir brukeren mulighet til å gi et meget personlig uttrykk for hvor godt gjenstanden er likt. Det er viktig å huske på at tekstlige kommentarer må analyseres før de kan brukes i anbefalingsprosessen. Ved bruk av implisitt feedback-måten vil brukerens handlinger bli overvåket og brukt som datakilde. For eksempel vil gjenstander der brukeren klikker seg inn for å lese en oppsummering av gjenstanden kunne anses som relevante, men dette må bestemmes i de forskjellige systemene. De trengs derfor ikke vurderes, men registrerte gjenstander er likte gjenstander. Hva en bruker liker kan dermed leses ut av den aktive brukerens historikk, og gjenstander representert i denne kan klustres for deretter å sammenlignes med potensielle gjenstander. Her kan for eksempel hva en bruker har kjøpt i en netthandel, eller hvilke bøker en bruker har i en boksamling på et sosialt nettsted brukes som historikk.

Til slutt vil Filtering Component bruke informasjonen innsamlet om gjenstander og brukere for å generere en anbefaling for den aktive bruker. I dette steget av prosessen handler det om å finne likhetskriteriet mellom nye gjenstander og tidligere gjenstander brukeren har gitt en form for positiv respons på. Målet er å rangere gjenstandene etter relevans og de høyest merkede gjenstandene vil bli gitt som anbefaling til brukeren. For handel av bøker på nettsteder som benytter en innholdsbasert metode vil dette bety at bøker ”likest” de bøkene brukeren allerede har kjøpt settes høyest på lista. Innholdsbaserte metoder er avhengig av at det finnes et ”innhold” å analysere i gjenstanden, derfor er metoden ofte brukt tekstlige dokumenter. Bøker, nettsider og dokumenter i en database er gode eksempler på denne type gjenstand.

Kjente utfordringer med innholdsbaserte metoder er cold start- problemet med nye brukere, datadekning og overspesialisering. Når en ny bruker registreres får den innholdsbaserte måten å gi anbefalinger på problemer fordi den ikke har en tilstrekkelig brukerprofil å basere anbefalinger på. Den nye brukeren må evaluere en viss mengde gjenstander for å generere anbefalinger som reflekterer hans preferanser. Datadekning i denne sammenhengen blir ofte referert til som et problem med et tynt eller spredt datagrunnlag. Dette gjenspeiler antall attributter som er registrert på gjenstandene og her er det viktig å registrere nok attributter slik at det blir lett å skille de forskjellige dokumentene. Hvis det foretas tekstanalyse må denne være grundig nok til å skille dokumenter fra hverandre. Det er også viktig at dokumentene kan grupperes, slik at anbefalingene til bruker blir gode.

Innholdsbaserte metoder er ofte utsatt for overspesialisering, at det foreslår for like gjenstander i sine anbefalinger. I forbindelse med denne utfordringen er begrepet serendipitet mye brukt. Serendipitet<sup>12</sup> kan kalles det motsatte av overspesialisering. Det stammer fra Sri Lanka, hvor det fortelles eventyr om de tre prinsene fra Serendip og deres evne til å gjøre oppdagelser uten å anstrenge seg. Selve ordet har mange definisjoner, og har blitt populært forskningsobjekt innenfor flere fagområder. Roberts (1989) forteller at oppdagelsen av uttrykket serendipitet tilhører Horace Walpole fordi han brukte det i et brev i 1745 hvor han fortalte om nettopp eventyret fra Sri Lanka. Kort kan man si at forklaringen for ordet serendipity er å gjøre en uventet men positiv oppdagelse. Dette forteller at vi har med flaks og det uventede å gjøre. Selve begrepet serendipitet brukes nå som en betegnelse når man snubler borti noe man ikke var på jakt etter, men som viser seg å være aktuelt.

I et anbefalingssystem for litteratur vil dette bety at brukeren sjelden eller aldri får utvidet sin litterære horisont. Siden anbefalingene baseres på likhet mellom innholdet i gjenstandene vil anbefalingene ligne mye på hverandre. Et eksempel på dette kan være en nettbutikk som anbefaler kunder bøker basert på tidligere kjøp. Brukerens første kjøp i den bestemte nettbutikken er alle de syv utgitte Harry Potter bøkene. Anbefalingssystemet vil automatisk tro at denne brukeren er overivrig interessert i fantasy litteratur og anbefaler bøker innenfor denne sjangeren i fremtiden når brukeren handler der. Virkeligheten kan være at brukeren har bokhyllen full og krim og spenningsbøker, men endelig hadde bestemt seg for å prøve seg på en annen sjanger. Derfor kan innholdsbaserte anbefalinger virke noe begrensende på brukers smak.

Sterke sider for innholdsbaserte metoder er i følge Lops et al. (2011) håndteringen av nye gjenstander, brukerselvstendighet og gjennomsiktighet. I og med at det er innholdet i nye gjenstander som blir analysert er det kjapt gjort å få dem inn i systemet etter innholdsanalysen. En bruker blir ikke sammenlignet med andre brukere og det er derfor ikke fare for en mismatch med andre brukere. Brukere liker ofte å få en forklaring på hvor for de blir anbefalt forskjellige gjenstander. Når det gjelder gjennomsiktigheten til anbefalingssystemer baser på innholdsbaserte metoder gjør den det lett for utvikler å si til brukeren at anbefalingen ligner på gjenstander han tidligere har likt. En bruker vil skjønne logikken i dette og det er ikke behov for å blande inn andre brukere. Det er lettere for en bruker å stole på seg selv enn hva ukjente liker.

---

<sup>12</sup> Søk på Serendipity på ordnett.no  
[http://www.ordnett.no/ordbok.html?search=serendipitet&search\\_type=&publications=17&publications=36](http://www.ordnett.no/ordbok.html?search=serendipitet&search_type=&publications=17&publications=36)

## 2.2.2 Kollaborative filtreringsmetoder

Kollaborativ filtrering handler kort fortalt om en type samarbeid eller menneskelig interaksjon som bruker personlige meninger. Schafer et al. (2007) mener at kollaborativ filtrering er et svar på den økende mengden tilgjengelig informasjon på nettet. Sammen med den økende mengden informasjon, og det at innholdsbaserte systemer som sorterer gjenstander kun på grunnlag av innhold ikke kan si noe om kvaliteten på gjenstanden, ble det sett etter andre metoder for å finne frem til de beste gjenstandene.

Vitenskapen fikk sitt første møte med kollaborativ filtrering i 1990-årene da denne metoden ble brukt for å filtrere epost ved Xerox Palo Alto Research Center. Der ønsket de en måte å sortere e-post slik at forskerne kunne finne de mest interessante e-postene uten å kaste bort tiden på de unyttige. Tanken på dette tidspunktet var at siden forskere innenfor samme felt ofte fant de samme artiklene nyttige ville informasjon om hva andre forskere innenfor samme felt leste og likte gjøre nytte for andre. Dette eksperimentet ble gjennomført med ”The Tapestry experimental mail system” og dermed var hovedtanken bak kollaborative filtrering sådd:

Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read. [...] Collaborative filtering is novel because it involves the relationship between two or more documents, namely a message and its reply, or a document and its annotations (Goldberg, Nichols, Oki & Terry 1992 s. 1).

Senere har denne formen for informasjonsfiltrering vist seg å være nyttig i flere forskjellige situasjoner. Kommersielle aktører bruker gjerne anbefalingssystemer for å fortelle en kunde hva andre kunder liker i håp om å selge mer.

For å gi en anbefaling med kollaborativ filtreringsmetoden bruker man i følge Massa & Avesani (2004) tre trinn: Første trinn er sammenligning av den aktive brukers vurdering og alle andre brukers vurderinger. Likhet mellom vurderingene blir regnet ut gjennom å bruke en algoritme og det er dermed en likhet mellom brukere som er i fokus. I det andre trinnet prøver anbefalingssystemet å gjette seg til hvordan den aktive brukeren ville vurdere gjenstander han ikke har vurdert. Dette gjøres ved at systemet tar for seg vurderingene som de andre brukerne, brukere som ligner den aktive brukeren, har gitt andre gjenstander. I det siste trinnet anbefaler systemet de gjenstandene som får den høyest antatte vurderingen for den aktive brukeren.

Et hovedmoment her er altså vurderinger. Flere forfattere deler vurderinger inn i tre kategorier; Skalerte vurderinger, liker/liker ikke og tekstelige kommentarer. (Schafer et al. 2007). En skalert vurdering vil ofte være stjerner, terningkast eller skala fra 1-10. En skalert vurdering kan også være verbal i form av; Helt uenig, litt uenig, nøytral, litt enig, helt enig. Amazon.com har stjernevurderinger der fem stjerner er best. Liker/liker ikke har to verdier; 1 eller 0, likte eller likte ikke, bra eller dårlig. I iTunes kan en bruker gi denne typen vurdering ved å trykke tommelen opp eller tommelen ned. Tekstelige kommentarer benytter seg av implisitt informasjon og kan hente informasjon fra andre steder enn ressursen. Et eksempel på dette kan være at en person har skrevet en omtale om en gjenstand de har kjøpt.

Ved kollaborativ filtrering skiller man mellom gjenstandsbasert og brukerbasert kollaborativ filtrering. Sarwar, Karypis, Konstan, & Riedl, (2001) og Jannach et al. (2011) skiller mellom disse begrepene slik: Brukerbasert kollaborativ filtrering er det som kalles den originale kollaborative filtreringen. Den aktive brukeren sammenlignes med andre brukere i databasen. Likhet mellom brukerne blir regnet ut på bakgrunn av tidligere vurderinger. Deretter foretas anbefalingen på bakgrunn av hva lignende brukere har vurdert og likt. Mens ved gjenstandsbasert kollaborativ filtrering er fokuset på gjenstandene og likhet mellom dem. Hva den aktive brukeren liker eller ikke liker er fortsatt viktig, men det er ikke denne informasjonen som står for selve likhetsmålingen.

Kollaborativ filtrerings algoritmer klassifiseres enten som memory-based, minnebaserte eller modell-based, modellbaserte algoritmer. Adomavicius & Tuzhilin (2005) og Su & Khoshgoftaar (2009) forklarer at ved minnebaserte algoritmer benyttes samlingen, hele eller deler av den, med gjenstander alle brukerne har vurdert. Dette ligger til grunn for å gi en prediksjon om hva den aktive brukeren kommer til å like. Den mest kjente minnebaserte algoritmen er neighborhood-baserte algoritmer. Der blir alle brukere blir delt inn i "grupper" med lignende interesser og smak og ved å identifisere "naboer" til den aktive brukeren kan en anbefaling for den aktive brukeren genereres. Ved modellbaserte algoritmer brukes en modell, som for eksempel maskinlæring og systemet lærer å gjenkjenne kompliserte mønster basert på et treningssett. Deretter foretar systemet en prediksjon basert på den brukte modellen og komplett datasett.

Schafer et al. (2007) mener at denne typen inndeling er foreldet siden de fleste algoritmer som nå benyttes innenfor kollaborative metoder nå bruker en form for for-databehandling. Som et

resultat at dette finnes det i praksis ingen algoritmer som er rene minnebasert eller rene modellbasert. Det er i virkeligheten nå snakk om hybride algoritmer som har lagret et utvalg av vurderinger i dataminnnet. De mener det er mer nyttig å se på om en algoritme er probabilistisk<sup>13</sup> eller ikke. Altså om algoritmen beregner sannsynlighet for hva en bruker vil vurdere en gjenstand til, eller om de beregner hvilke gjenstander en bruker vil like.

Utfordringer med kollaborative metoder er i følge Schafer et al. (2007) cold start-problemet for nye gjenstander, datadekning og det at en bruker forandrer smak. Cold start-problemet oppstår ved nye brukere som ikke har vurdert noe og nye gjenstander som ikke har blitt vurdert. Disse kan derfor ikke sammenlignes med andre brukere eller gjenstander. Flere anbefalingssystemer opererer med "falske" vurderinger for en bruker, til de har gitt tilstrekkelige vurderinger selv. De falske vurderingene kan for eksempel hentes fra hva den kollektive samlingen av brukere liker. Et dårlig datagrunnlag kan ha flere årsaker. Datasettet kan ha få vurderinger per gjenstand noe som gir unøyaktige anbefalinger. Årsaker til dette kan være få vurderinger generelt, for få brukere eller for stort datatilfang. Anbefalingene blir gitt på bakgrunn av hvilke vurderinger brukeren gir, men hvis brukeren forandrer smak vil systemet få problemer med å generere gode anbefalinger. Brukerens tidligere vurderinger kan da sees på som falsk informasjon, men det er ikke sikkert en bruker tar seg bryet med å endre dette selv eller at de er oppmerksomme på det.

Sterke sider for kollaborativ filtrering er i følge Herlocker, Konstan, Bochers & Riedl (1999) at man kan gjøre anbefalinger basert på kvalitet og smak, serendipitetsprinsippet blir bedre bevart og systemet kan håndtere ikke tekstlige gjenstander bedre enn et innholdsbasert anbefalingssystem. Brukers vurdering er med på å sikre kvalitet og smak, dermed er det mennesker som avgjør "interesseverdien" til en gjenstand. Serendipitet oppstår fordi brukere har forskjellig smak, og noen ganger vil det bli anbefalt gjenstander som den aktive brukeren ikke har hørt om før. Anbefalingene som blir presentert vil ikke bli anbefalt på bakgrunn av likhet mellom gjenstander som ved innholdsbaserte metoder, men fordi brukere som ligner den aktive brukeren har likt de. Det er likevel stor sannsynlighet for at de er aktuelle fordi andre brukere har vurdert det høyt nok til at anbefalingssystemet tar det med.

---

<sup>13</sup> Probabilistic: of or relating to or based on probability. (Wordnet)  
<http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=probabilistic&i=1&h=00#c>

### **2.2.3 Demografiske metoder**

Et anbefalingssystem som bruker demografiske metoder anbefaler gjenstander for brukere basert på en demografisk profil. Demografisk data karakteriserer mennesker, og blir ofte brukt innenfor statistikk. Demografisk anbefaling ser på brukere for å lære forholdet mellom en gjenstand og den typen mennesker som liker lignende gjenstander. Her er det likhet mellom brukere som i fokus og denne likheten blir gitt ved brukerkarakteristikk.

Alder, kjønn, utdannelsesnivå, yrke, inntekt, siviltilstand, religion med mer, er karakteristikk som ofte blir brukt på et menneske. I en demografisk profil samler man mennesker med flere felles trekk for eksempel er det trolig at kvinner i alderen 30-40 år er mer interessert i å få reklame for magasinet KK, enn menn i samme alder. I reklamebransjen har man benyttet seg av demografiske data lenge og har som et eksempel på dette tilpasset reklameinnslagene i et program deretter. På norsk tv vil man under visning av barneprogrammer, på reklamefinansierte kanaler, om morgenen finne et større innslag av reklamer for leker enn senere på dagen.

Pazzani (1999) mener at demografiske anbefalingsmetoder ikke har vært de mest utbredte da de er avhengig av en god del data fra brukeren, som brukeren kanskje ikke ønsker eller tar seg tid til å registrere. I sin forskning har han tatt for seg søkemaskiner og hentet demografiske data om brukerne som benytter seg av muligheten til å like/ikke like forskjellige nettsider. For de som likte en spesiell restaurant hadde nettsiden en positiv verdi, mens de som ikke likte den hadde nettsiden en negativ verdi. Forsøket viste at demografisk anbefalinger fungerer bra som et supplement til andre anbefalingsmetoder. Han anbefaler å bruke en hybrid av innholdsbaserte, kollaborative og demografiske metoder.

### **2.2.4 Kunnskapsbaserte metoder**

Ved noen former for anbefaling må man å gå dypere inn i hva brukeren trenger og hvilke forutsetninger som ligger til grunn. Dette gjelder spesielt store avgjørelser som kan innebære en større pengesum. I situasjoner hvor en bruker for eksempel skal kjøpe ny bil, hus eller PC, eller i situasjoner hvor brukeren skal investere penger trengs en annen form for informasjonskilde en den som brukes i tradisjonelle anbefalingssystemer. Informasjon samlet opp om avgjørelser som strekker seg lenger tilbake i tid vil være mindre relevant en det brukeren har behov for og ønsker der og da. Hvis man skal kjøpe nytt hus vil kjøpesum og beliggenhet trolig være de to første

betingelsene man setter til kjøpet. Til denne typen anbefaling brukes kunnskapsbaserte metoder. Likhets måles her mellom de forutsetningene og kravene en bruker har og hva systemet kan tilby.

Kunnskapsbaserte metoder vil i følge Burke (2007) og Jannach et al. (2011) på et generelt plan består av at brukeren aktivt gir fra seg et sett med relevante opplysninger på spørsmål tilpasset det bestemte systemet. Ut i fra dette vil systemet forsøke å finne et produkt som passer best mulig med brukerens svar. Hvis det ikke er samsvar mellom betingelsene brukeren setter kan svaret på forespørselen bli at et slikt produkt ikke finnes. Det finnes to forskjellige fremgangsmåter for kunnskapsbasert anbefaling; constraint-based anbefaling og case-based anbefaling. Ved constraint-based anbefaling vil man ha et sett med anbefalingsregler og anbefalinger blir generert ved å finne gjenstander som etterkommer reglene. Mens i case-based anbefaling brukes likhetsmål mellom brukerens betingelser og gjenstander som skal anbefales.

Jannach et al. (2011) skisserer kort hvordan en bruker interagerer med et kunnskapsbasert anbefalingssystem: Brukeren spesifiserer de betingelsene som ønskes for eksempel i et nettbasert skjema. Hvor detaljert skjemaet skal være varierer avhengende av hvilken ressurs som benyttes og hva som er formålet ved interaksjonen. Når nok informasjon om brukerens betingelser og formål er lagt inn vil brukeren bli presentert for et sett med anbefalinger. Det er opp til systemet om en forklaring følger med anbefalingene. Som oftest gis brukeren mulighet til å gå tilbake å endre enkelte svar. Deretter må brukeren velge om han skal benytte seg av anbefalingene eller ikke.

## **2.2.5 Kontekstuelle metoder**

Innenfor tradisjonelle anbefalingssystemer har man konsentrert seg om to hovedkomponenter, brukere og gjenstander. Den senere tiden har en tredje faktor gjort seg gjeldende, nemlig kontekst. Spesielt med utviklingen av allestedsnærværende applikasjoner for mobiltelefoner og annet mobilt digitalt utstyr, men også for mer tradisjonelle anbefalingssystemer er dette aktuelt. I forbindelse med denne anbefalingsmetoden snakker man om likhet til brukers kontekstuelle behov. Før jeg går nærmere inn på hvordan man kan utnytte kontekst i et anbefalingssystem er det viktig å ha klart for seg hva som legges i begrepet kontekst.

Et begrep som kontekst brukes innenfor mange fagområder, og har ikke overraskende flere forskjellige definisjoner. Den enkleste kan finnes i oppslagsverk som slår fast at kontekst er; sammenheng: den sosiale kontekst; sammenheng som et ord, uttrykk el. en setning står i: ordet

kan forstås ut fra konteksten<sup>14</sup>. De mer komplekse definisjonene er ofte utarbeidet innenfor respektive fagmiljøer og Adomavicius & Tuzhilin (2011) har samlet flere. Innenfor data mining kan kontekst defineres som hendelser som karakteriserer en brukers livsstadier og kan utgjøre en forandring i brukerens preferanser og status. Hendelser som er trukket frem er å få en ny jobb, få barn, gifte seg, skille seg og pensjonere seg. Ved netthandel skapes kontekst av brukerens, nå i egenskap av en kunde, grunner til å handle. Skal brukeren handle til seg selv? Er det en gave? Er det et kjøp i forbindelse med jobb eller hobby? Alle disse kriteriene vil avgjøre hva kunden er interessert i og hvilken kontekst gjenstanden skal brukes i. I et mobilt anbefalingssystem vil brukerens fysiske plassering spille en stor rolle. Kontekst i denne sammenhengen går på hvor brukeren befinner seg, hvilke andre ressurser eller personer som befinner seg i nærheten. En database brukes for å dekke et informasjonsbehov. Kontekst går i denne sammenheng ut på å identifisere dette informasjonsbehovet og bestemme på hvilket nivå det ligger. Med nivå så er det vanlig å skille mellom en bruker som har behov for noe informasjon om et emne og en bruker som skal ha all informasjon om et emne.

Dourish (2004) har konsentrert seg om menneske-maskininteraksjon, og har tatt for seg begrepet kontekst på en annen måte. Han går i dybden på konseptet og presenterer to syn på begrepet, representasjonelt og interaksjonelt syn. I det representasjonelle synet består kontekst av fire premisser. Kontekst er en form for informasjon, kontekst finnes, kontekst er stabil og kontekst er separat fra aktivitet. I det interaksjonelle synet er kontekst nærmest det motsatte. Kontekstualitet omfatter forholdet mellom objekt og aktivitet, kontekst defineres dynamisk underveis, kontekst avhenger av situasjonen og kontekst oppstår under aktivitet. Dourish fordyper seg i det interaksjonelle synet og mener bestemt at kontekst og innholdet i den aktiviteten brukeren er i gang med er viktig. Når det gjelder bruk av kontekst i anbefalingssystemer vil jeg trekke frem to tilnæringsmåter:

Adomavicius & Tuzhilin (2011) mener kontekst spesifiserer den kontekstuelle informasjonen som er assosiert med domenet anbefalingssystemet omfatter. De mener den kontekstuelle informasjonen kan deles inn i typer av kontekst, som tar for seg forskjellige aspekter. Eksempler på typer er tid, beliggenhet og meningen med interaksjonen. Den kontekstuelle typen kan i igjen ha en komplisert struktur slik at den deles inn med en hierarkisk struktur. Deres eksempel på dette er hentet fra et anbefalingssystem for kinofilmer. En

---

<sup>14</sup> [http://ordnett.no/ordbok.html?search=kontekst&search\\_type=&publications=23](http://ordnett.no/ordbok.html?search=kontekst&search_type=&publications=23)



kontekstuell type er *tid*. ”Time: Date → DayOfWeek → TimeOfWeek, Date → Month → Quarter → Year” (Adomavicius & Tuzhilin 2011 s. 225). De foreslår å legge til den kontekstuelle dimensjonen før tradisjonelle anbefalingsmetoder, etter tradisjonelle anbefalingsmetoder eller arbeide de inn i selve anbefalingsmetoden.

Anand og Mobasher (2007) bruker Dourish sitt interaksjonelle syn og har jobbet med anbefalinger på bakgrunn av dette. Deres modell bygger på hukommelsesteori innenfor psykologi og bruker begrepene langtidshukommelse og korttidshukommelse. De mener at en bruker trenger flere langtidshukommelser, fordi han eller hun vil interagere med systemet i forskjellige sammenhenger. I eksempelet brukt i teksten trekker de frem at en bruker av en nettbokhandel kan kjøpe bøker til seg selv i en arbeidssituasjon, en avslapningssituasjon og han kan kjøpe bøker til sin sønn. Hovedtanken bak deres fremgangsmåte er at systemet observerer brukerens handlingsmønster i den aktive sesjonen som de kaller korttidshukommelse og sammenligner det med tidligere sesjoner som de kaller langtidshukommelse. Etter en modell basert på menneskets hjerne. En bruker vil, ha mange forskjellige langtidshukommelser, som reflekterer tidligere mønstre. En bruker må oppgi hvilken langtidshukommelse den er i og på den måten kan brukeren ledes til den anbefalingskonteksten som er mest relevant innenfor den sesjonen. Det kan også tenkes at et anbefalingssystem for litteratur kan vurdere en lezers humør som en kontekstuell dimensjon. Systemet er da avhengig av at bruker opplyser om dette innenfor den aktive sesjonen.

## **2.2.6 Sosiale metoder**

Nettet er ikke gammelt men har allerede gjennomgått store forandringer. Fra nettets spede begynnelse med artikler som kun består av tekst og emneportaler, oppblomstringen av statiske hjemmesider og overgangen til dynamiske interagerende nettsamfunn med web 2.0 applikasjoner. Søkemaskiner må nå imøtekomme brukere med en enorm mengde teknologi og den digitale verden forsøker å integrere den sematiske webben. Det kan ikke lenger sies at nettet ikke kun befinner seg som et ”rom” der ute, eventuelt der inne på maskinen. Brukere er ikke lenger inaktive men er en aktiv del av nettet. De blogger, er med i et eller flere nettsamfunn, de tagger, de bidrar på wikier, de deler multimedia og de deler sin egen erfaring på nettet. Derfor er det viktig at anbefalingssystemer også benytter seg av denne type sosial informasjon.

Jannach et al. (2011) fokuserer på tre områder innenfor sosiale metoder. Brukerfelleskap og tillit, tagger og semantisk web. Brukerfelleskap og tillit blir ofte behandlet som to sider av

samme sak og forfatterne trekker frem at et brukerfellesskap er noe annet en at alle brukerne på et nettsted utgjør et fellesskap. Det er sannsynlig at en bruker vil stole mer på en anbefaling som er basert på brukerens ”faktiske” venner. Brukerens nettverk kan for eksempel utvides ved at brukeren velger hvilke andre brukere han ønsker å sammenligne seg med, eller informasjonen kan hentes eksternt. Bokelskere.no bruker denne funksjonen. Hvis man finner en bruker man ønsker å holde et øye med kan man ”følge” brukeren og på den måten vite for eksempel hva vedkommende leser og hvilke bøker de har i sitt bibliotek. Det er en teori at kvaliteten på anbefalingene vil gå opp, hvis man bruker kollaborative metoder for å regne ut likheten siden brukeren ofte sies å være venner med likesinnede.

En tag er et fritt emneord som en bruker setter på en gjenstand eller en ressurs, som for eksempel nettsteder. Brukergenererte klassifikasjonssystemer som baserer seg på tagger kalles folksonomier. Folksonomier er en nødvendighet på nettet da det uten brukerhjelp ville være for tidkrevende å klassifisere og systematisere innholdet som finnes der, og er nå veldig utbredt. Smith (2008) trekker frem fire karakteristikker ved folksonomier. Tagging blir gjennomført individuelt, tagger blir sammenkoblet, forhold blir sammensluttet og metoden for sammenslutning kan variere. Marinho et al. (2011) tar for seg anbefaling på nettsteder for sosiale bokmerker, som Flickr<sup>15</sup> og Delicious<sup>16</sup> og de sier at det utfordrende med denne type anbefaling er at brukere ikke bare interessert i å finne spennende innhold. Brukerne bruker nettstedene til å finne gode tagger og andre brukere. Utgangspunktet for anbefaling vil med en sosial metode kunne minne om anbefaling med kontekstuelle metoder. I stedet for de tradisjonelle datakildene brukere og gjenstander har man en tredje datakilde nemlig tagger.

Problemområder med å bruke tagger som en datakilde er i følge Milicevic, Nanopoulos & Ivanovic (2010) friheten folksonomier gir. De oppsummerer problemer med å samle åtte kategorier basert på andre forskere.

1. Tags have little semantics and many variations. [...]
2. Most tagging systems have their own specific way of functioning with and interpreting the meaning of tags. [...]
3. As an uncontrolled vocabulary that is shared across an entire system, the terms in a folksonomy have inherent ambiguity, [...]

---

<sup>15</sup> [www.flickr.com](http://www.flickr.com)

<sup>16</sup> [www.delicious.com/](http://www.delicious.com/)

4. Redundancy and ambiguity in the tag database both hinder the precision and recall of tagging systems. [...]
5. Some systems allow users to input tags separated by spaces. Problems arise when users would like to use phrases with multiple words [...]
6. The use of different word forms [...]
7. Some tags do not describe the document but gives judgment (“clever”). [...]
8. Many users in non-English-speaking countries tag documents in their own language[...] (Mathes; Shepitsen; Pluzhenskai; Gordon-Murname, gjengitt av Milicevic, Nanopoulos & Ivanovic 2010 s. 204).

Et anbefalingssystem må jobbe med disse utfordringene når tagger skal brukes som datakilde. En brukt metode er å begrense brukerne ved å legge inn forhåndsvalgte tagger i den foretrukkede formen for ressursen. Jannach et al. (2011) trekker blant annet frem hvordan man kan bruke leksikalske databaser som Wordnet for å møte folksonomiutfordringene. Et alternativ til leksikalske databaser er semantisk web. Disse metodene kan hjelpe til å spesifisere og standardisere det som skal måles i likhet.

Tim Berners-Lee (2001) redergjør for hvordan hans visjon om den semantiske webben kan lette den digitale hverdag for mennesker. Berners-Lee er fokusert på hvordan den semantiske webben bringer med seg struktur, mening og sammenheng. Han poengterer at nettet slik vi kjenner det nå er utviklet for menneske/maskin interaksjon. Han mener at i fremtiden vil maskiner ”forstå” data, uten at mennesker må være direkte involvert i prosessen. Ti år etter Berners-Lee presenterte sin visjon arbeides det fortsatt med å integrere den semantiske webben som en aktiv del av den digitale verden. Semantisk data blir publisert i en meget enkel struktur som kalles Linked Data og består av triples. En triples består som navnet tilsier av tre deler; subjekt, predikat (også kalt egenskap) og objekt. Sagt på enklest mulig måte representerer en triples; en ting som har en egenskap, den egenskapen kan være et objekt. Stolthet og fordom (subjekt) er skrevet av (egenskap) Jane Austen (objekt). Dermed er tanken at den semantiske webben kan hjelpe med klargjøring av begreper i et sosialt anbefalingssystem, og forsikre om at likhetskriteriet i et anbefalingssystem blir målt mellom riktige begreper. Problemet med Linked Data er at det ikke har kommet så langt ennå at den typen informasjon er tilgjengelig og dekkende for området.

## 2.2.7 Hybride metoder

Anbefalingssystemer streber etter å gi en bruker den beste mulige anbefalingen, men dessverre har de forskjellige metodene forskjellige svake sider. Jeg har tidligere nevnt overspesialisering, serendipitetsprinsippet, skalering og cold start-problemet som konkrete utfordringer. Jeg vil komme tilbake til forskning på problemområder senere i kapittelet. En potensiell måte å overkomme disse på er å velge en hybrid anbefalingsmetode som bruker flere av de tidligere nevnte anbefalingsmetodene. Dermed blir forskjellige mål på likhet kombinert, med en teori om at de forskjellige metodene skal utfylle hverandre.

For å forenkle forklaringen av hvordan hybride anbefalingsmetoder kan brukes ser jeg nå vekk fra forskjellige datakilder og bruker samlebetegnelsen innputt for brukerprofiler, sosiale data, data om gjenstandene eller en kunnskapsbaseregul. Selve anbefalingen finner sted, uavhengig av metode, i en anbefalingskomponent og produktet vi får ut er anbefalinger i en rangert liste. I denne sammenhengen er det kombinasjonen av anbefalingsmetoder som er spennende, ikke de forskjellige datakildene.

Burke (2007) har brukt tid på å gå igjennom hybride anbefalingsmetoder og kommet frem til at det finnes syv forskjellige typer hybrider. Jannach e. al (2011 s. 124) samlet disse syv metodene innenfor tre hybride design. *Monolithic-*, *parallelized-* og *pipeline design*. På norsk vil monolittisk, parallell- og seriedesign være dekkende begreper. Ved monolittisk design blir algoritmen som anbefalingsmetoden bruker manipulert. Algoritmer fra forskjellige anbefalingsmetoder kan slås sammen for å etterkomme datakilder og anbefalingssystemets mål bedre. Innputt sendes inn i én anbefalingskomponent fordi sammenslåingen skjer på algoritmenivå, resultatet er en rangert liste med anbefalinger.

Ved parallell- og seriedesign vil det finnes flere anbefalingskomponenter som foretar anbefalinger. Parallelldesign tar i mot og kjører samme innputt igjennom to eller flere anbefalingskomponenter. Der kan for eksempel det ene anbefalingskomponentet benytte en kollaborativ metode og det andre en innholdsbasert metode. Eller det kan være snakk om to metoder innenfor samme hovedmetode. Deretter blir resultatet slått sammen, et eksempel på sammenslåing er at de fem beste gjenstandene fra hver anbefalingskomponent blir dratt frem i en rangert liste. For seriedesign vil originalinnputt kjøres igjennom en anbefalingskomponent, resultatet fra første anbefalingskomponent vil være innputt i neste anbefalingskomponent. Deretter presenteres resultatet som en rangert liste. Eksempel på dette er å først bruke en sosial

metode, deretter bruker en innholdsbasert metode på resultatet før endelig anbefaling blir presentert for brukeren.

Fra dette kan man se at det finnes mange muligheter for å skape hybride metoder for anbefaling, og forskning på dette punktet er i skrivende stund et aktuelt tema. En utfordring på dette området er nok kunnskap. Kunnskap om originalmetodene man benytter seg av, hvordan de kan utfylle hverandre på best måte og hvor mange ledd man kan eller bør slå sammen. Den beste måten å finne ut av dette på er å prøve seg frem i forhold til de egenskapene som skal vektlegges i det aktuelle anbefalingssystemet innenfor det bestemte domenet. Først bør man skissere hva de forskjellige metodene gjør for seg, deretter hvordan de vil interagere for å se hva de gjør med anbefalingene. Det vil være forskjell på om utvikler benytter seg av serie eller paralleldesign, selv om det er de samme anbefalingsmetodene som slås sammen. For å vite hvordan de forskjellige anbefalingsmetodene påvirker et anbefalingssystem trengs informasjon om hvilke egenskaper som finnes i anbefalingssystemet.

## **2.3 Egenskaper i et anbefalingssystem**

Innenfor de syv forskjellige anbefalingsmetodene jeg har lagt frem finnes det sterke sider, utfordringer og problemområder. I tillegg er flere av anbefalingsmetodene fortsatt under utvikling og nye datakilder tas i bruk. Området er stort og jeg ønsker her å gi et overblikk over det som skjer på området, men går ikke inn for å gi en fullstendig oversikt. Det er vanlig at forskning setter søkelys på et problemområde eller en utfordring ved en anbefalingsmetode for å forbedre dette. Shani & Gunawarda (2011) går gjennom 13 egenskaper ved et anbefalingssystem som må evalueres. Da disse egenskapene i høyeste grad dekker forskning på området har jeg brukt dem som en mal på dette avsnittet, dermed konkretisert problemområder og kommet med forskning på det området.

### **2.3.1 Brukerpreferanse**

Å forstå seg på brukerne er en fundamentalsk nødvendighet for å utvikle et vellykket anbefalingssystem sier Picault, Ribière, Bonnefoy & Mercer (2011). Utvikling av brukerprofil og ønsket om å tilfredsstille brukeren gjør at utvikler må stille noen vesentlige spørsmål. Hvem er brukerne av ressursen? Hvilke forventninger og mål har brukeren ved å bruke denne ressursen? I hvilken kontekst brukes ressursen av brukerne? Vil brukernes forventninger forandre seg over tid?

Shani & Gunawarda (2011) mener at det ved oppstarten av et anbefalingssystem og evaluering underveis, er fullt mulig å gjennomføre brukerundersøkelser for å komme frem til hvilke preferanser brukerne har. De trekker frem at utvikler må være klar over at denne fremgangsmåten kan gi systematiske skjevheter og feil i forskningsresultatet. En av grunnene til dette er at brukere ikke er like. Velger utvikler å gjennomføre en undersøkelse og deretter å gå for det alternativet de fleste brukerne foretrakk, vil en del brukere nå bli dekket av et anbefalingssystem de ikke har full tiltro til. Derfor er det viktig å spørre brukerne hvorfor de foretrakk den ene fremgangsmåten foran den andre. Hvilke egenskaper er det den metoden brukerne foretrakk ivaretar, som er viktige for dem? Spørsmål som dette kan hjelpe utvikler å forstå brukerne.

Som tidligere nevnt er blant annet kjønn, alder, yrke og geografisk plassering brukt som brukerkaraktistikker. Dette gir utviklere en stereotypisk inndeling av brukere, et basisnivå. For et anbefalingssystem for litteratur er det ikke dekkende nok bare å ha tilgang til demografiske data om brukerne siden det ikke er slik at alle jenter på 12 kun er interessert i bøker om hester, selv om en viss mengde vil være det. Dermed må demografiske data brukes sammen med mer intrikat brukerdata. Utviklere trenger en forståelse for brukernes motivasjon, forventninger og mål. Er motivasjonen for brukerens interaksjon på vegne av seg selv, eller kan det tenkes at gjenstanden brukeren er på jakt etter skal brukes av noen andre? Hva brukerens mål med å bruke ressursen er avhenger sterkt av hva slags ressurs det er snakk om. Er det en nettbutikk, et sosialt nettsted, et bibliotek, en database eller en søketjeneste? Utvikler kan måle brukerpreferansen og det er nødvendig å forstå seg på brukerne, men undersøkelse på dette området må bearbeides med andre egenskaper i tankene. En utvikler må forstå hva som må forbedres hvis brukeren for eksempel syntes han får treff som ligner for mye på det han har lest.

### **2.3.2 Prediksjonsnøyaktighet**

Hvor nøyaktig et anbefalingssystem er til å forutsi hvilke anbefalinger det skal presentere for en bruker er et omdiskutert tema. Den allmenne oppfatningen er at en bruker foretrekker anbefalingssystemer som gir flest mulig relevante treff. Måling og evaluering av prediksjonsnøyaktighet er veldig utbredt og Shani & Gunawarda (2011) går igjennom tre måter å gjennomføre undersøkelser med dette som kriterier. Prediksjonsnøyaktighet i forhold til å forutsi vurderinger, forutsi bruk og forutsi rangering av gjenstander.

I anbefalingssystemer som baserer seg på vurdering ønsker man å forutsi hvilken vurdering en bruker ville gitt en gjenstand. Denne metoden kan brukes på ressurser der brukeren gir vurderinger i form av terningkast eller andre skalerte vurderinger. Bøker blir ofte vurdert på denne måten når brukeren er ferdig med å lese dem. Noen anbefalingssystemer bruker algoritmer som ønsker å forutsi gjenstander som brukeren kan komme til å bruke. Dette kan da brukes på gjenstander en bruker har vist interesse for men ikke har vurdert. Eksempel på dette kan være bøker som en leser har i sin boksamling, men ikke har vurdert. Velger man denne metoden er det vanlig å foreta offline undersøkelser som vurderes med kjente målekriterier hentet fra informasjonsgjenfinning. Presisjon, fullstendighet og spesifisitet er eksempler på dette. Undersøkelsen vil finne sted på et kjent datasett og Shani & Gunawarda (2011) bruker begrepene anbefalt og brukt (ekte positiv), anbefalt men ikke brukt (falsk positiv), ikke anbefalt men brukt (falsk negativ) og ikke anbefalt og ikke brukt (ekte negativ). Deretter kan vurderinger ”gjemmes” utvikler kan nå se hvor ofte anbefalingssystemet finner den rette vurderingen.

	Anbefalt	Ikke anbefalt
Brukt	Ekte positiv (ep)	Falsk negativ (fn)
Ikke brukt	Falsk positiv (fp)	Ekte negativ (en)

$$\text{Presisjon} = \text{ep} : (\text{ep} + \text{fp})$$

$$\text{Fullstendighet} = \text{ep} : (\text{ep} + \text{fn})$$

$$\text{Spesifisitet} = \text{fp} : (\text{fp} + \text{en})$$

*Tabell 1: Fullstendighet og presisjon.*

I noen ressurser vil det å forutsi en relevansrangert liste være viktigst. Denne metoden kan brukes hvis ressursen ønsker å vise anbefalingene i en rangert liste hvor den mest relevante anbefalingen vises øverst. Det finnes flere måter å behandle resultatet av anbefalinger på og hvordan det blir presentert for brukeren er viktig for bestemmelsesprosessen brukeren er igjennom når han velger å følge en anbefaling eller ikke.

Her er det viktig at utvikler har en klar formening om hvordan den rangerte listen skal se ut. Morville & Rosenfeld (2006) sier at det finnes to hovedavgjørelser når det gjelder å presentere resultatet. Den ene er hvor mange detaljer de forskjellige postene skal inneholde, mens det andre

er hvordan resultatet skal rangeres. Poster er i denne sammenhengen bøker i et bokanbefalingssystem. Hvor mange detaljer som skal vises om en bok er med på å bestemme hvor mange bøker man skal presentere samtidig. Det er viktig å huske at det visuelle er med på å avgjøre om en bruker velger en bok eller ikke. Morville & Rosenfeld (2006) poengterer at det er viktig at resultatlisten er ryddig og oversiktlig. Når det gjelder antall anbefaler de å vise 10 resultater, helst i et format som tilsier at brukeren ikke må rulle så langt nedover på siden. De poengterer at brukere ofte gir opp hvis de blir presentert for et stort resultat.

Sortering av en resultatliste kan gjøres på forskjellige måter. Morville & Rosenfeld (2006) trekker frem flere kriterier for sortering, alfabetisk, kronologisk, relevansrangert, klusteret og popularitet er blant disse. En relevansrangert liste viser bøker som er mest relevante for en bruker i forhold til den algoritmen som brukes. Denne metoden kan tenkes å bruke når man anbefaler litteratur. Det kan finnes mange titler som kan være aktuelle for den aktive brukeren, og erfaringsmessig vet man en bruker sjelden ønsker å bruke så langt tid på utvelgelsesprosessen. Dermed er det viktig at de mest relevante bøkene kommer høyest, og dermed tidligst i resultatlisten. Det kan tenkes at en kombinasjon av flere sorteringskriterier kan være et godt alternativ. En sortering på mest representerte sjangre og innbyrdes rangering innenfor et utvalg av sjangre kan være en mulighet.

Shani og Gunawarda (2011) mener det finnes to fremgangsmåter for å måle en anbefalingsliste. Utvikler kan undersøke om det forutsies en korrekt rangert liste, for deretter å sjekke hvor lik den rangerte listen er det brukeren faktisk gjorde i virkeligheten. Den andre metoden er å forsøke å måle nytteverdien av en rangert liste i forhold til brukeren. Velger man å undersøke likhet mellom den rangerte listen anbefalingsalgoritmen lager og den faktiske rangerte listen brukeren har satt opp blir dette gjennomført i et offline forsøk. Å måle nytteverdien i forhold til en bruker kan være en utfordring for da trenger man et mål på nytteverdi. Et populært alternativ er å anta at nytteverdien av en liste med anbefalinger er gitt ved å summere nytteverdien i de enkelte anbefalinger. Nyttverdien av hver anbefaling er nytteverdien av de anbefalte gjenstandene diskontert med en faktor som er avhengig av gjenstandens posisjon i listen av anbefalinger.

Adomavicius & Tuzhilin (2005) poengterer at de tradisjonelle måleenhetene for prediksjonsnøyaktighet trolig ikke er de mest egnede metodene for å evaluere et anbefalingssystem. Måling av fullstendighet og presisjon stammer fra informasjonsgjenfinning



og der er hovedmålet å finne dokumenter som passer en brukers behov, uttrykket ved en spørring. En annen grunn til at de ikke er egnet er at de kun tar for seg gjenstander en bruker har valgt å vurdere, og dette gir et skjevt inntrykk da brukere har en tendens til å vurdere gjenstander de liker oftere en gjenstander de ikke liker. Å gjennomføre forskning som tar for seg hele databasen sees på som for kostbart og gjøres sjeldent. Forfatterne etterlyser også evalueringer som tar for seg begreper som nytteverdi og kvalitet. De poengterer at det hjelper lite å anbefale lignende gjenstander hvis gjenstandene blir for like, eller hvis systemet anbefaler gjenstander som brukeren uansett ville benyttet seg av.

### **2.3.3 Dekning**

Begrepet dekning kan referere til flere forskjellige egenskaper ved et anbefalingssystem. Shani & Gunawarda (2011) sier man må vurdere om datasettet man har til rådighet har tilstrekkelig dekning i forhold til gjenstander og brukere. Det vanligste i denne sammenhengen er at begrepet dekning refererer til andelen av gjenstander som anbefalingssystemet kan anbefale, ofte kalt katalogdekning. Dekning kan også være andelen brukere som anbefalingssystemet kan anbefale gjenstander for. Cold start-problemet kan sees på som et problem innenfor dekning og rammer nye brukere i innholdsbaserte metoder og gjenstander i kollaborativ filtreringsmetoder.

Givon & Lavrenko (2009) gjennomfører et forskningsprosjekt for bokanbefalinger der hovedmålet er å bekjempe cold start-problemet innenfor kollaborativ filtreringsmetoder. De bruker 146 bøker fra Gutenberg-prosjektet og får dermed fulltekstdokumenter. For å lage ”falske” sosiale tagger bruker de en relevansmodell og hver bok får et sett med ekte og et sett med ”falske” tagger. Først undersøker de forskjellen mellom gjenfinning av bøker med falske og ekte tagger, den forskjellen viste seg å være ubetydelig. Deretter har de eksperimentert med tre forskjellige algoritmiske fremgangsmåter for å generere anbefalingene. En bruker-bruker algoritme, en gjenstand-gjenstand algoritme og en hybrid av de to algoritmene. De lager forsøk der 1, 3, 10, 20, 100, 300 og alle tagger er forbundet med dokumentet. De viser spesielt til den siste algoritmen hvor de bruker en hybrid algoritme og mener denne fungerer vesentlig bedre enn de to andre så lenge det er 10 eller mindre tagger tilgjengelig for dokumentet. Med dette mener de å ha funnet en fungerende metode å håndtere cold start-problemet.

Som et motstykke til den klassiske brukerbaserte kollaborativ filtreringsmetoden kom gjenstandsbasert kollaborativ filtreringsmetoder. Sarwar et al. (2001) legger frem forskning på dette området. De sier at ved denne metoden tar man de gjenstandene som en bruker har vurdert

og likt, deretter brukes disse gjenstandene til å finne gjenstander som ligner. På den måten leter man etter likhet mellom gjenstander og datasettet blir ikke så spredt. De testet flere algoritmer og er godt fornøyd med resultatet som de mener viser at gjenstandsbasert kollaborativ filtreringsmetoder klarer å håndtere store og spredte datasett bedre enn de tradisjonelle brukerbaserte kollaborativ filtreringsmetodene. Amazon.com bruker gjenstandsbasert kollaborativ filtrering nettopp på grunn av dette.

En teknikk som har blitt foreslått for å løse problemer med spedt datasett er å redusere dimensjonaliteten. Billsus & Pazzani (1998) forteller hvordan man kan fjerne uønskede eller uviktige brukere og gjenstander fra en brukermatrise for å redusere dimensjonaliteten. De brukte singular value decompression på et stort datasett og fikk positive resultater, i form av bedre anbefalinger, ved denne fremgangsmåten.

### **2.3.4 Tillit**

Shani og Gunawarda (2011) sier at tillit i denne sammenhengen går ut på om brukeren er villig til å stole på at anbefalingssystemet gir nyttige anbefalinger. De foreslår at et anbefalingssystem kan anbefale gjenstander som brukerne allerede har kjennskap til og liker. I et anbefalingssystem for litteratur kan dette være å anbefale flere bøker av samme forfatter, eller innenfor samme sjanger som brukeren allerede har i sin samling eller har lest. Det kan også være vanskelig å få brukeren til å stole på at anbefalingen er aktuell hvis han ikke har hørt om den anbefalte gjenstanden, da er det viktig å komme med en forklaring.

”Users will be more likely to trust a recommendation when they know the reasons behind that recommendation. Explanations will help users understand the process [...]” (Herlocker, Konstan & Riedl 2000 s. 2). Forfatterne er opptatt av at forkaringer skaper gjennomsiktighet til anbefalingssystemet. De mener at dette og bidrar til at brukere oppfatter situasjonen mer som en vanlig anbefaling, en venn som anbefaler en film vil si hvorfor han anbefaler nettopp den filmen.

LibraryThing og Amazon.com tilbyr brukere en forklaring ved å fortelle hvilken bok anbefalingen stammer fra, og de bruker anmeldelser fra andre brukere. Et anbefalingssystem som benytter seg av kollaborativ filtrering som metode vil ha større problemer med å forklare en anbefaling for brukeren. Det er vanskeligere å forklare noen på en kort og oversiktlig måte at de får anbefalinger fordi andre brukere har likt noe.

Mangel på tillit kan gjøre at en bruker ikke lenger benytter seg av anbefalingssystemet og forsterkes gjerne hvis gjenstandene som anbefales ikke blir likt av brukeren. En feilaktig anbefaling kan bryte ned mer tillit enn en riktig anbefaling skaper. Derfor er dette en viktig egenskap å ivareta. Shani og Gunawarda (2011) foreslår å undersøke egenskapen tillit med brukerundersøkelse. Man kan presentere brukere med en rangert anbefalingsliste og be de vurdere denne for så å spørre bruker i hvor stor grad de forskjellige anbefalingene var interessante og om brukeren stoler på at anbefalingssystemet gir de relevante anbefalinger.

Ben-Shimon, Tsikinovsky & Roka (2007) prøver ut en sosial anbefalingsmetode med MediaScout, en applikasjon som tilbyr personalisert media, filmer i dette forsøket, til mobil og tv. I MediaScout kan brukerne skaffe seg venner ved å sende venneforespørsel til andre brukere. En binær feedback mekanisme gjør at en bruker kan vurdere filmene i MediaScouter. Brukerprofilen består av tre lister, en for likte filmer, en for ikke likte filmer og en med brukerens venner. En bruker vil få anbefalinger som består av filmer vennene har likt. I eksperimentet brukte de et kjent antall brukere og filmer. Alle brukerne vurderte samme antall filmer, men ikke de samme filmene. Brukerne var studenter i samme klasse og fikk beskjed om å lage venneforbindelse med de andre i klassen de respekterte meningene til. Under selve forsøket ble fullstendighet brukt som måleenhet, på bakgrunn av at de mener alle filmene i datasettet er potensielle anbefalinger. Forfatterne mener de får et positivt resultat for anbefaling gjennom nære venner, men ønsker i fremtidige forsøk og benytte seg av mer omfattende måleenheter.

### **2.3.5 Selvtillit**

Hvor stor tro har systemet på egne anbefalinger, spør Shani & Gunawarda (2011). Dette går dermed ikke på hvor stor tillit brukerne har til systemet, men hvor stor tillit utvikler og systemet har til seg selv. Hvis et anbefalingssystem måler sannsynligheten for at en anbefaling vil være til nytte for en bruker i prosent, vil 100 prosent tilsi at det ikke er tvil om at brukeren vil ha nytte av anbefalingen. Det er ikke dermed sagt at et anbefalingssystem kun viser anbefalinger hvor det er 100 prosent sikkert at brukeren vil ha nytte av anbefalingen, siden dette vil begrense utvalget av anbefalinger mye.

Derfor må det i anbefalingssystemet settes en grense for hvor usikker man kan være på en anbefaling før man velger å vise den. Eventuelt kan man gjøre som LibraryThing gjør ved å legge ved informasjon om hvor sannsynlig det er at anbefalingen er treffende, dette tar jeg for meg i kapittel 3. Dette vil også gi en innvirkning på forrige egenskap. Bruker vil ikke nødvendigvis

miste tillit til systemet hvis det opplyses om hvor ”sikker” eventuelt ”usikker” anbefalingen er på forhånd. De vil heller ha en forståelse av at dette ikke er en perfekt anbefaling, men det kan tenkes at den faller i smak og kan være verdt å følge. Hvis boken ikke gjør det har anbefalingssystemet vært ærlig og kan ikke klandres for en mindre vellykket anbefaling. Denne egenskapen kan evalueres med et offline eksperiment som ligner på hvordan prediksjonsnøyaktighet måles.

### **2.3.6 Nyhet**

Å tilby brukeren en nyhet i anbefalingsøyemed er å tilby brukeren en gjenstand de ikke har kjennskap til mener Shani & Gunawarda (2011). Gjenstander brukeren ikke har sett, ikke har i sin samling eller ikke har vurdert. Den enkleste måten å påse at dette skjer er å filtrere ut gjenstander som brukeren har sett på, har i sin samling eller har vurdert. Det kan tenkes at det ikke alltid er tilstrekkelig, og at andre tiltak må gjennomføres. Et forslag er å tilføre nyheter ved å legge til gjenstander fra forfattere brukeren ikke har vurdert, eller har bøker av i sin samling. I et anbefalingssystem for nyheter vil denne egenskapen være ekstra viktig. Der er det et kort tidsrom hvor en nyhetssak er aktuell, før den blir så gammel at en bruker ikke trenger å vite om den. Denne typen anbefalingssystemer ser ofte på hvilke typer nyheter brukeren pleier å lese og kommer med anbefalinger innenfor spesifikke sjangre.

### **2.3.7 Serendipitet**

Denne egenskapen har jeg diskutert tidligere i kapittel 2, men det er en viktig egenskap. ”Serendipity is a measure of how surprising the successful recommendations are” (Shani & Gunawarda 2011 s. 286). Av dette kan man lese at det ikke er nok at anbefalingen er overraskende, den skal også være positiv i den forstand at brukeren har nytte av den. Serendipitet blir ofte brukt som et motstykke til prediksjonsnøyaktighet, da denne kan sies å gå ned når brukeren blir introdusert for noe nytt og uventet. Serendipitet kan legges inn i et anbefalingssystem enten med en algoritme, eller man kan tenke på det som et avvik fra de ”treffende” anbefalingene som oppnås ved å senke prediksjonsnøyaktighet.

Shani & Gunawarda (2011) mener at serendipitet i et anbefalingssystem for bøker kan økes ved å anbefale bøker fra forfattere den aktive brukeren har mindre kjennskap til fra tidligere av. De gjennomførte et forsøk hvor dette ble gjort på algoritmenivå ved å innføre en avstandsfunksjon mellom en bok og bøker i boksamlingen som brukeren allerede hadde lest.

Deretter lette de etter høyest representerte forfattere i samlingen og vektet ned disse bøkene. På denne måten blir bøker hvor forfattere er mindre kjent for brukeren rangert høyere på anbefalingslisten. Serendipitet er et aktuelt tema innenfor anbefalingssystemer og hvor nødvendig det er å vektlegge kommer sterkt an på domenet for anbefalinger. Dette blir knyttet til aktuelt domene og yterlige diskutert i kapittel 3.

### 2.3.8 Mangfold

Denne egenskapen sees på som motstykke til at noe skal være likt, og vil gå utover prediksjonsnøyaktighet. I enkelte anbefalingssystemer, deriblant reiseanbefalinger, ønsker man ikke å presentere mange tilsynelatende ”like” treff for brukeren, for eksempel fem treff til samme sted etter hverandre. Shani & Gunawarda (2011) bruker bokanbefalinger som et eksempel på et anbefalingssystem som ønsker å balansere mangfold og nøyaktighet på en god måte. De mener at innholdsbaserte metoder som kan skille gjenstander er mest brukt på dette området, da man har kontroll over hvor ”forskjellig” innholdet er.

Ziegler, McNee, Konstan & Lausen (2005) foreslår å bruke en metode for å øke emne variasjonen i en anbefaling. På denne måten ønsker de å presentere et anbefalingsresultat med større variasjon. Deres forsøk er foretatt på bøker og de mener at en bruker er tjent på at anbefalingssystemer som utvider en innholdsbasert metode trekker inn temaer som ikke er så populære. De så på anbefalingslistene som brukeren blir presentert med og undersøkte hvordan disse listene forandrer seg når de la inn en algoritme for å øke mangfoldet i anbefalingslisten. De mener at tradisjonell evaluering av anbefalingssystemer er for fokusert på prediksjonsnøyaktighet og at presisjon og fullstendighet ikke er gode nok målekriterier for å fange opp en brukers interessespekter. De tok for seg kollaborative filtreringsmetoder, nærmere bestemt en gjenstandsbasert og en brukerbasert metode. Data hentet de fra BookCrossing<sup>17</sup> og Amazon.com. Fra BookCrossing hentet de data om brukere, bøker og brukeres vurdering av bøker. Fra Amazon.com brukte forfatterne boktaksonomien de har utviklet der. Denne taksonomien brukte Ziegler et al. (2005) i algoritmen for mangfold, slik at det er klart definert hvilke bøker som er like og hvilke som ikke er det. De gjennomførte både offline eksperiment og online eksperiment.

Offline eksperimentet brukte de for å finne ut hvordan evalueringskriterier som presisjon og fullstendighet forandret seg når de vektet egenskapen mangfold med høyere og høyere

---

<sup>17</sup> <http://www.bookcrossing.com/>

mangfoldsfaktor, målt i prosent. Deres antagelser var at både presisjon og fullstendighet ville bli dårligere jo høyere mangfoldsfaktoren ble. Dette stemte med funnene som viste at både gjenstandsbasert og brukerbasert kollaborativ filtrering får lavere presisjon og fullstendighet. Eksperimentet viste også at gjenstandsbasert kollaborativ filtrering er mer mottagelig for mangfold enn brukerbasert kollaborativ filtrering.

Online eksperimentet ble foretatt med brukere av BookCrossing som informanter. Informantene ble kontaktet med e-post og i løpet av en treukers periode var det 2125 personer som deltok. Deltagerne ble presentert med en personlig topp-10 liste og stilt spørsmål om denne. Brukeren ble presentert for enten en gjenstandsbasert eller en brukerbasert kollaborativ filtreringsliste. Innenfor hver av de forskjellige kollaborative metodene brukte forskerne mangfoldsfaktor på 0, 30, 40, 50, 70 eller 90 prosent. Informantene ble bedt om å vurdere bøkene i listen, mangfoldet i listen, og hvor fornøyd de var med listen helhetlig sett. Når det gjelder vurdering av de ti bøkene på listen bekreftet denne undersøkelsen at prediksjonsnøyaktighet synker når mangfold øker. Informantene vurderte brukerbasert kollaborativ filtrering høyere enn gjenstands basert, på alle nivåer.

Forskerne brukte informantenes vurdering av listen som helhet for å undersøke brukertilfredshet. Deres hypotese var at økt mangfold også skulle øke en brukers tilfredshet med anbefalingssystemet. Samtidig vil den positive effekten ha en grense da både presisjon og fullstendighet synker når mangfoldet øker. For brukerbasert kollaborativ filtrering er ikke resultatet så oppløftende da tilføring av mangfold hadde en marginal og ubetydelig økning på brukertilfredsheten. Denne økningen ble etter hvert oppveid av den synkende prediksjonsnøyaktigheten. For gjenstandsbasert kollaborativ filtrering fikk de derimot et positivt resultat hvor brukertilfredsheten stiger, er på topp ved rundt 40 prosent og synker igjen. Igjen er det trolig lav prediksjonsnøyaktighet som gjør at brukertilfredsheten synker når mangfoldet stiger over 40 prosent. De konkluderer med at en algoritme for mangfold kan øke brukertilfredsheten når det gjelder gjenstands basert kollaborativ filtrering, men bare opp til et visst punkt.

### **2.3.9 Nytteverdi**

Det første man tenker på når det er snakk om nytteverdi er som regel økonomisk vinning. Et eksempel på dette er nettbutikker som ønsker å heve omsetningen. Dette er ikke aktuelt i forbindelse med alle typer anbefalingssystemer og vil avhenge sterkt på ressursen. Shani & Gunawarda (2011) definerer nytteverdi som den verdien enten ressursen eller brukeren sitter

igjen med på grunn av anbefalingssystemet. Nyttverdi i et anbefalingssystem for litteratur, hvor ingenting skal selges, vil for brukeren være at han får gode anbefalinger, noe som resulterer i at han er en fornøyd bruker. Fornøyde brukere er som tidligere nevnt lojale brukere og dette er nytteverdien for ressursen som trolig beholder brukerne sine.

### **2.3.10 Risiko**

Risiko går i følge Burke & Ramezani (2011) på hva brukeren risikerer ved å følge en anbefaling. Det er stor forskjell på å kjøpe en sang for 8 kroner på iTunes, kontra å kjøpe et hus til 2,3 millioner kroner. Risiko går direkte på brukerens toleranse for falske positive anbefalinger. En falsk positiv anbefaling vil lure brukeren i en nettbutikk til å kjøpe noe brukeren ikke trenger. I et anbefalingssystem som anbefaler brukere hvilke fond brukere skal plassere sparepengene sine vil derimot risiko være en meget viktig faktor. Det vil også være viktig å kartlegge hvor stor risiko en bruker er villig til å ta på forhånd. En bruker kan enten oppgi hvor villig han er til å ta risiko på forhånd, som ved plassering i fond. Eller så må systemet implementere en parameter som regner ut hvor mye risiko den aktive brukeren har tatt i sin historikk.

### **2.3.11 Robusthet**

Robusthet i anbefalingssystem gjenspeiler hvor flink et system er til å stå i mot angrep utenfra. Jannach et al. (2011) gjør rede for hvordan ondsinnede brukere går til angrep på et anbefalingssystem. Grunner til et angrep kan være å fremheve egne gjenstander, nedprioritere konkurrentens gjenstander eller rett og slett bare å sabotere systemet. Dette kan i all enkelhet gjøres ved å lage falske profiler og vurdere gjenstandene positivt eller negativt alt etter ønsket effekt. Forfatterne gir en kort innføring i de to mest brukte angrepene push- og nukeangrep. Pushangrep går ut på å heve sannsynligheten for at en gjenstand blir anbefalt, mens nukangrep er det motsatte og går inn for å forminske sannsynligheten for at gjenstanden blir anbefalt.

Mobasher, Bruke, Bhaumik & Williams (2007) gir et grundig innblikk i hvordan anbefalingssystemer kan motstå angrep. De tar spesielt for seg kollaborativ filtreringsmetoder, men også en hybridmetode hvor kollaborative og kunnskapsbaserte metoder kombineres. Angrepsmetoden med å lage og legge inn falske profiler kaller de profile injection-angrep, og det kan brukes til både push- og nukeangrep. De undersøker forskjellige algoritmers motstandsdyktighet i forhold til forskjellige profile injection-angrep. Det viser seg at brukerbaserte algoritmer er mindre motstandsdyktige enn gjenstands baserte algoritmer når det

gjelder pushangrep, mens gjenstandsbaserte algoritmer er utsatt når det gjelder angrep som går på en spesiell gruppe brukere. Hybride anbefalingsmetoder hadde en styrkende effekt på systemet og de mener det kan redusere omfanget til et profile injection-angrep. De avslutter med å gi en fremgangsmåte, med bestemte karakteristikk man skal se etter for å avsløre et profile injection-angrep.

### **2.3.12 Personvern**

Personvern er i høyeste grad et viktig punkt når det gjelder anbefalingssystemer da de fungerer bedre jo mer systemet vet om en bruker. En ressurs må verne om brukeropplysninger, og bør ikke gi ut opplysninger om en annen brukers preferanser. Hvis opplysninger om en bruker blir brukt i anbefalingsøymed skal brukeren anonymiseres. Shani og Gunawarda (2011) mener at nivået på personvernet kan gå utover nøyaktigheten i et anbefalingssystem. For å finne frem til den rette balansen kan en utvikler eksperimentere med algoritmer som er modifisert i personvernøymed.

Schafer et al. (2007) trekker frem hvordan anbefalingssystemer fungerer bedre jo mer informasjon det har om brukeren. Det er derfor viktig at en bruker kan stole på ressursen og gi fra seg informasjon uten å være redd for at opplysninger kommer på avveie. Brukere foretrekker personalisering av tjenester slår Kobsa (2007) fast, men samtidig ønsker de også å verne om privatlivet. Derfor blir det større og større press på personvernfokusert personalisering. Det er ikke kun personlige opplysninger som kan være komme på avveie, det er heller ikke trolig at en bruker ønsker å offentliggjøre mange års søkehistorikk. Forskning innenfor personvernfokusert personalisering ønsker å finne metoder for å gjennomføre den beste personaliseringspraksisen mulig innenfor visse personvernsgrenser.

Det er viktig for en leverandør av personaliserte tjenester å sette seg inn i personvernlover for det landet de opererer fra. Det finnes flere teknikker for å redusere faren for at brukerens personvern blir brutt. Kobsa (2007) samler noen velbrukte, og noen nyere teknikker for dette. Han begynner med å si at brukere bør ta i bruk pseudonymer og hvis det er mulig bør personopplysninger lagres på klientsiden. Han går igjennom hvordan kollaborative filtreringsmetoder ofte samler inn store mengder data i sentraliserte oppbevaringsmagasin. Denne metoden er spesielt utsatt for farer og Kobsa mener at kryptert data og bruk av pseudonymer er to av flere mulige teknikker her.



### 2.3.13 Tilpassningsdyktighet

I den digitale verden er det minimalt som er konstant over tid. Ressurser, meninger, tjenester og innhold er i kontant endring. I et anbefalingssystem vil både samlingen av gjenstander som skal anbefales og brukernes smak forandre seg. Derfor må en utvikler se på hvor sannsynlig dette er for sin ressurs og hvilke tiltak som må til for å møte denne utfordringen. For nyheter er det veldig viktig at ressursen oppdateres jevnlig og det en bruker er interessert i en dag kan være gammelt nytt dagen etter. En bok har en helt annen levetid enn nyheter. En bok vil ikke nødvendigvis tape seg over tid, men tilfanget av bøker vil øke. Samtidig er det ikke uvanlig at en bruker forandrer smak over tid. De er mulig å vekte en brukers vurderinger slik at de gjenstandene en bruker har vurdert nylig er de som teller mest. Mange brukere av Bokelskere.no oppgir at de har endret lesesmak siden de ble medlemmer på nettstedet. Mye av dette henger sammen med at de har oppdaget nye bøker gjennom nettstedet, gjennom anmeldelser av andre brukere, sitater og diskusjoner.

Shani og Gunawarda (2011) trekker frem at en bruker er opptatt av hvor fort en endring i vurdering eller det å legge til en ny gjenstand skal gi de en annerledes anbefaling. Hvis ikke denne endringen skjer slik at brukeren selv ser det, finner enkelte brukere det uvesentlig å gjøre endringer og mister tiltro til systemet. For et system som benytter seg av data som samles inn til faste tider vil forandring av anbefalinger ta litt tid, i hvert fall til neste oppdatering av systemet.

## **3 Anbefalingsdomenet**

Når et anbefalingssystem skal utvikles må systemet tilpasses det domenet det skal dekke. Burke & Ramezani (2011) sier at domenet til et anbefalingssystem er omfanget av gjenstandene som skal anbefales. En utviklers overordnede mål og en brukers grunn til å bruke systemet vil være viktig. Selve anbefalingens kontekst og hvilken ressurs anbefalingssystemet er en del av må også tas hensyn til, anbefalingssystemets rolle i ressursen og hva hvordan anbefalingssystemet skal implementeres. Jeg skal ikke gå dypt inn på implementering i denne oppgaven, men det er viktig å være klar over at valg i forbindelse med dette må tas. Anbefalingssystemets rolle vil derimot henge direkte sammen med domenet og gjenstandens om skal anbefales.

### **3.1 Anbefaling av litteratur**

Burke & Ramezani (2001) har arbeidet med å sammenligne anbefalingsmetoder og forskjellige domener, da de mener blant annet datakilder og formål er avgjørende i forbindelse med dette. Hva slags type informasjon utvikler har å arbeide med er i følge dem den enkleste måten å avgrense domenet på og derfor ser de på gjenstandene som skal anbefales. Gjenstandene jeg er interessert i å anbefale er bøker og domenet er litteratur. Derfor blir litteraturanbefalinger begrepet å bruke. Anbefalingene skal gjøre i en digital kontekst og det vil primært være nettbaserte ressurser, i form av sosiale nettsteder som står i fokus. En bok er et underholdningsmedium det er spesielt spennende å se hvordan en bok skiller seg fra andre underholdningsmedier som musikk og film. Med dette i tankene trengs et innblikk i boken som gjenstand og hvilke aspekter ved den som er spesiell i forhold til andre anbefalingssystemer for underholdningsmedier.

#### **3.1.1 En boks egenskaper**

En bok vil ha konkrete egenskaper som kan måles og brukes som sammenligningsgrunnlag for likhet innenfor et anbefalingssystem. En bok vil også ha subtile egenskaper som en leser bare kan oppdage ved å sette seg inn i boken. Jeg har tidligere sagt at jeg ikke skal fordype meg i litteraturteori i denne oppgaven, samtidig er det viktig å ha et overblikk over hvilke egenskaper konkrete og subtile, som kan appellere til en bruker. Hvilke egenskaper ved en bok gjør at nettopp denne blir likt, eller ikke likt av leseren og hvordan man skal sette ord på det? Er det kun gode bøker som skal anbefales og hva er en god bok?

Konkrete egenskaper uttrykkes ved bibliografisk data som, tittel, undertittel, sjanger, emneord, språk, omtalt tid, omtalt periode, forfatter, oversetter, redaktør, forlag, omtale, anmeldelse, og litteraturtype. Omtaler og anmeldelser kan struktureres gjennom tekstanalyse for å representere boken med for eksempel nøkkelord. De konkrete egenskapene er oversiktlige og lette å måle, analysere og bruke med bakgrunn i metoder for informasjonsvitenskap. En av de konkrete egenskapene brukere er komfortabel med å bruke er sjanger og noen ganger trekker de frem en sjanger som er mer aktuell enn andre sjangre. Sjanger blir også brukt for å spesifisere hva brukeren ikke vil ha. For eksempel en bruker som sier; jeg er ikke interessert i krimbøker.

Sosiale nettsteder legger også ofte til rette for at brukere selv kan sette emneord i form av tagger, noe som gir en annen dimensjon over emneordene. Smith (2008) trekker frem at tagger er fleksible. De kan brukes til å beskrive følelser, emner, om brukeren har lest en bok eller ikke og lar brukerne gi en form for vurdering av boken. Hvis et nettsted har et aktivt taggemiljø kan tagger sees på som kilde til egenskaper som sier noe om en boks appelleringsvevne. LibraryThing presenterer bøker ved hjelp av taggeskyer, hvor tagger som brukt ofte har større skrift enn de som blir brukt få ganger.



### *Bilde 1: Taggesky på LibraryThing.com*

En bruker kan da gå via taggeskyen som tilhører en bok til alle andre bøker som har den taggen brukeren velger å klikke på. Boken Shantaram har i LibraryThing følgende tagger: “adventure, Afghanistan, Australia, Australian, autobiographical, autobiography, biography, Bombay, crime, drugs, fiction, India literature, mafia, memoir, Mumbai, non-fiction, novel, own philosophy, poverty, prison, read, Roman, slums, TBR, to read, travel, unread, wishlist” (LibraryThing 2011). En bruker kan bruke denne metoden fra bok til bok, ved å følge et tema eller aspekt ved en bok med å følge taggesky etter taggesky.

Tagging har sterke kollaborative egenskaper og Smith trekker frem to viktige kollaborative punkter; “You can explore topics using tags of other users. [...] You might use tags to connect with other users who have shared interests” (Smith 2008 s. 27). Taggene i LibraryThing er med på å øke både muligheten til å utforske et emne, og oppdage andre brukere

med felles interesse. Tagger kan også brukes som tradisjonelle emneord, sammen med tradisjonelle emneord som en del av et dokumentets konkrete representasjon.

Dessverre er de konkrete egenskapene sjelden dekkende nok til alene til å beskrive en boks appell for en leser. De er nødvendig for å måle likhet mellom forskjellige bøker, men for å forstå en bok som gjenstand i et anbefalingssystem må en dykke dypere. De subtile egenskapene er vanskeligere å måle og analysere, men kan si mer om boken som gjenstand for brukeren. Saricks (2005) er interessert i hvordan man kan tilfredsstille brukere som ber om bokanbefalinger i folkebibliotek. Hun mener at det i bibliotek brukes to måter å snakke om en bok på, emneord og hvordan boken appellerer til brukeren.

Emneord er bibliotekets måter å katalogisere og konkretisere innholdet til den enkelte bok, og her brukes det konkrete egenskaper som nevnt ovenfor. Erfaringsmessig er ikke brukere så interessert i dette mener Saricks, og trekker frem at brukere ofte snakker om hvordan en bok får de til å føle. Brukere forteller hvorfor de likte en bok og hva som appellerte til dem. Å forstå hvorfor en bok appellerer til en bruker og vite hvilke subtile egenskaper i den bestemte boken gjør at brukeren likte den er viktig. Dette fordi et anbefalingssystem ønsker å presentere anbefalinger som er mest mulig relevante, og som brukeren vil velge å benytte seg av. Tilbake til Saricks (2005) og en boks appell til en leser som hun mener forfatteren konkretiserer gjennom fire områder; tempo, karakterisering, handling og ramme. Innenfor de fire områdene finnes det spesifikke spørsmål en kan stille en bruker for å finne ut hvordan han vil klassifisere boken. Saricks har utarbeidet spesifikke ord i anmeldelser og omtaler som kan brukes for å betegne en boks appell. Her trekker jeg ut ord brukt i forhold til subtile egenskaper:

**Pacing**

Breakneck, compelling, deliberate, densely written, easy, engrossing, fast paced, leisurely paced, measured, relaxed, states, unhurried

**Characterization**

Detailed, distant, dramatic, eccentric, evocative, faithful, familiar, intriguing secondary (characters), introspective, lifelike, multiple points of view, quirky, realistic, recognizable, series (characters), vivid, well developed, well drawn [...] (Saricks 2005 s. 66).

I en automatisert analyse av bokomtaler som Diskusjoner på Bokelskere.no kan denne typen ord oversatt til norsk være med å si noe om de subtile egenskapene. Her kan for eksempel ord som halsbrekkende, fengslende, overlatt, behersket, hurtig, lett og avslappet brukes om tempo.

### 3.1.2 Hva er en god bok?

Enkelt kan man si at en leser er interessert i den neste gode boken å stikke nesa ned i. Problemet er da å definere hva som ligger i begrepet en *god bok*? Saricks sier i forbindelse med anbefalinger i offentlige bibliotek. ”The term *good book* will be used in this book as a shorthand to mean fiction and nonfiction titles that are enjoyable to read and in which the writing meets a recognized quality standard” (Saricks 2005 s. 11). Begrepet god bok er lovende, men hva hun legger i begrepet er noe begrenset. Dette er trolig fordi personaliseringsaspektet ved anbefalinger gjør at en god bok ikke kan defineres på en enkel og universal måte.

Saricks snakker om litteraturanbefalinger generelt, samt metoder og teknikker for å yte denne servicen i et folkebibliotek og trenger derfor et konkret mål for anbefalingene. For enkelhets skyld defineres dette målet som en god bok, for å sikre brukertilfredshet. Som siste delen av sitatet hennes sier er en god bok en bok den bestemte leseren finner underholdende og som passer leserens kvalitetsnivå. Saricks sier også at en god bok kan være enten skjønnlitteratur eller faglitteratur. Siden blant annet bibliografier, selvhjelpsbøker og historiske bøker ikke er skjønnlitterære, men kan leses med underholdning for øye, ønsker hun ikke å skille mellom disse to litteraturtypene. Jeg vil gå dypere inn på skjønnlitteratur/faglitteratur diskusjonen senere i kapittelet.

Saricks går videre i sin søken etter å anbefale en god bok med å se på hvorfor en bruker velger en bok. Hun oppsummerer det slik; ”The elements of books to which a reader relates constitutes the appeal on the book for the reader” (Saricks 2005 s. 42). Videre legger hun frem hvilke elementer i en bok som sier noe om bokens appelleringssevne, nemlig bokens tempo, karakterisering, plott, ramme og stil. Informasjon som dette vil i biblioteket komme frem under en samtale mellom bruker og bibliotekar. På et sosialt nettsted kan man se på anmeldelser og dialoger i form av diskusjon, mellom brukere som en kilde til en brukers vurdering av bokens appelleringssevne. For å benytte seg av denne typen informasjon vil det være nødvendig å utarbeide et tekstanalyseprogram som kan trekke ut elementer som sier noe om hvordan boken, basert på brukers mening, er bygget opp. Dette kan være krevende, men det er mulig å gjennomføre.

Spørsmålet blir derfor hvordan man i et anbefalingssystem får tak i en brukers personlige oppfattelse av hva som er en god bok og hva som gjør en bok spesiell. I et intervju blir Tim

Spalding, mannen bak LibraryThing, bedt om å snakke om forskjellen ved å anbefale bøker og annen type underholdning.

[...] by and large, people buy the books they like. If they don't like them, there's still a reason they bought them—they thought they'd like them. In a way, people own the books they "deserve." Of course, you might have some gift book you hate but won't get rid of. In reality, those are a small part of anyone's library. In fact, most people rate books in the middle. The average is 3.8, and the standard deviation is low. We do a lot better looking at ownership patterns than using ratings. (Ritcher 2009 s. 4).

Nettopp fordi en bruker leser en bok for sin egen del, kan det være en idé å spørre brukerne hva de ønsker. Det er fullt mulig å gjennomføre en spørreundersøkelse på den aktuelle ressursen, eller i en omgivelse som ligner tilstrekkelig. I en undersøkelse som det kan man igjen støte på problemer. Fordi det som er en god bok for person A er ikke nødvendigvis en god bok for de 10.000 andre brukerne som er registrert på den aktuelle ressursen. På LibraryThing går skaleringen med stjerner fra 1 til 5. Spalding sier at gjennomsnittlig vurdering på LibraryThing er 3.8 fordi folk kjøper bøker de på forhånd tror de vil like, og dette gjenspeiler seg ved registrering av bøker på nettstedet. Han mener det er viktigere å se på en brukers boksamling som en helhet enn de forskjellige bøkene brukerne legger inn og vurderer. Det er trolig at dette også gjelder for Bokelskere.no hvor brukere kjøper, leser og legger inn bøker de i stor grad velger selv. Brukerens personlige "definisjon" av hva en god bok er vil gi seg til kjenne ved en analyse av den aktive brukerens boksamling.

En boks kvalitet og litterær kvalitet generelt et hett tema til diskusjon innenfor litteratursosiologi. I vår moderne tid er uttrykk som populærlitteratur, høyverdig litteratur, kiosklitteratur, bestselgere og klassikere med på å inndele bøker. Mange biblioteker har måttet ta et standpunkt om hva slags bøker de skal ta inn i boksamlingen. Er en god bok nødvendigvis en bok av høy kvalitet? Det finnes mange innfalsvinkler for å ta fatt på denne diskusjonen.

Robert Escarpit (1971) tok for seg det franske bokmarkedet på 1950 tallet og delte det inn i to forskjellige distribusjonskretsløp. Escarpits kretsløpsmodell har et kretsløp han kaller det dannede kretsløpet og et kretsløp han kaller det populære kretsløpet. Bøker fra det dannede kretsløpet blir solgt i middelstore bokhandlere til utdannede mennesker, borgerskapet, kunstnere og intellektuelle. Det er et lite kretsløp og omsetter færre titler og bøker fra færre forlag enn det populære kretsløpet. Det populære kretsløpet selger bøker til "massene" fra tobakksforetninger og aviskiosker, hvor boksalg ikke er hovedbeskjeftigelsen. Escarpit kaller dette en kommersiell

distribusjon til skjønnlitterær lesning som er tilgjengelig der leserne er. Han mener litteraturen i de to forskjellige kretsløpene er ikke skrevet av de samme forfatterne, for de samme leserne og den blir ikke gitt ut på de samme forlagene.

I tillegg til kvalitet snakker man ofte om popularitet blant bøker og forskjellige måter å måle dette. Sutherland (2007) gjør rede for bestselger fenomenet, og han poengterer at et bestselgerstempel på ingen måte er et kvalitetsstempel. Han setter bestselgere opp mot mer anerkjent litteratur, og mener at hvor mye en bok selger på lang sikt er et bedre tegn på kvalitet. Det at folk fortsatt er interessert i å kjøpe en bok 150 år etter den er utgitt mener han er et tegn på at boken innehar en litterær kvalitet. Sutherland mener også at hvor lett det er å trykke og gi ut bøker, prisnivået spesielt når man kan kjøpe heftede bøker, og lesingens formål er viktige faktorer for lesing i vår tid. At lesing i større grad er en underholdende aktivitet for leseren gjenspeiler seg i valg av litteratur og bestselgerfenomenet.

Den unike muligheten til å distribuere bøker i hele verden og digital litteratur gjør at ”hele” verden har tilgang til de samme bøkene samtidig. Han snakker om at bestselgere blir hypet opp og på denne måten selges det flere kopier av dem. Han mener forlag bevisst bruker aviser og egne bestselgerlister som en form for reklame, og som en form for kvalitet. Det er lettere å si at en bok er god når det er solgt mange kopier av den. Han poengterer at det på ingen måte må likestilles med kvalitetsbevisste analyser.

Et godt spørsmål i denne sammenhengen er om det er nødvendig å benytte seg av kvalitet i forbindelse med bokanbefalinger på et sosialt nettsted. Kan det være mer hensiktsmessig å se på andre kriterier, som for eksempel popularitet? På de sosiale nettstedene Bokelskere.no, LibraryThing og Visual bookshelf<sup>18</sup> finnes det en oversikt over hvilke bøker det finnes flest av i samlingene (jmf. Vedlegg 2). Bokelskere.no og LibraryThing oppgir også antall samlinger som inneholder bøkene, det har ikke Visual bookshelf. I listene kommer det klart frem at Rowling, med sine bøker om Harry Potter, er den mest representerte forfatteren. Med tre av 10 titler på Bokelskere.no, syv av ti på LibraryThing (de syv høyeste innføringene.) og syv av ti på Visual bookshelf. Hvis jeg skulle skrive masteroppgave om Harry Potter fenomenet hadde dette vært meget spennende, da dette viser at bøkene om trollmannsgutten har en enorm appelleringssevne til lesere i alle aldre.

---

<sup>18</sup> <http://www.facebook.com/apps/application.php?id=2481647302>

Når det gjelder begrepet en god bok mener jeg at hvis man skal bruk en boks kvalitet som en avgjørende egenskap vil det være viktigere å måle kvalitet i popularitet blant leserne. Denne populariteten kan man lese i hvilke bøker som er representert i en samling og hvilke vurderinger brukerne selv setter på bøkene i form av terningkast, stjerner eller tekstlige vurderinger som anmeldelser. Dermed vil ikke en ferdig definisjon av begrepet god bok være nødvendig, det som er viktig er å vite hvordan man finner frem til den aktuelle brukers mening om hva en god bok er. Det er der riktig valg av anbefalingsmetoder og egenskaper ved boken som skal vektlegges kommer inn i bildet.

### 3.1.3 Lesevaner

Spalding distanserer litteraturanbefalinger fra anbefalinger av andre typer underholdning i artikkelen om LibraryThing.

Books differ in a few ways. First, there are just *more* books—way more books than there are movies or CDs. Second, you experience books differently than you do movies. I see a lot of movies I don't like, but I mostly buy books I like. "Chick lit" is for chicks, but a "chick flick" is for guys to go to with their girlfriends, to suggest they're sensitive or to make up for making them sit through some action film. There are a couple of consequences of the differences. [...] For our purposes too, we're trying to entertain you most of all (Ritcher 2009 s. 4).

Spalding definerer heller ikke den gode boken, han prøver isteden for å forklare hvilke egenskaper en bok har i forhold til annen type underholdning og mener at det er de egenskapene som skal vektlegges. Spalding mener at lesingen er en individuell aktivitet, selv om mange velger å dele leseopplevelsen med andre i en sosial setting etterpå. Det er flere som deler denne oppfattelsen av lesing. "Å lese er å finne seg i en form for asosial ensomhet" (Claussen 2007 s. 177). Clausen mener at selv om man ser bort fra den fysiske gjennomlesingen ved det å lese en bok, er det en individuell opplevelse for en leser. Dette er fordi en leser reagerer på en bok på bakgrunn av personlige forutsetninger. Hvor leseren har vokst opp, hva han har lest tidligere, om han ble lest for i ung alder, hvor gammel han er, religion og sosiale egenskaper, er med på å forme ham som person og leser.

Long (2003) og Manguel (2007) befinner seg på den andre siden av denne diskusjonen og mener at lesing er en sosial aktivitet. Long snakker om "the social infrastructure of reading". Hun mener at det spesielt er to ting som gjør lesning til en sosial aktivitet. Det ene er at en person



lærer å lese med hjelp av andre mennesker, sammen med andre mennesker gjennom den sosiale omgivelsen en skole utgjør. Det andre er at lesere ofte i voksenlivet diskuterer bøker i blant annet bokklubber og på sosiale nettsteder. Manguel på sin side har tatt for seg lesingen fra dens spede begynnelse til dagens dato og trekker frem hvordan et barn blir lest for når det vokser opp. Gleden av å dele en bok og hvordan bøker har blitt analysert, nettopp for å dele innholdet med andre er med på å gjøre lesing sosialt.

Lesing som sosial eller individuell aktivitet kan diskuteres i det lengste, jeg mener derimot at lesing er en individuell opplevelse i det øye blikket det skjer. Jeg ser da vekk fra høytlesning til barn, og har den voksne leser i fokus. Den aktive leseren kan dele sin opplevelse med andre gjennom anmeldelser og diskusjoner på ett nettsted. Leseren kan sitere fra boken han leser til de som sitter rundt ham og dele boken på den måten. Uansett vil selve oppfattelsen av boken være leserens egen oppfatning. Den oppfattelsen skjer i det øyeblikket en leser foretar den faktiske lesingen. Det at lesing er en individuell opplevelse er med å ekskludere hensyn å ta til andre, boken som anbefales skal passe til en person og ikke flere.

I enkelte omgivelser som i en netthandel kan det tenkes at brukeren er på utkikk etter bøker til noen andre enn seg selv, men på et sosialt nettsted må vi regne med at brukeren er interessert i personlige anbefalinger. Dermed vil tidligere vurderinger, boksamling og dialog brukere i mellom være gode representanter for personlig smak. På et sosialt nettsted behøver man derfor ikke tenke på om den aktive brukeren også er den aktive leseren. Det kan likevel være et behov for å vurdere i hvilken kontekst en bruker trenger anbefaling. Kontekst i forbindelse med brukeren kan være hvor leseren er fysisk, formålet med lesing og leserens humør. Mennesker er følelsesstyrte i mange av valgene det foretar og det er ikke utenkelig at dette gjenspeiler seg i valg av lesestoff.

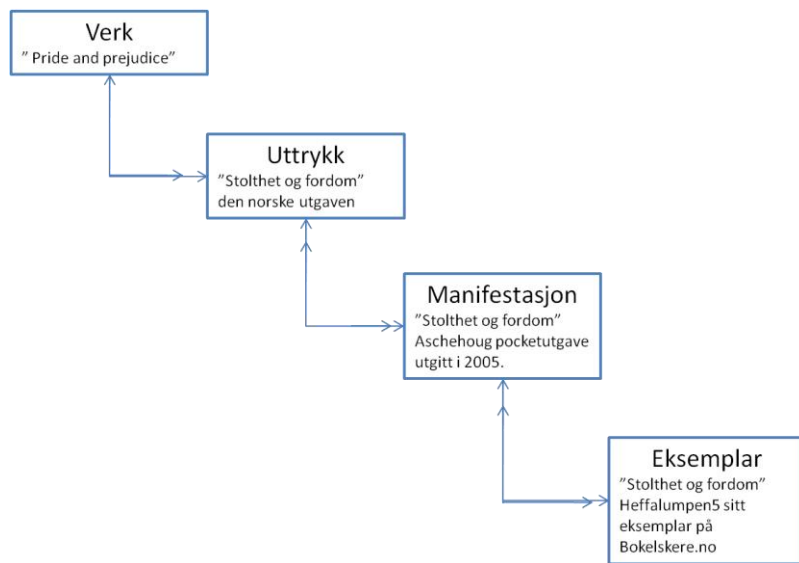
Brukere vil alltid være forskjellig og et nyttig spørsmål kan være: Liker lesere å utforske nye og ukjente leseropplevelser eller foretrekker de å lese mer av det samme? Dermed snakker man om serendipitet og mangfold i forbindelse med valg av litteratur. Serendipitet er som nevnt det å gjøre oppdagelser på bakgrunn av tilfeldige funn og vil i egenskap av sin definisjon omfavne mangfold. Lesere er som nevnt forskjellige av natur. Enkelte lesere liker formellitteratur som krim eller kjærlighetsromaner hvor historien og spenningstoppene bygges opp av gitte ”regler”. De liker å holde seg innenfor samme sjanger. Det finnes ikke noe fasitløsning på hva

slags brukere et sosialt nettsted har flest av, men det går an å gjennomføre forsøk for å balansere dette best mulig.

### 3.1.4 Utvalg av bøker og utgavekopier

Bokmarkedet er enormt. Det utgis bøker i hele verden, og oversetterbransjen jobber på spreng. Et eksempel på dette er Jean M. Auels siste bok hvor den engelske versjonen *The Land of Painted Caves* ble lansert 29. mars 2011. Den 30. mars samme år var boken å få kjøpt på norsk med tittel *De hellige hulers land*. Hvilket bringer oss inn på en velkjent utfordring innenfor bibliotek- og bokområdet, utgaveproblematikken. Forholdet mellom et verk, forskjellige utgavekopier og forskjellige formater har lenge vært et hodebry for bibliotekarer. Det er naturlig å se til bibliotekatalogen når en snakker om utgavekopier, selv om netthandel også må forholde seg til forskjellige verk er deres behov og bruk av verksidentifikator noe annerledes.

Tillet (IFLA 2000) har vært en stor bidragsyter til International Federation of Library Associations and Institutions [IFLA] utvikling av FRBR, som står for *Functional Requirements for Bibliographic Records*. Hun poengterer at hovedfunksjonen til bibliotekatalogen er å gi en bruker mulighet til å finne dokumentet brukeren leter etter så sant han kan tittel, eller forfatter eller emne. I håp om å imøtekomme dette bedre har FRBR måten å se på bibliografiske poster oppstått. FRBR modellen har fordelt de forskjellige entitetstyper forbundet med bibliografiske poster inn i tre grupper. I denne sammenhengen er det nærliggende å se på entitetstyper i gruppe 1: verk, uttrykk, manifestasjon og eksemplar. Verk er det selvstendige intellektuelle eller kunstneriske arbeidet som realiseres ved uttrykk. Uttrykk reflekterer det intellektuelle eller kunstneriske innholdet som finnes i en manifestasjon. En manifestasjon er den fysiske realiseringen av uttrykket i et verk som eksemplifiseres med et eksemplar. Eksemplaret er altså de forskjellige fysiske utgavene av et dokument, i vår sammenheng en bok. Verk og uttrykk er abstrakte definisjoner, mens manifestasjoner og eksemplarer er fysiske.



Bilde 2: FRBF modellen med eksempler fra Bokelskere.no.

Det har vært diskusjon rundt behovet for å skille mellom verk og uttrykk. "Oversettelser fra et språk til et annet, musikalske transkripsjoner og arrangementer, dubbede eller tekstede versjoner av en film er også sett på som forskjellige *uttrykk* av det samme originale *verket*" (International Federation of Library Associations and Institutions [IFLA] 2000).

### 3.1.5 Fag- og skjønnlitteratur

Vår litterære verden er sterkt farget av menneskers trang til å kategorisere og systematisere. Vi skiller konstant mellom fag- og skjønnlitteratur, men i våre dager har dette skillet blitt visket mer og mer ut. Nye sjangre dukker opp, forfattere blander fantasi og virkelighet og det er ikke like lett å vite hvilken type litteratur man sitter med i hånden eller på skjermen. Om et anbefalingssystem skal skille mellom fag- og skjønnlitteratur er et valg som må tas på bakgrunn av to forutsetninger. For det første om det er tilrettelagt for en slik type deling i den aktuelle ressursen. For det andre om blanding av fag- og skjønnlitteratur vil virke inn på hovedmålet med anbefalingssystemet.

Det er på ingen måte fornuftig bruk av tid og ressurser å diskutere om man skal tilby både fag- og skjønnlitteratur, eller velge en av typene, i anbefalingsresultatet hvis det ikke finnes en mulighet til å skille disse boktypene fra hverandre. Derfor må ressursen litteraturanbefalingen skal dekke undersøkes. Hvis det er mulig å skille mellom den type data står utvikler ovenfor tre potensielle muligheter. Tilby et blandet anbefalingssystem hvor både fag- og skjønnlitteratur er representert. Tilby et anbefalingssystem hvor brukeren kan velge enten skjønn- eller faglitteratur

som anbefalingsresultat, eller tilby begge i separate resultatlister. Tilby bruker kun den type anbefaling som utvikler har valgt er viktig, for eksempel kun skjønnlitteratur.

På nettstedet LibraryThing bruker de ikke sjanger, men bøkene tagges ofte med sjangerrelaterte termer. Det er fri tagging og det finnes flere nivåer en bruker kan velge å tagge på. "The Host" blir tagget med sjangerrelaterte termer som; Fantasy, Sci-fi, Science fiction, Fiction, Romance, Dystopia, Novel og Paranormal. Hvis emneord skal brukes for sjangerinndeling på denne måten er det avhengig av at dette er konsekvent gjennom hele nettsteder, altså alle bøker som er representert. I LibraryThing virker det absolutt som om denne type taggingen er tilstede på de fleste bøker. De har også en form for synonymkontroll som gjør at det de kaller alias av termer blir gruppert. For termen Fiction har de gruppert mer enn 180 termer.

På nettstedet Bokelskere.no som henter bøker fra flere forskjellige kilder er litteraturtyper en utfordring. Bøker hentet fra Bokdatabasen er utstyrt med litteraturtype inndeling, bøker hentet fra Amazon.com er ikke det. Det finnes dermed ikke muligheter i systemet til å dele inn bøker i litteraturtype på en konsekvent måte i skrivende stund. Dette må da foretas manuelt av de som driver nettstedet, noe som er ressurskrevende på et nettsted hvor det daglig legges inn nye utgavekopier av verk.

Valg om deling av litteraturtype kan kun besvares når man vet hva som er hovedmålet for anbefalingssystemet. Ved kommersielle anbefalinger i forhold til en nettbutikk er utvikler interessert i å selge mest mulig bøker. Utvalget av fagbøker er ofte stort og utgjør en betraktelig del av samlingen. En kunde som har kjøpt faglitteratur før vil trolig være interessert i å kjøpe det igjen. I et bibliotek vil leseropplevelsen og utlånsstatistikken konkurrere om oppmerksomhet. For bibliotek er utlånsstatistikken en måling av suksess, men den sier ingenting om utlånte bøker blir lest eller ikke. På et sosialt nettsted legger brukerne selv inn de bøkene de leser, og velger om de skal legge inn begge typer litteratur.

Aamotsbakken & Knudsen (2008) tar for seg sjangerutvikling og i den forbindelse trekker de frem Werlich som var interessert i teksttypologi. Werlich mener at teksttypologi er mer stabilt enn sjangerinndeling. Han skilte ikke mellom fag- og skjønnlitteratur i sine studier, men var mer interessert i de forskjellige skrivemåtene han fant gjennom litteraturstudiene. De fem måtene å skrive på, narrasjon, beskrivelse, argumentasjon, instruksjon eller eksposisjon, var viktigst.

Dermed flytter han fokus til hvordan teksten skal brukes, formålet med den. Skei (2006) tar direkte for seg tradisjonell inndeling av litteratur.

I våre dager er det stilt spørsmål ved holdbarheten i skillet mellom fiksjonsprosa (roman, novelle) og mange andre prosasjangere. [...] Det finnes svært mange blandingsformer, der det vi ellers ville kalle ”sakprosa” går inn i og spiller sammen med fiksjonsdelen av tekster (Skei 2006 s. 44).

Han trekker frem at dette skillet er viktigere å vurdere i en situasjon hvor man skal foreta en analyse av teksten enn ved lesing som gir spennende leseropplevelser. Selvhjelpsbøker er per definisjon faglitteratur, uten at jeg skal kaste meg inn i en lang diskusjon om hva som kvalifiserer til å få dette stempelen. Poenget er at selvhjelpsbøker er populært som aldri før og oftere og oftere fremstilles de som historier. Skrivemåter som narrasjon, beskrivelse og instruksjon blandes med resultat at bøkene leses både fordi brukeren vil både underholdes og lære noe nytt.

Zunshine (2006) tar i bruk kognitiv psykologi og bruker begrepet Theory of mind aktivt i sin bok *Why we read fiction*. Theory of mind er kort fortalt et menneskes egenskap til å forklare andres oppførsel, tanker, følelser, overbevisninger og begjær. Hun legger frem menneskers egenskap og trang til å forstå andre mennesker, også de fiktive. ”Intensely social species that we are, we thus read fiction because it engages, in a variety of particular focused ways, our Theory of mind” (Zunshine 2006 s. 162). Hvis Zunshines teori stemmer leser mennesker skjønnlitteratur for å tilfredsstille trangen til å forstå andre mennesker, også fiktive karakterer, engasjerer lesere. Sett i lys av denne teorien er det trolig at faglitteratur som forteller om mennesker også fenger på lik linje. Støtter man seg til dette minsker hensikten med å skille mellom fag- og skjønnlitteratur.

Ved anbefalinger innenfor sosiale nettsteder vil trolig leseropplevelsen være det viktigste. Brukere deler frivillig informasjon, legger inn den type bøker de er interessert i uavhengig av overordnet type og sjanger. De diskuterer aktuelle bøker og leter etter inspirasjon til neste bok de skal ta fatt på. Er det da nødvendig å skille mellom fag- og skjønnlitteratur? En bruker som er spesielt interessert i en litteraturtype eller sjanger vil trolig ha dette representert i sin samling. Anbefalinger vil bli generert enten på bakgrunn av en innholdsanalyse av den aktive brukerens samling, eller vurderinger brukeren setter på egne verk. Hvis en bruker konsekvent legger inn skjønnlitteratur er det lite trolig at et anbefalingssystem vil tilby faglitteratur, og motsatt. Hvis dette da skjer innimellom kan det hende det faller innunder serendipitetsprinsippet og at brukeren er tjent med det.

## 3.2 Anbefalingssystemer for litteratur

Innenfor anbefaling av litteratur finnes det flere aktive nettressurser i form av sosiale nettsted, nettbutikker og bibliotek. Jeg vil bruke det norske sosiale nettstedet Bokelskere.no som case og vil gå dypt inn på dette i kapittel 4, det er derfor ikke viet plass her. Jeg vil gå inn på LIBRA<sup>19</sup>, Amazon.com og LibraryThing for å se hva slags anbefalinger de gir en bruker. LIBRA er ikke lenger en fungerende ressurs, men et spennende forsøk på litteraturanbefaling og derfor aktuell å ta med. Amazon.com og LibraryThing er nettressurser som er aktive, og kun delvis åpne med hvordan de lager anbefalinger. I tillegg påpeker jeg kort noen andre ressurser som driver med litteraturanbefaling.

### 3.2.1 LIBRA

LIBRA ble utviklet sent på 1990-tallet og er som tidligere nevnt ikke lenger i funksjon, selv om en statisk versjon av nettstedet fortsatt finnes på nettet. Det er fortsatt spennende å se på da dette er et konkret eksempel på et innholdsbasert anbefalingssystem for litteratur. Nettstedet er utviklet av universitet i Texas og underlagt avdelingen for informatikk. Mooney & Roy (1999) legger frem hvordan LIBRA anbefaler bøker basert på en innholdsbasert metode hvor de bruker tekstkategorisering og henter halvstrukturert bokdata fra Amazon.com. LIBRA fungerer slik at en bruker må legge inn vurderinger fra 1-10 på et treningssett av boktitler. Deretter bruker systemet en Bayesian læringsalgoritme for å produsere en rangert anbefalingsliste. Brukeren kan gi relevance feedback på resultatlisten. På den måten kan systemet forbedre den rangerte listen underveis. Forfatterne mener dette er den beste metoden for et anbefalingssystem hvor brukerinformasjon er lav og er fornøyde med resultatet. Samtidig er de klar over hvilke begrensninger innholdsbasert metode har og derfor er de interessert i å bruke både innholdsbasert og kollaborative metoder i fremtidig forskning.

En svakhet i denne måten å generere anbefalinger på kan være måten brukerne velger ut bøkene de skal vurdere. Dermed er det trolig at brukeren bare søker opp bøker han eller hun liker. Dette resulterer i utelukkende positive anbefalinger som kan ha en innvirkning på anbefalingene som blir gitt.

---

<sup>19</sup> <http://www.cs.utexas.edu/users/libra/home.html>

### 3.2.2 Amazon.com

Amazon.com er utvilsomt en tungvekt innenfor netthandel. Nettbutikken er et amerikansk foretak med underavdelinger i mange land. Produktutvalget strekker seg fra bøker og musikk til mat, klær og verktøy. Det er også de som tilbyr det populære lesebrettet Kindle og de har i forbindelse med dette over 730.000 e-bøker<sup>20</sup>. De har gått hardt inn for personalisering av reklame og tilbud en kunde blir vist. De har lettet litt på sløret, men ikke helt, derfor starter jeg med en kort gjennomgang fra brukersiden.

Når en bruker kommer til Amazon.com sin forside er denne helt standard, men brukeren blir spurt om logge inn for personalisert anbefaling. Etter innlogging ønskes brukeren velkommen med et knippe anbefalinger. Brukeren kan endre på anbefalingene ved å klikke på lenken "fix this recommendation". Gjøres dette kommer det frem et vindu hvor brukeren får beskjed om at den anbefalingen er basert på et tidligere kjøp brukeren har foretatt. Deretter kan brukeren krysse av for at det tidligere kjøpet var en gave til andre, eller at den ikke skal brukes å generere anbefalinger. Den aktuelle anbefalte gjenstanden kan brukeren krysse av for om han eller hun eier, eller ikke liker. På den måten forsvinner den aktuelle anbefalingen, og andre anbefalinger basert på det kjøpet.

Før brukeren ser anbefalingen er det i midlertidig mye som skjer. Den faktiske algoritmen for hvordan anbefalingen blir generert gir ikke Amazon.com ut, men fremgangsmåten er allment kjent. For å unngå skaleringsproblemer og spede matriser, noe som fort kan oppstå med en nettbutikk med over 29 millioner kunder og flere millioner produkter for salg, har Amazon.com satset på det de kaller item-to-item kollaborativ filtrering. Linden, Smith & York (2003) forklarer metoden som blir brukt både i e-post kampanjer og på selve nettbutikken. Som tidligere nevnt går denne metoden ut på likhet mellom gjenstander. Anbefalinger på Amazon.com blir generert på bakgrunn av hva en kunde liker, likhet blir målt i kjøp og femstjernes vurderinger. Amazon.com har retningslinjer på hva de forskjellige stjernene betyr i forhold til den vurderte gjenstanden. 1 stjerne betyr at brukeren hatet den, 2 stjerner betyr at brukeren ikke liker den, 3 stjerner betyr at gjenstanden var ok, 4 stjerner betyr at brukeren likte den og 5 stjerner betyr at brukeren elsket den. Vurderingen kan en bruker sette på en gjenstand uansett om den er kjøpt på Amazon.com eller ikke. I stedet for å basere likhetsvurderingen på likhet mellom kundene, baseres det på kundens kjøpte og vurderte gjenstander og matchende gjenstander blir anbefalt.

---

<sup>20</sup> [http://www.amazon.com/Kindle-eBooks/b/ref=sa\\_menu\\_kbo8?ie=UTF8&node=1286228011](http://www.amazon.com/Kindle-eBooks/b/ref=sa_menu_kbo8?ie=UTF8&node=1286228011)

I stedet for en gjenstand-gjenstand matrise bruker de en algoritme som bygger en likhetstabell ved å finne gjenstander som kunder har en tendens til å kjøpe sammen. En av styrkene til Amazon.com er at de kan lage denne likhetstabellen offline, noe som er meget tid og pengebesparende. Metoden fungerer også bra på kunder med få kjøp, altså på kunder systemet har lite data om.

### **3.2.3 LibraryThing**

LibraryThing er en nettressurs som lar brukere katalogisere sine egne bøker. Siden oppstarten i 2005 har den utviklet seg til et stort sosialt nettverk. Med nesten 1,3 millioner brukere, over 60 millioner bøker, 73 millioner tagger og 1,3 millioner bokanmeldelser er det et eldorado av bokrelaterte data (Zeitgeist overview udatert). Spalding (Ritcher 2009) forteller at LibraryThing sine anbefalinger ikke er skapt for å selge noe, men for å hjelpe folk å oppdage nye bøker. Han presiserer at han i liten grad bruker en brukers vurdering av en bok til anbefalinger. Dette begrunner han ved å si at et menneske som regel kjøper de bøkene de liker. Derfor vil en brukers private bibliotek i hovedsak bestå av bøker brukeren har gitt positiv vurdering. LibraryThing som er en digital representasjon av en brukers bibliotek har et gjennomsnitt på 3,8 av 5 stjerner. Derfor har han satt fokus på eiermønster i stedet. Han presiserer at anbefalinger som er laget for at en bruker skal kjøpe noe som ligner på noe han eller hun har kjøpt før blir for kjedelige og det resulterer i en del støy. Som eksempel trekker han frem at en person som har kjøpt Harry Potter 1 vil bli anbefalt å kjøpe bok 2-7 også.

Når det gjelder selve algoritmen LibraryThing bruker er den ikke åpen. Derfor vil jeg gå igjennom anbefalingene fra en brukers side. Når en bruker logger inn blir han møtt med en side som samler opp brukerens siste handlinger på nettstedet deriblant nylig lagt til bøker og nylig mottatte anbefalinger. Velger man å klikke seg inn på anbefalinger får man opp en rangert liste med anbefalinger. På anbefalingssiden får brukeren anbefalinger både fra LibraryThing, det kaller de automastiske anbefalinger, og fra andre brukere. Anbefalingene fra LibraryThing blir delt inn i to grupper. En gruppe for anbefalinger til den aktive brukeren og en som de kaller read-alikes. Standardvisningen som dukker opp under anbefalinger er rangert under anbefalinger for den aktive brukeren. Under hver boktittel får brukeren muligheten til å klikke på "why" for å vite hvilken bok anbefalingen er basert på. Brukeren kan også velge å klikke på "no thanks", noe som gjør at anbefalingen forsvinner for alltid. Read-alikes er anbefalinger som er generert med



bakgrunn i bøkene som finnes i den aktive brukerens boksamling, uten at dette blir nærmere forklart.

Anbefalinger gitt av andre brukere deles også inn i to grupper, en for den aktive brukeren og en for den aktive brukerens bibliotek. anbefalinger for den aktive brukeren gir en rangert liste boktitler, med informasjon under titlene om hvilken bok anbefalingen er gitt på bakgrunn av. anbefalinger for den aktive brukerens bibliotek gir en oppdelt liste hvor boken anbefalingene blir gitt på bakgrunn av kommer først og selve titlene som blir anbefalt er rangert under. Ved denne fremgangsmåten kan brukeren velge hvilken bok han eller hun ønsker anbefalingen skal ligne.

En spennende ting på LibraryThing er at hvis brukeren klikker seg inn på en bestemt boktittel for å få mer informasjon om den boken kan han klikke på "Will you like it?". Dette resulterer i en visuell anbefaling hvor et punkt vises på en horisontal strek. Streken går fra rød på venstre siden til grønn på høyresiden. Under streken kommer anbefalingen, et eksempel er: LibraryThing thinks you **probably will like** The Lost Symbol (prediction confidence: **very high**). Det spennende med dette er at LibraryThing kommer med en algoritmisk gjetning for hvor mye brukeren vil like boken. Mulige resultater for anbefaling er; won't like, probably will like, will like og will love. Graderinger av gjetningen er; very low, low, medium, high og very high. På den får brukeren en bedre indikasjon om han skal følge anbefalingen eller ikke.

Som utenforstående virker det absolutt som om anbefalinger er gitt med bakgrunn i likhet av bøker, i strid med det Spalding sier. Spesielt siden man ved å spørre om å få en anbefaling forklart får opp bøker i egen samling anbefalingen er generert på. Det er mulig det er en forenkling gjort fra deres side for å kunne tilby en forklaring, det hadde derfor vært meget spennende og få et innblikk i algoritmer og tabeller de bruker for å generere anbefalinger.

### 3.2.4 Andre litteraturanbefalingssystemer

Det finnes mange nettsteder som tilbyr boker anbefalinger, og ressurser som i det hele er skapt for litteraturanbefalinger. What should I read next<sup>21</sup>, The Book Explorer<sup>22</sup> og Goodreads<sup>23</sup> er eksempler på dette. Felles for disse nettressursene er at en bruker danner seg en profil ved å registrere bøker. Bøkene kan vurderes og anbefalinger blir gitt på bakgrunn av brukerens bibliotek og vurdering av bøker. Et samfunn av brukere som liker å lese, diskuterer og anmelder

---

<sup>21</sup> <http://www.whatshouldireadnext.com/>

<sup>22</sup> <http://www.thebookexplorer.com/>

<sup>23</sup> <http://www.goodreads.com/>

bøker blir derfor kalt sosiale nettsteder og Bokelskere.no faller innunder denne typen nettressurser. Jeg vil gå dypere inn på denne typen nettressurser i kapitel 4. Metodologi. Et annet eksempel på en ressurs som tilbyr anbefalinger for brukerne er Facebook applikasjonen Visual Bookshelf hvor en bruker kan registrere sitt private bibliotek digitalt. Brukeren kan gi stjerner i skalaen 1-5 og legge inn anmeldelser av leste bøker. For hver bok en bruker legger inn dukker det opp nye anbefalinger basert på denne boken. De anbefalte bøkene kan legges rett inn i brukerens samling under fanene; Already read, Want to read eller Reading now. Dessverre finnes det ikke noen dokumentasjon på hvordan anbefalinger blir generert. Det eneste de sier om anbefalinger er at de er basert på bøker i brukers boksamling og de er unike for bruker.

## 4 Metodologi

Med denne oppgaven ønsker jeg å finne frem til en eller flere anbefalingsmetoder for litteratur som kan benyttes på et sosialt nettsted. Oppgaven er utformet som en mulighetsstudie med vekt på hvilke muligheter en utvikler har når et anbefalingssystem skal utarbeides og hvilke konsekvenser eventuelle valg får. Jeg jobber med en problemstilling som lyder:

*Hvilke anbefalingsmetoder kan vise seg mest hensiktsmessig å benytte i et nettbasert anbefalingssystem for litteratur, der brukeren står i sentrum?*

Hensiktsmessig<sup>24</sup> er i denne sammenhengen brukt for å skape anbefalinger som blant annet er nyttige, gode, fruktbare, riktig og velegnet for brukeren, i den konteksten anbefalingssystemet befinner seg. Det kommer klart frem i problemstillingen at det er brukeren som er den viktige parten og det er dermed ikke lagt vekt på implementariske og økonomiske valg i forbindelse med oppgaven. Når det gjelder implementering og økonomiske valg er dette er valg som hver utvikler må ta i forbindelse med konkrete ressurser. Dette mulighetsstudie er dermed begrenset til tekniske valg, men det er viktig å være klar over at denne typen valg også må vurderes i forbindelse med utarbeidelse av et anbefalingssystem.

Denne oppgaven har en teoretisk tilnærming og er sterkt forankret i tidligere litteratur på området. Siden denne litteraturen er meget omfattende har jeg gjort rede for forskjellige anbefalingsmetoder som finnes i skrivende øyeblikk. Jeg har vektlagt hva en utvikler kan gjøre med forskjellige datakilder og hvilke som vil være tilgjengelig for sosiale nettsteder. Målet for oppgaven er å drøfte de mulighetene som finnes når en skal utvikler et anbefalingssystem. Jeg ønsker å drøfte disse valgene og hvilke konsekvenser de vil få brukerne. Jeg ønsker gjennom denne drøftingen å legge frem de mest aktuelle anbefalingsmetoder for litteratur. På den måten kan utviklere av litteraturanbefalinger støtte seg til denne oppgaven ved utvelgelse av anbefalingsmetode og fremgangsmåte.

---

24

[http://www.ordnett.no.ezproxy.hio.no/ordbok.html?search=hensiktsmessig&search\\_type=&publications=2&publications=3](http://www.ordnett.no.ezproxy.hio.no/ordbok.html?search=hensiktsmessig&search_type=&publications=2&publications=3)

Jeg har utarbeidet fire underspørsmål som skal hjelpe meg å besvare problemstillingen min:

- Hvilke forskjellige anbefalingsmetoder finnes?
- Hvilken anbefalingsmetode kan en utvikler benytte seg av i forhold til tilgjengelig data fra et sosialt nettsted?
- Hvordan skiller det å anbefale en bok seg fra anbefaling av andre gjenstander?
- Hvilke egenskaper er viktig å legge vekt på når en skal utvikle anbefalingsmetoder for litteratur og hvilke konsekvenser får dette?

I analysedelen går jeg gjennom de tre hovedfaktorene som inngår i min drøfting og har dermed valgt å gå dele opp analysedelene i tre trinn. Trinn en er viet anbefalingssystem i lys av anbefalingsdomenet, og jeg presenterer de domenerelaterte valgene en utvikler står ovenfor. I trinn to settes de forskjellige egenskapene et anbefalingssystem innehar opp mot tilgjengelig data fra sosiale nettsteder for på den måten å diskutere hvilke egenskaper som er viktig å vektlegge. I trinn tre drøfter jeg de valgte metodene for litteraturanbefalinger nøye, hvilke likhetskriterier en utvikler kan benytte seg av og hvilke konsekvenser dette får. Dette trinnet er direkte knyttet til data hentet fra valgte case Bokelskere.no.

## 4.1 Et mulighetsstudie

Jeg ønsker med oppgaven å kartlegge anbefalingsområdet for å finne ut hvilke anbefalingsmetoder som kan fungere best for å gi litteraturanbefalinger. Et mulighetsstudium kalles også prosjektstudie og er studie av gjennomførbarheten til en ide eller et prosjekt. Dermed har jeg gjennomført et mulighetsstudie, med fokus på teknisk gjennomførbarhet. Jeg ønsker å kartlegge hvilke valg som må tas når et anbefalingssystem skal utvikles for et sosialt nettsted. Samtidig ønsker jeg å drøfte de mulighetene som finnes for deretter å presentere de valgene jeg mener er best for sosiale nettsteder. ”Technical feasibility is a measure of how practical solutions are and whether the technology is already available within the organization” (Whitten, Bentley & Ditterman 2004 s. 425). Som Whitten et al. (2004) sier vil jeg undersøke om gjennomføring av anbefalinger er teknisk mulig med de datakildene som finnes på sosiale nettsteder, eller om det vil være behov for å videreutvikler hovedressursen. Dette gjøres på et teoretisk plan og mine drøftinger vil resultere i en konsekvensanalyse av anbefalingsmetoder for Bokelskere.no.

For å få en god oversikt over anbefalingsmetoder har det vært viktig å gå grundig igjennom tidligere forskning på området. Kapittel 2 er i tråd med dette et omfattende kapittel, hvor jeg har lagt frem de anbefalingsmetodene som finnes. Jeg har lagt spesielt vekt på forskning på litteraturanbefaling, og tatt for meg flere ressurser som tilbyr litteraturanbefalinger. Denne kunnskapen vil jeg benytte meg av i analysen.

Materialet til dette er funnet gjennom en omstendelig litteratursøkeprosess i aktuelle baser for bibliotek- og informasjonsvitenskap. Det arbeidet med anbefalingssystemer innenfor mange fagområder og jeg har gått inn for å få et grundig overblikk over situasjonen, men ikke for en fullstendig oversikt da omfanget er meget stort. Jeg har tatt for meg Library and Information Science Abstracts<sup>25</sup>, The Collection of Computer Science Bibliographies<sup>26</sup>, Communications of the ACM<sup>27</sup> og flere fagrelaterte tidsskrifter. Innenfor landegrensene har jeg sett etter prosjekter og forskning om litteraturanbefalinger i biblioteksektoren. Biblioteklaboratoriet<sup>28</sup> som er et samarbeid mellom ABM-Utvikling og Norsk Digitalt Bibliotek, og avdeling for journalistikk, bibliotek- og informasjonsfag ved Høgskolen i Oslo. Jeg har lest i norske tidsskrifter som *Bok og bibliotek*<sup>29</sup> og *Bibliotekaren*<sup>30</sup> for å få et overblikk over situasjonen her til lands. Jeg har aktivt brukt litteraturlister til relevante artikler og informasjon fra internasjonale konferanser for å forsøke å samle inn mest mulig aktuell informasjon.

Med bakgrunn i det omfattende litteraturutvalget har jeg også vurdert hvilken tilnærming til stoffet jeg skulle bruke. Det er ikke foretatt noen konkrete eksperimenter i forbindelse med denne oppgaven, da den kan sees på som en omfattende konsekvensanalyse. I den avsluttende analysen bruker jeg drøfting for å forklare hva som skjer ved de forskjellige metodene jeg har valgt på bakgrunn av teorien i kapittel 2 og 3.

---

<sup>25</sup> [http://csaweb107v.csa.com/ids70/quick\\_search.php?SID=02gohj7lvhk860qijj72km8jp7](http://csaweb107v.csa.com/ids70/quick_search.php?SID=02gohj7lvhk860qijj72km8jp7)

<sup>26</sup> <http://iinwww.ira.uka.de/bibliography/index.html>

<sup>27</sup> <http://cacm.acm.org/>

<sup>28</sup> <http://www.biblab.no/blog/>

<sup>29</sup> *Bok og bibliotek* utgis av ABM-media og inneholder faglige artikler om teknologi, litteratur- og kulturformidling og bibliotekforskning i Norge og Norden.

<sup>30</sup> *Bibliotekaren* utgis av Bibliotekarforbundet på månedsbasis.

## 4.2 Case

Målet for oppgaven, å finne frem til en god metode for litteraturanbefalinger for brukere av en nettbasert ressurs må samsvare med valg av et case. Krav til caset er at det må være interessant i en kontekst for litteraturanbefalinger og det må ha tilgjengelige datakilder jeg kan jobbe med. Ringdal (2009) sier at valg av enkeltstående case kan gjøres på bakgrunn av teoretisk eller strategisk utvelgelse. Valg av case sto for meg mellom en nettbutikk, et bibliotek eller et sosialt nettsted for bokinteresserte. I en nettbutikk er utviklerne interessert i å selge varer og dermed vil ikke brukerne være i sentrum. Slik som bibliotek er nå til dags vet vi mye om brukernes lånevaner, men har på grunn av personvernmessige årsaker ikke lov til å lagre og bruke dette. Selv om historikk i et bibliofilbibliotek kan gi en oversikt over lånte bøker er det lite å hente i form av respons fra brukerne. Dermed har jeg foretatt et strategisk valg og tatt for meg litteraturanbefalinger i forbindelse med sosiale nettsider.

Siden det var et ønske fra Bokelskere.no som hjalp til å trigge min nysgjerrighet for anbefalingssystemer bestemte jeg meg for å undersøke dette nettstedet nærmere. Jeg har undersøkt om det finnes tilsvarende norske nettsteder, men har ikke funnet noen. Det finnes uttallige blogger og mange bibliotek har litterære forum, men Bokelskere.no har utnyttet de sidene som sosiale medier har å tilby og kan kalles et sosialt nettsted. Derfor bestemte jeg meg tidlig i prosessen å bruke Bokelskere.no som case og har derfor hentet datakilder derfra. Jeg har brukt nettstedet for å få en fullstendig oversikt over oppbygningen av det og hvilke muligheter en bruker har. Jeg har gått igjennom diskusjonsforumet der inne for å samle inn data om nettstedet og om hva brukerne av nettstedet mener i forhold til litteratur og lesing.

Målet for oppgaven er å utarbeide forslag for anbefalingsmetoder som passer for sosiale nettsteder dermed blir ikke dette et tradisjonelt casestudium. Jeg mener dette studiet kan sees på som et mulighetsstudie hvor målet er å se på hvilke konsekvenser de forskjellige anbefalingsmetodene og valg i utviklingsprosessen vil få for et anbefalingssystem. Jeg bruker Bokelskere.no til å finne sammenhengen mellom bestemte aspekter ved en bok, om bruker og litterære anbefalinger. Jeg bruker data hentet derfra for å vise reelle situasjoner en utvikler vil møte og hvordan de kan løses i den sammenhengen. På denne måten kan jeg tolke dette i en anbefalingssammenheng, og trekke ut de viktigste faktorene og komme med konkrete meninger om gjennomførbarheten til de forskjellige metodene i virkeligheten.

## 4.2.1 Det sosiale nettstedet Bokelskere.no

Bokelskere.no er et sosialt nettsted av og for bokentusiaster. Arbeidet med Bokelskere.no ble startet opp i 2008 av privatpersonene André Nesse og Tore Renberg. Nettstedet åpnet for brukere høsten 2009 og har siden den gang fått over 12.600 brukere, og det er registrert i overkant av 122.000 utgavekopier av bøker, som vist på siden *Bokelskere:Tall* (Bokelskere.no). En bok kan ha flere utgavekopier alt etter hvilket språk den er utgitt på, når den er utgitt og hva slags cover den har. Bokelskere.no henter norsk bokdata fra Den norske Bokdatabasen<sup>31</sup> og internasjonale bokdata fra Amazon.com (Bokelskere.no).

**Bokelskere.no**  
Vi elsker bøker. Akkurat som deg. Velkommen hjem.

logg inn bli en bokelsker! hjelp

finn en bok

**hjem** til forsiden   **finn bøker** søk eller bla   **diskusjoner** se hva vi snakker om   **bokelskere** finn og følg   **sitater** godt sagt   **lister** vi har dem

### Del dine leseopplevelser

*Fordi hodet ditt koker. Fordi du har lest en roman, fordi du har lest en diktsamling, fordi du har lest en kokebok om asiatiske retter. Og nå har du vann i munnen. Du vil dele det med noen. Du vil se om det finnes noen der ute som har lest det du har lest, som har kjent det du har kjent, og du vil ha nye anbefalinger, for du vil ha nye leseopplevelser.*

### Hos bokelskere.no kan du

- ✓ lage din egen boksamling
- ✓ sortere bøkene dine i bokhyller
- ✓ finne andre med samme lesesmak som deg
- ✓ se hva andre leser akkurat nå
- ✓ delta i diskusjoner om bøker

eller ta en titt på [andre bokelskere](#)

**Bla i bøker**   **Nyeste lister**   **Sist sagt**   **Nyeste sitater**

Bilde 3: Forsiden på Bokelskere.no.

<sup>31</sup> <http://www.bokbasen.no/id/11023538>

En potensiell bruker melder seg inn, finner på et brukernavn og lager en profil. I profilen legger brukeren inn demografiske data som kjønn, fødselsår og bosted. Denne informasjonen brukes til statistisk data og det er i dag 26 prosent menn på nettsiden, så kvinnene er sterkt representert. En bruker velger selv om han vil være anonym. Deretter er brukeren klar til å legge til bøker i boksamlingen sin. Bøkene kan søkes opp på tittel, forfatter eller emneord. Resultatlisten viser passende treff, og brukeren kan deretter velge riktig bok. Etter å ha valgt riktig utgavekopi av den aktuelle boken kan en bruker klikke seg inn på boken for å se på informasjonen om den eller han kan velge "legg til bok". Når brukeren har valgt å legge til en bok, blir boken markert som hans og tilhører etter dette brukerens boksamling. Skjermbildet som boken da finnes på er den aktive brukerens "hovedside" for den bestemte boken (jmf. Vedlegg 3). En bruker kan slette en bok fra boksamlingen sin hvis han ønsker det.

På hovedsiden for en bestemt bok kan brukeren strukturere boksamlingen på flere måter. En bok har en tilstand som representerer bokens "syklus" hos brukeren. Boken kan ha seks tilstander, *skal lese*, *leser*, *har lest*, *avbrøt*, *ønsker deg*, *oppslagsverk*. De seks tilstandene sier noe om brukerens forhold til boken ved det aktuelle tidspunktet. Tilstandene *leser*, *har lest* og *avbrøt* indikerer at brukeren har eller har hatt det fysiske dokumentet. Tilstanden *ønsker deg* indikerer at dette er en bok brukeren skal skaffe seg i fremtiden. Når det gjelder tilstanden *skal lese* kan den i egenskap av tilstanden *ønsker deg* indikere at brukeren har boken, men ikke har lest den enda. Tilstanden *oppslagsverk* er mer en forklaring på hvordan en bok blir brukt. Bøker som får denne tilstanden tilknyttet seg er antageligvis ikke bøker som leses fra perm til perm, det kan for eksempel være en fagbok. Inndeling i tilstander gir brukeren en mulighet til å bruke nettstedet som en liste over bøker de har hørt om og ønsker å lese i fremtiden. I tillegg til en oversikt over hvilke bøker han har lest, og når han leste dem.

En brukers boksamling struktureres slik (Heffalumpen5 er brukt som eksempel og antall bøker for denne brukeren står i parentes):

**Alle** (490). Alle utgavekopiene en bruker har i boksamlingen sin. Uavhengig om de er lest eller ikke. Hvis en bruker har både den engelske og den norske versjonen av en bok telles det som to bøker, med mindre utgavekopiene er lenket sammen.

**Favoritter** (16). De bøkene som er i boksamlingen og brukeren har satt de som favorittbok. Uavhengig av om de er lest eller ikke.



**Leser** (3). De bøkene en bruker leser i øyeblikket. Utgjør 3 av 490 bøker i dette eksempelet. Hvis en bok ligger aktiv under lesing er den ikke en del av Har lest, selv hvis boken har vært lest før.

**Skal lese** (23). Her ligger bøker brukeren har tenkt å lese i fremtiden.

**Avbrøt** (1). Bøker brukeren har valgt å avslutte uten å lese ferdig. 1 av 490 i dette eksempelet.

**Har lest** (431). Verk brukeren har lest minus de andre kategoriene.

**Ønskeliste** (17). Verk brukeren regner med å lese og derfor har lagt til.

**Oppslagsverk** (3). Denne betegnelsen brukes ofte på fagbøker og oppslagsverk som en bruker ikke vil lese fra perm til perm.

En bok kan plasseres i en bokhylle som brukeren selv navngir. Brukeren kan selv velge antall bokhyller han ønsker å operere med, og øke dette når han finner behov for det. Det finnes over 37.000 forskjellige bokhyller konstruert av de forskjellige brukerne. Bokhyller har fått navn som kan representere blant annet sjangere, fysiske steder, brukerens vurdering av bøker, tidsepoker og forfattere. Konkrete eksempler på bokhyllenavn er Favoritter, Fantasy, Biblioteks bøker, Hylla mi, Skjønnlitteratur, Tidsfordriv, Lånt av venn, Beste av de beste, Bøker som beveger og Chick-litt. Ved å se på bokhyllenavn kommer det klart frem at hver enkelt bruker kan navngi bokhyllene fritt for å fordele bøkene i boksamlingen sin.

Brukeren kan tagge bøker i sin egen boksamling, de omtales på nettstedet som emneord. Emneordene er i likhet med bokhyllenavn frie og brukerne bestemmer selv hva de skal representere som emnet i boken, følelse det ga å lese den, sjanger eller tidsepoke. Konkrete eksempler på emneord for Knut Nærums ”Monster” er; Fremtidsvisjon, Hotellkonge, Humor, Isbjørn, Miljøverninteresser, Norsk litteratur, Polaris, Skjønnlitteratur og Turistnæring. Emneordene skilles ved kommategn og det finnes ingen begrensinger på antall emneord en bruker kan sette. Det er registrert over 13.000 emneord på de bøkene som er registrert på Bokelskere.no. En bruker velger også selv om han vil bruke entall eller flertallsform på emneordet. Dette gjør at det finnes mange overlappende emneord.

En bruker kan registrere lesedato for når boken er lest ferdig, og hvis brukeren leser samme bok flere ganger så representeres dette med forskjellige datoer og nettstedet teller antall ganger boken blir lest. Brukeren kan velge å se på boksamlingen sin på flere måter, en måte er å velge et år for deretter å se hvilke bøker han leste det året. En annen måte er at brukeren kan velge å klikke på en tilstand, for eksempel *skal lese* for å se hvilke bøker han skal lese i fremtiden.

Brukeren kan starte en diskusjon om en bok, som fungerer som brukerens skriftlige vurdering av den bestemte boken. Diskusjonen kan være så kort eller lang som brukeren selv ønsker. Diskusjonen starter en diskusjonstråd om boken på nettstedets diskusjonsside, som er åpen for alle registrerte medlemmer. Terningkast er en annen måte for brukeren å vurdere en bok. Terningskalaen går fra 1-6, hvor 1 betyr at brukeren ikke likte en bok, og 6 betyr at brukeren likte en bok veldig godt. I tillegg til terningkast kan en bruker sette en bok som en favoritt, uavhengig av terningkast.

Nederst på en boks hovedside finnes det en oppsummering av den bestemte boken, som kalles boktall. Det er en statistisk oversikt over, gjennomsnittlig alder på eiere, prosentandel av eierne som er hannkjønn, prosentandel av lesere som avbrøt boken og gjennomsnittlig terningkast. Samme type statistikkoversikt finnes om medlemmene av Bokelskere.no, på nettsiden *Din side* og fanen *Hva skjer?* (jmf. Vedlegg 4). Under overskriften *Tall* kan man blant annet finne bøker i boksamlingen, notater skrevet, terningkast trillet, gjennomsnittlig terningkast og totalt antall sider lest. Denne informasjonen er kun tilgjengelig for hver enkelt bruker om seg selv. De viser brukeren en del morsom informasjon om sin egen lesing og karaktersetting, i tillegg til å vise noe av det en utvikler har mulighet til å benytte seg av.

#### **4.2.2 Likhet mellom brukere på Bokelskere.no**

På nettstedet Bokelskere.no er det lagt til rette for at en bruker kan velge å følge en eller flere andre brukere. Grunner en bruker kan ha for å velge å følge en annen bruker kan være forskjellige. Brukeren som følges kan ha lest interessante bøker. Brukeren som følges kan ha lest mange av de samme bøkene som den aktive brukeren. Brukerne har likhet i lesersmak. Brukerne kjenner hverandre. Når en bruker velger å følge andre brukere får han lettere informasjon om de brukerne han følger. På siden *bokelskere: finn og følg* kan den aktive brukeren finne frem til brukere han ønsker å følge (jmf. Vedlegg 5).

Det spennende med denne utvelgelsen av brukere å følge er i anbefalingsøyemed hvordan Bokelskere.no gir den aktive brukeren mulighet til å sammenligne seg med andre brukere. Under brukernavnet til en bestemt bruker finner man informasjon om hvor mange som følger den bestemte brukeren, og hvor mange bøker den aktuelle brukeren har. Det finnes også mulighet til å finne ut hvor mange felles bøker en har med den bestemte brukeren og å finne likhet i lesesmak (jmf. Vedlegg 5).

Likhet i lesesmak er interessant og klikker man på denne lenken får man et desimaltall til svar. Likhet i lesesmak mellom Heffalumpen5 og Tore Reneberg. Likhet: -0,71, basert på felles vurdering av 8 verk. For den vanlige bruker vil jeg tro at dette kan virke noe uforklarlig da det ikke er gitt noen forklaring på hvordan tall-likheten skal tolkes. Nesse forklarer dette i en diskusjonstråd på nettsiden.

Har lagt ut en ny funksjon som du finner på lista over bokelskere. Der er det en link til "Finn likheter i lesesmak". Den sammenligner dine terningkast med en annen bokelsker, og returnerer et nummer mellom -1 og 1. -1 beskriver totalt forskjellig smak, og 1 totalt lik. Tallet regnes ut ved hjelp av Pearsons korreleringsalgoritme (Nesse 2010).

Pearsons korreleringsalgoritme brukes ofte i forbindelse med anbefalinger. Den er avhengig av at brukerne har vurdert de samme gjenstandene og regner ut likhet mellom brukere basert på dette. Det kan også gjøres en normalisering av algoritmen for å ta høyde for den individuelle brukerens personlige vurderingspreferanser. En brukers gjennomsnittlige vurdering blir da brukt og på den måten er det lettere å sammenligne brukere som er "snille" og "strenge" med sine terningkast. Det kommer ikke klart frem om Nesse bruker normalisering i sin utregning.

Når en bruker har valgt å følge andre brukere har han det fortrinnet at mer informasjon om den andre brukeren er lettere tilgjengelig. På siden *din side* under fanen *hva skjer?* Finnes overskriften *De du følger leser nå*. Dermed får man en oversikt over hvilke bøker de brukerne den bestemte brukeren følger leser for øyeblikket (jmf. Vedlegg 6). Dette kan sees på som en måte å gi lesertips på da det trolig er interessant for den aktive leseren hva brukerne han følger leser.

### **4.2.3 Anbefalinger på Bokelskere.no**

Det er mulig for den aktive brukeren å få bokanbefalinger basert på vurderinger, altså terningkast, av bøker i egen boksamling. På hovedsiden til en bok finnes det en lenke som heter; *Likte du denne? Vi har bokanbefalinger!* (jmf. Vedlegg 7). Denne lenken vil ikke dukke opp på alle bøker, det er satt en minimumsgrense i forhold til antall vurderinger av den aktuelle boken. Følges lenken får den aktive brukeren bokanbefalinger basert på at brukeren liker den bestemte boken. Denne anbefalingsmåten ble endret i april 2011. Jeg vil først gå igjennom den gamle og forklare utfordringer den sto ovenfor. Deretter vil jeg kort forklare den anbefalingsmetoden som er aktiv. Den gamle metoden for anbefalinger vist som en liste, med forklaringer på hvorfor de

forskjellige bøkene ble anbefalt. Forklaringen på anbefalingene for den gamle måte var oppgitt denne måten: ”De 4 bokelskere som likte **Shantaram** og har vurdert **Kristin Lavransdatter** liker **Kristin Lavransdatter 20,50%** mer enn andre bokelskere. (6,25)” (Bokelskere.no udatert b). Bokelskere er i denne sammenhengen brukere av Bokelskere.no. Nesse forklarer hvordan disse anbefalingene blir generert i e-post korrespondanse.

- 1) henter alle terningkast for alle utgavene av et verk, for eksempel "twilight". (Hvis noen har satt en utgave av verket som favorittbok, regner jeg dette som terningkast 7).
- 2) lager en liste over bokelskere som var \*positive\* til utgivelser av "twilight". (Jeg regner terningkast 5, 6 eller 7 om positivt, 4 som nøytralt og tre og lavere som negativt.)
- 3) undersøker hvilke \*andre\* bøker de som liker twilight har vurdert. Legger verk som har tre eller flere vurderinger fra disse bokelskerne i en ny liste.
- 4) deretter går jeg gjennom disse verkene og sammenligner gjennomsnittsvurderingen gjort av dem som \*liker\* twilight med de \*øvrigte\* bokelskerne. Når det er et \*avvik\* i vurderingene mellom dem som likte twilight og alle andre bokelskere, har vi en anbefaling. (Anbefalingen kan være i positiv eller negativ retning.) (André Nesse, e-post, 23. november 2010).

Ut i fra denne informasjonen kan jeg lese at anbefalingene blir generert ved å benytte en kollaborativ filtreringsmetode. Likheter mellom brukere blir målt i form av vurdering av felles verk. Anbefalingsmetoden er begrenset til å gjelde en og en bok. Det vil si at den ikke tar hensyn til andre bøker i den aktive brukerens boksamling. Dermed vil ikke informasjon om hvordan den aktive brukeren har vurdert andre verk tas med når anbefalingene blir generert og dette kan gi unøyaktige anbefalinger for brukerne. Positive anbefalinger vises på grønn bakgrunn. Ruller man nedover Heffalumpen5 sin liste med anbefalinger basert på Shantaram vil det etter hvert komme frem negative anbefalinger, som vises på rød bakgrunn. ”De 6 bokelskere som likte **Shantaram** og har vurdert **Da Vinci-koden** liker **Da Vinci-koden -17,38%** mer enn andre bokelskere. (3,50)” (Bokelskere.no).

Dette er hva Nesse kaller et negativt avvik og det tydeliggjør svakheten ved Nesses metode, i tillegg til en språklig utfordring for utvikleren. Heffalumpen5 har selv vurdert Da Vinci-koden til terningkast 4, noe som tilsier en positiv verdi i Nesses terningkastvurdering. Dette betyr at Da Vinci-koden burde vært et eksempel på positiv anbefaling hvis den skulle være korrekt for Heffalumpen5. Siden denne måten å anbefale litteratur ikke tar med brukerens tidligere vurderinger mangler personaliseringsaspektet. Eksempelet påpeker også en teknisk svakhet at brukere får anbefalt verk de allerede har vurdert som et resultat av utgavekopi

utfordringen Bokelskere.no står ovenfor. De har ikke klart å koble sammen alle utgaver av et verk da brukere kan søke opp utgavekopier som allerede er representert på nettsiden, fra Amazon.com eller legge de til manuelt. Det dannes ikke en automatisk kobling mellom utgavekopier.

Den nye anbefalingsmetoden presenterer anbefalingene i en rangert liste uten den forklarende teksten som den gamle metoden hadde. Listen består av 100 verk og vises med ti og ti verk per side (jmf. Vedlegg 7). Den nye anbefalingsmetoden forklares slik:

Tidligere har vi basert bokanbefalingene våre på terningkast, men dette har vi nå endret. Brukerne på bokelskere.no har i fellesskap lagt mer enn 1.125.971 bøker i boksamlinger, og vi har nå tatt i bruk en algoritme som tar hensyn til samtlige av disse. For å finne ut hvilke bøker vi bør anbefale til dem som f.eks liker Alkymisten av Paulo Coelho, gjør vi slik:

- 1) Først finner vi ut hvilke andre bøker som finnes i bokhyllene til dem som har denne boka blant bøkene sine.
- 2) Deretter går vi gjennom alle disse bøkene og regner ut populariteten deres i prosent blant eierne av Alkymisten. (Vi kaller dette PE)
- 3) Deretter sjekker vi populariteten på disse bøkene blant dem som ikke har Alkymisten i sin boksamling. (Vi kaller dette PI)
- 4) Så deler vi PE på PI og får et eierfrekvensavvik som vi kaller A.
- 5) Så sorterer vi verkene etter A, og den boka som er på toppen av lista er den med høyest avvik, og dermed er dette den med størst anbefalingsgrad (Nesse 2011, april b).

Jeg mener de to fremgangsmåtene ligner på hverandre siden det fortsatt tas utgangspunkt i en bestemt bok, ikke den aktive brukeren. De regner ut en form for avvik som minner om den første metoden. Forskjellen ligger i at de nå bruker boksamlinger som sammenligningsgrunnlag i stedet for stjerner.

På diskusjonsforumet hvor Nesse forklarer metoden er det spørsmål om hvorfor de har gått vekk fra å bruke terningkast. Nesse mener selv den gamle anbefalingsmetoden ikke ga gode nok anbefalinger, men er usikker på hvorfor. Han trekker frem at det kunne være algoritmen de brukte eller datagrunnlaget (Nesse 2011, april c). I forklaringen ovenfor sier Nesse at de trekker ut alle bøker i boksamlingene som inneholder Alkymisten, dette gir et stort omfang bøker. Deretter regner de ut populariteten til disse bøkene, dette gjøres ved at de sjekker hvor mange som har de forskjellige bøkene. På den måten får de en liste over hvor ofte bøkene er representert. Deretter går de igjennom brukere som ikke har Alkymisten og sjekker hvor populære bøkene i boksamlingene blant disse brukerne, altså hvor mange ganger bøkene er representert der. Deretter

deler de antallet bøker hos de brukerne som ikke eier Alkymisten i forhold til antallet bøker hos de eierne som har Alkymisten og sitter igjen med det de kaller eierfrekvensavvik som de rangerer anbefalingslisten etter.

Det finnes flere svakheter ved denne fremgangsmåten og den mangler fortsatt personaliseringsaspektet jeg etterlyste i den første fremgangsmåten. Anbefalingene tar utgangspunktet i en bok og den eneste som er personlig er at den aktive brukeren liker den boken som brukes. Det tas ikke hensyn til bøker i den aktive brukerens boksamling derfor risikerer brukeren å få anbefalt bøker han allerede har i samlingen. Dette vises spesielt godt innenfor sjangre som er sterkt representert hos den aktive brukeren. Heffalumpen5 ber om anbefalinger basert på boken Twilight. Dette er en Fantasy bok og Fantasy bøker er sterkt representert i Heffalumpen5 sin boksamling. I den rangerte listen, som viser ti og ti treff er samtlige på første side alt representert i Heffalumpen5 sin boksamling. Men problemet stopper ikke der fordi Bokelskere.no fortsatt har et problem med kobling av utgavekopier. Av de ti titlene som vises på første side er to representert to ganger. Dermed står det åtte forskjellige titler, fordelt på to forfattere igjen på listen. Den andre siden viser seks titler brukere allerede har lest, men ingen av titlene ble representert flere ganger. Dette gir oss totalt fire potensielle anbefalinger der det burde vært 20. På en side viser det jo at de er inne på en spennende måte å lage anbefalinger, men det finnes også en utfordring forbundet med utregning av popularitet.

Nesse sier at de tar for seg hele boksamlingen, altså de benytter seg av alle bøkene en bruker har lagt inn. Dette er uavhengig av terningkast brukeren har satt eller hvilken tilstand en bok har hos brukeren. En bruker poengterer nettopp dette i en diskusjon med Nesse.

Men da vil man jo risikere at en bok som har fått generelt dårlige karakterer blir satt på toppen av anbefalingslisten. Hvis 300 av oss har en bok og vi har gitt den terningkast 2 så er det vel ikke så kjekt for bokelskere.no å anbefale denne boken? Så blir det tatt hensyn til andre faktorer i tillegg til hvor mange som har boken i samlingen sin? Jeg har også lagt inn bøker jeg skal lese. Blir disse tatt med i tellingen? (Ankie 2011).

Nesse er enig med brukeren om at dette kan skje i teorien, men at det ikke vil skje i praksis. Dette begrunner han med at gjennomsnittlig terningkast på nettstedet er over 5, og at dette tyder på at brukere kun legger inn det de liker på nettstedet. Bøker som skal leses blir også tatt med (Nesse 2011, april b).

#### 4.2.4 Data hentet fra Bokelskere.no

Under gjennomgangen av nettstedet Bokelskere.no ble det klart at det er mye informasjon tilgjengelig. På nettstedet Bokelskere.no har utvikler tilgang til eksplisitt data som brukeren selv har lagt inn, i tillegg til implisitt data som kan samles inn gjennom handlinger brukeren foretar seg på nettstedet. Informasjonen brukt i oppgaven er hentet fra data jeg har mottatt i XML format fra Bokelskere.no sin relasjonsdatabase (jmf. Vedlegg 8). Denne informasjonen har jeg supplert med data hentet ut gjennom Bokelskere.no sin API<sup>32</sup> som gir meg en enkel og oversiktelig måte å hente ut bestemt data fra nettstedet. I tillegg har jeg brukt nettsidene på en omfattende måte slik at jeg har en bred forståelse for nettstedet informasjonsarkitektur og brukermulighetene det tilbyr. For enkelhetens skyld snakker jeg om data hentet derfra med dagligdagse navn i stedet å bruke markeringsspråket XML. Dette gjør jeg for at det skal bli enkelt å skjønne bruken av attributtnavnene i den flytende teksten og fordi jeg i denne oppgaven ikke skal programmere.

Jeg har valgt å dele inn dataene fra Bokelskere.no i de tre kategoriene til Ricci et al. (2011) nevnt i kapittel 2. Dermed fordeler jeg data knyttet til brukere, gjenstander og transaksjoner på følgende måte.

##### **Data knyttet til en bruker**

Alder, Kjønn, Sted, Navn, Stjerner gitt, Diskusjoner, Antall bøker, Favorittbøker, Utgaver av verk, Emneord, Bokhyller, Gjennomsnittskarakter gitt, Favorittsitat, Følger, Følges av, Boktilstand og Dato lest.

##### **Data knyttet til et verk**

Brukere, Emneord, Utgave, Tilstander, Bokhyller, Tittel, Undertittel, Forfatter, Forlag, Sideantall, Sjanger, Lesenivå, Serie, Omtale, Utgivelsesdato, Diskusjoner, Sitater, Satt som favoritt, Omtalt person, Omtalt sted, Omtalt tidsepoke, Terningkast (Hvor mange), Terningkast (Gjennomsnittlig), Boksamlinger.

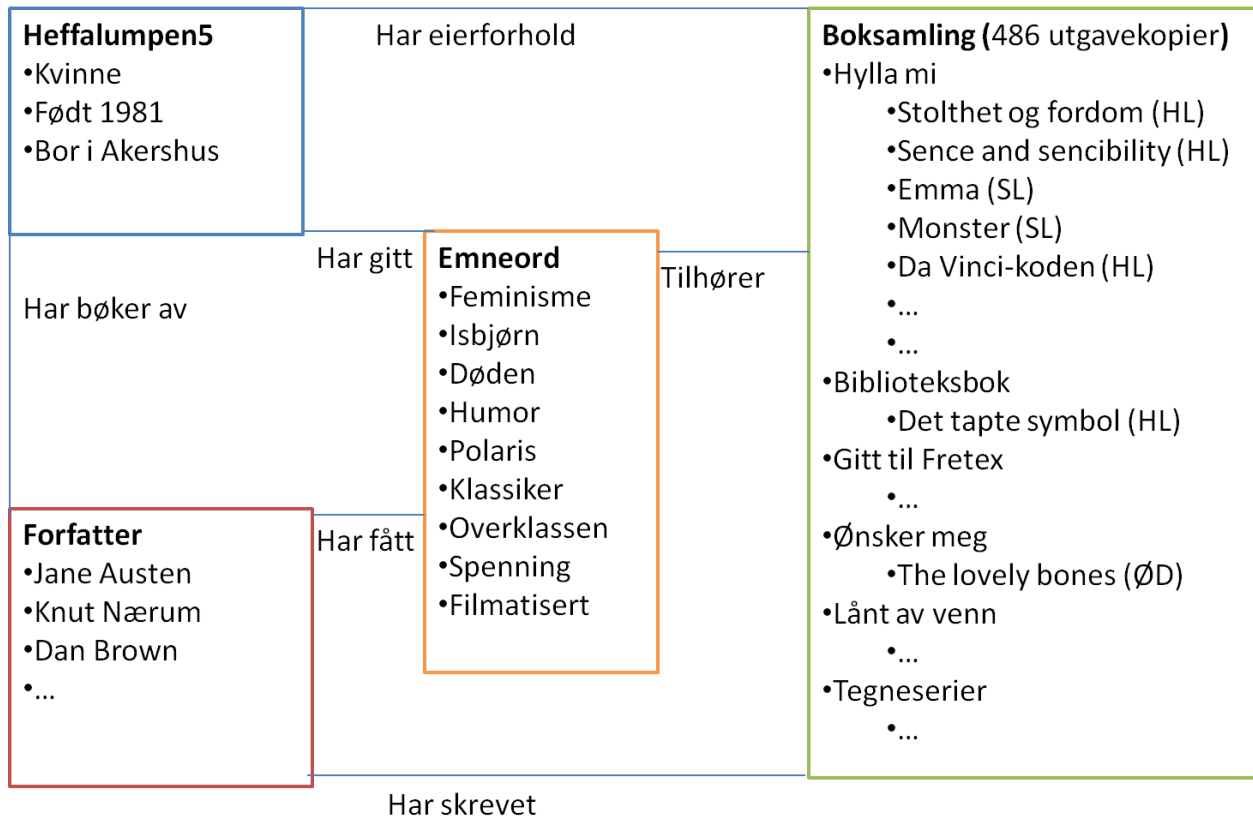
##### **Data som skapes av transaksjoner**

Denne type data vil inkludere bruker-gjenstand, gjenstand-gjenstand og bruker-bruker forhold og kan med fordel uttrykkes ved bruk av eksempler og skjematisk fremstilling. Det eksisterer mange transaksjonsforhold i ressursen Bokelskere.no, derfor vil jeg trekke frem de som kan være spesielt gunstige i anbefalingsøyemed. Bruker-gjenstand forhold er; har eierforhold, har gitt, har

---

<sup>32</sup> <http://bokelskere.no/api/1.0/>

bøker av, har bokhyller. Gjenstand-gjenstand forhold er; samme forfatter, samme emneord, finnes i samme boksamling. Bruker-bruker forhold er; bruker følger og bruker følges.



Bilde 4: Oversikt over transaksjoner på Bokelskere.no.

Brukeren Heffalumpen5 har et eierforhold til Stolthet og fordom siden hun *har lest* (HL) den. Hun har bøker av forfatterne Jane Austen, Knut Nærum, Dan Brown og flere. Hun har gitt bøker forskjellige emneord. Hun har forskjellige bokhyller med navn som Hylla mi, Biblioteksbok, Gitt til Fretex og mer. Stolthet og fordom er skrevet av Forfatter Jane Austen og Heffalumpen5 har vurdert verket til Terningkast 5. Emneord brukt er Filmatisert, Kjærlighet, Klassiker og Overklassen, og den norske utgaven av verket er plassert i Heffalumpen5 sin bokhylle som heter Hylla mi. Heffalumpen5 *skal lese* (SL) Monster skrevet av forfatter Knut Nærum, i Heffalumpen5 sin boksamling finnes det fire verk av Knut Nærum. Heffalumpen5 *har lest* (HL) en av disse: Madonnagåten som har fått terningkast 4. Heffalumpen5 *skal lese* (SL) totalt tre verk til av Knut Nærum, de er satt til bokhylla Hylla mi og som betyr at hun eier de tre verkene. Bøker som er lagt til under kategorien *skal lese* er et godt tegn på hva brukeren ønsker å lese i



fremtiden. Det kan derfor være lurt å vekte disse bøkene slik at de blir regnet som like relevante som de bøkene brukeren leste sist.

Transaksjoner mellom gjenstander kan eksemplifiseres ved å trekke frem verk av samme Forfatter. Flere verk som har samme Emneord og det kan også tenkes at det er nyttig å se på andre samlinger som inneholder det bestemte verket. Transaksjoner mellom brukere kan eksemplifiseres med følge funksjonen; Heffalumpen5 følges av 31 bokelskere, hun følger 7 bokelskere. Det kan forekomme diskusjoner mellom brukere, om bøker eller temaer som interesserer de. En bruker kan gi og få stjerner på innlegg eller sitarer de legger ut.

## 5 Analyse med drøfting av hovedfaktorer

Når et anbefalingssystem skal utvikles må utvikleren ta flere viktige valg. Dette er ikke nødvendigvis en lineær prosess, og den består ofte av prøving og feiling. Jeg vil i dette kapittelet gi en fremstilling av denne prosessen og de konkrete valgene som må tas. Jeg prøver å holde et nivå på fremstillingen slik at denne informasjonen enkelt kan brukes på andre tilsvarende sosiale nettsteder som ønsker å tilby anbefalinger, i tråd med problemstillingen. I selve modelleringen må jeg være så spesifikk som mulig så den delen er utarbeidet for bruk med data fra Bokelskere.no. Jeg konkretiserer med andre ord modellene med eksemplene fra Bokelskere.no for å simulere hvordan de kan virke på et sosialt nettsted som dette.

Jeg har valgt å dele inn analysen i tre hovedtrinn. Trinn en: Anbefalingsdomenet, hvor konkrete valg i forhold til hovedressursen som skal betjenes med et anbefalingssystem, litteratur som domene og bok som anbefalingsgjenstand blir drøftet. Trinn to: Utvelgelse av egenskaper i et anbefalingssystem, der jeg drøfter konkrete egenskaper som finnes i anbefalingssystemer generelt i forhold til egenskaper som bør vektlegges for det spesifikke litteraturdomenet. Trinn tre: Drøfting av anbefalingsmetoder, der forskjellige anbefalingsmetoder blir drøftet med fokus på tilgjengelig data for sosiale nettsteder.

Trinn en og to vil foregå på et mer generelt nivå og forteller om litteraturanbefalinger på sosiale nettsteder, mens i trinn tre trekker jeg det ned på casenivå. I tredje trinnet vil jeg bruke resultatet fra drøftingen i trinn en og to, og data hentet fra Bokelskere.no. Dette oppsettet vil gi en ryddig fremstilling av prosessen, men i virkeligheten vil arbeidet med disse trinnene ofte finne sted parallelt.

Målet for analysearbeidet er å finne frem til en eller flere metoder for litteraturanbefaling, hvor hovedmålet med anbefalingen er å tilfredsstille brukeren av systemet. Hovedfunksjonen for metodene er å bevisstgjøre utvikler av anbefalingssystemer for litteratur hvilke valg som må tas. Jeg gir en oversikt over anbefalingsprosessen og valg av anbefalingsmetoder men jeg har ikke gått inn for å presentere potensielle algoritmer. Jeg har brukt tiden til å drøfte de valgene som må tas innenfor de valgte anbefalingsmetodene og konsekvensene ved å ta disse valgene.

## **5.1 Trinn en: Litteraturanbefaling som domene**

Domenet litteraturanbefaling har bøker som fokusgjenstand, men er ikke manifestasjonsavhengig, jeg tar dermed ikke stilling til om en bok er et fysisk eksemplar, en e-bok eller lydbok. Jeg tar med andre ord ikke hensyn til hva slags type bok det er så lenge den kan registreres på sosiale nettsider for bokinteresserte som Bokelskere.no og LibraryThing. I kapittel 3 la jeg frem spesielle aspekter ved en bok som må tas til etterretning når en utvikler et anbefalingssystem. Et anbefalingssystem er en kompleks tjeneste og kan kreve mye av utvikler både med tanke på valg av beste anbefalingsmetode, tilgjengelig data og de tekniske kravene det setter til programvaren. Jeg setter i dette trinnet anbefalingsgjenstanden, boken, sammen med generelle valg en utvikler står ovenfor tidlig i denne prosessen. Jeg vil diskutere utviklers valg med bakgrunn i Picult et al. (2011) sine synspunkter for å utvikle et godt anbefalingssystem for et bestemt domene. De deler inn valg som skal tas i forhold til anbefalingssystemet i tre områder. Det første går på anbefalingssystemet som del av en ressurs, det andre er brukerne av anbefalingssystemet og det tredje tar for seg tilgjengelige data. Enkelte avgjørelsers implementariske natur er nevnt da det vil være unaturlig ikke å kommentere dem, men er ikke vektlagt.

### **5.1.1 Ressursrelaterte valg**

Under ressursrelaterte valg mener Picult et al. (2011) at utvikler må se på den større konteksten og på tekniske løsninger. De mener at ressursrelaterte valg går ut på anbefalingssystemet rolle, hvordan anbefalingene skal vises, hva som er spesielt med det bestemte domenet og hvilken digital arena ressursen inngår. Det første som må klargjøres er anbefalingssystemets rolle i forhold til ressursen den betjener. Er anbefalingssystemet ressursens hovedfunksjon, en ekstra funksjon eller har den en bestemt rolle? For netthandel er ofte anbefalingssystemet en av hovedfunksjonene og må være aktiv og tilgjengelig til en hver tid. For sosiale nettsteder vil anbefalingssystemet være en ekstrarfunksjon som utvikler legger til for å hjelpe og tilfredsstillere brukere. Brukere på sosiale nettsider som Bokelskere.no er ofte aktivt på utkikk etter nye leseropplevelser og da kan et anbefalingssystem være et godt verktøy for dette.

Siden litteraturanbefalinger vanligvis ikke er hovedfunksjonen på sosiale nettsteder vil dette ofte gå frem i hvordan anbefalingene vises grafisk. Det kan være nok å la anbefalinger være en egen lenke på brukerens profilside, i tilknytning til visning av bøker eller utvikler kan lage en egen fane som er viet til anbefalinger. Dette er et valg som utviklere av hver enkel ressurs må

bestemme og som må passe med informasjonsarkitekturen på nettstedet. Jeg mener sosiale nettsteder for bokentusiaster bør gjøre anbefalinger lett tilgjengelig for brukeren av nettstedet og en logisk tilknytning er brukerens egne boksider. Resultatet for litteraturanbefalinger bør vises i en rangert liste hvor de mest aktuelle bøkene kommer øverst for ved å gi brukeren flere valg samtidig er det større sjanse for at han blir fornøyd. Visuelt sett kan det være lurt å benytte seg av små coverbilder, men dette er et valg som også må passe med nettsidens grafiske design.

En utvikler må ta hensyn til om det er behov for å begrense gjenstandsomfanget i et anbefalingssystem. For litteratur er det naturlig å ta stilling til om systemet skal anbefale skjønnlitteratur, faglitteratur eller begge deler. For å avgjøre dette må det sees på brukere og hvorfor de ønsker anbefalinger. Finnes det spesielle grunner til å velge vekk en av litteraturtypene? Det er viktig å tenke igjennom hvilken arena anbefalingssystemet skal dekke. Et anbefalingssystem gir anbefalinger på bakgrunn av brukeres smak, uavhengig av hvordan denne måles. Jeg arbeider med å utvikle et anbefalingssystem hvor brukerne er i fokus og har ingen skjulte interesser av å anbefale den ene typen litteratur fremfor den andre, dermed virker det ikke logisk å velge vekk noen former for litteratur. Et anbefalingssystem som tar hensyn til brukerens smak vil anbefale bøker som skal tilfredsstille brukeren uansett litteraturtype. Hvis brukeren leser mye faglitteratur er det trolig at han også setter pris på å få dette anbefalt, og motsatt. Hvis en bruker har lest og vurdert en god blanding av faglitteratur og skjønnlitteratur er det god grunn til å tro at anbefalinger innenfor begge litteraturtypene vil falle i smak.

Når det gjelder arena for sosiale nettsider er det selvforklarende at de skal fungere på nettet og bør tilpasses slik at det kan brukes uavhengig av brukers valgte nettleser. Jeg har i kapitel 3 drøftet lesing og mener dette er en individuell aktivitet. På bakgrunn av dette mener jeg at anbefalingene skal tjene enkeltstående brukere. Selv om brukere av sosiale nettsteder deler sine leseropplevelser vil anbefalinger tjene de som enkeltstående brukere. Når det gjelder infrastrukturen til en nettressurs er det et implementarisk valg som tas av utvikler i forbindelse med hver enkelt ressurs. Jeg går dypere inn på denne typen avgjørelser i oppgaven. Ressursrelaterte valg gjennomgått ovenfor er samlet i en enkel tabell hvor mine forslag for sosiale nettsteder som tilbyr litteraturanbefalinger er tatt med på en enkel og oversiktlig måte. Tabellen vises på neste side.

Ressursrelaterte valg	Mulige valg	Forslag for litteraturanbefaling
Anbefalingssystemets rolle i ressursen	<ul style="list-style-type: none"> <li>•Hovedfunksjonen?</li> <li>•Ekstrafunksjon?</li> <li>•Spesifikke oppgaver?</li> </ul>	En ekstrafunksjon for hovedressursen.
Visning av anbefalingsresultat	<ul style="list-style-type: none"> <li>•Presentasjonen av en og en bok?</li> <li>•Liste med flere bøker?</li> <li>•Sekvens av bøker som hører sammen?</li> </ul>	Rangert liste, siden dette gir brukeren en bedre oversikt over potensielle bøker.
Resultatpresentasjon og plass på skjermen	<ul style="list-style-type: none"> <li>•Integrert på siden?</li> <li>•Brukeren velger å be om det?</li> <li>•Bruker kan velge det bort?</li> </ul>	Integrert på brukers boksider.
Begrensning av gjenstandsomfanget	<ul style="list-style-type: none"> <li>•Skjønnlitteratur?</li> <li>•Faglitteratur?</li> <li>•Begge?</li> </ul>	Ikke begrenset. Dette vil bli justert av brukers boksamling. Eventuelt at bruker kan reservere seg.
Arena for anbefalingssystemet	<ul style="list-style-type: none"> <li>•Nettside?</li> <li>•Mobilapplikasjon?</li> <li>•Flere arenaer samtidig.?</li> </ul>	Nettbasert tjeneste. Tilgjengelig via nettlesere som Firefox, Internet Explorer og Safari.
Bruk	<ul style="list-style-type: none"> <li>•En spesifikk bruker?</li> <li>•En brukergruppe?</li> </ul>	Skal levere anbefalinger til en enkeltstående bruker.
Infrastrukturen i anbefalingssystemet	<ul style="list-style-type: none"> <li>•Tjenerbasert?</li> <li>•Klientbasert?</li> </ul>	Ikke aktuell for meg å avgjøre.

Tabell 2: Ressursrelaterte valg for litteraturanbefalinger.

### 5.1.2 Brukerne i en ressurs

En ressurs vil alltid ha en brukergruppe den skal tjene, og anbefalingssystemet må tilpasses dette. Brukere er ofte forskjellige og kan ha forskjellige formål. Det er dermed viktig å innhente nok data om anbefalingssystemets brukere, deres formål og forventninger, for å kunne betjene dem på beste mulige måte. Paradoksalt nok holdes innsamling av demografisk data om brukeren ofte til et minimum og representeres med kjønn, alder og stedstilknytning. Som vist tidligere i oppgaven kan brukere være restriktive med hva slags type data de ønsker å dele. Det er derfor viktig for en utvikler ikke å spørre om for mye demografisk data så en bruker velger å la være å bruke ressursen. Et alternativ kan i denne sammenhengen være å gjøre mindre viktige punkter valgfrie.

Denne restriktive adferden hos brukere er synd når en skal utvikle anbefalingssystemer fordi jo mer et anbefalingssystem vet om brukerne, jo bedre anbefalinger kan det tilby. Picault et al. (2011) trekker frem at teknikker som stereotyping av brukere kan brukes til å skille eller samle

brukere på en enkel måte. De trekker frem at en på bakgrunn av demografisk data blant annet kan hjelpe en ny bruker å få anbefalinger uten å ha vurdert, kjøpt eller sett på gjenstandene i et anbefalingssystem. En utvikler bør tilby brukeren å være anonym gjennom et unikt brukernavn. Dette gir brukerne en større følelse av anonymitet og de føler ofte at det er lettere å bidra til fellesskapet når de er anonyme. Samtidig gjør det at brukere ikke kan gjenkjennes av søkesystemer som Google, som henter inn nettsider fra nettet.

Et sosialt nettsted for litteraturinteresserte er trolig åpent for dem som ønsker det, men en restriksjon kan være alder. Dette gjør at brukerne er meget forskjellige, og at deres fellestrekk er lesing. Det viser seg at når lesere kommer til tenårene er det ikke så nyttig lenger å se på alder. Boksamlinger på både Bokelskere.no og LibraryThing viser at klassikere og fantasybøker finnes representerte skjønnt forening uten sammenheng med brukerens alder. Det er på sosiale nettsteder for litteraturinteresserte lesing som er aktiviteten som deles i felleskap.

En mulighet for å utnytte lesing som likhetstrekk kan være å be brukerne legge inn foretrukket sjanger, eventuelt sjangre de ikke liker og dermed ikke ønsker å bli anbefalt. En brukers mål ved å bruke et anbefalingssystem for litteratur vil være å finne en eller flere gode bøker som inspirer til videre lesning. Lesing og bøker forbundet med sosiale nettsteder kan kalles fritidslesing. Dermed er det ikke nødvendigvis en spesiell bok, eller emne en bruker er på jakt etter. Et anbefalingssystem knyttet til en fagrelatert database vil ha brukere som i stor grad er på jakt etter bestemte dokumenter eller temaer og gjerne alle potensielle gjenstander.

Anbefalingssystemer for fritidslesning derimot, og trenger ikke å være fullstendige i sine anbefalinger. I et anbefalingssystem for litteratur vil det ikke være nødvendig for en bruker å få frem alle bøker som kan være aktuelle. De er trolig mer interessert i noen potensielle bøker, eller å leke litt rundt. Det er vanlig at en brukers smak som leser forandrer seg over tid og det må tas høyde for dette under utviklingen av systemet. Tabell for brukerrelaterte valg, mulige valg og mine forslag for anbefalingssystemer for litteratur finnes på neste side.

<b>Brukerrelaterte valg</b>	<b>Mulige valg</b>	<b>Forslag for litteraturanbefalinger</b>
Demografisk data	<ul style="list-style-type: none"> <li>•Hvilke muligheter er det til å samle inn brukerdata?</li> <li>• Hvor mye data trengs?</li> </ul>	<ul style="list-style-type: none"> <li>•Bruker legger inn et minimum i profilen, alder, kjønn og bosted.</li> <li>• Mye data om brukere skapes gjennom transaksjoner.</li> <li>• Er det mulighet kan brukere få valgfrihet til å legge inn supplerende data som sivilstand, yrke, fritidsinteresser og religion.</li> </ul>
Hvem er brukerne?	<ul style="list-style-type: none"> <li>•Finnes det sterke fellestrekk?</li> <li>•Kan og bør brukerne stereotyperes?</li> </ul>	<ul style="list-style-type: none"> <li>•Brukere i alle aldre, med stor variasjon i yrke og bakgrunn.</li> <li>•Fellestrekk er lesing.</li> <li>•Dekning avgjør om anbefalingssystemet kan dra nytte av stereiotyping.</li> </ul>
Brukers mål	<ul style="list-style-type: none"> <li>•Hvilket mål har brukeren?</li> <li>•Hva er formålet?</li> </ul>	Det overordnede målet er å finne frem til en eller flere gode bøker.
Brukers forventning	<ul style="list-style-type: none"> <li>•Forventer brukeren å finne alle potensielle gjenstander, noen potensielle gjenstander eller er det for gøy.</li> </ul>	Anbefalingssystemet er en tilleggsfunksjon og blir brukt i fritiden for å finne noen potensielle gjenstander eller for gøy.
Brukerstabilitet	<ul style="list-style-type: none"> <li>•Vil bruker forandre smak?</li> </ul>	<ul style="list-style-type: none"> <li>•Ja brukerne vil forandre smak etter hvert som de leser.</li> <li>•Dette vises gjennom terningkast og innhold i boksamling.</li> <li>•Anbefalingssystemet må ta hensyn til dette.</li> </ul>
Personvern	<ul style="list-style-type: none"> <li>•Hvor sensitiv er informasjonen som finnes om brukerne?</li> <li>•Hvordan kan personlig data missbrukes?</li> <li>•Er det trolig at personlig data fra ressursen missbrukes?</li> </ul>	<ul style="list-style-type: none"> <li>•Informasjonen som lagres om brukerne på sosiale nettsteder er ikke spesiell sensitiv.</li> <li>•Nettsteder operer vanligvis med brukernavn.</li> <li>•Det er ikke trolig at noen ønsker å missbruke informasjonen.</li> <li>•Brukerne skal være anonyme i en anbefalingssituasjon.</li> </ul>

*Tabell 3: Brukerrelaterte valg for litteraturanbefalinger*

### 5.1.3 Datagrunnlaget

Gjenstandene et anbefalingssystem skal anbefale, hvordan data om de blir hentet inn, lagret og brukt er meget relevant for en utvikler. Jeg kommer til å bruke mye tid på å drøfte datakilder som finnes på Bokelskere.no, så dette blir kun en kort gjennomgang av hva en utvikler bør tenke på. Picult et al. (2011) poengterer at datakildene for gjenstandene ofte finnes på ressursen som anbefalingssystemet skal tjene og derfor kan det hende man må jobbe med data i en annen form enn ønsket. Hva slags data som er tilgjengelig vil ha mye å si for hva slags anbefalingsmetoder som kan benyttes. Hvilke datatyper som er tilgjengelige, hvordan de hentes inn og hvordan de er strukturerte må gjennomgås og eventuell ustrukturert data må struktureres før de kan brukes. En brukervurdering gis på sosiale nettsider eksplisitt gjennom for eksempel terningkast eller leses implisitt gjennom for eksempel boksamling. Nettressursens datakvalitet sier mye om hvordan utvikler kan bruke dataene, dette blir nøye diskutert i trinn tre. Derfor blir denne gjennomgangen en kort presentasjon av utfordringer og mulige valg.

Antall gjenstander som skal anbefales i forhold til antall brukere som benytter ressursen må vurderes, da det kan gi utslag i hvilken algoritme utvikler bør velge. Gjenstandsmengdens stabilitet må vurderes i forhold til om samlingen øker eller er stabil. En gjenstands levetid er også viktig, er det slik at gjenstander blir uaktuelle? På en sosial nettside for litteraturinteresserte vil det være et konstant tilslag av nye titler, men gamle titler blir ikke nødvendigvis uaktuelle. Jeg har laget en tabell på dette, men her er det mange valg å ta og konsekvensene kan slå tydelig ut på hvordan anbefalingssystemet fungerer. Dette vil jeg drøfte omfattende i tredje trinn. Tabellen gir altså ingen konkrete forslag, men en kort oppsummering og kan finnes på neste side.



<b>Datarelaterte valg</b>	<b>Mulige valg</b>	<b>Litteraturanbefalinger</b>
Datatype	<ul style="list-style-type: none"> <li>•Ustrukturert data?</li> <li>•Halvstrukturerte data?</li> <li>•Strukturerte data?</li> </ul>	Nettsider har ofte både halvstrukturert og ustrukturert data.
Datakvalitet	<ul style="list-style-type: none"> <li>•Hvordan er kvaliteten på dataen?</li> <li>•Kan gjenstander skilles fra hverandre?</li> </ul>	Må diskuteres med det spesifikke datagrunnlaget for hånden.
Gjenstandsmengde	•Inneholder ressursen mange gjenstander i forhold til brukere.	Sosiale nettsteder for litteratur har ofte flere gjenstander enn brukere registrert
Stabilitet	<ul style="list-style-type: none"> <li>•Er gjenstandsmengden stabil?</li> <li>•I hvilket tempo forandrer den seg?</li> <li>•Blir gjenstandene uaktuelle?</li> </ul>	Bøker gir en forholdsvis stabil samling. Det kommer konstant nye, men de gamle blir ikke uinteressante for brukerne.
Brukervurdering	<ul style="list-style-type: none"> <li>•Implisitt?</li> <li>•Eksplisitt?</li> </ul>	Nettsteder har ofte både eksplisitt og implisitt data.
Type vurdering	<ul style="list-style-type: none"> <li>•Binære vurderinger?</li> <li>•Skalerte vurderinger?</li> <li>•Unære vurderinger?</li> </ul>	Må diskuteres. Nettsteder har vanligvis mulighet til å bruke skalerte vurderinger, tekstlige kommentarer eller boksamling.

*Tabell 4: Datarelaterte valg for litteraturanbefalinger.*

## **5.2 Trinn to: Egenskaper i et anbefalingssystem**

Egenskaper ved anbefalingssystemer ble gjennomgått i kapittel 2, nå er det på tide å diskutere dem. Det er viktig å gå dypere inn i de egenskapene et anbefalingssystem har for å kunne si noe om hva som bør vektlegges ved utvikling av et anbefalingssystem tilpasset et bestemt domene. Jeg vil drøfte egenskapene i lys av litteraturanbefalinger, hva som er spesielt med bøker som gjenstand og lesere som brukere. Jeg vil også trekke frem de viktigste egenskapene for domenet og si hvorfor jeg mener disse bør vektlegges. Avslutningsvis oppsummerer jeg egenskapene i en tabell med konkrete retningslinjer for litteraturdomenet.

### **5.2.1 Drøfting av potensielle egenskaper**

Flere av de 13 egenskapene blir nevnt som evalueringskriterier hos Jannach et al. (2011), Herlocker et al. (2004) og Adomavicius & Tuzhilin (2005). På bakgrunn av dette regner jeg med at jeg dekker de egenskapene som er viktige å tenke igjennom i forhold til anbefalingssystemer. Det er viktig å være klar over at noen av egenskapene er omvent proporsjonale. Et eksempel på dette er at prediksjonsnøyaktigheten vil synke når serendipitet eller mangfold øker. Derfor må en utvikler tenke nøye igjennom hvilke egenskaper som passer for det aktuelle anbefalingsdomenet. Det finnes ikke nødvendigvis en riktig sammensetning av egenskaper å vektlegge, det vil variere innenfor forskjellige ressurser og på bakgrunn av tilgjengelig data. Utvikleren må derfor velge ut de viktigste. Det kan også variere hvilke egenskaper som skal vektlegges ved en oppstartsfasen og hvilke som kan bearbeides når anbefalingssystemet er etablert.

#### **Brukerpreferanse**

Som tidligere nevnt er brukerpreferanse viktig fordi utviklere trenger en forståelse for brukernes motivasjon, forventninger og mål. Hva brukerens mål med å bruke ressursen er vil være avhengig av hva slags ressurs det er snakk om. Anbefalingssystemet vil i forbindelse med denne oppgaven være en ekstraraksjon på et nettsted. Sosiale nettressurser er som oftest åpne for alle som ønsker det, dermed vil det finnes brukere i alle aldre der. Bakgrunn, erfaring, lesernivå, lesepreferanser og bruk av nettstedet vil variere kraftig og det finnes ikke ett fasitsvar på hvem brukeren er. Lesing er som diskutert en subjektiv handling og dette er med på å vanskeliggjøre definisjonen av brukerpreferanse innenfor denne ressursen. Dermed vil det være mer fruktbart å se på andre egenskaper enn brukerpreferanser når viktige egenskaper i et anbefalingssystem skal kartlegges. Når anbefalingssystemet er aktivt og har brukere kan utvikler foreta brukerundersøkelser for å kartlegge hvilke egenskaper som kan justeres i forbindelse med en forbedring av anbefalingssystemet. Denne typen brukerundersøkelser kan sjekke brukers preferanser innenfor anbefalingssystemets design og resultat. Innenfor resultat kan egenskaper som prediksjonsnøyaktighet kontra serendipitet være spennende å se på.

#### **Prediksjonsnøyaktighet**

Anbefalingssystemer generelt arbeider ut fra en teori om at en bruker vil ha flere gjenstander som ligner på gjenstander brukeren allerede liker. Systemet bruker forskjellige metoder for å lære seg hva en bruker foretrekker og kommer med anbefalinger på bakgrunn av dette. Prediksjonsnøyaktighet forteller en utvikler hvor nøyaktig et anbefalingssystem er til å forutsi

hvilke anbefalinger det skal presentere for en bruker. Det er denne egenskapen som forteller en utvikler hvor mye anbefalingene som genereres i systemet ligner på bøker brukerne ville valgt selv. Den vil for litteraturanbefalinger bygge på tidligere vurderinger og/eller boksamling. Undersøkelser på dette gjøres ofte gjennom offline forsøk hvor deler av datasettet ”gjemmes” og målekriterier som prediksjon og fullstendighet benyttes ofte. Problemet med høy prediksjonsnøyaktighet er at anbefalingene kan bli for like, og brukere kan føle at anbefalingssystemet ikke gir de noe nytt. For like anbefalinger kalles overspesialisering og er det motsatte av serendipitet.

Dette er et omdiskutert tema og på den ene siden er oppfatningen at brukere foretrekker anbefalingssystemer som gir flest mulig relevante treff, mens på den andre siden ønsker brukere også å finne noe nytt. Det er trolig at brukere ønsker en viss mengde likhet med tidligere valg, bøker som er lest og likt. Prediksjonsnøyaktighet er omvendt proporsjonal med egenskapene serendipitet, mangfold og til en viss grad nyhet. Det er dermed viktig å undersøke virkningsgraden under utviklingen av anbefalingssystemet. Spørsmålet er hvordan utvikler skal finne ut hvor grensen går for å tilfredsstille både lesere som foretrekker bøker som ligner mye på tidligere likte bøker og lesere som foretrekker bøker som tilfører dem noe nytt. Dette problemet er størst blant innholdsbaserte anbefalingsmetoder og velger utvikleren å benytte seg av dette bør teknikker diskutert under egenskapen serendipitet vurderes. For å undersøke dette i det aktuelle anbefalingssystemet kan en brukerundersøkelse gjennomføres for å kartlegge hvordan anbefalingene som genereres mottas. Dermed kan de enkelte ressursene se på om prediksjonsnøyaktighet skal justeres.

## **Dekning**

Begrepet dekning kan betegne flere utfordringer som finnes i forbindelse med mengden data og kvaliteten på datasettet man har til rådighet. I forbindelse med litteraturanbefalinger på sosiale nettsteder vil kvaliteten på gjenstandsdata være en stor utfordring. I dagens digitale samfunn ønsker man ikke finne opp hjulet på nytt, det å bruke data fra andre kilder er svært utbredt. Mange nettbaserte tjenester er mashup-drevet<sup>33</sup> og har ikke kontroll over originalkilden men adopterer og adapterer poster til sitt bruk, og det er ikke uvanlig å hente poster fra flere kilder.

---

<sup>33</sup> Mashup: a web application that uses content from more than one source to create a single new service displayed in a single graphical interface. [www.mashups.web2learning.net/glossary](http://www.mashups.web2learning.net/glossary)

To store utfordringer finnes i forbindelse med dette. Dataene ikke blir representert på en konsekvent måte og systemet kan ha flere forskjellige utgaver av samme gjenstand. Et godt eksempel på dette er Bokelskere.no som har en stor utfordring når det gjelder datagrunnlag, spesielt med tanke på utgavekopi. Bokelskere.no henter norsk bokdata fra Den norske bokdatabasen og internasjonal bokdata gjennom Amazon.com. Dette gjør at de har mange utgaver av hver bok. Forskjeller på utgaver kan være språk, forskjellig cover, lydbokutgaver på cd, lydbokutgaver på mp3, storskriftsutgaver og digitale utgaver. De forskjellige utgavene av en bok blir ikke nødvendigvis koblet sammen.

De norske utgavene av en bok og bøker av kjente forfattere har i hovedsak blitt koblet sammen. Søker man opp *Stolthet og fordom* får man tre treff, ett av dem er en samlepost med alle de tilgjengelige utgavene av boken. Klikker man seg innover på *Stolthet og fordom* ser man at Bokelskere.no har fått til skillet mellom, verk, uttrykk, manifestasjon og eksempel (jmf. Vedlegg 9). Søker man opp *Shantaram* får man ni treff. Den første er en samlepost med fire utgaver, som ved den første posten i søket på *Stolthet og fordom*. Deretter følger 8 forskjellige utgaver av boken med forskjellig innbinding, format og språk. Her har de ikke fått til sammenkoblingen og boken ligger inne med mange utgavekopier som ikke er koblet sammen. Det jobbes med denne utfordringen på Bokelskere.no, men det er et omfattende problem som ikke kan løses over natten og de må derfor ta hensyn til dette i en anbefalingssituasjon.

Det finnes også brukere som er motstandere av at denne typen data kobles sammen. De hevder at å lese en bok på norsk, å lese en bok på engelsk eller å få den opplest er tre forskjellige leseropplevelser. På Bokelskere.no er det slik at hvis en diskusjon startes i forbindelse med en utgavekopi vil den være synlig gjennom alle de andre utgavekopiene av det verket som er koblet sammen. Dermed mener enkelte brukere at diskusjoner som handler om en utgavekopi av et verk ikke nødvendigvis vil handle om noe som angår en annen utgavekopi. Det er under hver utgavekopi mulig å starte en egen diskusjon og den første setningen i diskusjonen vil være synlig for andre brukere, dermed kan forskjellige temaer om et verk diskuteres parallelt under de forskjellige utgavekopiene og være synlig for alle som har registrert de. Jeg mener det må være bedre å samle alle utgaver og at det i en diskusjon vil komme frem hva som skal diskuteres og i hvilken kontekst. Hvis noen diskuterer personene i en bok vil det være like aktuelt for meg selv om jeg har lest boken på engelsk i stedet for på norsk. Hvis noen diskuterer fortellerstemmen på en lydbok så kan brukeren som ikke er interessert i dette la den spesifikke diskusjonstråden ligge.

## Norsk utgave

Shantaram

av Gregory David Roberts (forfatter) og John Erik Frydenlund (oversetter).

2008 (Innbundet) Forlaget Press

Likte du denne? Vi har boker anbefaling!

Kjøp hos bokkilden.no

Kjøp hos bokdykk.no

Finn denne boka på ditt nærmeste bibliotek:

Ullensaker bibliotek

Omtale fra Den Norske Bokdatabasen:

I 1978 ble Greg Roberts dømt til 20 års fengsel for å ha ranet en rekke banker. To år senere rømte han fra et av Australias sikreste fengsler og ble landets mest ettersøkte mann de neste ti årene. Åtte av disse årene tilbrakte han i Bombay. Denne boken tar utgangspunkt i forfatterens opplevelser fra livet på rømmen; et liv i randsonen av hva et menneske kan tåle.

Omtale fra forlaget:

I 1978 var Gregory David Roberts heroinmisbruker og ble dømt til 19 års fengsel for en serie væpnede ran. Etter to år rømte han fra et høyrisikofengsel på høylys dag og ble Australias mest ettersøkte mann. Roberts søkte tilflukt i Bombay, der han slo seg ned i slummen: først med en improvisert praksis som doktor blant slumbeboerne, senere som Bombay-mafiaens våpenpusher, smugler og falskner. Shantaram er Roberts egen medrivende beretning om et liv helt i ytterkanten av det menneskelig mulige, om frihetens pris, om betingelsene for en ny begynnelse, om kampen mellom det gode og det onde.

Emneord: Bombay, Dokumentar-roman, India, Mafia, Outsidere

## Engelsk utgave

Shantaram

av Gregory David Roberts (forfatter).

2005 (Paperback) Abacus

Likte du denne? Vi har boker anbefaling!

Kjøp hos bokkilden.no

Kjøp hos bokdykk.no

Finn denne boka på ditt nærmeste bibliotek:

Ullensaker bibliotek

Se på denne utgaven hos amazon.co.uk

- ISBN-13: 9780349117546
- Sider: 944

Emneord: India, selvbiografisk, Skjønnlitteratur (engelsk), slum

Finn denne boka på ditt nærmeste bibliotek:

Ullensaker bibliotek

rediger | mindre bokdata

Dine emneord for denne boka: India, selvbiografisk, slum (rediger)

Bilde 5: Forskjellige utgavekopier av Shantaram.

Siden Bokelskere.no som diskutert ovenfor henter bokdata til en utgavekopi fra flere kilder vil mengden bokdata på utgavekopi variere deretter. Dette gjør at sammenligningsgrunnlaget blir mer usikkert, og kommer tydelig frem i forskjellen mellom den norske og den engelske versjonen av *Shantaram*. På bildet ovenfor ser man at bokdata på den engelske versjonen kun består av Tittel, Forfatter, Emneord, Sideantall, ISBN og Diskusjon. Den norske versjonen er utstyrt med Omtale fra både Den norske bokdatabasen og fra forlaget, i tillegg til omfattende bibliografisk informasjon. Likheter ved bruk av Omtale vil derfor ikke være en enkel løsning for disse to bøkene.

En annen naturlig kilde til likhet er emneord. Den engelske utgaven av *Shantaram* er utstyrt med emneordene India, Selvbiografisk, Skjønnlitteratur (engelsk) og Slum. Den norske utgaven har emneordene Bombay, Dokumentar-roman, India, Mafia og Outsidere. Av åtte mulige emneord har de to bøkene kun et emneord felles. Med disse eksemplene kommer det klart frem at dekning er en viktig egenskap for utvikler å arbeide med, slik at datasett blir utarbeidet slik at det representerer ressursen på best mulig måte. Emneordene satt av brukere på Bokelskere.no kommer bare frem på en bok som ligger i brukerens samling. Dette er trolig forbundet med at en

bruker kun har mulighet til å gi emneord på bøker han har en tilknytning til gjennom boksamlingen sin.

Cold start-problemet rammer både nye brukere i innholdsbaserte metoder og gjenstander i kollaborativ filtreringsmetoder. Dette gjelder for hvilke metode som velges for litteraturanbefalinger også. I kapittel 2 gikk jeg igjennom forsøk på hvordan dette problemet kan løses. Det er i følge Shani & Gunawarda (2011) viktig å kartlegge dette problemet i det aktuelle anbefalingssystemet. Deretter må en utvikler vurdere hvor mye tid og penger han skal bruke på å forbedre de brukerne eller gjenstandene som blir rammet. Alternativer er blant annet hybride metoder og ”falske” vurderinger hvor en gjennomsnittsvurdering blir brukt. Saricks (2005) snakker om at metoden med å tilby bøker fra en forhåndsdefinert liste i forbindelse med folkebibliotek. Hun legger blant annet bøker fra populære forfattere ved det bestemte biblioteket til grunn. Det er derfor trolig at dette kan benyttes digitalt som falske positive vurderinger. På Bokelskere.no finnes det forskjellige lister med bøker som kan tenkes å brukes til nettopp dette. Lister over populære bøker, bøker som leses for øyeblikket, bøker som er satt som favorittbøker og bøker med høy favorittprosent vil være aktuelle i denne forbindelsen.

### **Tillit**

Egenskapen tillit som Shani & Gunawarda (2011) snakker om går ut på om brukeren er villig til å stole på at anbefalingssystemet gir nyttige anbefalinger. De foreslår at et anbefalingssystem kan anbefale gjenstander som brukerne allerede har kjennskap til og liker. På denne måten vinner anbefalingssystemet brukers tillit, og brukeren velger å følge deres anbefalinger. En annen måte å skape tillit på er å forklare ovenfor brukeren hvorfor den og den boken blir anbefalt. På sosiale nettsteder kan utvikler begrunne anbefalinger som dannes på grunnlag av lignende innhold, eller fordi andre brukere liker bøkene ute at det byr på store problemer. Det er også slik at en anbefaling som ikke faller i smak ikke koster brukeren så mye økonomisk sett, men det vil jo være snakk om å ”kaste bort” litt tid. Siden brukers tap ved å følge en litteraturanbefaling ikke er så stort vil terskelen for å stole på anbefalingssystemet også være lavere enn i systemer hvor økonomisk tap er en realitet. En annen måte er å forklare hvilke andre brukere som liker gjenstanden som anbefales. Hvis dette er personer som brukeren stoler på, som for eksempel brukere den aktive brukeren følger, blir det som å høre på en ”venn”.

## Selvtillit

Selvtillit til egne anbefalinger kan være en aktuell funksjon å legge til i anbefalingsøyeblikket. Det kan tenkes at sosiale nettsteder som Bokelskere.no er tjent på presenterer hvor stor tillit de har til anbefalingene for brukerne, da det er lettere å tilgi anbefalinger som ”bommer” i en slik situasjon. Det kan også tenkes at Bokelskere.no bør se på hvor stor treffliste de presenterer og hvis listen er stor nok til ikke å ta med mindre sikre anbefalinger lar de det være. Slik som Bokelskere.no presenterer anbefalinger nå så vises det 100 anbefalinger hver gang brukeren klikker på *Vi har bokeranbefalinger* lenken. De vises med ti anbefalinger per side med tittel, forfatter, bilde og kort informasjon om hva andre bokelskere som et fellesskap syntes om den i form av terningkast og antall ganger den er satt som favorittbok. Presentasjon og begrensning er viktig og en bokelsker på nettstedet sier hun ikke orker å sjekke når en anbefalingsliste inneholder hundrevis av anbefalinger (Hilde S H 2011). Dette er i tråd med det Morville & Rosenfeld (2006) sier om resultatlistene. Utvikler må også tenke igjennom hvordan denne typen informasjon blir uttrykket. Det anbefales ikke å bruke prosentbetegnelse da det kan virke mot sin hensikt. LibraryThing har hatt suksess med sin skalerte vurdering hvor de forteller både hvor sannsynlig det er at boken faller i smak, og hvor mye de stoler på egen vurdering som fortalt i avsnittet om LibraryThing i kapittel 3.

Jeg mener dette kan være hensiktsmessig å gjennomføre i et anbefalingssystem, men det er ikke like nødvendig i alle sammenhenger. Hvis brukeren får presentert en rangert liste vil de bøkene som anbefalingssystemet har funnet ut passer best ligge på toppen av listen. Dermed er det også de bøkene systemet mener passer brukeren best. Det kan i denne sammenheng være best for utvikleren å sette en grense for hvor mye en bok skal matche før den blir med i resultatlisten, i stedet for et bestemt antall bøker.

## Nyhet

Å tilby brukeren en nyhet i anbefalingsøyet er å tilby brukeren en gjenstand de ikke har kjennskap til, i den forstand at de ikke har sett den, ikke har den i sin samling eller ikke har vurdert den. Denne egenskapen sammen med egenskapene mangfold og serendipitet er med på å trekke prediksjonsnøyaktigheten ned. Som diskutert tidligere har også egenskapene nyhet, mangfold, serendipitet og prediksjonsnøyaktighet mye å si i forhold til hvilken lesetype brukeren av anbefalingssystemet er. Her settes det med å anbefale noe kjent og trygt opp mot å anbefale noe nytt og ukjent. Det er et viktig spørsmål som utvikler må ta stilling til. På Bokelskere.no går

det mange diskusjoner om hvorfor og hvordan man velger litteratur. Mange er interessert i å utvide sin litterære horisont og poengterer at dette har skjedd etter de ble brukere på nettstedet. Både det å velge nye typer bøker og det å eksperimentere litt med sjanger er nevnt.

Poenget med dette innlegget er vel egentlig å si at innkjøpene mine nok er preget av mitt medlems[s]kap her på BE. For bare ett år siden ville nok impuls kjøp som dette vært av en lettere art, såkalt "chick lit" "feel good" og nyere populære utgivelser. Litt sånn "minste motstands vei" Ikke at jeg ikke har lest seriøs litteratur tidligere, men vektskålen har vel hellet mot det litt enklere. Nå er det mer jevnt fordelt (Ann Christin 2011 mars).

Dette støtter teorien om at det er lurt å sørge for at brukere av anbefalingssystemet blir servert nyheter i resultatlisten. Her mener jeg vektning av forfattere som ikke er representert i den aktive brukerens samling kan være en god mulighet.

### **Nytteverdi**

Hvordan måle nytteverdi i et anbefalingssystem for litteratur, hvor ingenting skal selges? Da må nytteverdien være hva brukeren får ut av systemet, nemlig at han får gode anbefalinger og i egenskap av dette er en fornøyd bruker. Fornøyde brukere er lojale brukere og dette er nytteverdien for ressursen som beholder brukerne sine. Brukerne er fornøyde fordi de får litteraturanbefalinger som passer dem. For Bokelskere.no vil denne egenskapen være mindre viktig siden anbefalingssystemet er en ekstraraksjon. Det er ikke trolig at noen vil velge å slutte å bruke nettstedet på grunn av anbefalinger de mottar.

### **Risiko**

Risiko sier noe om hvor mye en bruker er villig til å satse i forhold til å følge anbefalinger. Siden jeg har konsentrert meg om anbefaling av litteratur vil produktet som brukeren skal ta stilling til være bokanbefalinger. Risikoen i den sammenhengen er liten, prisen av en bok eller tiden det tar å låne den på biblioteket. De fleste brukere har en viss toleranse for dette, men anbefalinger som ikke er til nytte kan svekke en brukers tillit til ressursen anbefalingssystemet er knyttet til. Flere av brukerne poengterer selv at de er mer vågale når det gjelder å låne bøker enn å kjøpe bøker. "Om jeg kjøper noe er det som regel alltid en jeg har bestemt meg for på forhånd, med mindre det er kjempesalg eller noe. Jeg bruker biblioteket flittig, og ettersom det er gratis er terskelen for å eksperimentere langt lavere" (Aina 2011 mai). Derfor er det viktig med brukertilfredshet, og ikke å gjøre anbefalingene veldig risikable å følge, selv om ressursen ikke skal selge noe.



## **Robusthet**

Enhver nettressurs trenger å utvise en viss robusthet mot angrep fra ondsinnede mennesker. På den andre siden er det lite trolig at et anbefalingssystem på et sosialt nettsted som ikke skal selge noe, men kun gi litteraturanbefalinger blir angrepet. Derfor anser jeg ikke dette som det viktigste punkt i denne sammenhengen.

## **Personvern**

Denne egenskapen er viktig og en het politisk potet. Brukere kan ofte være veldig beskyttende og skeptiske mot å gi fra seg personlige opplysninger. Et diskusjonsinnlegg på Bokelskere.no som etterlyste en mer utfyllende profil skapte heftig diskusjon. ”Hva med å utvide profilen litt. Uten at det skal skremme liv av folk... Faste karakteristikk som yrke, sivilstand, (religion kan være et noget dodgy område) og andre hobbyer” (Heffalumpen5 2011). Dette fikk mange negative reaksjoner mer eller mindre begrunnet.

Nei dette er jeg veldig uenig i. Hva i alle dager skulle poenget med det være? Dette er ikke Facebook, men en nettside for de som elsker bøker. Jeg har lagt ut hele bokhyllen min her og logger meg på for å diskutere bøker. (Hedvig 2011).

Som videre setter spørsmålstegn ved Heffalumpen5 sin seriøsitet med å fremme forslaget. Noen trekker frem at hvis folk ikke bruker sitt sanne navn så er det tåpelig å legge inn mer profildata. Noe som beviser at vanlige brukere ikke helt skjønner at denne typen data ikke brukes på hver og enkelt som person, men kan brukes i anbefalinger eller for at utvikler kan lage statistikk om brukerne sine. Noe en av utviklerne er enig i.

Det hadde vært utrolig kjekt å vite alt om leserne av bøker! F.eks: Liker lærere krim bedre eller dårligere enn annen skjønnlitteratur? Liker bibliotekarere erotisk litteratur mer enn frisører? Hvis jeg hadde hatt tro på at det gikk an å høste inn denne informasjonen fra brukerne våre uten at noen tok anstøt av det, så hadde jeg synes det var utrolig spennende å leke med slik statistikk. Realiteten er at lese-mønstre er blant de mest følsomme tingene man kan dele, og jeg føler strikken er strukket langt nok allerede. Jeg tror man ville fått en litt uggen følelse av å bli spurt om enda flere personlige spørsmål (Nesse 2011, april d).

Dette er et godt eksempel på at brukere er skeptiske til å gi ut informasjon om seg selv, og at man må trå varsomt frem. Et interessant eksperiment kunne være å legge inn felter for yrke, sivilstand og avkrysning for hobbyer i profilen og se hvor mange som fyller ut disse i fremtiden. Jeg mener at brukere tjener mer på å bruke brukernavn til sin profil enn på å nekte å legge inn demografisk

data. Denne vegringen mot å legge inn demografisk data kan minne om en misforståelse i forhold til hva denne type informasjon skal brukes til. Det viktigste for utvikler er å ikke be om mer enn brukerne er villig til å gi, for da kan han miste dem.

### **Tilpassningsdyktighet**

I alle typer anbefalingssystemer er tilpassningsnøyaktighet viktig, men for noen gjenstander er levetiden lenger enn andre. For nyheter er det veldig viktig at ressursen oppdateres jevnlig og det en bruker er interessert i en dag kan være gammelt nytt dagen etter. En bok har en helt annen levetid enn nyheter. En bok vil ikke nødvendigvis tape seg over tid, men tilfanget av bøker vil øke. Samtidig er det ikke uvanlig at en bruker forandrer smak over tid. De er mulig å vekte en brukers vurderinger slik at de gjenstandene en bruker har vurdert nylig er de som teller mest. Mange brukere av Bokelskere.no oppgir at de har endret lesesmak siden de ble medlemmer på nettstedet. Mye av dette henger sammen med at de har oppdaget nye bøker gjennom nettstedet, gjennom anmeldelser av andre brukere, sitater og diskusjoner. De nye bøkene behøver ikke være nye i forhold til utgivelsesdato, men nye for brukerne.

### **5.2.2 Sammendrag av utvalgte egenskaper for litteraturanbefalinger**

På bakgrunn av gjennomgangen av egenskaper i anbefalingssystemer og vektlegging av litteraturdomenet trekker jeg frem de egenskapene jeg anser som spesielt viktige ved utviklingen av et anbefalingssystem: Prediksjonsnøyaktighet, dekning, tillit, serendipitet, tilpassningsdyktighet og til en viss grad personvern. Disse egenskapene bør gjenspeiles i valg av anbefalingsmetode. Prediksjonsnøyaktighet viser utvikler hvor nøyaktig systemet er til å forutsi anbefalinger som er relevante for brukeren. Siden den allmenne oppfatningen er at en bruker foretrekker anbefalingssystemer som gir flest mulig relevante treff, er dette en viktig egenskap for anbefalingssystemet. Et anbefalingssystems dekning er viktig på flere måter. Det må være tilstrekkelig datadekning for å gjennomføre de tekniske trinnene som inngår i et anbefalingssystem. Det må også være tilstrekkelig data tilgjengelig for gjenstandene, slik at de kan representeres på en måte hvor man kan måle likhet mellom dem på en tilfredsstillende måte. Dessuten må utvikler være klar over cold start-problemet.

Utvikler skal kunne skille mellom forskjellige utgavekopier av en bok slik at bruker ikke blir anbefalt gjenstander han har fra før. En brukers tillit til anbefalingssystemet må opparbeides slik at en bruker benytter seg av anbefalingene. Serendipitet i et anbefalingssystem vil tilfredsstillende en brukers trang til å oppdage nye ting. Tilpassningsdyktighet er viktig for et hvert

anbefalingssystem som brukes over tid. En bruker vil utvikle seg, og i mange tilfeller forandre smak over tid. Derfor må anbefalingssystemet reflektere denne utviklingen. Personvern er viktig for et hvert anbefalingssystem. Informasjon om brukere skal anonymiseres og en bruker skal aldri føle seg utlevert. Utvikler har også et ansvar for å passe på at utenforstående ikke får tak i personopplysninger om brukere.

Ikke prioriterte egenskapene er: Brukerpreferanse, selvtillit, nyhet, mangfold, nytteverdi, risiko og robusthet. Disse egenskaper ble ikke prioritert fordi jeg anser dem som mindre viktig for et anbefalingssystem innenfor litteratur, eller fordi de dekkes opp av andre inkluderte egenskaper. Brukerpreferanse er en egenskap som blir dekket opp hvis prediksjonsnøyaktighet og serendipitet blir overholdt. Det er viktig for en utvikler å være klar over egenskapen brukerpreferanse og evaluere anbefalingssystemet for å se om brukere mener denne egenskapen blir overholdt etter hvert. Siden domenet er bestemt og målet for å generere litteraturanbefalinger er å gi brukeren den neste gode leseropplevelsen mener jeg at det på nåværende tidspunkt ikke er nødvendig å bruke tid på denne egenskapen. Systemets selvtillit til egne anbefalinger vil indirekte vises i en rangert liste, hvor de mest relevante bøkene vises øverst. Denne egenskapen er mer relevant i et system hvor tillit er vanskelig å bygge opp og risikoen for hva en bruker ”taper” ved å følge en feilaktig anbefaling høyere enn innenfor litteraturanbefaling. Siden egenskapen risiko er lav er det heller ikke nødvendig å med et stort kartleggingsarbeid av dette. Nyhet og mangfold er egenskaper som jeg mener blir dekket av egenskapen serendipitet. Nytteverdi er i liten grad målbart innenfor litteraturanbefalinger hvor det ikke selges noe. Det er lite trolig at et litteraturanbefalingssystem blir utsatt for angrep utenfra og derfor trenger utvikler ikke bruke mye tid på egenskapen robusthet. Hvis utvikler på forhånd vet at dette systemet skal utvikles for store enorme mengder data vil denne egenskapen være mer fremtredende. Tabell med egenskaper finnes på neste side.

Egenskaper	Mulige satsninger	Forslag for litteraturanbefaling
<b>Brukerpreferanse</b>	Kan evalueres underveis med brukerundersøkelser.	Ikke viktig i oppstarten. Kan foreta brukerundersøkelser etter at anbefalingssystemet er implementert og fungerer.
<b>Prediksjonsnøyaktighet</b>	Fullstendighet, presisjon og spesifisitet. Kan forutsi vurderinger, forutsi bruk og forutsi rangering av gjenstander.	Viktig. Bør vektlegges fordi det er et mål på "korrekte" anbefalinger. Bør foretas som off-line eksperiment når man prøver ut algoritmer.
<b>Dekning</b>	Dekning på bruker og gjenstandssiden er viktig. Mange ressurser "låner" data fra andre kilder. Det kan hende at datasett må bearbejdes. Det må ikke ta for lang tid for brukere å laste siden.	Viktig. Må jobbe med utgavekopier. Avgjørende ved valg av anbefalingsmetode. Cold start-problemet må vurderes i forhold til anbefalingsmetode. Det er viktig at det ikke tar lang tid for bruker å laste anbefalingene. I ressursen er det mulig at det bør skaleres grunnet skjevhet i vurderinger gitt.
<b>Tillit</b>	Vinne brukers tillit med forklaringer. Vinne brukers tillit ved å gi anbefalinger brukeren har kjennskap til.	Viktig. Det bør foreligge en forklaring på hvordan anbefalingene blir generert som brukeren har forutsetning for å forstå.
<b>Selvtillit</b>	Hvor stor tro anbefalingssystemet har på egne anbefalinger. Valg av grense for anbefaling og brukere kan opplyses om dette.	Ikke viktig på dette tidspunktet. Resultatlisten med anbefalinger bør rangeres etter hvilke verk som er mest relevante.
<b>Nyhet</b>	Vekte inn "ukjente" gjenstander.	Ikke viktig siden det dekkes av serendipitets-prinsippet.
<b>Serendipitet</b>	Muligheten for brukeren å gjøre overraskende, men positive oppdagelser.	Viktig. Leser deles ofte inne de som liker å oppdage nye ting og de som liker å lese like ting. Serendipitet kan vektes inn eller er et resultat på at prediksjonsnøyaktigheten senkes.
<b>Mangfold</b>	Skal sikre at anbefalingene ikke blir for like. Det er en påstand om at for like anbefaling gjør at bruker blir lei.	Ikke viktig siden det dekkes av serendipitets-prinsippet.
<b>Nytteverdi</b>	Økonomisk vinning. Det brukeren sitter igjen med etter å ha fulgt en anbefaling. Det utvikler sitter igjen med etter å ha gitt gode anbefalinger.	Ikke viktig i denne sammenhengen siden det ikke selges noe og anbefalingssystemet er en ekstrafunksjon til ressursen.
<b>Risiko</b>	Hvor stor risiko det utgjør for brukeren å følge en anbefaling.	Ikke viktig i denne sammenhengen da bøker er billige varer og gratis på biblioteket. Men bør sjekkes etter hvert fordi falske anbefalinger kan gi misstro til systemet.
<b>Robusthet</b>	Anbefalingssystemets robusthet mot angrep.	Ikke aktuelt da det trolig ikke er noen økonomisk vinning å påvirke anbefalingene på sosiale nettsider
<b>Personvern</b>	Brukeropplysninger skal anonymiseres og sikres mot missbruk.	Viktig. Utvikler må velge et nivå på personopplysninger som ikke gjør at brukere vegrer seg for å bruke ressursen. Opplysninger skal sikres.
<b>Tilpassningsdyktighet</b>	Brukere forandrer mening. Nye gjenstander bli lagt til i samlingen.	Viktig. En bruker vil forandre lesesmak over tid, kan vekte nyere leseropplevelser høyere enn gamle. Det må være tilrettelagt for nye gjenstander i systemt.

Tabell 5: Egenskapene i et anbefalingssystem med til litteraturanbefalingssystemer.

### 5.3 Trinn tre: Potensielle anbefalingsmetoder

I dette trinnet tar jeg utgangspunkt i forskjellige anbefalingsmetoder og data som er tilgjengelig fra det sosiale nettstedet Bokelskere.no. Jeg bruker aspektene jeg har diskutert om bøker innenfor litteraturanbefalingsdomenet og egenskaper ved et anbefalingssystem. Jeg har valgt å dele trinn tre i to separate drøftinger med bakgrunn i forskjellige anbefalingsmetoder. Jeg har presentert innholdsbaserte metoder og kollaborative metoder som potensielle metoder for sosiale nettsteder og har brukt Bokelskere.no som min konkrete kilde for tilgjengelig data. Anbefalingsmetode utvalget er gjort fordi jeg mener at de to metodene er mest aktuell på bakgrunn av tidligere forskning på området, i forhold til dette bestemte anbefalingsdomenet og tilgjengelig data fra sosiale nettsteder. Den grundige gjennomgangen av forskjellige anbefalingsmetoder i kapittel 2 har gitt meg en god bakgrunn for valg av de metodene jeg presenterer her. Jeg vil underveis drøfte hvilke konsekvenser disse valgene får. Etter drøfting av fremgangsmåter innenfor de forskjellige anbefalingsmetodene oppsummerer jeg dette i en punktvis fremstilling.

Sosiale nettsteder gir fra seg tre typer data. Brukerdata, gjenstandsdata og data som skapes av transaksjoner. De beste metodene i den sammenhengen tar for seg minimum to av denne typen data for å generere anbefalinger. I denne forbindelsen er det kun demografiske metoder og kunnskapsbaserte metoder som ikke er spesielt aktuelle. Disse metodene brukes i hovedsak i andre sammenhenger enn sosiale nettsteder. Kontekstuelle metoder kan brukes for sosiale nettsteder men siden disse metodene er opptatt av brukerens kontekst blir denne metoden noe begrenset av at ressursen er nettbasert. Det kan tenkes at for et sosialt nettsted for filmer, ved en bestemmelse av hvor man skal gå på kino, at denne metoden tilfører noe vesentlig. For sosiale nettsteder mener jeg at innholdsbaserte metoder, kollaborative metoder, sosiale metoder og hybride metoder er mest aktuelle.

Innenfor de syv anbefalingsmetodene er det spesielt fire anbefalingsmetoder som utmerket seg. Innholdsbaserte metoder fordi bøker er en type gjenstand som har mye innholdsforklarende data knyttet til seg. Kollaborative metoder siden nettstedet har aktive brukere som kommenterer og vurderer gjenstandene. Jeg vurderte sosiale metoder fordi nettstedet generer mange typer transaksjoner mellom brukere. Jeg slo den siste metoden fra meg fordi jeg mener de mulighetene som finne på nettstedet ble godt dekket opp av kollaborative metoder. Det er i fremtiden en mulighet å se om sosiale metoder er et område det kan være verdt å revaluere. Hybride metoder bør ses på som en metode for å forbedre eller utfylle eventuelt valgte metoder.

Før denne metoden velges bør utvikler se nærmere på hva de to mest aktuelle metodene gjør, og eventuelt å evaluere hvilke konsekvenser bruken av dem er. De tre resterende metodene blir forbigått grunnet data de trengte eller målet ved anbefalingsmetoden.

### **5.3.1 Drøfting av litteraturanbefalinger med innholdsbaserte metoder**

Innholdsbaserte metoder er tradisjonelt sett mye brukt på tekstlige dokumenter og er en god måte å finne like gjenstander på, forutsatt at systemet som skal betjenes er utviklet med tilstrekkelig datadekning. Rene innholdsbaserte metoder benytter seg av fremgangsmåter innenfor informasjonsgjenfinning og lar gjenstandene i systemet være eneste kilde til likhet. Trenden innenfor anbefalingssystemer har etter hvert blitt slik at det i tillegg til informasjonsgjenfinningsteknikker også benytter seg av brukervurderinger i brukerprofiler. Uansett er det gjenstanden som ligger som utgangspunkt, og må analyseres. Bøker er tekstlige gjenstander og vil ha både attributter med fastsatte verdier og attributter med løpende tekst som kan analyseres, kalt ustrukturert data. Potensiell attributter for gjenstandene som skal anbefales på Bokelskere.no er: Tittel, Undertittel, Forfatter, Emneord, Språk, Sjanger, Serie, Litteraturtype, Kategori, Lesnivå, Omtalt person, Omtalt sted, omtalt tid, Omtale og Diskusjon. På andre nettsteder vil det finnes andre potensielle attributter på gjenstandene som skal anbefales. Det første trinnet vil være å velge hvilke attributter som skal være meningsbærende i en gjenstandsrepresentasjon.

Det andre trinnet ved en innholdsbasert anbefaling er å bestemme hvordan man skal måle likhet. Innholdsbaserte metoder baserer seg på likhet mellom gjenstander, i dette tilfellet bøker, men utvikler må bestemme hvilke likhetskriterier en gjenstand skal representeres med. Skal likheten måles mellom den gjenstanden som er aktuell i akkurat det bestemte øyeblikket eller skal det tas hensyn til historikk? En klar kilde til historikk vil i forbindelse med et sosialt nettsted som Bokelskere.no være brukerens boksamling. Dermed må valget tas mellom å basere anbefalinger på en aktuell bok som kan være den boken brukeren leser for øyeblikket, eller på en representasjon av brukerens boksamling.

Det tredje trinnet er innebærer at utvikler finner en måte å måle hva brukeren liker. Dette er brukerens smak i bøker og skal uttrykkes gjennom brukerprofilen. En utvikler kan velge å tro at brukeren liker en gjenstand fordi den er benyttet eller han kan bruke brukerens vurderinger av gjenstander. På et sosialt nettsted som Bokelskere.no må utvikler velge mellom å bruke en brukers boksamling eller vurderinger i form av terningkast. Dette gjenspeiler seg i hvordan

brukerprofilen blir bygget opp, og avgjør hva utvikler skal måle likhet til potensielle bøker med. Når de tre trinnene nevnt ovenfor er gjennomgått og utvikler har tatt de valgene som må til kan systemet generere anbefalinger og presentere disse på ønsket måte.

### **5.3.1.1 Gjenstandsrepresentasjon**

Utvelgelsen av attributtene knyttet til en bok er viktig for å få frem bokens innhold. Derfor vil det første trinnet være å bestemme attributter tilstede i gjenstandsrepresentasjonen utvikler bruker til å finne likhet mellom bøker. Det er viktig å huske at denne gjenstandsrepresentasjonen skal representere det boken handler om. Dette er ikke informasjonsgjenfinning hvor en bok skal finnes igjen ved å bruke bestemte søkekriterier. Lancaster (2003) poengterer at det er dette som skiller å lage en representasjon som går på å konkretisere et emne i et dokument og en representasjon av innhold i det han kaller et imaginert verk. Det er ikke selve verket men innholdet som er imaginert, altså fra fantasien. Dette kan gjøres ved å velge ut bestemte attributter og eventuelt gjennomføre en analyse av ustrukturerte data. Poenget med denne prosessen er å finne frem til gjenstandsrepresentasjoner som kan skille de forskjellige gjenstandene fra hverandre, men også samle de som er like nok til at de bør samles gjennom klustering.

Klustering er kort fortalt å se på gjenstandsrepresentasjonene i en samling og gruppere de som ligner på hverandre. Denne grupperingen kan skje gjennom å se på hvilke av gjenstandsrepresentasjonene, i denne sammenhengen bøker, som har flest like nøkkelord. Jeg vil komme nærmere tilbake til dette. For å konkretisere hva en bok handler om er det viktig at attributt og tekstanalyse gjøres på en tilfredsstillende måte. Her har utvikler to ulike fremgangsmåter å bygge opp en gjenstandsrepresentasjon. Utvikler kan velge å kaste de forskjellige verdiene fra alle attributtene inn i gjenstandsrepresentasjonen som en ”bag of words” eller analysere de forskjellige attributtene med forskjellige teknikker. I denne sammenhengen er bokrelaterte data for Bokelskere.no et godt eksempel på at nettstedet er en ressurs som har omfavnet den friheten internett bringer med seg. Bokelskere.no har noe data som er bearbeidet av profesjonelle kilder som Den norske bokdatabase og Amazon.com. På den andre siden har de noe bruker skapt data, og dataene brukes sammen uten en klar form for kontroll. Dette gjør at de forskjellige utgavekopiene av et verk representeres forskjellig. Jeg tror på bakgrunn av dette at ”bag of words” metoden blir for upresist. Før jeg diskuterer utfordringer og konsekvenser forbundet attributtene i gjenstandsrepresentasjonen vil jeg kort gå igjennom teknikker for å fjerne tvetydighet forbundet med dette arbeidet.

Stemming er å beholde, og noen ganger endre stammen av et ord på bakgrunn av regler som er tilpasset den bestemte ressursen, slik at bøker med ord som sykkel, sykling og sykler behandles likt. Stopp ord er ord som ikke gir en betydning for innholdet av en bok, som for eksempel preposisjoner og artikler. Det hjelper lite å klustre bøker etter ord som; er, og, en, for, ved og andre ikke meningsbærende or. Derfor fjerner stopp ordlister denne typen ord fra løpende tekst. Ved synonym kontroll samles ord som har samme betydning, men skrives ulikt og ved homonym kontroll skilles ord som skrives likt men har forskjellig betydning. En synonymkontroll kan utformes som et register og hvis et av emneordene i eksempelet brukes blir alle ordene hentet frem som potensielle treff. En bok som handler om båter og en bok som handler om skip handler om det samme og dette kan fikses med et synonymregister. Når en analyserer tekst fjernes i tillegg til ofte brukte ord gjennom stopp lister også ofte ord som sjelden brukes. Dette er mye brukt i fulltekst dokumenter, men det er usikkert om det er nyttig i forbindelse med tekstanalyser av korte tekster som sammendrag og omtaler da de inneholder få ord i seg selv. Jeg vil her drøfte de forskjellige attributtene, hva som må gjøres med dem og hvilke konsekvenser dette får for anbefalingssystemet.

### **Tittel**

Siden det i hovedsak er snakk om lesing i underholdningsøyemed, og skjønnlitteratur er sterkt representert, kan det diskuteres om attributtet Tittel vil være aktuelt. Titler for skjønnlitteratur trenger på ingen måte å fortelle noe om innholdet av boken. Selv med faglitteratur er det ikke nødvendigvis slik at tittelen sier noe om innholdet selv om det er mer trolig med denne typen litteratur. På den andre siden kan tittel være viktig når det er flere bøker som handler om samme person, som for eksempel Harry Potter. Her må en vurdering foretas om bruk av tittel vil være forstyrrende eller informativt. Hvis Tittel velges hvordan kan den tilføre gjenstandsrepresentasjonen informativ data? Jeg ser på dette med et eksempel *En dåre fri* er en tittel som kan gi mening som tekststreng hvis den analyseres semantisk. Da må dåre analyseres som substantivet i setningen og sammenlignes med andre ord som betyr det samme. I Ordnett sammenlignes dåre med ordet dumrian<sup>34</sup> som gir oss en massiv liste med synonymer for dumrian. Fri må analyseres som det å være fri, men ikke å fri. Dermed sier denne Tittelen at boken handler om en fri dumrian. Men gjør dette datasystemet som skal finne ”like” bøker noe klokere? Trolig vil ord fra bokomtalen som; psykoser, lidelser, psykiatrisk, omsorg og uforstand være bedre

---

<sup>34</sup> <http://www.ordnett.no/ordbok.html?search=dumrian&publications=3>



representasjoner om at denne boken handler om en kvinne som er psykisk syk. I andre tilfeller som for eksempel *Ascend* vil ikke tittelen si noe meningsbærende i det hele tatt. Selv hvis boken ble oversatt med *Å stige opp* vil det ikke gi noen indikasjon på hva den handler om. På bakgrunn av dette mener jeg at attributtet Tittel vil være et forstyrrende element i denne type gjenstandsrepresentasjon.

### **Undertittel**

Undertitler kan være mer informative enn hovedtitler for skjønnlitteratur, spesielt nå i nyere tid hvor det er populært å skrive bokserier. Boken fra forrige eksempel poengterer dette *Ascend: Trylle Trilogy, book 3* er bokens fulle tittel på Bokelskere.no. Undertittel gir mer informasjon. Ja, vi vet nå at dette er tredje bok i en serie og at den kanskje ikke bør anbefales hvis ikke leseren har lest de to første. Men sier den noe mer om hva boken handler om? Jeg mener at den ikke gjør det. Hvis Undertittelattributtet blir med i en gjenstandsrepresentasjon bør det være for å kartlegge denne typen informasjon, men ikke for å sammenlignes for likhet.

### **Serie**

Som en fortsettelse på forrige attributt må serietittel diskuteres. Noen serietitler blir på Bokelskere.no oppgitt som undertitler, mens noen serietitler blir oppgitt under attributtet Serie. Derfor kan det være aktuelt å bruke begge disse attributtene i en gjenstandsrepresentasjon. Å vite om en bok er en del av en serie er viktig i forhold til om utvikler skal anbefale flere bøker i denne serien. I et anbefalingssystem for litteratur er det trolig at brukeren er klar over at den aktuelle boken er del av en serie, og et anbefalingsresultat som inneholder mange oppfølgere innenfor samme serie kan virke støyende. Jeg foreslår at begge disse attributtene brukes som et trinn i sortering av anbefalingsresultatet men ikke for å sammenligne likhet i boken.

### **Litteraturtype**

Dette er et attributt som gir begrenset informasjon når det gjelder å finne likhet mellom dokumenter. Verdier innenfor dette attributtet er begrenset til skjønnlitteratur eller faglitteratur, dermed kan attributtet gi en mulighet til å skille mellom dette. Så må avgjørelsen tas om dette er et skille utvikler ønsker å foreta i anbefalingssystemet. Som diskutert i kapittel 2 mener jeg at dette er et valg utvikler ikke behøver å ta for brukeren. En bruker som har mange fagbøker i boksamlingen vil være interessert i å få anbefalt flere fagbøker. Bøker med attributtverdi fagbøker kan i dette tilfellet prioriteres. Det motsatte gjelder for brukere som har registrert mange skjønnlitterære bøker. Hvis en bruker har registrert bøker av begge litteraturtypene kan utvikler

se bort fra hele attributtet, eller vekte anbefalingene i forhold til den litteraturtypen som er sterkest representert. En annen mulighet er å se vekk fra hele attributtet å la innholdet i gjenstandsrepresentasjonen avgjøre resultatlisten uavhengig av litteraturtype. På Bokelskere.no er skjønnlitteratur sterkest representert noe som er i tråd med at dette er et nettsted for privatpersoner, jeg anser på grunn av dette at det ikke er nødvendig å benytte seg av litteraturtypeattributtet.

### **Kategori og sjanger**

Kategoriattributtet kan tenkes på som en sjangerantydning som representeres med verdier som, Ungdomsbøker 10 +, Tekstbøker, Krim / spenning, Utenlandsk, Kokebøker, Håndarbeid, Romaner og Fantasy / Science fiction. Denne kategoriseringen virker litt rotete på Bokelskere.no i og med at de ikke holder seg til ett nivå. Med nivå mener jeg hva verdien i attributtet gjenspeiler. Roman er en sjanger, men Krim / spenning er også en sjangerantydning og alle verk som er romaner er ikke merket Romaner. Verdier som Krim / spenning og Fantasy / Science fiction er derimot med på å skille verk på sjangernivå, med veletablerte sjangertyper. På noen få bøker dukker også attributtet Sjanger opp, men det virker som om dette oftest blir brukt i forbindelse med verdien Fantastisk litteratur. Fantastisk litteratur er en fornorsking av Fantasy begrepet, som på nettstedet blir brukt som mulig verdi i Kategoriattributtet. Hvis Sjanger og Kategori skal brukes bør det brukes en form for synonym og homonym kontroll.

### **Emneord**

Attributtet Emneord blir ofte brukt som kilde til å finne likhet mellom gjenstander, siden de representerer innholdet med forklarende ord utvalgt for det dokumentet. Emneord kan settes av forlaget, av utvikler av ressursen eller av brukerne. I forbindelse med Bokelskere.no sine Emneord, som er tagger, finnes det flere utfordringer. De er brukerskapt og er derfor utenfor utviklerens kontroll. Dette er godt for kreativitet og brukerengasjement, men som diskutert i kapittel 2 så gir dette noen utfordringer. Hvordan kan man hente ut informasjon om en boks innhold når mennesker bruker forskjellige ord for å forklare innholdet?

Det finnes som nevnt over 13.000 emneord registret på nettstedet både enkle ord og tekststreng. En kan argumentere for at emneordene; Broderi, Broderi og billedveving, Brodering vil betegne det samme innholdet i en bok. Derfor bør gjenstandsrepresentasjonen inneholde en stemmet versjon av ordet Broderi, og tekststrengen Broderi og billedveving deles opp og Billedveving stemmes fordi det finnes et emneord Billedvev. Et annet alternativ er å sette opp

synonymordslister og hvis et en bok har emneordet Broderi vil det sammenlignes med bøker hvor ordet Brodering er brukt. Et tredje alternativ for å forsøke å bringe kontroll over emneord er kontrollert tagging. Utvikler setter på forhånd opp eksempler på hva som er tillat og tidligere brukt, disse vil komme opp som forslag når bruker taster inn potensielle emneord. Siden dette kan være dempende på kreativiteten og friheten en bruker har vil jeg ikke anbefale det. Jeg vil i denne situasjonen utarbeide en form for synonymkontroll for å bruke den i gjenstandsrepresentasjonen.

For de tre forrige attributtene Kategori, Sjanger og Emneord er det viktig å få frem meningen i av den verdien attributtene uttrykker. Det er en mulighet å utvikle en ontologi for Bokelskere.no. Göker & Davies (2009) sier at en ontologi er en formal kunnskapsrepresentasjon som viser en modell av et utdrag av verden. En ontologi kan brukes som et semantisk rammeverk og siden de bygges opp av et formalt språk kan de brukes til å forbedre meninger som ikke er direkte uttrykt i dataene. Siden en ontologi ikke er tilgjengelig for øyeblikket er det mulig å bruke et annet alternativ for å få frem meningen i begrepene brukt nemlig en leksikalsk database. Det finnes flere leksikalske databaser men en av de store og godt brukte er WordNet<sup>35</sup> som tar sikte på å forklare ord og uttrykk på et generelt nivå i det engelske språket. Der støter jeg på nok et problem i oppgaven, språklige utfordringer i utgavekopiene som jeg vil komme tilbake i slutten av dette avsnittet. Hvis Bokelskere.no var en internasjonal ressurs basert på engelsk språk ville jeg dratt nytte av WordNet med rundt 150.000 termer som er ordnet i synset, altså synonymsett (Princeton University 2006). Da norsk ikke er representert på WordNet vil ikke denne leksikalske databasen hjelpe meg i denne situasjonen. Et samarbeid med den norske språk- og ordboktjenesten Ordnett<sup>36</sup> kan være en mulighet.

### **Forfatter**

Dette attributtet vil være aktuelt da brukere ofte er interessert i flere bøker av samme forfatter. Jeg har sagt at en gjenstandsrepresentasjon skal si noe om innholdet i en bok, det vil ikke Forfatterattributtet direkte gjøre. Her vil jeg bruke Saricks teori om at forfatteren står for de subtile egenskapene som trekker en bruker til en bok. Man kan ikke si at en bruker som liker en forfatter vil like alt denne forfatteren har skrevet, men ofte kan man si at forfattere vil ha en spesiell skrivestil som brukeren liker eller misliker. Det er også mange forfattere som gjennom

---

<sup>35</sup> <http://wordnetweb.princeton.edu/perl/webwn>

<sup>36</sup> <http://ordnett.no/ordbok.html>

sitt forfatterskap holder seg til en type sjanger. Derfor bør Forfatter være en del av gjenstandsrepresentasjonen.

### **Omtale og Diskusjon**

Disse to attributtene er ustrukturerte data, derfor vil det være nødvendig med en tekstanalyse av disse. Når disse dataene er strukturert kan de brukes og likhet mellom verkene kan måles med for eksempel cosinus-likhet. I forbindelse med tekstanalysen bør teknikker som stemming, bruk av stopp ord, synonymkontroll og homonymkontroll benyttes.

Når det gjelder å bruke Diskusjonsattributtet i en analyse for likhet vil den skille seg fra analysen av Omtale. Omtale er standardisert fra forlaget og/eller Den norske Bokdatabasen, mens Diskusjoner finnes det ingen regler for. Det finnes en maksimumslengde for selve diskusjonsposten, men den er så stor så at de fleste som skriver innlegg ikke vil merke denne. Diskusjonene er brukerprodusert og har ingen overordnede regler for hvordan de skal bygges opp. Diskusjoner knyttet til en utgavekopi kan være en kort kommentar fra brukeren om hvorfor han likte boken, hvorfor han mislikte boken eller hvorfor boken bør leses. Det kan også være lange anmeldelser med oppsummering av handling og sitater fra teksten. Noen brukere bruker Diskusjoner til å komme med tankerekker om hva boken får de til å tenke på, eller de mimrer over den første gangen de leste boken. Det er derfor vanskelig å forsikre seg om at Diskusjoner inneholder informasjon som gjør at verket kan sammenlignes med andre verk. Jeg anser dette attributtet som usikkert til å bygge anbefalinger på. Det kan tenkes at Diskusjon kan brukes som en representant for personlige meninger og interaksjon, dette vil jeg komme tilbake til under drøftingen av kollaborative metoder.

Omtaleattributtet er korte omtaler av boken, noen ganger inneholder den både omtale fra forlaget og fra Den Norske Bokdatabasen, men noen ganger bare en av dem. Dette gjør at omtaleattributtet kan bestå av omtaler som strekker seg fra fem linjer til en halv side (jmf. Vedlegg 10). Med korte tekster er det viktig å ta vare på alle ord som kan være meningsbærende, derfor foreslår jeg å bruke stopp ordliste for ofte brukte ord. I fulltekst søk er også vanlig å fjerne sjelden brukte ord, slik at man bare sitter igjen med ord i mellomstikket. Dette vil jeg ikke anbefale i denne sammenhengen nettopp fordi disse tekstene er korte vil de ordene som kun finnes en eller to ganger i teksten være viktige.

## **Språk**

Tittel og undertittel av en bok vil bli representert i det språket boken er skrevet på. Dette vil kun være et problem for de utgavekopiene som ikke er lenket sammen. Jeg har tidligere valgt vekk Tittel i forhold til gjenstandsrepresentasjonen, men Undertittel vil i forhold til ikke norske bøker kun gi treff på undertitler av samme språk. Derfor kan det argumenteres for at disse ikke skal brukes hvis Språkattributtet er satt til noe annet enn norsk. Det er også slik at noen emneord er mer populært på engelsk og dette gir enda en grunn til å innføre et synonymregister for denne typen attributter.

## **Utgavekopiproblematikken**

En stor utfordring i forbindelse med Bokelskere.no er utgavekopier, at det samme verket har flere utgaver som ikke kobles sammen. Siden Bokelskere.no henter bokdata til en utgavekopi fra flere kilder vil mengden bokdata variere deretter. Dette gjør at sammenligningsgrunnlaget blir mer usikkert, og det kommer tydelig frem i forskjellen mellom den norske og den engelske versjonen av *Shantaram*. Den norske versjonen er utstyrt med Omtale fra både Den norske Bokdatabasen og fra forlaget, det er ikke den engelske. Likhet ved bruk av Omtale og Diskusjoner kan derfor ikke være en komplett løsning for Bokelskere.no før de har løst utgavekopi problematikken. Den engelske utgaven av *Shantaram* er utstyrt med emneordene India, Selvbiografisk, Skjønnlitteratur (engelsk) og Slum. Den norske utgaven har emneordene Bombay, Dokumentar-roman, India, Mafia og Outsidere. Av åtte mulige Emneord har de to utgavekopiene kun et Emneord felles. Kobling av utgavekopier ville dermed løst en del usikkermomenter ved arbeidet med attributter. Nå er det slik at jeg er klar over denne utfordringen og må prøve å arbeide rundt den.

## **Gjenstandsrepresentasjon på Bokelskere.no**

For Bokelskere.no mener jeg en gjenstandsrepresentasjon bør bestå av Forfatter, Undertittel, Serie, Sjanger, Kategori, Emneord, Omtalt sted, Omtalt person og Omtale. De forskjellige attributtene vil som diskutert ovenfor måtte behandles forskjellig for å få frem hva som er meningsbærende i forhold til bokens innhold.

### 5.3.1.2 Bok kontra boksamling

Hvis en bok alene skal ligge til grunn for likhetsvurderingen må valg av bok vurderes nøye. I denne sammenhengen er boken den aktive brukeren leser for øyeblikket eller en bok som nettopp er ferdiglest. En annen mulighet er å la den aktive brukeren velge en hvilken som helst bok i den kollektive samlingen av bøker og basere anbefalingen på dette. Personaliseringsaspektet ved denne fremgangsmåten er at brukeren får anbefalinger basert på en bok han har signalisert at han liker. Hvis en bok skal brukes vil jeg som utvikler velge å basere anbefalinger gjort på bakgrunn av en bok, på den sist leste boken for en leser. Da vet jeg at brukeren har likt boken godt nok til å fullføre boken. På Bokelskere.no kan dette settes av brukeren ved å endre status på boken til *har lest*.

Det andre alternativet er å basere likhetsvurderingen på boksamlingen. Velger utvikler å basere likhetsvurderingen på bakgrunn av brukers boksamling er det flere faktorer som må tas stilling til deriblant tidsaspektet og hvordan bøkene i boksamlingen bør klustres for å kunne angi hvilke attributter som skal vektlegges. Her vil den aktive brukerens historikk, altså boksamlingen han har bygget seg opp være kilden til informasjon. Det finnes fordeler og ulemper ved begge metodene, men før jeg diskuterer de vil jeg gå kort igjennom hvordan en bok og en boksamling kan representeres.

Velger utvikler en brukers boksamling som likhetsgrunnlag blir jobben noe utvidet. Først vil prosessen med utvelgelse av gjenstandsrepresentasjon gjennomgås, deretter må boksamlingen som en helhet sees på. Boksamlingen har to viktige faktorer knyttet til seg, tidsaspektet og hvordan gjenstandsrepresentasjonene for bøkene klustres. En boksamling vil hos de fleste brukere øke i omfang etter hvert som tiden går, og bøker som blir lagt til eller lest vil si noe om brukerens smak. En lesers smak er ofte under utvikling og dermed er det ikke sikkert at alle bøkene i boksamlingen er like gode representanter for leserens smak ved det aktuelle øyeblikket. På Bokelskere.no kan brukeren legge inn dato for når en bok er lest ferdig. Det kan legges inn en vektfunksjon som vokter bøker med dato nærmest det aktuelle tidspunktet for anbefaling høyere enn bøker med eldre dato. Samtidig kan bøker med tilstandsverdi *leser* vektet like høyt som de bøkene som var sist lest, siden dette forteller hva leseren leser på det aktuelle tidspunktet. Det kan diskuteres om bøker i boksamlingen som ikke har blitt lest enda skal være med som grunnlag for anbefaling. Bøker med tilstanden *ønsker deg* og *skal lese* indikerer at dette er bøker brukeren er interessert i, men vi vet ikke om de faller i smak eller ikke. Brukeren har helt tydelig plukket ut

de bøkene med en plan om å lese de og det er på grunn av dette noe ved de bøkene som tiltaler en bruker.

Når det kommer til smak og boksamling så finnes det ikke en garanti for at bøkene som finnes i boksamlingen er bøker som brukeren har likt. Der kommer Spalding sin teori om at brukere av slike typer nettsider legger inn bøker de liker inn i bildet. Som nevnt i kapittel 2 ligger gjennomsnittlig vurdering på LibraryThing på 3,8 av 5. En mulig fremgangsmåte for vekting er å vektlegge det som leses i det aktuelle øyeblikket og det som har blitt lest de siste tre månedene sammen. Deretter å sette en grense for det siste halve året, det siste året og en ny grense år etter år bakover i tid. En utvikler må også klustre bøkene i en boksamling på en måte som gjør at det går an å lese seg frem til brukerens lesesmak gjennom dette. Ved å klustre like verk for deretter å se hvilke klustre som er størst vil jeg som utvikler kunne bruke dette for å se etter denne type verk som potensielle anbefalinger.

Fordeler og ulemper ved valg av likhetskriterier er viktige å ha tenkt igjennom. Ved å bruke en bestemt bok som likhetskriterium, for eksempel sist likte bok vil en bruker få anbefalt bøker som ligner på kun den boken. Dette kan føre til overspesialisering og dette bør rettes opp ved å tilføre serendipitet. Hvis boksamlingen brukes som likhetskriterium vil dette gi et større spekter av bøker for å finne lignende bøker. Dermed vil trolig de potensielle anbefalingene variere mer i forhold til innhold. Denne måten tar også hensyn til hvordan en brukers smak endrer seg over tid. På den andre siden igjen så får anbefalinger basert på en bestemt bok frem hva brukeren er interessert i akkurat det øyeblikket. Fordelen ved dette er i situasjoner hvor brukeren vil ha mer av det samme, men ulemper er hvis brukeren er i en situasjon hvor han vil finne noe nytt.

### **5.3.1.3 Brukerprofil**

For å anbefale bøker en bruker kommer til å like må utvikler vite hvilke bøker det er brukeren allerede har likt. Jeg har allerede gått grundig igjennom anbefalinger basert på brukernes tidligere leste bøker og ren innholdsbasert anbefaling. Flere systemer som bruker innhold mellom gjenstander benytter også brukers vurderinger. I stedet for å basere seg på en analyse av brukerens boksamling kan en brukers vurdering av bøker benyttes for å avgjøre hva brukeren liker. Da trenger utvikler eksplisitte data om dette. På Bokelskere.no har brukeren mulighet til å legge inn terningkast på de forskjellige bøkene de har lest. Dette er en skalert vurdering med verdier 1-6. I tillegg har Bokelskere.no en funksjon hvor man kan sette et verk som favoritt. Det

er ikke nødvendigvis et verk de har gitt terningkast 6, derfor kan det vurderes om det skal gis en høyere verdi til disse bøkene som for eksempel 6 eller 7. Det kan være et verk som brukeren ikke syntes er ”kvalifisert” til terningkast 6 på grunn av skjønnlitterært innhold, men som brukeren på tross av dette elsker å lese. Det kan vurderes om bøker som er merket med favoritt uansett skal vurderes til terningkast 6. Denne prosessen kan normaliseres for å ta hensyn til om brukeren er snill eller streng med terningkastene.

Cold start-problemet for nye brukere er et velkjent problem innenfor innholdsbaserte anbefalingsmetoder og kan løses ved å gi en bruker ”falske” vurderinger. Hvis en bruker har færre enn 10 bøker i samlingen eller har vurdert mindre enn for eksempel 10 bøker kan systemet trekke ut 10 bøker som har fått høye vurderinger av mange brukere. Bokelskere.no har data om gjennomsnittlig vurdering, antall brukere som har registrert verket og hvor mange som har den som favoritt. På siden som heter Populære bøker finner man de bøkene som er representert i flest boksamlinger. På siden som heter Favorittbøker finner man de bøkene som brukerne har gitt favorittstatus, det verket som er favoritter hos flest brukere står øverst. I tillegg finnes siden Høy favorittprosent.

Dette er en justert liste over bøkene som er markert som favoritter av bokelskere. Den er sortert etter hvor mange prosent av eierne som har markert den som favoritt. Kun verk der flere enn 25 bokelskere har markert det som favoritt er med (Bokelskere.no ).

Dermed finnes det flere kilder som kan brukes for å overkomme cold start-problemet som nye brukere rammes av. Jeg vil diskutere bruk av vurderinger grundig i avsnittet om kollaborative metoder, men det er viktig å vite at dette er en mulighet. Jeg har trukket frem vurderinger fordi det er en mulighet å bruke de sammen med innholdsbaserte metoder, og jeg ønsker å peke på de mulighetene som finnes. Jeg vil i min fremstilling av innholdsbaserte metoder ikke benytte meg av vurderinger, men går for en ren innholdsbasert fremgangsmåte.

#### **5.3.1.4 Sammenligning og rangering**

Innholdsbaserte anbefalingsmetoder generer anbefaling på bakgrunn av et likhetskriterium mellom verkene en bruker har lest og likt, og potensielle anbefalte verk. Det er på dette stadiet i prosessen selve sammenligningen foretas. Et nytt verk som blir lagt til på nettstedet vil bli analysert for deretter å sammenligne gjenstandsrepresentasjonen med en brukers profil for å se om det er tilstrekkelig likt den typen verk brukeren liker at det blir anbefalt. Samtidig er det



viktig å huske at en bruker vil forandre mening over tid, slik at man gir mulighet til å forandre brukerprofilen og ”malen” for hva en bruker liker etter hvert. I denne delen av prosessen må tiltak som går direkte på utfordringer ved innholdsbaserte metoder iverksettes. Her finnes det forskjellige algoritmer for å bestemme likhet og en utvikler må velge den ut i fra hvilke data han velger å benytte seg av. Det er viktig for utvikler å tenke på de problemområdene som finnes med innholdsbaserte metoder og kompensere for disse.

Innholdsanalysen er veldig viktig siden den skal gi utvikler en mulighet til å skille ulike bøker, og sammenligne like bøker. Hvis ikke innholdsanalysen gjennomfører denne oppgaven på en tilfredsstillende vil det gå utover prediksjonsnøyaktigheten til anbefalingssystemet. Tidligere nevnte operasjoner som stemming, fjerning av stoppord, synonym- og homonymkontroll er viktig. Det kan også tenkes at det ville være til stor hjelp å utvikle en domeneontologi, eller bruke et leksikalsk oppslagsverk som Wordnet for å hjelpe til med å lage gjenstandsrepresentasjoner.

Overspesialisering grunnet likt innhold kan bekjempes med å innføre serendipitet i anbefalingen. Dette er et stort problem for innholdsbaserte anbefalingsmetoder og sett i sammenheng med litteraturdomenet har jeg vurdert dette til en viktig egenskap. Lesing av litteratur er en selvstendig aktivitet og det finnes ikke noe fasitsvar på hva som vil falle i smak hos en bruker. Det som er vanskelig å vite er hva slags type lesere brukerne er. Den typen leser som har funnet en sjanger, skrivestil og nivå de liker og vil holde seg innenfor denne. Eller den typen leser som stadig er på jakt etter en ny opplevelse og som gjerne hopper mellom sjanger, skrivestil og nivå. Derimot finnes det flere eksempler på at hvis gjenstandene som anbefales blir for like over tid blir brukerne lei. Basert på forskning på området, meninger og observasjoner fra Bokelskere.no vil jeg ved bruk av innholdsbaserte metoder tilføre anbefalingsresultatet serendippe funn.

Når det gjelder nye brukere vil det være nødvendig å hjelpe de litt på vei i starten før de har bygget opp et tilstrekkelig grunnlag med likte bøker i egen boksamling. De overnevnte listene med populære bøker, eller bøker som leses for øyeblikket vil være aktuelle å bruke i den sammenhengen. En utvikler kan også be brukerne om å vurdere et utvalg av bøker med en standardmelding om at dette må gjøres for å generere gode anbefalinger. Et sammendrag av hovedtrinnene og valgmulighetene for innholdsbaserte metoder vises i punktvis fremstilling.

## Punktvis fremstilling av innholdsbasert metode:

- **Gjenstandsrepresentasjon.**
  - Valg av attributter som representerer innholdet i gjenstanden.
    - Innholdsanalysen.
      - ◇ Stemming, stopp ordliste, synonymkontroll, homonymkontroll.
  
- **Brukerprofil.**
  - Basert på en bok.
    - Bok til bok sammenligning går direkte på innholdet, men ekskludere personaliseringsaspektet ved anbefalingen.
  - Basert på boksamling.
    - Utvikler må bruke god tid på å klustre boksamlingene slik at det kan lese klare trender om hva brukeren liker.
    - Anbefalingene blir personlige fordi hele brukerens boksamling vil representere brukerens smak.
  - Basert på terningkast.
    - Utvikler må normalisere terningkast for å ta høyde for personlige vurderinger.
    - Utvikler må vurdere om det er gitt tilstrekkelig terningkast så dette vil gi et korrekt bilde av brukerne.
  
- **Sammenligningen.**
  - Gjenstandsrepresentasjon og bruker sammenlignes med algoritmer.
  - Hjelpeteknikker.
    - Serendipitet.
      - ◇ Introduksjon av ”ukjente” forfattere.
      - ◇ Senke prediksjonsnøyaktigheten.
    - Datadekning.
      - ◇ Valg av likhetskilde må være tilstrekkelig.
      - ◇ Ontologi eller leksikalsk oppslagsverk.

- Nye brukere må ha hjelp.
  - ◊ Innføring av ”falsk” vurdering på bakgrunn av hva brukere i fellesskap liker.
- Resultat.
  - Bok for bok.
    - ◊ Du har lest X, det er derfor trolig at du liker Y som ligner på X.
  - Rangert liste hvor det som kommer høyest på lista er mest anbefalt.
    - ◊ Resultat for sammenligning av boksamling kan forklares ved å si du har tidligere lest X og Y, det er derfor trolig at du liker Z som ligner på X og Y.
    - ◊ Resultat for sammenligning med terningkast kan forklares ved å si at brukere som har lest det samme som deg liker X det er derfor trolig at du også gjør det.

### **5.3.2 Drøfting av litteraturanbefalinger med kollaborative metoder**

Kollaborative anbefalingsmetoder oppstod for å dekke et behov for kvalitet i anbefalinger og fungerer godt i sosiale ressurser hvor brukerne gir tilbakemeldinger på gjenstandene. Det finnes forskjellige fremgangsmåter innenfor kollaborative metoder en utvikler må velge om han skal benytte seg av, deriblant velge mellom brukerbaserte eller gjenstandsbaserte metoder.

Jeg var kort inne på utfordringer forbundet med datasettet utvikler har til rådighet tidligere i kapittelet. Desrosiers & Karypis (2011) legger frem fem konkrete egenskaper som utvikler bør vurdere når han velger mellom brukerbaserte og gjenstandsbaserte metoder:

Prediksjonsnøyaktighet, systemets effektivitet, stabilitet, rettferdiggjøring og serendipitet.

Prediksjonsnøyaktighet er som kjent hvor nøyaktig et anbefalingssystem er til å forutsi hvilke anbefalinger det skal presentere for en bruker. I tilfeller hvor antall brukere er høyere enn antall gjenstander mener forfatterne at gjenstandsbaserte metoder er å foretrekke, mens ressurser som har flere gjenstander enn brukere drar større nytte av brukerbaserte metoder. Systemets effektivitet avhenger av forholdet mellom brukere og gjenstander. I et system hvor der er flere brukere enn gjenstander mener forfatterne at å bruke en gjenstandsbasert metode vil spare tid.

Stabilitet kan sammenlignes med egenskapen tilpassningsdyktighet og Desrosiers & Karypis (2011) mener at hvis ressursen har et fast antall, eller små forandringer på gjenstandssiden vil gjenstandsbaserte metoder være å foretrekke. Ressurser hvor gjenstandskatalogen er i stadig forandring vil derimot brukerbaserte metoder være mer stabile. Rettferdiggjøring vil falle under diskusjonen om tillit. Det er lettere å forklare en bruker at han får en anbefaling på grunn av tidligere vurderinger enn at han får de fordi andre brukere liker de gjenstandene. Brukerbaserte metoder tilfører ofte anbefalingene mer serendipitet enn gjenstandsbaserte metoder. Dette er fordi ved gjenstandsbaserte metoder vil gjenstandenes likhet ligge til grunn for anbefalingen, men med brukerbaserte metoder kommer brukeres smak bedre frem.

På Bokelskere.no finnes det over 115.000 utgavekopier og over 12.000 brukere, og både antall utgavekopier og antall brukere øker. Dette betyr at sannsynligheten for at noen bøker har et veldig lavt antall vurderinger er høy. Samtidig er det slik at på et nettsted for litteraturinteresserte vil man alltid finne brukere som er interessert i det som kalles smal litteratur. Dette er litteratur som ikke er omtalt mye i media og er mindre representert enn bøker som klassikere og bestselgere. Det finnes heller ikke et fasitsvar på om brukere ønsker at anbefalingene er serendippe eller ikke. Dermed må utvikler ta hensyn til disse forholdene når han velger anbefalingsmetode og i denne sammenhengen mener jeg utvikler bør velge brukerbasert anbefaling.

Jeg har i denne oppgaven valgt å konsentrere meg om å bruke en fremgangsmåte som på norsk kan kalles brukerbasert nærmeste nabo. Dette har jeg gjort fordi det sosiale nettstedet tilbyr flere måter å måle likhet mellom brukere på, som vil bli diskutert omfattende. Desrosiers & Karypis (2011) formaliserer nærmeste nabo metoden ved å si at den aktive brukeren vil komme til å vurdere en ukjent gjenstand med en vurderingsgrad som ligner på den vurderingsgraden en annen bruker har benyttet seg av, hvis den aktive brukeren og den andre brukeren tidligere har vurdert andre gjenstander på lignende vis. Det er viktig å ha tungen rett i munnen for å få med seg hva som menes med forrige setning. For å bryte ned det de sier vil jeg strekke forklaringen over flere setninger. Et anbefalingssystem kan forutse hvilken vurdering den aktive brukeren kommer til å sette på en uvurdert bok. Dette klarer anbefalingssystemet ved å se på hva den aktive brukeren og andre brukere har gjort tidligere. Hvis den aktive brukeren og den brukeren han sammenlignes med har vurdert de samme bøkene tidligere, og har vurdert de med ganske like

vurderingsgrad er de som brukere like nok til kalle de naboer. Dermed kan vi lese ut i fra hva naboene til den aktive brukeren har gjort om den aktive brukeren vil like den uvurderte boken eller ikke. Det som er viktig i denne sammenhengen er hva som brukes som likhetskriterium og hvem som brukes som naboer.

Potensielle kilder til kollaborative data på Bokelskere.no vil være data om brukere og transaksjoner: Alder, Kjønn, Sted, Navn, Stjerner gitt, Diskusjoner, Antall bøker, Favorittbøker, Utgaver av verk, Emneord, Bokhyller, Gjennomsnittskarakter gitt, Favorittsitat, Følger, Følges av, Boktilstand og Dato lest. Bruker-gjenstand forhold er; har eierforhold (tilstand), har gitt (emneord) (terningkast), har bøker av, har bokhyller. Gjenstand-gjenstand forhold er; samme forfatter, samme emneord, finnes i samme boksamling. Bruker-bruker forhold er; bruker følger og bruker følges. På andre nettsteder kan det finnes andre potensielle datakilder.

Det første trinnet når man skal foreta kollaborative anbefalinger er å bestemme hvordan man skal måle likhet mellom brukere. Ofte måles dette i form av en vurdering brukeren gir fra seg. Denne metoden er dermed avhengig av interaksjon med brukere, noe som ligger i definisjonen av å være et sosialt nettsted og man kan her se på de sosiale tjenestene nettsteder tilbyr. Tilbakemeldingene kan komme i form av implisitte data hvor det observeres hvordan brukeren interagerer med ressursen eller fra eksplisitt data som brukeren selv legger inn.

Kollaborative metoder for litteraturanbefaling kan oppsummeres i tre trinn. Først må utvikler definere likhetsmål for å sammenligne den aktive brukerens mening og smak om bøker, med andre brukeres smak. Det andre trinnet er å forutsi hva den aktive brukeren kommer til å like av de bøkene han ikke har kjennskap til, bøker han ikke har lest tidligere basert på hva andre brukere syntes om bøkene. Det siste trinnet er å velge ut og presentere en rangert liste med anbefalte bøker for den aktive brukeren, hvor den boken brukeren trolig liker mest rangeres først på listen. I virkeligheten er det mange valg å ta før en liste med anbefalinger blir presentert for brukeren. Jeg vil her gå dypere inn på faktiske muligheter med den type kollaborative data jeg har med Bokelskere.no som utgangspunkt.

### **5.3.2.1 Likhetsmål, hva brukerne mener**

Kollaborative metoder går ut på å høste menneskers samlede meninger og handlingsmønster for å bruke det i anbefalingssammenheng. Den mest brukte måten er å samle en brukers mening gjennom vurderinger. Det at kollaborative metoder ofte benytter seg av brukerskapte vurderinger som terningkast eller stjerner betyr ikke at de er begrenset det. Innenfor kollaborative metoder

finnes det også sosiale data å ta hensyn til. Dermed grenser denne metoden tett inntil sosiale anbefalingsmetoder som poengtert i kapittel 1. Jeg vil senere i avsnittet gå inn på alternative kilder til en brukers mening, men starter med å se på vurderinger på Bokelskere.no. Vurderinger deles som nevnt i tre kategorier Liker/likes ikke, Skalerte vurderinger og Tekstelige kommentarer (Alag (2009; Herlocker et al. 2004; Lopes et al. 2011; Schafer et al. 2007).

### **Liker/likes ikke**

I forbindelse med sosiale nettsteder kan en boksamling være en kilde for utvikler å bestemme om brukeren liker en bestemt bok. Spalding sin teori, som diskutert i kapittel 3, sier at en bruker i hovedsak eier de bøkene han ”fortjener” altså bøker han kjøper selv eller anskaffer for eget mål. Videre sier han at disse bøkene kjøpes på kriterier at en bruker tror han vil like dem, Spalding ser vekk fra bøker som er mottatte gaver, da han mener de utgjør en liten del av et menneskes boksamling. Hvis Spaldings teori holder mål, er boksamlingen en god kilde til bøker brukeren liker. Nesse (2011, mai) sier at gjennomsnittlig terningkast på Bokelskere.no er over 5, på de bøker det er gitt terningkast på. Dette kan sees på som en støtte av Spalding sin teori og en brukers boksamling fortjener på bakgrunn av dette å blir vurdert som likhetskilde.

De bøkene som tilhører en brukers boksamling står da for bøker en bruker liker og utvikler kan sammenligne boksamlinger. Hvilke andre boksamlinger inneholder mange av de bøkene den aktive boksamlingen inneholder og hvilke andre bøker i tillegg inneholder de andre boksamlingene? Da blir boksamlingene uttrykket for en brukers smak og en bruker blir sammenlignet med andre brukere som har samme type mønster i sin boksamling. Mønster i en boksamling kan være flere ting. Det kan være bøker fra samme forfatter, sjanger eller emneord. Det kan også være at man kan lese et bestemt mønster i de forskjellige boksamlingene. Når en bruker legger til en bok i sin samling, hvilke brukere følger opp med å legge til de samme bøkene? Finnes det en rekkefølge for når brukere leser bestemte bøker annet en kronologisk rekkefølge basert på utgivelse? Finnes det grupperinger av sjangre hvis man ser kronologisk på leste bøker? Denne typen mønster mellom brukere og boksamling kan være spennende å se på. Jeg mener at en boksamling kan representere bøker brukere liker hvis det ikke finnes annen konkret informasjon for dette. Finnes derimot eksplisitte vurderinger er det mindre rom for at utvikler tolker disse feil. Hvis en bruker gir en bok terningkast fem er det ikke tvil om at han liker den, men det kommer jeg tilbake til.

## **Tekstlige kommentarer**

Tekstlige kommentarer som Diskusjoner vil på Bokelskere.no representere brukerne smak på en veldig personlig måte. Sosiale nettsider legger ofte til rette for at brukeren skriver anmeldelser og starter diskusjoner om bøker. Hvis et attributt som Diskusjoner skal tas i bruk som en vurdering må den igjennom en tekstlig analyse tilpasset nettopp dette attributtet og det må kjøres et program som kan tolke om Diskusjonen er positiv eller negativ. Dette betyr en semantisk analyse som jobber med blant annet ordnærhet for å kunne tolke fortelle forskjellen på: ”Dette er en superflott bok, jeg anbefaler alle å lese den.” og ”Dette er ikke en god bok. Jeg vil på ingen måte anbefale den.” Semantiske programmer som kan tolke setninger og meninger kan utarbeides men det er en krevende prosess, og krever ofte at systemet prøves ut på tekster fra den aktuelle ressursen før programmet har lært å trekke ut riktig mening. Når dette er gjennomført på en tilfredsstillende måte kan Diskusjoner brukes til å finne andre bokelskere som er positive og negative til bøker i den aktive brukerens boksamling.

Det har vist seg at en god del av brukerne på Bokelskere.no er restriktive med å sette terningkast til bøker i sin boksamling. Enkelte brukere bryr seg rett og slett ikke med denne muligheten. Mange syntes at det er vanskelig å forholde seg til terningkast. De er usikre på hvordan de skal kaste terningen og de liker ikke tanken på å skulle favne alt en bok inneholder med et eneste terningkast. Noen brukere etterlyser en skalering som gir flere vurderingsgrader, som for eksempel 1-10 stjerners vurdering. Summen av dette utgjør at kun en av fire brukere på Bokelskere.no gir terningkast. Mange som ikke gir terningkast er derimot inne og skriver bokomtaler eller diskuterer de forskjellige bøkene. Dermed gir de sine vurderinger i form av tekstlige kommentarer. Jeg mener det på Bokelskere.no vil være nyttig å se på tekstlige kommentarer nettopp fordi det er et godt supplement til terningkast.

## **Skalert vurdering**

Skalerte vurderinger er terningkast, stjerner eller en annen form for gradert vurdering. På LibraryThing har de stjernevurdering med spenn fra 1-5, men brukeren kan velge å gi halve stjerner. Derfor har de i teorien 10 vurderingsgrader som strekker seg fra 0,5 til 5. På Bokelskere.no har de terningkast med spenn fra 1-6, her finnes det ingen halve terningkast å gi så dermed har de 6 vurderingsgrader. Schafer et al. (2007) poengterer at med flere vurderingsgrader i anbefalingssystem vil det dannes mer informasjon om hver brukers smak. De presiserer at dette kun stemmer opp til et vist punkt, deretter vil det ha en forstyrrende effekt på systemet. Hvis du

tilbyr en bruker en skala som strekker seg fra 1-100 og ber de vurdere de samme gjenstandene to ganger men med et tidsrom i mellom vil ikke gjenstandene vurderes til samme poengsum de to gangene. Dette skjer fordi det i den situasjonen er for mange vurderingsgrader slik at brukeren ikke klarer å huske hvordan han vurderer fra gang til gang.

### **Terningkast på Bokelskere.no**

Terningkast melder seg som en klar kilde for brukervurdering på Bokelskere.no, men det er fortsatt valg som må tas i forbindelse med dem. Som diskutert ovenfor har Bokelskere.no en funksjon som heter Favoritt og det kan diskuteres om bøker som settes som favoritt burde vurderes høyere enn andre bøker. Nesse benyttet seg av dette i den første anbefalingsmetoden for nettstedet, som forklart i kapittel 4. I denne anbefalingsmetoden setter han terningkast verdi 7 på de bøkene som brukerne klassifiserer som favoritter. Dette vil gi en skala med vurderingsgrader 1-7. Er det dermed sagt at det kan leses mer informasjon om brukeren ut av denne handlingen? For å bestemme dette bør man se på de bøkene som får favorittstempelet. Det er logisk å tro at de bøkene som brukere har satt som favoritter har fått terningkast 6, og kanskje noen har fått terningkast 5. Så vil det utgjøre noe forskjell om de får terningkast 7 når utvikler allerede vet at de er likt?

Å bruke terningkast 7 på bøker som er favoritter har to fordeler. Det første er forbundet med brukers smak og den andre omhandler datamaterialet. Favorittstempelet viser at brukeren mener det noe spesielt ved nettopp de bøkene som gjør at brukeren har dem som favoritter. Det kan tenkes at en bruker er så reflektert i sin vurdering at han ikke klarer å sette terningkast 6 på en bok hvor det finnes en logisk brist i handlingen eller språket i boken ikke er det beste men brukeren faller for boken uansett. Dermed settes den med terningkast 5 i favorittthylla, og brukeren vil trolig sette pris på bøker som ligner denne. I den situasjonen vil det være nyttig å gi favoritter terningkast 7. På en sosial nettside hvor poengskalaen er begrenset til 1-6 mener jeg det vil være nyttig å tildele favoritter terningkast 7. Dette gir en større poengskala og bøker som har fått favoritt er spesielle for brukeren som setter favorittstatusen. Det er dermed trolig at flere brukere vil sette pris på bøker med favorittstempelet.

Den andre fordelen dette gir er flere terningkast. Dette er frivillig å gi terningkast, og alle brukere benytter seg ikke av den muligheten. Da det viser seg at en del bøker som ligger i forskjellige brukeres boksamling ikke er vurdert med terningkast kikket jeg nærmere på nettopp



dette. Det finnes en del data om de forskjellige utgavekopiene som vises blant annet som når en bok søkes opp.



1.  **Shantaram**  
av Gregory David Roberts  
2008 (Innbundet) [Se alle utgaver \(4\)](#)

[ditt eksemplar](#)

Likte du denne? Vi har bokanbefalinger!

 7	 1287	 5	 135 (10,5%)	 5,14 (313)	 23
---	--	---	---	--	---

Bilde 6: Informasjon om en utgavekopi i en søkesituasjon.

For data som vises under utgavekopien og for denne utgaven av *Shantaram* kan jeg lese dette: Det finnes 7 brukere som ønsker seg verket. Boken finnes totalt i 1287 boksamlinger og det er fem brukere som leser den i skrivende stund. 135 brukere har denne boken som favorittbok, det tilsier 10,5 prosent av alle brukerne her inne. Den har et gjennomsnittlig terningkast på 5,14 og det er 313 brukere som har gitt terningkast på denne utgaven. Det finnes 23 diskusjonstråder om denne utgaven. Det som er spennende i denne sammenhengen er at det kun er 313 av 1287 brukere som har gitt den terningkast. Det betyr at det kun er 24 prosent av brukerne som har kastet terning på boken.

Dette er spennende og det kan være på sin plass å undersøke hvordan det fordeler seg på hele boksamlingen, siden dette vil ha betydning for om man kan bruke terningkast i en brukers profil for å bestemme hva brukeren liker. Jeg tok for meg aktive bokdiskusjoner og valgte ut de ti sist diskuterte bøkene. Dermed fikk jeg et utvalg av aktuelle bøker for øyeblikket representert på nettstedet. Det kan diskuteres om utvalget er representativt for hele nettstedet, og for endelige konklusjoner bør et større utvalg bearbeides. Men for å få et innblikk i situasjonen holder det med et lite utvalg som vist i tabellen på neste side.

Tittel	Brukere som har boken	Brukere som har gitt terningkast	Prosent brukere som har gitt terningkast
Etter oss	4	1	25
Little bee	441	126	29
Huset ved moskeen	947	251	27
Jenter som kommer	68	13	19
Alle sjelers natt	42	12	29
Innsirkling	854	194	23
Historien om fru Berg	92	16	17
Checkout	86	21	24
Shantaram	1287	313	24
Engleslakt	4	1	25
<b>Totalt</b>			24

Tabell 6: Prosentandel som gir terningkast på bøker i Bokelskere.no.

Tabellen viser at prosentandelen brukere som setter terningkast på bøkene er lav. Spennet strekker seg fra 17 til 29 prosent, med et gjennomsnitt på 24 prosent. Det betyr at omtrent tre av fire brukere ikke setter terningkast. For en utvikler som ønsker å benytte seg av terningkast i brukerprofilen er ikke dette positiv lesning. Den lave terningkastprosenten viser at det ved å velge terningkast begrenser utvikleren både datagrunnlaget for å finne likhet betraktelig og han ekskluderer den delen av brukere som ikke setter terningkast fra anbefalingsprosessen. Dette vil trolig gi et skjevt bilde av den egentlige tilstanden på nettstedet, og resultere i dårligere anbefalinger. Bokelsker har selv diskutert og reflektert over dette på nettstedet i en diskusjon om gjennomsnittskarakterer. ”Jeg får ikke til å gi terningkast. Synes ikke de sier noen ting, sjøl om jeg tror jeg skjønner intensjonen med de[m]. Hvordan velger du mellom 3 og 4 når en bok er sånn midt på treet?” (Tao 2010). Noen brukere syntes det er problematisk å vite hva de skal gi og noen velger det klart vekk.

Grunnen til at gjennomsnittet mitt er så høyt, er at jeg sjelden triller terningen for bøker jeg har lest. Faktisk er det bare yndlingsbøkene mine som er blitt utsatt for denne æren(?), og heller ikke alle dem har jeg satt karakter på. Vet ikke hvorfor det er blitt slik. Jeg har selvsagt en klar mening om alt jeg leser og jeg vet hva jeg liker, men så er det alle disse andre ytre påvirkningene som spiller inn... Om en bok er en 5 eller 6 for meg avhenger ikke nødvendigvis bare av teksten, men blant annet også av sinnsstemningen jeg er i og stedet jeg befinner meg på når jeg leser (AnjaE 2010).

Denne usikkerheten ser ut til å være det største hinderet for brukere å sette terningkast, men ved å sette en bok til favoritt viser klart at de liker den. Dermed vil bruk av favoritt som terningkast 7 øke mengden terningkast. Jeg vil i forbindelse med Bokelskere.no foreslå bruke den muligheten det bringer å vurdere en favoritt til terningkast 7.

Skalaen for terningkast strekker seg som diskutert fra 1-6 og det er allment forstått at 1 er dårligst og 6 er best. Det er ikke dermed sagt at brukere av nettstedet tolker skalaen likt. Hvor grensene til de forskjellige brukerne går i forhold til dette er subjektivt. Det betyr at en bruker kan syntes en bok er fin og gi den terningkast 4, mens for en annen betyr en fin bok terningkast 5. Det er heller ikke sikkert at brukeren benytter seg av hele skalaen. Hos enkelte brukere kan terningkast 3 være dårlig, og de benytter seg sjelden av den laveste delen av skalaen. Dette vil resultere i en høyere gjennomsnittskaracter for disse brukere. Andre brukere igjen kan bruke hele skalaen og vil få en lavere gjennomsnittskaracter.

En utvikler kan velge å sette bestemte grenser for hvilken verdi som er negativ, nøytral og positiv. For eksempel 1-2 er negativt, 3-4 er nøytralt og 5-6 er positive verdier som forklart i den første anbefalingsmetoden for Bokelskere.no i kapittel 3. Anbefalingssystemet vil i denne sammenhengen gå glipp av nyanser i brukerens vurdering, og det vil overse en brukers subjektivitet. Dermed vil det være bedre for utviklere av anbefalingssystemet å benytte seg av en normaliseringsprosess. Hovedoppgaven med en normaliseringsprosess er å fjerne usikkerheten ved hva en bruker mener med sine terningkast i forhold til terningkast gitt i den kollektive samlingen.

For å normalisere terningkast brukes ofte gjennomsnittlig vurdering. Koren & Bell (2011) forklarer dette med å bruke grunnlinjeprediktor. De skisserer et eksempel hvor utvikler er interessert i å få vite hvordan en bruker ville vurdert en bestemt gjenstand. For å finne frem til dette bruker de gjennomsnittlig vurdering og avvik. De sier at en grunnlinjeprediktor for en bruker om en bestemt gjenstand er brukernes samlede vurdering, den aktive brukerens avvik og gjenstandens avvik. På Bokelskere.no er den gjennomsnittlige vurderingen på samtlige bøker 5,0. Heffalumpen5 sin gjennomsnittlige vurdering av leste bøker ligger på 4,0 og gjennomsnittlig karakter på *Stolthet og fordom* er 5,3. Den antatte vurderingen for Heffalumpen5 blir da  $5 + (4-5) + (5,3-5) = 4,3$ . Denne metoden tar hensyn til både den aktive brukerens vurderingspreferanser og den kollektive vurderingen av gjenstanden. Dermed vil meningene til både den aktive brukeren og brukerne som en helhet influere anbefalingen. Det finnes flere fremgangsmåter innenfor

normalisering og utvikler må velge den som passer tilgjengelige datakilder og algoritmen valgt til anbefalingssystemet. I anbefalingssystemer som benytter seg av brukervurderinger mener jeg det er meget viktig å normalisere terningkast.

### **Vennenettverk**

Sosiale medier er en stor del av menneskers hverdag og brukerfellesskap er blitt en kraft som er verdt å benytte seg av. Victor, De Cock & Cornelis (2011) har brukt tid til å se på hva de kaller tillitsfremmende anbefaling. I denne sammenhengen går tillit ut på å stole mer på anbefalinger fra mennesker man kjenner eller verdsetter meningen til. Dette springer ut fra tanken om at hvis man lurer på om man skal se en film eller lese en bok spør man noen man kjenner om de har sett filmen eller lest boken. For å bruke dette på nettet trengs en sosial ressurs hvor brukere enten blir venner eller blir bedt om å vurdere i hvor stor grad de stoler på andre brukeres dømmekraft. Disse brukerne kalles ”web of trust”, altså et tillitsnettverk, på norsk kan det kalles et vennenettverk. Forfatterne legger til at det i tillegg til vennenettverk kan finnes blokkeringslister hvor den aktive brukeren legger brukere han på ingen måte ønsker å bli sammenlignet med.

Funksjonen som heter *finn og følg* på Bokelskere.no kan sees på som en måte å uttrykke vennskap mellom brukerne. Det finnes flere grunner til å velge å følge en bruker, som poengtert i kapittel 4. Brukerne kan ha lik lesesmak og en bruker kan velge å følge en annen bruker fordi han legger merke til at de leser mange av de samme bøkene. Brukerne kan kjenne hverandre fra før, og er derfor ekstra interessert i å følge med på hva den andre leser. Når en bruker velger å følge andre brukere på Bokelskere.no får han lettere informasjon om hva de han følger leser. Dermed kan følger funksjonen sees på som muligheter til å danne vennenettverk innenfor et større nettverk. Der er trolig slik at en bruker verdsetter meninger fra venner høyere enn fra ukjente, dermed kan brukere som den bestemte brukeren har valgt å følge være en kilde for anbefalinger.

På Bokelskere.no bruker de nettopp dette under fanen *Hva skjer* til å vise frem bøker som leses av de brukerne den aktive brukeren følger. Dette kan sees på som en måte å anbefale litteratur som brukere den aktive brukeren stoler på liker. Denne måten vil ivareta personaliseringsaspektet ved en anbefaling fordi den aktive brukeren har valgt å følge disse personene. Det er på den andre siden en svakhet knyttet til dette siden bøkene ikke er ferdigleste derfor vil det ikke foreligge informasjon om brukerne som leser den aktuelle boken liker den eller ikke. En måte å redusere denne usikkerheten kan være å velge ut bøker som er ferdig lest og vurdert. Bøkene kan eventuelt vises med terningkast fra den fulgte brukeren, men da er utvikler

igjen avhengig av terningkast. Følger funksjonen kan også brukes som et supplement til vanlig nærmeste nabo vurderinger. Hvis utvikler er usikker på om resultatet fra denne fremgangsmåten er tilfredsstillende grunnet sparsomt bruk av terningkast kan bøker som finnes i resultatlisten og i samlingene til en eller flere av de brukerne den aktive brukeren følger vektet slik at de kommer høyere opp på den endelige resultatlisten.

For å utvikle et større vennenettverk på Bokelskere.no kan det tenktes at utvikler kan bruke Diskusjoner. Trolig vil brukere som svarer hverandre ofte dele mange meninger, og like å kommunisere på denne måten. Den andre siden av dette er at brukerne absolutt ikke går sammen og er uenige i alle meninger den andre kommer med. Det kan være lurt å bruke en form for semantisk tekstanalyse for å avgjøre om meningsutvekslingen er positiv eller negativ.

### **Emneord**

Jeg har tidligere snakket om emneordene på Bokelskere.no som i realiteten er brukerskapte tagger. Tagger er et uttrykk for hva brukerne mener om en bok og det vil også være en mulighet for å bruke tagger for å finne samsvar mellom brukere. Emneord som er satt til en bestemt bok vil vises i forbindelse med bokens hovedside. De emneordene en aktiv bruker har satt selv skiller ut på brukerens egen bokside for boken. Det vil si at en bruker kan velge å sette emneord på en bestemt bok selv om de alt er satt, men fordi brukeren selv ønsker å forbinde de emneordene med boken. Dermed kan man på nettstedet se hvilke emneord en bruker ofte benytter seg av og hvilke brukere som benytter seg av samme emneord. Det kan være givende for utvikler å se på denne forbindelsen mellom brukere. Går emneordene på innholdet i bøkene eller brukes de til personlige vurderinger eller plasseringer av bøkene. Emneordet 1001 bøker settes på bøker nevnt i *1001 bøker du må lese før du dør* som er et referanseverk med innflytelsesrike romaner som er valgt ut som ”må leses” bøker. Brukere som har brukt dette emneordet er trolig genuint interessert i å lese mange av disse bøkene og det kan være positivt for utvikler å koble denne typen brukere sammen. For å forsikre seg om dette er en kobling utvikler bør gjøre kan han se på Diskusjoner for å se om det finnes diskusjonstråder om dette og hvilke brukere som er aktive på de trådene.

### **Bokhyller**

På samme måte som emneord er en god representant for hvordan brukere tenker om en bok er bokhyllene en måte å vise hvordan brukerne liker å strukturere bøkene sine. Det kan tenkes at brukere som har mange bøker innenfor samme kategori er fordelaktig å sammenligne enn andre

brukere. Men det er et enormt spenn over navn brukere setter på bokhyllene sine fra fysisk inndeling etter hvor dokumentet befinner seg, emne, format og matvareassosiasjoner. Hvis tagger eller bokhyller skal brukes er det viktig å innføre en synonymkontroll for å få kontroll over bokhyller som trolig har samme type innhold (jmf. Vedlegg 11). Slik at bokhyller som heter noe med vampyrer blir samlet i samme gruppe. Det er også interessant å se hvordan brukere ordlegger seg. Hyllene Vampyrkjærleik og Vampyrkliss har 4 bøker hver, tre av disse bøkene er de samme. Kliss er ofte et mer negativt ladet ord enn kjærleik og det hadde vært meget spennende å se på de forskjellige brukerne og hvordan de har vurdert bøkene i de to bokhyllene, hvis de har gjort det. Etter dette kan man se om de forskjellige bokhyllenavnene bare er en måte å ordlegge seg, eller om det er en mening bak ordvalget. Å bruke bokhyller for å sammenligne brukere byr som diskutert ovenfor en del utfordringer til tolkning av brukere. Det er derfor trolig bedre for utvikler å se på annen data om brukeren som gir mindre rom for syensing.

### **Lister**

Brukere på Bokelskere.no kan sette opp egne lister over boktitler under fanen Lister. Disse listene er personlige og kan handle om alt fra gode bøker innenfor et emne, bøker en bruker skal lese, bøker en bruker har lest, bøker som inneholder spesielle ord i tittelen eller bøker som skal selges. Her er det bare fantasien som setter stopper for hva slags type liste man vil lage. Jeg er usikker på hvordan disse listene kan brukes i anbefalingssammenheng, men de vil kunne si noe om brukeren. Gi informasjon om hvordan den aktive brukeren som har laget listen tenker og brukerne som velger å kommentere på den. Jeg mener de tre kildene til informasjon, lister, bokhyller og emneord kan være spennende å se på for å bygge ut et vennenettverk, men det er ikke det første en utvikler bør se på for å bestemme likhetskilder.

### **Stjerner**

Bokelskere.no har en enkel måte for brukere å vise at de liker en Diskusjon eller et Sitat. Under disse to datakildene finnes det et stjernemerke som en bruker kan krysse av for og som viser hvor mange stjerner Diskusjonen eller Sitatet har fått. Hver bruker har en oversikt side over hvor mange stjerner de gitt og for hva, samtidig har de en oversikt over hvor mange stjerner de har gitt og til hvem for hva. Ut fra dette kan en utvikler lese hvilke sitater den aktive brukeren liker og hvilke bøker sitatene er fra. Det vil være spennende å se om brukere liker sitater hentet fra bøker eller forfattere de liker, på bakgrunn av terningkast og boksamling. En annen mulighet er å se på hvilke brukere som gir hverandre stjerner. Finnes det trender mellom brukere? Finnes det brukere

som aldri gir hverandre stjerner? Her vil det være spennende å se om brukere som også gir hverandre stjerner er aktive i de samme diskusjonene. Hvis utvikler allerede har bruker semantisk tekstanalyse av diskusjoner vil Stjerner og Diskusjoner sammen kunne gi et klarere bilde av hvilke brukere som er enige i litterære diskusjoner.

### **Valg av likhetskilder på Bokelskere.no**

Terningkast er en naturlig kilde til brukerbaserte vurderinger på Bokelskere.no derfor vil jeg i første omgang konsentrere meg om disse. Men med de utfordringene som finnes grunnet lav prosentandel som gir terningkast vil denne metoden måtte bearbeides en del. Dette kommer jeg tilbake til senere i avsnittet.

#### **5.3.2.2 Sammenligning av naboer**

For å regne ut hvilken vurdering en bruker vil sette til en uvurdert bok trengs det som diskutert over vurdering, og en sammenligning av den aktive brukeren og andre brukere. Dermed må de mest like brukerne altså naboene finnes. Desrosiers & Karypis (2011) deler inn utvelgelsen i naboer i to trinn. Det første trinnet ekskluderer de brukerne som ikke har vurdert de samme verkene, eller som har vurdert så få av de samme verkene slik at en anbefalings vil være meget usikker. På Bokelskere.no generer de en liste med de mest like brukerne en gang i døgnet. Denne listen tar for seg brukere som har flest like utgavekopier med den aktive brukeren (jmf. Vedlegg 12). Denne listen kan brukes som et utgangspunkt for å finne naboer, da de naboene som har flest like bøker som den aktuelle brukeren samles. Den viser også at Bokelskere.no kan være sårbar når det gjelder denne typen vurderinger. Heffalumpen5 hadde da listen ble generert 555 bøker i sin boksamling. De ti brukerne med flest like bøker som den aktive brukeren hadde fra 260 til 178 like utgavekopier. Dette understreker at Bokelskere.no er et lite nettsted, og at en utvikler ikke må glemme utgaveproblematikken der.

Det neste trinnet er i følge Desrosiers & Karypis (2011) å undersøke hvilke av de utvalgte naboene som tidligere har vurdert gjenstander likest den aktive brukeren. Her må utvikler se på tidligere vurderinger og bestemme hvor mange nærmeste naboer som skal være med i det endelige utvalget. Her argumenter forfatterne med at det er en balansegang å finne riktig antall nærmeste naboer. Hvis det brukes under 20 naboer blir utvalget for snevert og resultatet unøyaktig. Hvis det blir valgt over 50 naboer vil utvalget bli stort og meningene for ulike slik at resultatene igjen blir unøyaktige. Dermed bør en utvikler bruke mellom 20 og 50 naboer, men for å finne den beste balansen bør dette undersøkes når anbefalingssystemet er oppstartet.

Etter å ha funnet naboer må vurderinger normaliseres. Schafer et al. (2007) sier at den enkleste måten å finne ut dette på er å ta gjennomsnittet av alle brukeres terningkast. Problemet med dette er at da blir alle brukeres mening vektlagt like mye, selv om deres vurdering har forskjellig likhetgrad i forhold til den aktive brukerens vurderinger. Derfor er det nødvendig å vektlegge vurderinger fra brukere som ligner den aktive brukeren. Utvikler bør også ta hensyn til om brukeren som vurderer er snill eller streng med terningkastene med å normalisere brukervurderinger med gjennomsnittlig vurdering som forklart ovenfor. Vurderingene normaliseres for å ta hensyn til hvordan både den aktive brukeren vurderer og hvordan de utvalgte naboene vurderer.

Til denne typen anbefalinger trenger utvikler å sette opp en vurderingsmatrise med brukere og bøker. Jeg har satt opp en matrise med de fire brukerne som har flest like bøker som brukeren Heffalumpen5, denne matrisen konkretiserer noen av utfordringene for nettstedet. Eksempelmatrise:

	Livsglede	Harry Potter og de vises stein	Hulebjørnens klan	Dødslelene	Dødsjakt i Louisiana	Saman er ein mindre aleine	Enders spill	Vindens skygge	Ved daggry	Jane Eyre
Heffalumpen5	?	5	6	5	5	5	5		6	5
Hulderjenta	4	5	5	?	4	6	?	5		5
Bokelskerinnen		6		6		?	?			
Bokoholiker	?		4	6	?	?	5		?	4
Mona Elizabeth		6		?	?	?	?	5	?	

Tabell 7: En eksempelmatrise med brukere og vurderinger.

Matrisen har veldig mange ”hull”. De rosa feltene er bøker brukerne har i samlingen men ikke har lest. De blå feltene er bøker brukere har lest, men ikke vurdert. Spørsmålstejn indikerer bøker brukeren ikke har i samlingen og ikke har lest. Det finnes flere måter å forbedre dekningen i matrisen Desrosiers & Karypis (2011) trekker frem at man kan bruke gjennomsnittlig vurdering for å fylle inn tomrom. I forhold til Bokelskere.no har jeg fylt inn de blå områdene i matrisen med gjennomsnittlig terningkast på boken fra samtlige brukere av nettstedet.



	Livsglede	Harry Potter og de vises stein	Hulebjørnens klan	Dødslekene	Dødsjakt i Louisiana	Saman er ein mindre aleine	Enders spill	Vindens skygge	Ved daggry	Jane Eyre
Heffalumpen5	?	5	6	5	5	5	5		6	5
Hulderjenta	4	5	5	?	4	6	?	5	4,7	5
Bokelskerinnen		6	4,7	6	4,3	?	?		4,7	5,1
Bokholiker	?	5,3	4	6	?	?	5	5	?	4
Mona Elizabeth	3,8	6	4,7	?	?	?	?	5	?	

Tabell 8: En eksempelmatrise med brukere og brukere, med gjennomsnittlig vurdering.

I den andre matrisen har jeg nå fylt inn gjennomsnittlig terningkast for hele nettstedet på de bøkene brukerne har lest men ikke vurdert selv. Dermed får jeg en mer fullstendig matrise. Ulempene er at brukervurderingene kan være mer unøyaktige i forhold til de utvalgte naboene. Dette må veies opp mot hvor unøyaktige en matrise med flere ”hull” vil være. For å finne ut av dette vil den beste fremgangsmåten være å gjennomføre offline eksperimenter som måler prediksjonsnøyaktigheten. Når naboene er valgt ut gjenstår det å bestemme egenskaper som skal vektlegges og hvordan listen skal rangeres.

### 5.3.2.3 Resultat og rangering

I det siste trinnet anbefaler systemet gjenstanden eller de gjenstandene som får den høyest antatte vurderingen for den aktive brukeren. Før resultatet presenteres må bøker brukeren allerede har lest filtreres ut. Utvikler må bestemme om han skal ta enkelte valg for å manipulere hvordan listen rangeres, trengs det mer serendipitet? Det kan også diskuteres om det skal finnes en form for kontroll hvis det er mange bøker av for eksempel samme forfatter, serie eller sjanger. Det finnes som diskutert tidligere flere måter å vise frem resultatet. Siden dette handler om å vise en bruker alternative bøker han trolig vil like lønner det seg å vise resultatet i en rangert liste. På den måten vil brukeren se flere mulige aktuelle bøker samtidig, noe som er nyttig for brukeren i denne sammenhengen.

Visuelt sett er det alltid å foretrekke å vise en bok med bilde av cover, og ikke ha for mange treff på en side. Utvikler må tenke igjennom hvor grensen går for bøker som skal anbefales eller ikke anbefales. Slik som Bokelskere.no fungerer nå velger de å vise 100 anbefalinger totalt fordelt på 10 sider. Det kan tenkes at de siste anbefalingene på denne listen ikke er de beste. Grensen for hva som skal være med kan settes av en grenseverdi i forhold til algoritmen, men det kan hende at en brukerundersøkelse må til for å finne den passende

grenseverdien. En generell oppsummering av brukerbasert nærmeste nabo kan finnes i oppført som punktoversikt.

### **Punktvis fremstilling av brukerbasert nærmeste nabo metode:**

- **Utvelgelse av likhetsvurdering.**
  - Valg av vurdering. Liker/like, skalert, tekstelige kommentarer.
  - Valg av supplerende data. Vennennettverk, emneord, diskusjoner, boksamling.
- **Utvelging av naboer.**
  - Først foretas utvalg av potensielle naboer med en sammenligning av alle brukere for å velge ut brukere som kan betegnes naboer basert på vurderinger av felles bøker. Dette sparer systemet for lagringsplass og begrenser antall nabokandidater.
    - Utvelgelsen kan gjøre på bakgrunn av et bestemt antall bøker både den aktive brukeren og potensielle naboer har vurdert.
    - Utvelgelsen kan gjøre på bakgrunn av hvor likt den aktive brukeren og potensielle naboer har vurdert.
- **Utrekning av likhetsvekt mellom naboer.**
  - De potensielle naboene gjennomgås slik at de naboene som er ”likest” den aktive brukeren velges ut.
    - Utvelgelsen gjøres på bakgrunn av hvilke naboer som har likhetsvekt med størst omfang.
- **Normalisering av brukervurderinger.**
  - Vurderingene normaliseres for å ta hensyn til hvordan både den aktive brukeren og naboene vurderer.

## 6 Oppsummering

Den digitale verden har eksplodert og litteraturutvalget sammen med den. Både i form av formater som digitale lydbøker og e-bøker, samtidig har utvalget av bøker aldri vært større derfor trenger lesere hjelp til å finne frem til bøker de liker å lese. Siden lesere er mer og mer involvert i den digitale hverdagen er det viktig å møte dem der de er, her kommer anbefalingssystemer inn i bildet. Hovedspørsmålet mitt i forbindelse med denne oppgaven er hvilke anbefalingsmetoder for litteratur viser seg mest hensiktsmessig for en bruker i en nettbasert ressurs?

Jeg har gjennom en teoretisk tilnærming tatt for meg anbefalingsmetoder og det aktuelle domenet. Gjennom grundig drøfting har jeg kommet frem til flere potensielle metoder, satt fingeren på mange problemområder og pekt på noen mulige løsninger for anbefalingssystemer på sosiale nettsteder. Jeg ønsker i dette kapittelet å oppsummere dette arbeidet å legge frem de anbefalingsmetodene jeg har kommet frem til og hvilke konsekvenser eventuelle valg får for brukeren. Jeg peker på mulige svakheter ved gjennomføringen av oppgaven og utviklers ansvar.

### 6.1 Oppsummering av underspørsmål

Jeg utarbeidet fire underspørsmål jeg har jobbet med i forbindelse med oppgaven og vil gi en oppsummering av disse. Å jobbe med disse underspørsmålene har fungert som en mal gjennom oppgaven.

#### 1. Hvilke forskjellige anbefalingsmetoder finnes?

Gjennom litteraturstudiet har jeg kommet frem til at det ved skrivende stund finnes syv forskjellige anbefalingsmetoder som en utvikler kan velge å benytte seg av basert på domenet han jobber innenfor og datakilder han har tilgjengelig. De syv anbefalingsmetodene er innholdsbaserte metoder, kollaborative metoder, demografiske metoder, kunnskapsbaserte metoder, kontekstuelle metoder, sosiale metoder og hybride metoder. Innenfor de forskjellige metodene vil det finnes forskjellige fremgangsmåter og algoritmer alt etter hvilket behov utvikler arbeider utifra som gjennomgått i kapittel 2.

## **2. Hvilke anbefalingsmetoder kan en utvikler benytte seg av i forhold til tilgjengelig data fra et sosialt nettsted?**

Ved å undersøke de forskjellige anbefalingsmetodene som finnes og arbeidet med case mener jeg det er fire potensielle metoder som har utmerket seg. Disse fire metodene fungerer godt innenfor formålet for litteraturdomenet og med de datakildene sosiale nettsteder har tilgjengelig. De fire metodene er Innholdsbaserte metoder, Kollaborative metoder, Sosiale metoder og Hybride metoder. Direkte knyttet til data fra Bokelskere.no har jeg valgt å drøfte innholdsbaserte metoder og kollaborative metoder.

## **3. Hvordan skiller det å anbefale en bok seg fra anbefaling av andre gjenstander?**

En bok skiller seg fra andre underholdningsmedier som musikk og film gjennom at den kun trenger å ta hensyn til den aktive brukeren. En film kan sees sammen med andre mens lesing er en individuell og personlig aktivitet hvor boken som anbefales skal passe leserens og ingen andre sin smak. Opplevelsen av en bok kan deles, men en voksen leser vil som oftest lese den alene. En bok har både konkrete og subtile egenskaper som ligger til grunne når en leser vurderer bokens appell. En leser kan ikke alltid sette fingeren på hvorfor han liker eller ikke liker en bok som diskutert i kapittel 3. Derfor må utviklere som benytter seg av innholdsbaserte metoder passe på at gjenstandsrepresentasjonen er omfattende nok til reflektere hva en bok handler om. En bruker vil alltid tolke en bok fra sine litterære preferanser derfor må anbefalingsmetoden som velges gjenspeile flest mulig av brukernes smak.

## **4. Hvilke egenskaper er viktig å legge vekt på når en skal utvikle anbefalingsmetoder for litteratur?**

Jeg mener den store utfordringen for å tilfredsstillende en leser er å finne den riktige balansen mellom prediksjonsnøyaktighet og serendipitet. Lesere kan fordeles på en skala som strekker seg fra to ytterpunkter. De som foretrekker mer av den samme typen litteratur og er fornøyd med det, og de som foretrekker å stadig bli overrasket og prøve nye sjangre. Mange av leserne vil befinne seg et sted på denne skalaen. Derfor er det viktig å undersøke dette å prøve å balansere resultatet slik at begge typer litteratur blir representert.

## 6.2 Anbefalingsmetoder for litteratur

Gjennom arbeidet med denne oppgaven har det kommet frem at det ikke finnes en perfekt anbefalingsmetode som vil omfavne alle brukere. Dette er fordi brukere ikke er like og hva en bestemt bruker liker vil alltid være subjektivt. Samtidig er anbefalinger basert på informasjon som er tilgjengelig om de forskjellige brukerne. En anbefaling blir mer og mer nøyaktig jo mer informasjon det er innsamlet om en bruker. Kanskje det er mer riktig å si det finnes ingen perfekt anbefalingsmetode som vil omfavne alle brukere, foreløpig. Det hele er et spørsmål om tilgang, innsamling av data og hvordan dataene kan manipuleres. Dette gjør anbefalingssystemer meget spennende å jobbe med. Gjennom arbeidet med oppgaven har jeg sett på mange løsninger for hvordan man kan høste mest mulig informasjon om de forskjellige brukerne og gjenstandene for å presentere de beste anbefalingene for brukere. Jeg har sett på hvilke metoder en utvikler har å velge mellom med de dataene de har tilgjengelig. Dette arbeidet har vist meg at noen anbefalingsmetoder for litteratur er mer aktuelle å benytte seg av enn andre. Det handler ikke om å finne en perfekt metode, men metoder som kan gi flest mulig brukere gode anbefalinger.

De mest lovende metodene for litteraturanbefalinger jeg har gjennomgått i forbindelse med rammene for denne oppgaven er innholdsbaserte og kollaborative metoder. De to metodene skiller av hvordan likhet måles. Det finnes fordeler og ulemper forbundet med begge, derfor vil jeg oppsummere og sette de to metodene opp mot hverandre. På den måten ønsker jeg å konkretisere hva konsekvensene av å velge mellom disse to metodene utgjør for brukerne. Jeg har trukket ut prediksjonsnøyaktighet, dekning, tillit, serendipitet og tilpassningsdyktighet som viktige egenskaper i et anbefalingssystem for litteratur. Jeg har lagt vekt på at bøker har både konkrete egenskaper som kan leses direkte fra bokens attributter og subtile egenskaper som sier noe om en boks appell på leseren. Drøftingen i kapittel 5 viser at det i forbindelse med Bokelskere.no finne så mange kilder til usikkerhet at de to metodene trolig ikke vil levere hensiktsmessige anbefalinger hver for deg. Dermed vil en hybrid løsning med fremgangsmåter hentet fra de to metodene være et bedre alternativ.

### 6.2.1 Innholdsbaserte litteraturanbefalinger

Ved å velge innholdsbaserte metoder blir gjenstandene kilde for likhet og gjenstandsrepresentasjonene blir den viktigste faktoren. Derfor er bibliografisk data og hvordan brukerens litterære preferanser kan leses ut fra den viktig med hensyn til datadekning. Jeg har vist hvor problematisk datadekning kan være i en ressurs som har en blanding av egenprodusert og innhentet data. Dette problemet vil ikke forsvinne i den digitale verden, men det arbeides kontinuerlig med løsninger. Den semantiske webben gjennom Linked Data er et område hvor det arbeides med å konkretisere hva data betyr og hvordan de henger sammen. Dette arbeidet har gått saktere enn antatt, men det universelle problemet om å vite hva noe betyr har flere forskningsområder. Bruk av leksikalske databaser som WordNet og Wikipedia er under utvikling, det trengs derimot at vi får med norske ressurser på banen. For Bokelskere.no må en forholde seg til de datakildene som finnes og bearbeide de på best mulig vis. På skrivende tidspunkt vil en innholdsbasert metode ekskludere enkelte bøker fordi de ikke er koblet sammen som utgavekopier av samme bok og fordi de ikke har samme mengede data tilgjengelig. Dette vil gå utover prediksjonsnøyaktigheten, og kan resultere i at brukere får anbefalinger som inneholder bøker de allerede har lest, representert av en annen utgavekopi.

Brukerens smak kan i denne sammenhengen baseres på en bestemt bok som brukeren har likt, eller baseres på brukerens boksamling. Velger utvikler å se på en bestemt bok som kilde for likhetskriterium vil de anbefalte bøkene ligne den aktuelle boken som var utgangspunktet. Dette gir høy prediksjonsnøyaktighet men lav sjans for serendipitet. Dette betyr at anbefalingene blir meget spesialisert og at brukeren ikke vil få nye input i sin litterære horisont. Anbefalingene kan for brukeren virke ensformig og tar ikke hensyn til hvilke bøker leseren tidligere har lest. Velger utvikler å se på boksamlingen til en bruker vil fortsatt anbefalingene ligne på tidligere bøker som brukeren har likt. Så sant brukeren har størst andel av tidligere likte bøker i sin boksamling vil prediksjonsnøyaktigheten fortsatt være høy. Resultatet vil være litt mer variert enn anbefalinger gjort på en bestemt bok siden det er basert på forskjellige bøker, men det vil fortsatt resultere i lav serendipitet. Det vil ikke plutselig dukke opp noe nytt og uventet i anbefalingene fordi den typen bøker som anbefales per definisjon av en innholdsbasert anbefalingsmetode allerede må finnes i den aktive brukerens boksamling. Velges boksamling vil en brukers historikk ivaretas og bøker som ligner den sist leste kan vektes høyere enn bøker som ligner de som brukeren leste for et år siden.

Begge de innholdsbaserte metodene jeg har skissert her er tilpassningsdyktige innenfor sine rammer. De tar hensyn til hva brukeren har lest i det siste og vil fange opp forandringer i brukerens smak. Den metoden som baserer seg på boksamling vil bruke lenger tid enn den som baserer seg på en bestemt bok fordi den trenger å se en ny trend i bøker brukeren har lest før utslaget av dette vises på resultatlisten. En tankevekker med tilpassningsdyktighet i denne sammenhengen er at en bruker må hente inspirasjon til valg av bøker å lese fra et annet sted enn anbefalingssystemet for at de skal få bøker som er annerledes. Hvis de ikke leser og legger inn bøker som er annerledes trenger de heller ikke at anbefalingssystemet er tilpassningsdyktig.

Brukere har lettere for å stole på seg selv en andre så derfor er det lett å vinne en brukers tillit for innholdsbaserte anbefalingsmetoder. Systemet sier: Du har tidligere lest og likt bok X, på bakgrunn av dette tror vi du vil like bok Y og Z også fordi de ligner på bok X. Det eneste som kreves av brukeren er at han stoler på sin egen smak og siden brukeren faktisk har likt bok X vil han være villig til å lese mer som ligner på det.

Jeg mener den største konsekvensen ved å velge en ren innholdsbasert metode at de er avhengig av gode gjenstandsrepresentasjoner og at resultatet overspesialiseres. Brukeren fortsetter å bli anbefalt bøker innenfor den samme typen og vil miste muligheten til å utvide sin litterære horisont. Selv for en bruker som liker en bestemt type bøker vil denne typen anbefalinger virke repetitive og mindre interessante over tid. På Bokelskere.no kan en utvikler kompensere for dette ved å tilføre et element av serendipitet. Et eksempel på dette er å hente ut titler som er populære på nettstedet, deretter legge inn bøker skrevet av forfattere som ikke er representert i den aktive brukerens boksamling. Denne fremgangsmåten går vekk fra en ren innholdsbasert metode og kan betegnes som en hybrid løsning av innholdsbasert metode og brukernes kollaborative meninger.

## **6.2.2 Kollaborative litteraturanbefalinger**

Ved kollaborative litteraturanbefalinger ønsker utvikler å høste informasjon om brukeres bokpreferanser og presentere anbefalinger på bakgrunn av dette. Dermed er brukerdatabasen den viktigste informasjonskilden. Hvordan en utvikler kan lære mest mulig om hva den aktive brukeren liker, hvilke andre brukere han ligner og hvilke brukere han interagerer med.

Bokelskere.no har eksplisitte terningkast som brukere gir på bøker de har lest. Dessverre benytter ikke alle brukere seg av denne muligheten og dette resulterer i et tynt grunnlag å basere anbefalinger på. Som vist i kapittel 5 kan gjennomsnittlig terningkast hjelpe til å bygge ut

datasettet, men anbefalingene blir basert på mer generell data. Det bør derfor gjennomføres tester for å se hvilke av de to fremgangsmåtene som gir best mulig prediksjonsnøyaktighet. Siden prediksjonsnøyaktighet er omvendt proporsjonal med serendipitet gir kollaborative metoder utvikler en god mulighet til å benytte seg av metoder som ikke gir ”nøyaktig” resultat. Det at kollaborative metoder baserer seg på hva andre brukere liker gir automatisk et grunnlag for serendipitet i motsetning fra innholdsbaserte metoder. Konsekvensene av dette er at utvikler må gjennomføre eksperimenter, gjerne med brukere for å se hvor balansen mellom prediksjonsnøyaktighet og serendipitet bør gå.

Alternative muligheter på Bokelskere.no er å benytte seg av andre kilder enn terningkast. Her har utvikler flere alternativer å velge mellom og det kan tenkes at en kombinasjon av flere brukergenererte data bør ligge til grunn for denne type anbefaling. Stjerner, Lister, Emneord, Bokhyller og Diskusjoner er kilder for brukermeninger. En kombinasjon kan være å sammenligne boksamlingene til de brukerne som en bruker følger og se hvilke bøker forskjellige brukere har i boksamlingen sin. Deretter kan bøker anbefales basert på antall i de forskjellige samlingene. En annen mulighet er å se på trender mellom bruker ved å benytte seg av Emneord, Lister og Bokhyller, finne brukere som har lignende Emneord eller Bokhyllenavn og se hvilke bøker de liker å lese for deretter å anbefale denne typen bøker for den aktive brukeren.

Å bygge tillit for kollaborative anbefalinger krever litt mer enn for innholdsbaserte metoder. I denne sammenhengen ber utviklere den aktive brukeren om å stole på andre brukere og ikke bare egne erfaringer. Noen brukere er skeptiske til å stole på meninger fra brukere de ikke kjenner. De tror ikke nødvendigvis at andre menneskers meninger kan reflektere deres smak. Dermed kan vennenettverk være positivt. Hvis brukere får vite at de anbefalingene som vises er på bakgrunn av brukere de selv har valgt å følge kan dette øke deres tillit til systemet.

Den største konsekvensen ved å benytte kollaborative metoder på Bokelskere.no er at den er avhengig av mer omfattende data enn kun terningkast. Dette er fordi det er for få brukere som gir terningkast til at denne datakilden kan brukes alene. Ved å benytte terningkast vil mange brukere få vurderinger som genereres på sparsomme datasett, og de vil være unøyaktige. Dette gjør at terningkast alene ikke er den mest hensiktsmessige anbefalingsmetoden for Bokelskere.no



### **6.2.3 En hybrid løsning**

For Bokelskere.no mener jeg det kan være nyttig å se på hybride løsninger basert på fremgangsmåter hentet fra både innholdsbaserte og kollaborative metoder. Innholdsbaserte metoder blir begrensende med tanke på serendipitet, mens kollaborative metoder basert på terningkast blir unøyaktige med den dataen som er tilgjengelig. Som vist i kapittel 2 kan hybride metoder settes sammen på forskjellige måter. Utvikler trenger konkret informasjon om de metodene han velger å sette sammen og gjennomgangen av de to overnevnte metodene har gitt grunnlag for dette. Det finnes flere hybride løsninger for Bokelskere.no

Utvikler kan benytte seg av parallelldesign med to separate metoder for deretter å velge de beste resultatene fra begge metodene. Her kan innholdsbasert metode hvor den aktive brukers sist leste bok brukes som likhetskriterier som første metode. Mens terningbaserte vurderinger kan fungere som den andre metoden. En rangert liste hvor anbefalingene slås sammen blir presentert for brukeren. På den måten vil resultatet inneholde både bøker som ligner på tidligere bøker brukeren har lest og bøker som brukeren ikke har noe kjennskap til.

En annen mulighet er å bruke seriedesign. Her benyttes en semantisk tekstanalyse av Diskusjoner for å undersøke brukers meninger om bøker, deretter forsterkes meningene med Stjerner den aktive brukeren har gitt. Bøker diskutert i denne prosessen kan deretter brukes som et grunnlag for innholdsbasert analyse. På den måten vil de bøkene som velges ut for en likhetsvurdering være bøker brukeren er interessert i. Resultatlisten vil være bøker som ligner på de utvalgte bøkene og derfor være aktuelle. Poenget med å velge en hybrid løsning er å sette sammen den mengden data som er tilgjengelig for å kunne presentere de mest hensiktsmessige anbefalingene for brukerne.

## **6.3 Refleksjoner om oppgaven**

Et hvert vitenskapelig arbeid skal være oppmerksom på egne mindre sterke sider og jeg vil her trekke frem sider ved min oppgave som virke inn på det jeg har funnet. Jeg vil også kort diskutere det eventuelle negative sider forbundet med personalisering og anbefalinger, og utviklers ansvar forbundet med dette

### **6.3.1 Den teoretiske fremstillingens begrensninger**

Dette er en teoribasert oppgave hvor jeg har jobbet med teori fremsatt av tidligere forskning og brukt drøfting av potensielle metoder basert på valgte case. Jeg har arbeidet med dette som en mulighetsstudie og trukket frem metoder og skissert løsninger for tekniske utfordringer. Det er dermed usikkerhetsmomenter knyttet til arbeidet. Utfordringer jeg har pekt på og forsøkt å arbeide rundt kan tenkes være mindre eller større en det jeg har portrettert, siden denne oppgaven i stor grad benytter seg av mine evner til logisk drøfting. Oppgaven med å kartlegge anbefalingsområdet og anbefalingsdomenet ble meget omfattende. Drøftingen og metodene jeg har valgt ut som potensielle anbefalingsmetoder kan derfor ikke sees på som pålitelige kilder, men som forslag for fremtidig forskning.

### **6.3.2 Hva mister den anbefalte verden?**

Jeg har snakket mye om hvordan man kan anbefale, hvorfor utvikler bør anbefale og hvordan dette kan hjelpe brukerne. Som med mange tekniske nyvinninger kommer det en makt som gjør at forskere må spørre seg selv om dette kan ha en negativ effekt på verden. Derfor trenger jeg også å spørre; Bør man anbefale og hva kan brukere eventuelt miste ved å følge anbefale? Utvikler har her et ansvar. Personalisering og anbefalinger på nettet gir utviklere av programmer og systemer en makt til å begrense hva en bruker får vite om. Teknologien er utviklet for å hjelpe brukere å overkomme information overload-problemet, men resultatet er at brukeren går glipp av potensielle gjenstander. Være det; bøker, musikk, filmer, nyhetsartikler, eller nettsider.

Dette har lenge vært diskutert, spesielt i forbindelse med søkemotorer som Google. Denne typen søkemotorer har regler for hvordan de skal operer ved søk i tillegg til at de har evnen til å ”huske” hvem brukeren er og hva han tidligere har sett på. Det er også et tankekors at nettaviser nå kan sortere nyheter etter hvor mange som leser de. VG<sup>37</sup> har på sin nettside et filter man kan bruke for å se hvilke artikler som er mest lest for øyeblikket. Dette blir en form for ikke personlig anbefaling basert på hvilke artikler andre brukere leser.

Pariser (2011) har tatt for seg det han kaller ”The filter Bubble” som handler om konsekvenser av den stadig voksende personaliseringen som foregår i den digitale verden.

---

<sup>37</sup> <http://www.vg.no/>

Most of us assume that when we google a term, we all see the same results-the ones that the company's famous Page Rank algorithm suggests are the most authoritative based on other pages' links. But since Desember 2009, this is no longer true. Now you get the result that Google's algorithm suggests is the best for you in particular-and someone else might see something entirely different (Pariser 2011).

Pariser (2011) tar for seg hvordan personalisering som Google benytter seg av kan endre en brukers syn på verden ved å snevre inn hva brukeren får vist i sin nettleser. Han mener at mennesker blir fanget i en filtreringsboble. Det som slipper gjennom hvert menneskes boble er tilpasset hva dette mennesket allerede er interessert i. Problemet med dette slik Pariser ser det er at det vi slipper inn i vår boble er ting vi allerede er interessert i eller mennesker vi allerede kjenner. Han mener at på den måten er ikke internett lenger en brubygger mellom kulturer, men en vegg som holder det nye og ukjente ute.

Det populære sosiale nettverket Facebook får gjennomgå for sin filtrering av brukernes statuser.

At first, the News Feed showed nearly everything your friends did on the site: But as the volume of posts and friends increased, the Feed became unreadable and unmanageable. [...] Facebook's solution was EdgeRank, the algorithm that powers the default page on the site, the Top News Feed. EdgeRank ranks every interaction on the site (Pariser 2011).

Ved hjelp av denne algoritmen begrenses statusoppdateringer brukeren ser på sin side til de brukerne som den aktive brukeren interagerer mest med. Men Pariser er interessert i at Facebook skal bestemme at han ikke er hva som skjer i livet til en kamerat fordi han ikke har kommentert statusen hans på en stund, eller har vært inne på hans profilside. Det handler om makt. For hvem skal bestemme hva vi ser på internett. Internett som ble skapt som den ultimate frisonen, hvor alle som hadde en mening kunne komme med den.

Jeg vil avslutte denne tankerekken med en skjønnlitterær fremtidsvisjon. I boken Matched møter vi en ung kvinne som vokser opp i en verden hvor valgmulighetene er begrenset til et minimum. Samfunnet styrer hva de spiser basert på hvilken ernæring de trenger for å leve sunnest mulig i de 80 årene de får tildelt. Alle dør nemlig på sin 80 årsdag fordi Samfunnet har bestemt at dette er den optimale alderen. Et menneske har ved fylte 80 opplevd det livet har å by på, men har enda sin helse. Samfunnet styrer hva de skal jobbe med, basert på informasjon lagret om dem gjennom oppveksten får de tildelt et yrke som passer til dem som person. Samfunnet styrer hvem

de skal gifte seg med gjennom en verdensomspennende matching prosess, basert på kompatibilitet og sannsynligheten for at de fremtidige barna får best mulig helse.

There is no luck in the Society. I nod. Of course. I should know better than to use such an archaic, inaccurate term. There's only probability now. How likely something is to occur, or how unlikely (Ally Condie 2010 s. 18).

Jeg er tilhenger av anbefalingssystemer og fordelene de bringer, jeg tror fast på at de kan være nyttige for brukerne. I dagens digitale samfunn trenger brukere hjelp med den massive mengden data som finnes. Men når vet vi at nok er nok? Hvem sitter med makten? Jeg tror det er viktig å se på i hvilke kontekst personalisering og anbefalinger blir foretatt. Et anbefalingssystem for litteratur vil være et supplement til hvordan en brukers leseropplevelser og jeg ser ikke noe galt med det. Et anbefalingssystem for nyheter er også ikke problematisk så lenge brukeren er klar over at nyhetene er tilpasset ham. En søkemotor som gir forskjellige brukere forskjellige bilder av verden er jeg mer skeptisk til. En avis som vektlegger og uthever artikler brukere leser og deler over nettet uten å opplyse brukerne om dette mener jeg ikke er akseptabelt. Jeg mener at hvis en ressurs benytter seg av personalisering og anbefalingssystemer har utvikler et ansvar om at brukeren må få beskjed om det, og de bør få muligheten til å velge det bort. Et anbefalingssystem skal hjelpe en velger i en digital verden, ikke ta valgene for ham.

## **7 Videre forskning**

Det finnes flere mulige forskningsprosjekter knyttet til dette arbeidet. Det kan være aktuelt å fortsette arbeidet med nettstedet Bokelskere.no, se på litteraturanbefalingssystemer generelt eller som en digital bibliotekstjeneste.

### **7.1 Videre forskning på Bokelskere.no**

I denne oppgaven har jeg teoretisk tatt for meg de metodene som er mest hensiktsmessig i forhold til anbefalingsmetoder for litteratur. Det neste trinnet for Bokelskere.no er å gjennomføre eksperimenter med de valgte metodene. I disse eksperimentene kan man først se hvilken metode som får de beste resultatene ved bruk av egenskapen prediksjonsnøyaktighet. Jeg foreslår å prøve ut innholdsbaserte metoder, kollaborative metoder og en hybrid løsning av de to metodene. Eksperimentet kan foretas som et offline eksperiment som forklart av Shani & Gunawarda (2011) i kap 2. Når den mest hensiktsmessige metoden er valgt må utvikler fortsette eksperimentet med vektning i forhold til de involverte partene i algoritmen.

Når en algoritme er utviklet tilstrekkelig vil implementering og en brukerundersøkelse være et spennende skritt. En brukerundersøkelse kan gjøres med interesse i å se hvordan anbefalingsmetoden blir mottatt på nettstedet, kanskje med en sammenligning med tidligere anbefalingsmetoder. Eventuelt kan en forsker kan velge å sette opp en brukerundersøkelse i et datalaboratorium som kartlegger hvordan brukeren faktisk benytter seg av systemet og brukers første reaksjoner på resultatet som blir presentert. Det er også mulig å presentere brukeren for resultat basert på variasjoner av algoritmen, en algoritme som vektlegger prediksjonsnøyaktighet og en som vektlegger serendipitet. Deretter kan brukers reaksjon høstes for å finne den rette balansen på disse egenskapene. Det er viktig å huske at brukere er subjektive så uansett vil ikke det valgte anbefalingssystem tilfredsstille alle brukere, utvikler må velge det som tekker flest mulig.

#### **7.1.1 Refleksjoner for Bokelskere.no**

Bokelskere.no er i en unik posisjon i Norge. Det finnes ingen andre nettsamfunn for bokinteresserte som fungerer som de gjør. Det finnes mange norske blogger tilknyttet enkeltpersoner, biblioteker eller andre institusjoner men ingen som gir de mulighetene til katalogisering av eget bibliotek. Det er en forholdsvis ung nettside og utviklere viser stadig

velvilje til forbedring og utvikling, noe som gjør at det har vært veldig spennende å jobbe med. Gjennom arbeidet med nettstedet har jeg gjort meg konkrete tanker om utviklingsmuligheter som jeg har samlet i dette avsnittet. Jeg vil samtidig legge til at jeg syntes nettstedet er et godt og berikende tiltak innenfor litteratur på nettet i Norge.

Hovedutfordringen til Bokelskere.no ligger som jeg ser det i å koble sammen utgavekopier. De sliter fordi de har omfavnet den friheten nettbaserte samfunn er kjent for å omfavne. På den ene siden ønsker de å gi brukerne den beste mulige måten til å personalisere sin boksamling slik brukeren selv ønsker. Dette kommer til uttrykk i hvordan bruker kan velge forskjellige utgaver av samme verk, ikke bare ta til takke med valg mellom språk og format. I tillegg til å hente inn bøker fra Amazon.com kan brukere også legge inn bøker automatisk, og i den prosessen trenger ikke brukeren å oppgi ISBN nummer engang. På den andre siden ønsker de å systematisere dataene sine på en oversiktlig måte slik at de kan brukes effektivt.

I dag finnes det som jeg har snakket om tidligere 20 forskjellige utgavekopier for *Stolthet og fordom* og forskjellen på enkelte av dem går kun på coverbildet. Disse utgavekopiene er i midlertidig slått sammen med hjelp utenfra.

Å slå sammen utgaver til verk er en oppgave som er vanskelig å gjøre automatisk. Våre "sammenslåinger" blir gjort manuelt av et knippe fagfolk i Den Norske Bokdatabasen, og de fleste utgaver av et verk du finner her inne er slått sammen til rett verk. Men ikke alle (Nesse 2011 april a).

Nesse mener dermed at det ikke finnes noe lett måte og fikse dette på, og jeg må si meg enig i dette hvis de skal fortsette med denne metoden. Det betyr også at de aldri vil komme helt à jour, fordi det alltid er tilsig av bøker i samlingen. Samtidig har Bokelskere.no en mulighet som tilbyr brukere å legge inn sitt eget cover på en utgave som finnes på nettstedet. En mulighet er å opplyse om dette når brukere legger inn nye bøker. Hvis en bruker får treff på en bok fra Amazon.com som allerede finnes på Bokelskere.no kan det komme opp en lenke som sier at denne finnes og at brukeren senere kan gå inn å endre cover. Dermed vil man få færre utgavekopier å forholde seg til, og disse kan variere på grunn av språk og format i stedet for den visuelle variasjonen et bokcover utgjør.

Terningkastsetting er som diskutert et opphav til mye frustrasjon på Bokelskere.no. Enkelte brukere liker det ikke i seg selv å velger å se bort fra det, mens andre brukere syntes det blir for vanskelig å sette terningkast fordi de ikke er sikre på hvilke kriterier de skal bruke. Noen

brukere er ikke fornøyde med terningkast og ønsker seg en skala med flere vurderingsgrader. Her kan det være flere ting utvikler kan gjøre. Det enkleste kan være å legge inn en hjelpetekst som kan klikkes på ved behov. Hjelpeteksten kan oppfordre til frihet og si det at brukeren står fritt til å sette egne kriterier for terningkastene han legger inn. Hjelpeteksten kan også være formanende å komme med retningslinjer for hvordan brukeren skal sette terningkastene. Det viktigste i en slik hjelpetekst mener jeg vil være å spille på at brukerens terningkast kan hjelpe andre brukere å plukke ut potensielle bøker. Brukere liker ofte å føle at de hjelper andre brukere. En annen mulighet er at det kommer opp en lenke på bokens side når brukeren legger inn tilstanden *har lest* som stiller spørsmålet: Vil du gi terningkast nå? Det er også en mulighet at utvikler bruker litt av plassen på en brukers innstillingsside til å oppfordre til å gi terningkast.

Skalagraden ved vurdering strekker seg fra liker/liker ikke til hvor mange en utvikler vil ha. Som oftest brukes liker/liker ikke, fem stjerner, terningkast eller ti stjerner. Det betyr at på Bokelskere.no kan de gå ned å be brukeren sette liker/liker ikke, eller gå opp å bruke ti stjerner. Jeg tror ikke liker/liker ikke er veldig informativt i en litterær kontekst, da dette er en veldig snever måte å se på verden. Jeg vil heller foreslå at Bokelskere.no ser på skalaen til Internett movie database<sup>38</sup>, og det ti stjerners systemet de bruker. Brukere klikker enkelt på et stjernebarometer fra en til ti. Vurderingen for hver enkelt film vises med desimaltall ved siden av filmen. Dette gir brukere en litt større frihet til å vurdere, men ikke så stor at de bruker skalaen på en inkonsekvent måte.

## 7.2 Videre forskning på litteraturanbefalinger

Gjennomgangen av andre ressurser som tilbyr litteraturanbefalinger har vist at utviklere av anbefalingssystemer velger forskjellige metoder å generere anbefalinger på. Da jeg undersøkte de forskjellige sosiale nettstedene som tilbyr personlige anbefalinger av litteratur på slo det meg at kunne være spennende å utføre en komparativ design studie hvor flere av de sosiale nettstedene med anbefalingssystem, ble valgt som case. Her er Amazon.com, LibraryThing, Bokelskere.no, What should I read next, Visual Bookshelf og GoodReads potensielle kandidater.

For å se på disse kan man lage identiske profiler og sammenligne anbefalingene man får i de forskjellige systemene etter for eksempel 10, 50 og 100 bøker i samlingen. På denne måten kan en undersøke hvordan de forskjellige nettressursene anbefaler litteratur og forsøke å finne

---

<sup>38</sup> <http://www.imdb.com/>

likheter og ulikheter. Ut i fra dette kan man forsøke å se hvilke kriterier som vektlegges. Her er det spesielt spennende å se på forskjellen på anbefalingssystemene i forhold til hva slags ressurs de skal betjene. Er det forskjell på hva en netthandel og hva et sosialt nettsted vektlegger? Er det enkelte av ressursene som anbefaler mange bøker av samme forfatter, innenfor samme sjanger eller oppfølgere i en serie? Denne forskningen kan suppleres med brukerundersøkelser for å undersøke hva de aktive brukerne faktisk mener om anbefalingene som blir generert for dem.

### **7.3 Videre forskning tilknyttet bibliotek**

På et bibliotekarisk plan vil et samarbeid med Bibliofil, og bruk av deres funksjon Mappa mi være meget spennende. Bibliofil er ledende innenfor biblioteksystemer for folkebibliotek og de er godt i gang med å utvikle den private brukerdelen av det digitale biblioteket. Mappa mi er en funksjon som alle brukere har muligheten til å benytte seg av og som gjør at de kan administrere egne lån. Brukerne kan bestille, reservere og fornye bøker selv. Hvis brukeren selv ønsker det kan han huke av for at biblioteket skal huske lånene som foretas gjennom bibliofilsystemet, dette kalles historikk. Dermed er Bibliofil godt på vei til å samle inn informasjon til et anbefalingssystem. Neste trinn i prosessen vil være å generere anbefalinger basert på brukers historikk. Siden det ikke er automatisk at en bruker liker alt han låner bør det tilrettelegge for å samle inn brukeres meninger.

En spennende utvikling i en slik prosess kan være å la brukere sette terningkast på bøker de har lest, dette er en utbredt form for vurdering og noe vi som mennesker møter på ofte i vår digitale hverdag. Det kan tenkes at den informasjonen som er samlet inn om brukeren kan brukes i stedet for eller i tillegg til den karusellen de allerede har som er basert på lånetall for biblioteket generelt. De gitte vurderingene kan også brukes for å velge ut bøker som anbefalingssystemet bruker til å finne likhet til andre dokumenter. Bibliotekskatalogen er utstyrt med mye bibliografisk data og gjenstandsrepresentasjoner basert på denne informasjonen kan bli meget omfattende. Dermed er innholdsbaserte metoder et godt alternativ for å finne bøker som ligner på de bøkene brukeren har lest og likt. Å bruke denne type anbefalinger forutsetter at brukeren er logget inn i forbindelse med søkeprosessen.

Dette viser er at det er utviklingsmuligheter på bibliotekfronten. Arbeidet vil kreve større involvering fra brukere, men kan gi dem en god mulighet til å finne gode bøker. Det vil kreve at



bibliotekene får større frihet med håndtering av persondata slik at dette blir gjennomførbart i vanlige biblioteker.

## 8 Litteraturliste

Fullstendig litteraturliste over skriftelige kilder og muntlig kommunikasjon brukt i oppgaven.

### 8.1 Kilder

- Aamotsbakken, B., & Knudsen, S. V. (2008). *Å tenke teori: om lese teorier og lesing*. Oslo: Gyldendal.
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17 (6). (s. 734-749).
- Adomavicius, G., & Tuzhilin, A (2011). I: Introduction to recommender systems handbook. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 217-253). Boston: Springer.
- Aina. (2011 31. mai). Nytt & ukjent eller vellest & kjent. [Svar på Heffalumpen5s melding datert april 2011]. Hentet 4. juni fra <http://bokelskere.no/tekst/91496/>
- Alag, S. (2009). *Collective intelligence in action*. Greenwich: Manning
- Amatriain, X., Jaimes, A., Oliver, N., & Pujol, J. M. (2011). Data mining methods for recommender systems. . I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 39-72). Boston: Springer.
- Anand, S.S., & Mobasher, B. (2005). Intelligent techniques for web personalization. *Intelligent Techniques for Web Personalization* (s. 1–36). Berlin: Springer.
- Anand, S. S., & Mobasher, B. (2007). Contextual Recommendation. I: Berendt, A. Hotho, D. Mladenic & G. Semeraro (Red.), *From Web to Social Web: Discovering and Deploying User and Content Profiles* (s. 142-160). Berlin: Springer.
- Ankie. (22. Mai 2011 ). Nye boka anbefalinger. [Svar på André Nesses melding datert 22.mai 2011]. Hentet fra <http://bokelskere.no/tekst/94030/>
- Ann Christin. (2011 mars). Har du endret lesevaner, evt oppdaget noen skatter etter å ha blitt medlem her?. [Melding til diskusjonsforum]. Hentet fra <http://bokelskere.no/tekst/84041/>

- Ben-Shimon, D., Tsikinovsky, A. & Roka, L. (2007). Recommender system for personal social network. *Advances in intelligent web mastering: proceedings of the 5<sup>th</sup> Atlantic web intelligence conference AWIC 2007*. (s. 27-55).
- Berners-Lee, T. (2001). *The Semantic Web*. Hentet fra [http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf)
- Billsus, D., & Pazzani, M. J. (1998). Learning collaborative information filtering. *Proceedings of the 15<sup>th</sup> international conference on machine learning*.
- Bokelskere.no. (udatert). *Bokelskere.no*. Hentet fra [www.bokelskere.no](http://www.bokelskere.no)
- Burke, R. (2007). Hybrid web recommender systems. I: P. Brusilovsky, A. Kobsa & W. Nejdl (Red.), *The adaptive web: methods and strategies of web personalization* (s. 377-408). Berlin: Springer.
- Burke & Ramezani (2011). Matching recommendation technologies and domains. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 367-386). Boston: Springer.
- Claussen, J. M. (2007). *Kunsten å lese*. Oslo: Bokvennen.
- Condie, A. (2010). *Matched*. New York: Penguin.
- Datatilsynet. (2011 17. februar). *Datalagringsdirektivet på høring i Stortinget*. Lokalisert 26. april 2011 på nettet: [http://www.datatilsynet.no/templates/Page\\_\\_\\_\\_3778.aspx](http://www.datatilsynet.no/templates/Page____3778.aspx)
- Desrosier, C., & Karypis, G. (2011). A comprehensive survey of neighborhood-based recommender methods. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 107-144). Boston: Springer.
- Dourish, P. (2004). What do we talk about when we talk about context. *Personal and Ubiquitous Computing* 8 (1). (s.19–30).
- Escarpit, R. (1971). *Litteratursosiologi*. Oslo: Cappelens.
- Givon, S., & Lavrenko, V. (2009). Predicting social-tags for cold start book recommendations. *RecSys' 09*. (23-25). New York.
- Goldberg, D., Nichols, D., Oki, B., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70.

- Göker, A., & Davies, J. (2009). *Information retrieval: Searching in the 21<sup>st</sup> century*. Chippenham: Wiley.
- Hedvig. (April 2011). Utfyllende profil. [Svar på Heffalumpen5s melding datert 6. april 2011]. Hentet 9. mai fra nettet: <http://bokelskere.no/tekst/91286/>
- Heffalumpen5. (6. april 2011). Utfyllende profil. [Melding til diskusjonsforum]. Hentet 9. mai fra nettet: <http://bokelskere.no/tekst/91286/>
- Herlocker, J., Konstan, J. Borchers, A., & Riedl, J. (1999) An algorithmic framework for performing collaborative filtering. *SIGIR '99 Proceedings of the 22<sup>nd</sup> annual international ACM SIGIR conference on Research and development in information retrieval* (s. 230-237.) New York.
- Herlocker, J., Konstan, J. A., & Riedl, J. T. (2000). Explaining collaborative filtering recommendations. *Proceedings of ACM 2000 Conference on Computer Supported Cooperative Work* (s. 241–250).
- Herlocker, D., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *Transactions on information systems*, 22 (1). (s. 5-53).
- Hilde S H. (Mai 2011). Anbefalinger. [Svar på Heffalumpen5s melding datert 6. mai 2011]. Lokalisert 9. mai 2011 fra nettet: <http://bokelskere.no/tekst/98098/>
- International Federation of Library Associations and Institutions. [IFLA] (2000). Funksjonskravene til bibliografiske poster. L. A. Holm. (Ove.). Lokalisert 6. April 2011 på nettet: [www.nb.no/content/download/2019/16496/version/1/.../FRBR.pdf](http://www.nb.no/content/download/2019/16496/version/1/.../FRBR.pdf)
- Jannach, D., Zanker, M., Felfering, A., & Fredrich, G. (2011). *Recommender systems: An introduction*. Cambridge.
- Kobsa, A. (2007). Privacy-enhanced web personalization. . I: P. Brusilovsky, A. Kobsa & W. Nejdl (Red.), *The adaptive web: methods and strategies of web personalization* (s.628-670). Berlin: Springer.
- Koren, Y., & Bell, R. (2011). Advances in collaborative filtering. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 145-186). Boston: Springer.
- Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice*. (3 ed.). Illinois: Universitetet av Illinois.

- LibraryThing. (2011). Shantaram. Lokalisert på nettet 6. april 2011 på: [www.librarything.com/work/5152](http://www.librarything.com/work/5152)
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE computer society*.
- Long, E. (2003). *Book clubs: women and the uses of reading in everyday life*. Chicago: University of Chicago.
- Lops, P., Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: state of the art and trends. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 73-106). Boston: Springer.
- Manguel, A. (2007). *En historie om lesning*. (2 utg.) Oslo: Aschehoug.
- Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., & Symeonidis, P. (2011). Social Tagging Recommender Systems. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 615-644). Boston: Springer.
- Massa, P., & Avesani, P. (2004). Trust-Aware Collaborative Filtering for Recommender Systems, *Lecture Notes in Computer Science*, (3290) Springer.
- Milicevic, A. K., Nanopoulos, A., & Ivanovic, M. (2010). Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions. *I: Artificial Intelligence Review*, 33 (3). (s. 187-209). Springer.
- Mobasher, B., Bruke, R., Bhaumik, R., & Williams, C. (2007). Toward trustworthy recommender systems: An analysis of attack models and algorithm. *ACM transactions on internet technology*, 7 (4).
- Mooney, R. J., & Roy, L. (1999). Content-based book recommending using learning for text categorization. Lokalisert 14. januar 2011 på nettet: <http://portal.acm.org/citation.cfm?id=336662>
- Morville, P., & Rosenfeld, L. (2006). *Information architecture: for the world wide web*. (3.ed.). Sebastopol: O'Reilly.
- Nesse, A. (2011, april a). For få søkemuligheter på "Finn bøker". [Svar på Pegasines melding datert april 2011]. Hentet 6. juni fra <http://bokelskere.no/tekst/90601/>
- Nesse, A. (2011, april b). Nye bokanbefalinger. [Melding til diskusjonsforumet]. Hentet 15. mai 2011 fra <http://bokelskere.no/tekst/94030/>

- Nesse, A. (2011, april c). Nye bokanbefalinger. [Svar på Ankis melding datert 23. april 2011]. Hentet 15. mai 2011 fra <http://bokelskere.no/tekst/94030/>
- Nesse, A. (2011, mai). Nye bokanbefalinger. [Svar på Ankis melding datert 29. mai 2011]. Hentet 1. juni 2011 fra <http://bokelskere.no/tekst/94030/>
- Nesse, A. (2010). Ny funksjon. [Melding til diskusjonsforum]. Hentet 6. mai 2011 fra <http://bokelskere.no/diskusjoner/finn/?finn=ny+funksjon>
- Nesse, A. (2011, april d). Utfyllende profil. [Svar på Heffalumpen5s melding datert 6. april 2011]. Hentet 9. mai fra <http://bokelskere.no/tekst/91286/>
- Papagelis, M., Rousidis, I., Plexousakis D., & Theoharopoulos, E. (2005). Incremental Collaborative Filtering for Highly-Scalable Recommendation Algorithms. *Lecture Notes in Computer Science*, (s. 7-17).
- Pariser, E. (2011). *The filter bubble: what the internet is hiding from you*. New York: Penguin.
- Pazzani, M., J. (1999). A Framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13 (5-6). (s.393-408).
- Picault, J., Ribière, M., Bonnefoy, D., & Mercer, K. (2011). Hoe to get the recommender out of the lab? I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 333-366). Boston: Springer.
- Pijpers, G. (2010). A lot of information. I: Pijpers, G. *Information overload: a system for better managing everyday data* (s. 19-34). Honoken: Wiley.
- Princeton University. (13. april 2010). *About WordNet*. Princeton. Hentet fra <http://wordnet.princeton.edu/wordnet/>
- Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 1-35). Boston: Springer.
- Ridderstrøm, H. (2011). Resepsjonsteori. *Bibliotekarstudenters nettleksikon om litteratur og media*. Hentet fra <http://home.hio.no/~helgerid/litteraturogmedieleksikon/resepsjonsteori.pdf>
- Ringdal, K. (2009). Enhet og mangfold. (2. utg.). Bergen: Fagbokforlaget.
- Ritcher, A. (2009). LibraryThing: an interview with Tim Spalding. *Crossroads*, 15 (4). (s. 3-5). Hentet fra <http://delivery.acm.org/10.1145/1560000/1558899/p3->

ritchie.pdf?key1=1558899&key2=1121706031&coll=DL&dl=ACM&ip=84.49.109.91&CFID=24541137&CFTOKEN=17156902

- Roberts, R. M. (1989). *Serendipity: accidental discoveries in science*. New York: John Wiley.
- Saricks, J. G. (2005). *Readers advisory service in the public library*. (3 ed.). Chicago: American Library Association.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based collaborative filtering recommendation algorithms. Hentet fra [http://www.grouplens.org/papers/pdf/www10\\_sarwar.pdf](http://www.grouplens.org/papers/pdf/www10_sarwar.pdf)
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. I: P. Brusilovsky, A. Kobsa & W. Nejdl (Red.), *The adaptive web: methods and strategies of web personalization* (s. 291-324). Berlin: Springer.
- Shani, G., & Gundawardana, A (2011). Evaluating recommendation systems. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 257-297). Boston: Springer.
- Skei, H. M. (2006). *Å lese litteratur* (2. utg.). Oslo: Gyldendal akademiske.
- Smeaton, A. F., & Callan, J. (2005). Personalisation and recommender systems in digital libraries. *International Journal on Digital Libraries*, 57 (4). (s. 299-308).
- Smith, G (2008). *Tagging: People-powered metadata for the social web*. Berkley: New riders.
- Solveig Strand. (2011 mai). Nytt & ukjent eller vellest & kjent. [Svar på Heffalumpen5s melding datert april 2011]. Hentet fra: <http://bokelskere.no/tekst/91286/>
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial intelligence*. Hentet fra <http://www.hindawi.com/journals/aai/2009/421425.html>
- Sutherland, J. (2007). *Bestsellers: a very short introduction*. Oxford : Oxford University Press.
- Tao. (2010 mai). Gjennomsnittlig terningkast. [Svar på Tone Lilleenget Bys melding datert mail 2010]. Hentet fra <http://bokelskere.no/tekst/44437/>
- Vesanen, J. (2007). What is personalization? A conceptual framework. *European journal of marketing*, 41(5/6), 409-418.

- Victor, P., De Cock, M., & Cornelis, C. (2011). Trust and recommendations. I: F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (Red.), *Recommender systems handbook* (s. 645-704). Boston: Springer.
- Whitten, J., L., Bentley, L., D., & Ditterman, K., C. (2004). *Systems analysis & design methods*. (6 ed.). Boston: Irwing.
- WordNet Search - 3.0. (2006). Serendipity. New Jersey: Princeton. Hentet 14. Februar fra <http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&s=serendipity&i=1&h=100#c>
- Zeitgeist overview (udatert). LibraryThing. Hentet 7. mars 2011 fra <http://www.librarything.com/zeitgeist>
- Ziegler, C. N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *World Wide Web conference 2005*. Japan.
- Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Ohio: Columbus.



## **8.2 Personlig kommunikasjon**

- André Nesse, nettredaktør, Bokelskere.no (e-post, 23. november 2010)

## 8.3 Tabeller

S. 39 Tabell 1: Fullstendighet og presisjon.

S. 85 Tabell 2: Ressursrelaterte valg for litteraturanbefalinger.

S. 87 Tabell 3: Brukerrelaterte valg for litteraturanbefalinger.

S. 89 Tabell 4: Datarelaterte valg for litteraturanbefalinger.

S. 100 Tabell 5: Egenskapene i et anbefalingssystem med til litteraturanbefalingssystemer.

S. 122 Tabell 6: Prosentandel som gir terningkast på bøker i Bokelskere.no.

S. 128 Tabell 7: En eksempelmatrise med brukere og vurderinger.

S. 129 Tabell 8: En eksempelmatrise med brukere og brukere, med gjennomsnittlig vurdering.

## 8.4 Bilder

S. 51 Bilde 1: Taggesky på LibraryThing.com

S. 59 Bilde 2: FRBF modellen med eksempler fra Bokelskere.no.

S. 71 Bilde 3: Forside på Bokelskere.no.

S. 80 Bilde 4: Oversikt over transaksjoner på Bokelskere.no.

S. 93 Bilde 5: Forskjellige utgavekopier av *Shantaram*.

S. 121 Bilde 6: Informasjon om en utgavekopi i en søkesituasjon.

# 9 Vedlegg

## Vedlegg 1: Ønskebok.no

Hovedsiden for Ønskebok.no:



Brukeren velger kriterier:



Resultat på bakgrunn av valg:

**ønskebok.no**

glad  trist   
morsom  alvorlig   
trygg  truende   
forventet  uventet   
storslagen  Jon Inær   
nydelig  motbydelig   
mild  brutal   
lett  krevende   
uten sex  mye sex   
tradisjonell  uvanlig   
optimistisk  dyster   
kort  lang   
prøv igjen

skift til hovedperson, handling og sted  
lydbok nynorsk film

**Kjærlighet ved første treff**  
av Jon Michelet

lån

**Leserkommentar**  
Gjett om ting begynner å skje når unggutten Nelson fra Rio de Janeiro seiler inn i Oslo havn og bestemmer seg for å dra på hopprenn i Holmenkollen! At en furten og kranglevoren jente bombarderer ham med snøballer er bare begynnelsen! Er det kjærlighet ved første blikk? Tøft og morsomt om forelskelse og kulturforskjeller - og med mange hertlige misforståelser når Nelson og Hildegun prøver å krangle (og flørte) på en blanding av engelsk, norsk og portugisisk!

**Les utdrag fra boka**

**Parallell**

- > [Vår afrikanske eksplosjon av Jon Michelet](#)
- > [Sinnssykt forelsket av Katarina von Bredow](#)
- > [Ærlighetsminuttet av Bjørn Sortland](#)

**Olive tok en øks**  
av Minette Walters

lån

## Vedlegg 2: Antall representerte bøker

Utvalget er foretatt 12. april 2011.

### **Bokelskere.no** (Totalt 11.921 brukere registrert)

1. Menn som hater kvinner av Stieg Larsson (2940) 24,4 %
2. Harry Potter og de vises stein av J. K. Rowling (2885)
3. Drageløperen av Khaled Hosseini (2865)
4. Da Vinci-koden av Dan Brown (2740)
5. Beatles av Lars Saabye Christensen (2644)
6. Jenta som lekte med ilden av J. K. Rowling (2637)
7. Harry Potter og mysteriekammeret av J. K. Rowling (2561)
8. Harry Potter og fangen fra Azkaban av J. K. Rowling (2545)
9. Naiv. Super. av Erlend Loe (2516)
10. Luftslottet som sprengtes av Stieg Larsson (2491)

### **LibraryThing** (Totalt 1.335.000 brukere registrert)


1. Harry Potter and the Sorcerer's Stone av J. K. Rowling (59,989) 4,3 %
2. Harry Potter and the Half-Blood Prince av J. K. Rowling (54,541)
3. Harry Potter and the Order of the Phoenix av J. K. Rowling (52,328)
4. Harry Potter and the Prisoner of Azkaban av J. K. Rowling (52,281)
5. Harry Potter and the Chamber of Secrets av J. K. Rowling (51,927)
6. Harry Potter and the Goblet of Fire av J. K. Rowling (51,185)
7. Harry Potter and the Deathly Hallows av J. K. Rowling (47,753)
8. The Da Vinci Code av Dan Brown (41,950)
9. The Hobbit av J. R. R. Tolkien (39,624)
10. Nineteen Eighty-Four av George Orwell (38,205)

**Visual Bookshelf** (Det finnes ikke opplysninger om antall brukere)

1. Harry Potter and the Deathly Hallows av J. K. Rowling
2. Harry Potter and the Chamber of Secrets av J. K. Rowling
3. Harry Potter and the Prisoner of Azkaban av J. K. Rowling
4. Harry Potter and the Half-Blood Prince av J. K. Rowling
5. Harry Potter and the Goblet of Fire av J. K. Rowling
6. The Da Vinci Code av Dan Brown
7. Twilight av Stephenie Meyer
8. Harry Potter and the Order of the Phoenix av J. K. Rowling
9. Harry Potter and the Sorcerer's Stone av J. K. Rowling
10. To Kill a Mockingbird av Harper Lee



## Vedlegg 3: Hovedsiden for en brukers bok

Hovedsiden for boken ”Monster” i boksamlingen til brukeren Heffalumpen5



# Bokelskere.no<sup>beta</sup>


Vi elsker bøker. Akkurat som deg. Velkommen hjem.

★ 42   innstillinger [logg ut](#) [hjelp](#)

---

<a href="#">hjem</a> til forsiden	<a href="#">din side</a> 481 bøker	<a href="#">finn bøker</a> søk eller bla	<a href="#">diskusjoner</a> se hva vi snakker om	<a href="#">bokelskere</a> finn og følg	<a href="#">sitater</a> godt sagt	<a href="#">lister</a> vi har dem	<a href="#">inviter</a> en bokelsker
--------------------------------------	---------------------------------------	---	---	--	--------------------------------------	--------------------------------------	---

---



### Ditt eksemplar

## Monster

skrekkroman  
av [Knut Nærum](#) (forfatter).  
318 sider

Likte du denne? Vi har bokanbefalinger!

Kjøp hos [bokkilden.no](#)

Emneord: fremtidsvisjon, hotellkonge, Humor, isbjørn, miljøverninteresser, norsk litteratur, polaris, skjønnlitteratur, turistnæring

Finn denne boka på ditt nærmeste bibliotek:


[rediger](#) | [mer bokdata](#)

oppgi emneord for denne boka

Bokhylle:  + opprett ny

Ditt terningkast:

★ en favorittbok?

 du har ikke lest denne (rediger).

### Aktive bokelskere





Ruller en bruker seg nedover fortsetter hovedsiden med diskusjoner om den aktuelle boken

## Diskusjoner om denne boka

emne	nyeste post av	oppdatert
» <a href="#">Jeg kan ikke si ...</a> (0)	<a href="#">Tuva-Kristin Furehaug</a>	09 Jan 09:42
» <a href="#">Hadde det ikke vært ...</a> (3)	<a href="#">Rose-Marie</a>	24 Feb 14:17
» <a href="#">Dette er en fremtidsvisjon ...</a> (0)	<a href="#">Rose-Marie</a>	18 Feb 20:24

## Si hva du mener om boka

Dette starter en ny diskusjon.

Lagre

Forhåndsvisning

avansert | tekstformatering

Ruller en bruker lenger ned på siden følger sitater brukere har lagt inn fra boken

## Sitater fra denne boka



### Iselin siterer **Monster** (side 72)

5 dager siden. Ingen har det som favoritt.



-Det har ingenting å si hvilket språk man snakker, sa Professoren, -så lenge folk ikke vil forstå.



godt sagt (2)



legg til i dine favorittsitater



svar

varsle



### Frida Tørring siterer **Monster** (side 131)

1 år siden. Ingen har det som favoritt.



Aksel gjorde noen kjappe boksefinter. - La dem komme. Jeg er ikke redd for dem. Det var akkurat dette som gjorde meg redd for ham. Han var ikke redd for noe annet enn mettede fettsyrer.



godt sagt (1)



legg til i dine favorittsitater



svar

varsle

En bruker kan selv legge inn sitater fra boken. Nederst på siden om den aktuelle boken finnes det bokdata om den aktuelle boken.

### Sitér denne boka:

Skriv kun inn det ordrette sitatet her. Du trenger ikke sette inn anførselstegn.

Fra side: (valgfritt)

Lagre

### Boktall

---

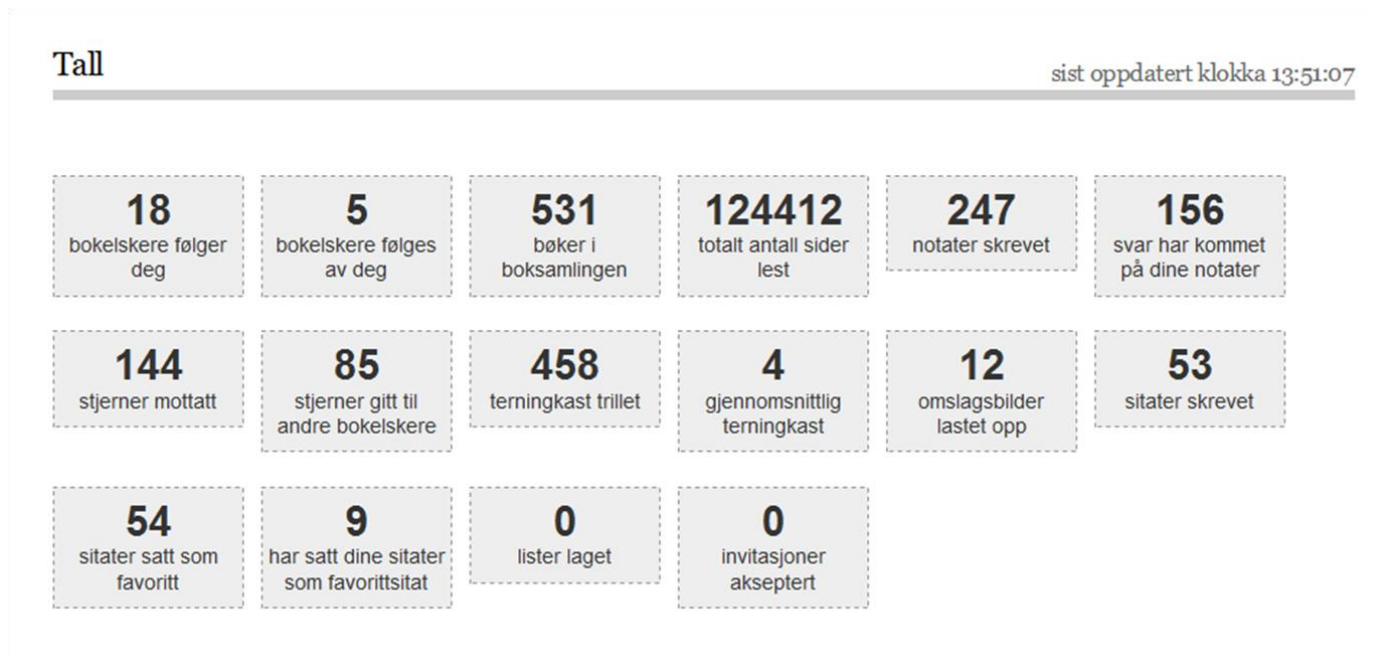


Vil du være med å påvirke disse tallene? Oppgi fødselsår og kjønn på [innstillinger](#), trill terningkast og markér en tilstand på bøkene dine.

Bilder er hentet 14. mars 2011 fra Heffalumpen5 sin brukerprofil på Bokelskere.no fra <http://bokelskere.no/bok/monster/37122/>

## Vedlegg 4: Tall om den bestemte bokelskeren

Tall om brukeren Heffalumpen5:



<http://bokelskere.no/nytt/> (Lenken virker kun når en er logget inn som Heffalumpen5 grunnet personvern)

## Vedlegg 5: Å følge en annen bruker

Hvordan en bruker kan finne andre brukere å følge.

<b>finn bøker</b> søk eller bla	<b>diskusjoner</b> se hva vi snakker om	<b>bokelskere</b> finn og følg	<b>sitater</b> godt sagt	<b>lister</b> vi har dem
------------------------------------	--	-----------------------------------	-----------------------------	-----------------------------

### Bokelskere:

de du følger de som følger deg alle

Søk etter bokelskere; navn eller brukernavn

Fylke:  Sorter etter: popularitet

Viser 1 til 20 av 10949 bokelskere:

	<b>Tore Renberg (Hekkan)</b> Har vært en bokelsker i 1 år, 8 måneder. Følges av 566 bokelskere. Bøker: 322. Finn likheter i lesesmak   Hvor mange felles bøker?	<input type="button" value="Følg"/>
	<b>Stig Ove Voll (bokhandleren)</b> Har vært en bokelsker i 1 år. Følges av 284 bokelskere. Bøker: 136. Finn likheter i lesesmak   Hvor mange felles bøker?	<input type="button" value="Følg"/>

Likhet og antall felles verk mellom brukere en bruker kan velge å følge.

<b>finn bøker</b> søk eller bla	<b>diskusjoner</b> se hva vi snakker om	<b>bokelskere</b> finn og følg	<b>sitater</b> godt sagt	<b>lister</b> vi har dem
------------------------------------	--	-----------------------------------	-----------------------------	-----------------------------

**Bokelskere:**

de du følger de som følger deg alle

Søk etter bokelskere; navn eller brukernavn

Fylke:  Sorter etter: popularitet

Viser 1 til 20 av 10949 bokelskere:

	<b>Tore Renberg (Hekkan)</b> Har vært en bokelsker i 1 år, 8 måneder. Følges av 566 bokelskere. Bøker: 322. Likhet: -0,71, basert på felles vurdering av 8 verk.   Du og Tore Renberg har 20 bøker felles.	<input type="button" value="Følg"/>
	<b>Stig Ove Voll (bokhandleren)</b> Har vært en bokelsker i 1 år. Følges av 284 bokelskere. Bøker: 136. Likhet: 0, basert på felles vurdering av 8 verk.   Du og Stig Ove Voll har 10 bøker felles.	<input type="button" value="Følg"/>

Bilder er hentet 14. mars 2011 fra Heffalumpen5 sin brukerprofil på Bokelskere.no på nettet:

<http://bokelskere.no/bokelskere/>

## Vedlegg 6: De du følger Leser nå

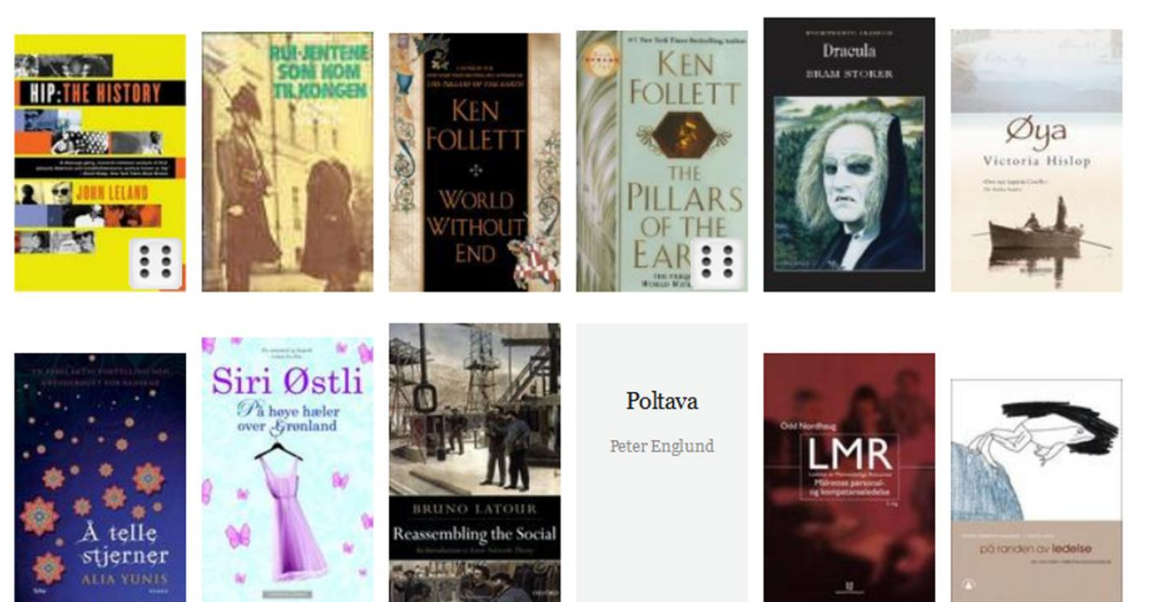
Bokelskere Heffalumpen5 følger Leser disse bøkene.

boksamling   hva skjer?   profil   organiser bøker

---

De du følger Leser nå

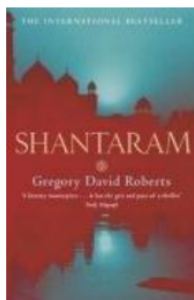
---



Bildet er hentet 7. mai 2011 fra Heffalumpen5 sin brukerprofil på Bokelskere.no på nettet: <http://bokelskere.no/nytt/> (Lenken virker kun når en er logget inn som Heffalumpen5 grunnet personvern)

## Vedlegg 7: Bokanbefalinger

Utgangsposisjon for bokanbefalinger som bygger på boken ”Shantaram”.



har lest

Legg inn omslagsbilde  
Se alle omslagsbilder (1)

finnes i  
**109**

nevnt  
**11**

Ditt eksemplar

**Shantaram**

av Gregory David Roberts (forfatter).

Se på denne utgaven hos [amazon.co.uk](https://www.amazon.co.uk)

944 sider

Likte du denne? Vi har bokanbefalinger!

Kjøp hos [bokkilden.no](https://www.bokkilden.no)

Kjøp hos [bokdykk.no](https://www.bokdykk.no)

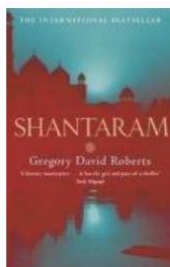
Kjøp hos [cdon.com](https://www.cdon.com)

Emneord: Skjønnlitteratur (engelsk)

Finn denne boka på ditt nærmeste bibliotek:

### Den gamle metoden.

Etter å ha klikket på lenken [Likte du denne? Vi har bokanbefalinger!](#)



har lest

Legg inn omslagsbilde  
Se alle omslagsbilder (1)

finnes i  
**109**  
boksamlinger

nevnt  
**11**  
ganger

Bokanbefalinger for deg som **liker** Shantaram

Ditt terningkast for Shantaram:



Hvis du er logget inn legger vi boka til boksamlingen din når du triller terningkast.

fjern terningkast



Legg til bok

Gå til bokdetaljside

**Kristin Lavransdatter: kransen**

av Sigrid Undset

Ditt terningkast for Kristin Lavransdatter:



Hvis du er logget inn legger vi boka til boksamlingen din når du triller terningkast.

De 4 bokelskere som likte Shantaram og har  
vurdert Kristin Lavransdatter liker Kristin  
Lavransdatter

**20,50%**

mer enn andre bokelskere. (6,25)

## Den nye metoden.

Etter å ha klikket på lenken Likte du denne? Vi har bokanbefalinger!



**SHANTARAM**  
Gregory David Roberts

Legg til bok

Legg inn omslagsbilde  
Se alle omslagsbilder (1)

finnes i **1274** | nevnt **69**

### Bokanbefalinger for Shantaram

Viser 1 til 10 av 100 bøker:

- 

1. **Noen kjenner mitt navn**  
av Lawrence Hill  
2009 (Innbundet) Se alle utgaver (6)  
Likte du denne? Vi har bokanbefalinger!

1 378 2 28 (8,1%) 5,05 (115) 10
- 

2. **Vinter i Madrid**  
av C.J. Sansom  
2008 (Heftet) Se alle utgaver (4)  
Likte du denne? Vi har bokanbefalinger!

2 397 0 15 (4%) 4,32 (107) 5

Bildene er hentet 14. mars og 15. mai 2011 fra Heffalumpen5 sin brukerprofil på Bokelskere.no på nettet: <http://bokelskere.no/bok/shantaram/200645/anbefalinger/>

## **Vedlegg 8: Datamodellen til Bokelskere.no**

### **Utgavekopier:**

'id', 'bokhylle\_id', 'tilstand\_id', 'utgave\_id', 'user\_id', 'favoritt'

### **Terningkast:**

'id', 'utgave\_id', 'user\_id', 'terningkast'

### **Users:**

'user\_id', 'fodselsaar', 'kjonn', 'fylke'

### **Tilstander:**

'id', 'navn'

### **Utgaver:**

'id', 'skaper\_id', 'isbn13', 'utittel1', 'hovedtittel', 'verk\_id'

**Aktører** (skaper, altså som regel forfatter av en utgave):

'id', 'fullt\_navn', 'bdnavn', 'rolle'

### **Omslagsbilder:**

'utgave\_id', 'url\_stor', 'url\_medium', 'url\_liten'

### **Emneord:**

'utgave\_id', 'user\_id', 'tag'

### **Bokhyller:**

'user\_id', 'navn', 'id'



## Vedlegg 9: Utgavekopi-håndtering av *Stolthet og fordom*

stolthet og fordom

bokelskere.no

Viser 1 til 3 av 3 treff på stolthet og fordom.

- 

**Pride and prejudice**  
av Jane Austen  
2003 (Heftet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

1 933 7 94 (10,1%) 5,37 (169) 3
- 

**Stolthet og fordom**  
av Jane Austen  
2008 (Innbundet)

0 36 1 4 (11,1%) 5,25 (12) 1
- 

**Min mangel på stolthet og fordom i jakten på Mr. Darcy**  
av Ingrid Schibsted Jacobsen  
2011 (Innbundet)

0 4 1 0 (0%) 6,00 (1) 1

**Stolthet og fordom**  
*Verket*  
20 forskjellige utgaver  
i en felles post.

**Stolthet og fordom**  
*Emma*  
*Overtalelse*  
Samleutgave med  
tre bøker


**Min mangel på**  
**stolthet og**  
**fordom i jakten på**  
**Mr. Darcy**

Post 1 viser *Verket* ”Stolthet og fordom”.


Etter å ha klikket seg inn på: ”alle utgaver” lenken:

Utgaver av verk

Viser 1 til 10 av 20 utgaver:

- 

**Stolthet og fordom**  
av Jane Austen  
2003 (Innbundet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

2 982 3 97 (10,1%) 5,35 (181) 4
- 

**Pride and prejudice**  
av Jane Austen  
2003 (Heftet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

2 982 3 97 (10,1%) 5,35 (181) 4
- 

**Stolthet og fordom**  
av Jane Austen  
2005 (Heftet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

2 982 3 97 (10,1%) 5,35 (181) 4

- 

**Stolthet og fordom**  
av Jane Austen  
2006 (Compact disc) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

2 982 3 97 (10,1%) 5,35 (181) 4
- 

**Pride and prejudice**  
av Jane Austen  
2005 (Innbundet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

2 982 3 97 (10,1%) 5,35 (181) 4
- 

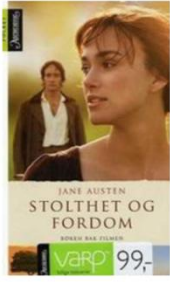
**Pride and prejudice**  
av Jane Austen  
2007 (Heftet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

2 982 3 97 (10,1%) 5,35 (181) 4
- 

**Stolthet og fordom**  
Elizabeth  
av Jane Austen  
1991 (Heftet) Se alle utgaver (20)  
[Likte du denne? Vi har bokanbefalinger!](#)

Til sammen vises 20 poster med *Uttrykk* av ”Stolthet og fordom” på Bokelskere.no kalt utgaver.

Etter å ha klikket seg inn på en utgave:



**Stolthet og fordom**  
av [Jane Austen](#) (forfatter) og [Merete Alfsen](#) (oversetter).  
362 sider

Likte du denne? Vi har bokanbefalinger!

Kjøp hos [bokkilden.no](#)

Kjøp hos [bokdykk.no](#)

Emneord: 1001 bøker, biblio1, filmatisert, Kjærlighet, klassiker, Overklassen, oversatt, Roman, verdensbiblioteket

Finn denne boka på ditt nærmeste bibliotek:

[rediger](#) | [mer bokdata](#)

[oppgi emneord for denne boka](#)

Legg inn omslagsbilde  
Se alle omslagsbilder (2)

finnes i **982** boksamlinger | nevnt **14** ganger

Aschehous 2005 utgave av ”Stolthet og fordom” er en *manifestasjon* på Bokelskere.no kalt utgavekopi.

Boken legges inn i brukeren Heffalumpen5 sin boksamling:



**Stolthet og fordom**  
av [Jane Austen](#) (forfatter) og [Merete Alfsen](#) (oversetter).  
Likte du denne? Vi har bokanbefalinger!

Finn denne boka på ditt nærmeste bibliotek:

Emneord: 1001 bøker, biblio1, filmatisert, Kjærlighet, klassiker, Overklassen, oversatt, Roman, verdensbiblioteket

Finn denne boka på ditt nærmeste bibliotek:

[rediger](#) | [mer bokdata](#)

[oppgi emneord for denne boka](#)

Bokhylle:  + [opprett ny](#)

Ditt terningkast: (fjern)

har lest

Legg inn omslagsbilde  
Se alle omslagsbilder (2)

finnes i **982** boksamlinger | nevnt **14** ganger

Heffalumpen5 har nå et *eksemplar* av ”Stolthet og fordom” i sin boksamling, på Bokelskere.no kalt utgavekopi.

## Vedlegg 10: Omtale

Omtale for *Da-Vinci koden*:

### Omtale fra Den Norske Bokdatabasen:

Når museumsкуратор på Louvre, Jacques Sauniere, blir brutalt myrdet, tilkalles Robert Langdon, som er professor i religiøs symbolologi ved Harvard-universitetet. Ved liket er det en uforståelig kode, og sammen med den franske kryptologen Sophie Neveu finner Langdon spor som fører dem til Da Vincis verker og den hemmelige Sion-ordenen. Boka er filmatisert. Dette er den andre boka om Robert Langdon.

Omtale for *Menn som hater kvinner*:

### Omtale fra Den Norske Bokdatabasen:

Finansjournalisten Mikael Blomkvist trenger en pause, og bestemmer seg for å ta et avbrekk fra jobben i tidsskriftet Millennium. Samtidig med dette, får han et oppdrag fra Henrik Vanger, tidligere en av Sveriges fremste industriledere. Han ønsker at Blomkvist skal skrive Vanger-slektens historie, men det viser seg raskt at dette er et skalkeskjul for at han ønsker å finne ut hva som skjedde med sin unge slektning, Harriet, som har vært sporløst forsvunnet i snart førti år. Sammen med Lisbeth Salander graver Blomkvist i familien Vangers fortid, og de avdekker mørkere og blodigere hemmeligheter enn noen av dem kunne ane. Dette er den første romanen om Blomkvist og Salander.

### Omtale fra forlaget:

Stieg Larssons første kriminalroman hylles av en enstemmig svensk presse. Økonomijournalisten Mikael Blomkvist trenger et avbrekk. Han er nettopp dømt til tre måneders fengsel for å ha ærekrenket finansmannen Hans-Erik Wennerström. Han bestemmer seg for å ta en pause fra jobben som redaktør i tidsskriftet Millennium. Omtrent samtidig får han et uvanlig oppdrag: Henrik Vanger, tidligere en av Sveriges fremste industriledere, vil at Blomkvist skal skrive Vanger-slektens historie. Men det viser seg snart at familiekroniken bare er et dekke for Blomkvists virkelige oppdrag: å finne ut hva som hendte med Henrik Vangers unge slektning Harriet, som har vært sporløst forsvunnet i snart førti år, noe Henrik Vanger aldri har kunnet slutte å gruble over. Blomkvist påtar seg motvillig oppdraget og får hjelp av Lisbeth Salander, en mutt, tatovet kvinne på 24, med piercing og et stort talent for å irritere medarbeidere. Men Salander er også en fremragende researcher og en av landets fremste hackere. Sammen graver Blomkvist og Salander stadig dypere i familien Vangers fortid og finner en historie som er både mørkere og blodigere enn noen av dem kunne ane. «Genuin og genial» er blant de uttrykkene som brukes av en enstemmig svensk presse om Stieg Larssons debut som kriminalforfatter. *Menn som hater kvinner* er den første i en serie på tre kriminalromaner om Mikael Blomkvist og Lisbeth Salander.

## Vedlegg 11: Bokhyllenavn

Oversikt over bokhyller som har vampyrrelaterte navn på Bokelskere.no:

---

vam	Vamp	Vamp & Were
Vampire	Vampire academy	vampire and fantasy
Vampire books	Vampire Diaries	Vampire Diaries, The
Vampire Game	Vampire Love <3	Vampire Queen
Vampire romance	Vampire Series	Vampire stuff
Vampire/Immortal	Vampires	Vampires :)
Vampires n' such	Vamps	Vampyr
Vampyr & Fantasi	Vampyr & varulv	Vampyr bøker
Vampyr bøker <3	Vampyr etc.	Vampyr hylla mi
Vampyr litt.	Vampyr litteratur ++	Vampyr o.l.
Vampyr og kjæleik	Vampyr-bøker	Vampyr-litteratur
Vampyr-romantikk	Vampyr, etc.	Vampyr(ish)
Vampyr/fantasy o.l.	vampyr/mystery	Vampyr+
Vampyrbøker	vampyrbøker/Twilight	Vampyrcrap
vampyrdritt	Vampyrer	Vampyrer & Demoner
Vampyrer & Horror	Vampyrer & Vandøde	Vampyrer og monstre
Vampyrer og slikt	Vampyrer og snacks	Vampyrer og stæsj
Vampyrer og varulver	vampyrer ol	Vampyrer ol.
Vampyrer vihss	Vampyrer, overtro osv	Vampyrer/Zombier
Vampyrgreier	vampyrgøy	vampyrhylla
vampyrisme	Vampyrkjærleik	vampyrkliss
Vampyrlektyre	Vampyrlitteratur	Vampyrmoro
Vampyrmorro	Vampyrrelatert	Vampyrroman
Vampyrromantikk	Vampyrserier	Vampyrteori
Vandrehistorier	vandringsbøker	Vanlig, 0 fantasy

Bilder er hentet 16. juni 2011 fra bokhyller på Bokelskere.no fra <http://bokelskere.no/bokhyller/>

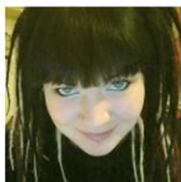
## Vedlegg 12: Brukere med flest like bøker

Bokelskere med flest like bøker i sin samling

hva er dette?

Dette er en liste over andre bokelskere som har utgaver av de samme verkene som du har, sortert etter hvor mange felles verk dere har i boksamlingen.

Lista genereres en gang i døgnet.



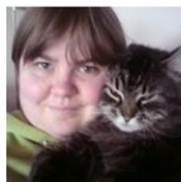
Tone Lilleenget By  
verk felles: 260



bokelskerinnen  
verk felles: 243



Bokoholiker  
verk felles: 224



Linn Charlotte  
Pettersen  
verk felles: 211



Mona Elizabeth  
verk felles: 204



Hege  
verk felles: 204



Hilde H H  
verk felles: 197



Anne-Kirstin Haugen  
verk felles: 192



Marielle  
verk felles: 186



Lene Andresen  
verk felles: 181



Sol-Angelica  
Michelle Skipnes



Christine Meisterlin  
verk felles: 178

Bildet viser de brukerne som har flest like utgavekopier med Heffalumpen5.

Bildene er hentet 14. juni 2011 fra Heffalumpen5 sin brukerprofil på Bokelskere.no.