

Joachim Rustad Fugleberg

**Automatisk klassifikasjon av bøker
basert på brukeranmeldelser**
Et konsept

Sammendrag

Bøker har mange egenskaper utover de konkrete egenskapene, som bl.a. tittel, forfatter og emneord. Bøker vekker også følelser og reaksjoner hos leseren. Denne oppgaven tar for seg hvordan følelsene og reaksjonene - også kalt de subtile egenskapene - kan brukes til å anbefale litteratur. Videre er de subtile egenskapene implementert i et gjenfinningssystem for litteratur, slik at de kan søkes opp på samme måte som man slår opp emneord. Det har blitt benyttet et datasett med representasjoner av 2,8 millioner bøker med både ordinære bokdata og brukergenerert innhold fra henholdsvis Amazon.com og LibraryThing.com. Brukergenerert data er ment å skulle avdekke de subtile egenskapene ved bøkene. Undersøkelser har likevel vist at brukergenererte data som anmeldelser, tagger og tallkarakterer i liten grad formidler følelser og reaksjoner. Derimot er brukeranmeldelser bedre egnet for å avdekke bøkens appell. Jeg har basert meg på Joyce G. Saricks rammeverk for bøkens appell og utviklet et konsept ved hjelp av verktøyene Solr og OpenNLP.

Abstrakt

There are many qualities to books besides their specific characteristics, such as title, author and subject. Books also evoke the readers emotions. This thesis explores how emotions and other subtle qualities can be discovered in user generated data and subsequently used in a system for automatic classification of books. Furthermore, I have developed a system based on Joyce G. Saricks model for classifying the appeal of books by using the tools Solr and OpenNLP. The thesis is based on a data set of 2,8 million books, including their regular characteristics as well as user generated data. These data sets have been collected from Amazon.com and LibraryThing.com. Results indicate that user generated data such as tags and ratings only to a limited extent provide a good picture of a books appeal. On the other hand, reader's reviews seem to be better suited to describe the emotions evoked by a certain book.

Innholdsfortegnelse

1	INNLEDNING	6
1.1	BAKGRUNN	6
1.2	PROBLEMSTILLING	7
1.3	OPPGAVEN I KORTE TREKK	8
2	TEORETISK FORANKRING OG TIDLIGERE FORSKNING	10
2.1	HUMØR OG KONTEKST	10
2.2	BØKER OG EGENSKAPER	10
2.2.1	<i>Andre domener</i>	13
2.3	FORSKJELLEN PÅ EGENSKAPER (KORT DRØFTING)	14
2.4	BRUKERGENERERTE DATA	15
2.4.1	<i>Brukeranmeldelser</i>	15
2.4.2	<i>Tagger</i>	17
2.4.3	<i>Brukervurderinger</i>	18
2.5	ET RAMMEVERK FOR APPELL	19
2.5.1	<i>Appell</i>	20
2.5.2	<i>Appellelementer</i>	21
2.5.2.1	Tempo	21
2.5.2.2	Karakterisering	23
2.5.2.3	Ramme	25
2.5.2.4	Handling	26
2.5.2.5	Oppsummering av appellelementer	27
2.5.3	<i>Vokabular, termer og eksempler</i>	27
2.6	TEKSTKLASSIFIKASJON	30
2.6.1	<i>Klassifikasjon og kategorisering</i>	31
2.6.1.1	Trening	31
2.6.1.2	Modellbygging	32
2.6.1.3	Overtilpasning	33
3	METODISK TILNÆRMING	34
3.1	DATASETT	34
3.2	VERKTØY	36
3.2.1	<i>Solr</i>	36
3.2.1.1	Skjema	38
3.2.1.2	Konfigurasjonsfil	40
3.2.2	<i>OpenNLP</i>	40
3.2.2.1	Setningsidentifikasjon	40
3.2.2.2	Tekstkategorisering med OpenNLP	41
3.3	TILPASSNING AV SARICKS RAMMEVERK	41
3.3.1	<i>Tilpassning av elementet tempo</i>	42
3.3.2	<i>Tilpassning av elementet karakterisering</i>	42
3.3.3	<i>Tilpassning av elementet ramme</i>	43
3.4	UTARBEIDELSE AV EN MODELL	43
3.4.1	<i>Modell for tempo</i>	44
3.4.2	<i>Modell for karakterisering</i>	45
3.4.3	<i>Modell for ramme</i>	45

4	ANALYSE AV UNDERSPØRSMÅL 1	45
4.1	HVILKE TYPER BRUKERGENERERTE DATA ER HENSIKTSMESSIG Å BENYTTE FOR Å AVDEKKE UNDERLIGGENDE SUBTILE EGENSKAPER I ET BOKVERK?	45
4.1.1	Tagger.....	45
4.1.2	Brukervurderinger.....	48
4.1.3	Brukeranmeldelser.....	50
5	GJENNOMFØRING	51
5.1	FELTER	52
5.2	KONFIGURASJON	53
5.3	UTVIDELSE AV SOLR.....	54
5.3.1	Integrering av OpenNlp i Solr.....	54
6	OPPSUMMERING	55
6.1	RESULTATER OG DISKUSJON	56
6.1.1	Hvilke typer brukergenererte data er hensiktsmessig å benytte for å avdekke underliggende egenskaper i bøker?.....	56
6.1.2	Hvordan kan automatisk klassifisering av underliggende egenskaper knyttet til underliggende egenskaper i bøker realiseres?.....	57
7	FORSLAG TIL VIDERE UNDERSØKELSE	57
8	LITTERATURLISTE	59
9	VEDLEGG	64
9.1	VEDLEGG 1.....	64
9.2	VEDLEGG 2.....	70
9.3	VEDLEGG 3.....	72
9.4	VEDLEGG 4.....	73
9.5	VEDLEGG 5.....	74
9.6	VEDLEGG 6.....	75
9.7	VEDLEGG 7.....	77
9.8	VEDLEGG 8.....	92
9.9	VEDLEGG 9.....	99
9.10	VEDLEGG 10.....	104
9.11	VEDLEGG 11.....	108
9.12	VEDLEGG 12.....	109
	Tabell 1: Invertert indeks (Grainger & Potter, 2014, s. 54).....	37
	Tabell 2: Indikatorer for tempo	42
	Tabell 3: Indikatorer for karakterisering.....	43
	Tabell 4: Indikatorer for ramme	43
	Tabell 4: Ti mest brukte tagger i datasett fra INEX.....	46
	Tabell 5: Fordeling av tagger.....	48
	Tabell 6: Brukervurderinger	49

Figur 1: The Vocabulary of Appeal (Saricks, 2005, s. 66).....	29
Figur 2: Stadiene i utviklingen av et klassifikasjonssystem (Owen, 2012, s. 233).....	32
Figur 2: Xml-elementer i datasett fra INEX (Koolen, Kazai, Preminger, & Doucet, udatert)	35
Figur 3: Dokumenter i MySQL-database	35

1 Innledning

Motivasjonen for denne oppgaven var i utgangspunktet litteraturanbefalinger basert på følelser og stemninger som et alternativ emneord og sjangerbetegnelser for lesere. Bøker består av mer enn bare tittel, sjanger og andre kvantifiserbare data. De vekker følelser og reaksjoner hos lesere og innehar således underliggende egenskaper som ikke er umiddelbart identifiserbare. En bok kan eksempelvis være skummel, morsom eller så engasjerende at den er umulig å legge vekk. Noen spørsmål gjorde seg raskt gjeldende. Lesere er ikke like og bruker forskjellige kriterier når bøker skal velges og har lesere i det hele tatt en klar strategi. Det finnes ikke et felles språk for hvilke erfaringer lesere har med en bok og hvordan bøker beskrives er sentralt når det skal avdekkes affektive egenskaper. Etter en kartlegningsperiode ble det klart at brukergenererte data, også kjent under betegnelsen sosiale data, kunne være en kilde til å avdekke hvordan lesere oppfatter bøker på et emosjonelt plan. For å kunne utnytte det potensialet som muligens finnes i brukergenererte data trengs det et datamateriale. Ved skolen er det tilgang til et større datasett med både ordinære bokdata og innhold skapt av lesere. Brukergenererte data finnes i mange former og alle er ikke nødvendigvis like anvendelige. De må finnes i en viss mengde, utfylle de ordinære bokdataene og gi uttrykk for de underliggende egenskapene ved en bok. En gjennomgang av de tilgjengelige brukergenererte dataene er derfor påkrevd. Noen kan utelukkes fordi mengden ikke er tilstrekkelig. Andre data, tagger, brukeranmeldelser og brukervurderinger i form av tallkarakterer finnes i rikelig antall om må undersøkes nærmere.

Det var opprinnelig tenkt å anvende brukergenererte data i en anbefalingstjeneste men det ble fort tydelig at arbeidet måtte begynne langt tidligere i prosessen enn ved det sluttproduktet en anbefalingstjeneste kan sies å være. Det var likevel et ønske at oppgaven skulle ha et praktisk element for å utforske hvordan de egenskapene som eventuelt avdekkes gjennom en undersøkelse av brukergenererte data kan systematiseres og brukes til å merke bøker slik at de kan brukes til gjenfinning eller anbefalinger?

1.1 Bakgrunn

Utvalget og tilgangen på bøker har aldri vært større og lesere kan glede seg over ikke bare flere bøker men også over informasjonen om dem som finnes tilgjengelig både på spesialiserte nettsamfunn og kommersielle webbaserte bokhandlere. Samtidig har det kanskje aldri vært vanskeligere å velge en bok. Det er ikke lenger nok å se etter bøker innen et viss

sjanger eller med et visst emne for utvalget er uoverkommelig. Paradoksalt nok fører for mye informasjon til at vi i liten grad utvider vår egen horisont. I en tid hvor teknologien setter oss i stand til holde oss informert har vi samtidig problemer med å holde oss oppdatert (Andrejevic, 2013, s. 2). Det er umulig å holde følge og noen valg må tas. Søkemotorer gir en enkel inngang til å finne informasjon, men fordi mengden informasjon er uoverkommelig er vi tvunget til å innskrenke søket. Kildene vi er vant med og liker har blitt mer tilgjengelig og oppsøkes uten at vi ser etter nye (Baez, Birukou, Casati, & Marchese, 2010, s. 31). Det samme er tilfelle for nettsamfunn og anbefalingstjenester når brukere kan finne spesialiserte brukergrupper med like interesser som en selv og får anbefalt et snevert utvalg basert på egne preferanser, den såkalt bobleeffekten (Trieu, 2011). I en amerikansk undersøkelse fra så langt tilbake som 2008, en evighet i dagens informasjonssamfunn, svarte en urovekkende høy andel av respondentene at de stolte mer på alternative nettbaserte kilder enn på tradisjonelle informasjonskanaler (Andrejevic, 2013, s. 3).

For lesere innebærer dataoverbelastningen at det er lettere enn noen gang å holde seg innenfor sin nisje og unngå å bli eksponert for nytt materiale. Tim Spalding fra nettsamfunnet LibraryThing.com bekrefter dette ved å forklare at lesere velger bøker de tror kommer til å falle i smak (Syversen, 2011, s. 54). Til tross for dette er det egne forum hos nettsamfunn for bøker hvor brukere etterspør ny litteratur som kan falle i smak. Ekstra interessant er det at brukere ofte etterspør bøker med en viss stemning eller følelse.

1.2 Problemstilling

I oppgaven er det arbeidet med følgende problemstilling:

Kan brukergenerert data brukes til å automatisk avdekke underliggende elementer ved en bok og brukes i søk etter litteratur?

Problemstillingen er åpen og lite spisset og skyldes at oppgaven først og fremst er utforskende. Målet er ikke å komme frem til absolutte konklusjoner, men å se nærmere på forholdet mellom bøker og brukergenererte data. Den overordnede problemstillingen har ledet til to underspørsmål som må besvares:

- Hvilke typer brukergenererte data er hensiktsmessig å benytte for å avdekke underliggende egenskaper i bøker?
- Hvordan kan automatisk klassifisering av underliggende egenskaper knyttet til underliggende egenskaper i bøker realiseres?

1.3 Oppgaven i korte trekk

I kapittel 2 gjøres det rede for oppgavens teoretiske forankring og det blir en gjennomgang av tidligere forskning. Spesielt er det sett på tidligere forskning knyttet til andre aspekter ved bøker og hvordan forskjellige kilder kan brukes for å utnytte disse. Det blir videre en utforskning av bøkens ulike egenskaper med kort drøfting. Brukergenererte data blir definert og tagger, brukervurderinger og brukeranmeldelser gjennomgått for å identifisere hvilke områder som kan være problematiske eller fordelaktige. I kapittel 2 blir det også gitt en utfyllende oversikt over et utvalgt rammeverk som kan brukes til å avdekke egenskapene presentert tidligere. Mot slutten av kapittelet blir det sett nærmere på automatisk tekstklassifisering for å danne et grunnlag for den praktiske komponenten i oppgaven.

I kapittel 3 blir det gitt en oversikt over datasettet med en gjennomgang innholdet. Deretter blir verktøyene Solr og OpenNLP brukt i den praktiske gjennomføringen presentert. Fremgangsmåten for hvordan rammeverket i kapittel 2 kan tilpasses den praktiske gjennomføringen av underspørsmål to blir lagt frem. Tilslutt i kapittelet blir det gjort greie for utarbeidelsen av modellene som er brukt for klassifisering i den praktiske gjennomføringen.

I kapittel 4 blir underspørsmål en besvart med en diskusjon av fordeler og ulemper ved de utvalgte brukergenererte dataene fra kapittel 2.

I kapittel 5 blir en gjennomgang av den praktiske utførelsen av underspørsmål to. Gjennomgangen forklarer hvordan Solr er konfigurert, indekseringsprosessen og hvordan hvert enkelt dokument klassifiseres av OpenNLP. Det forklares ikke i detalj hvordan selve programmeringen er utført eller går inn på hvordan java fungerer. Selve programmeringen er ikke en sentral og benyttes kun som verktøy for å besvare oppgavens problemstilling. Skjønt programmeringen hadde fortjent større plass med tanke på tidsbruk og læringskurve. Det anbefales ikke å gjennomføre et lignende prosjekt uten solide ferdigheter innen programmering. Alle programmering er lagt ved oppgaven og blir referert til når det er aktuelt. Det er også satt opp en server på skolen med fullt administrasjonsgrensesnitt hvor det er mulig å se resultatet av gjennomføringen. Følgende adresse kan benyttes:

<http://bibin.hioa.no:8090/solr-s162452/#/nlpCat2/query>. Lenken leder til grensesnittet for søk og er forholdsvis intuitiv. Enklest er det å trykke på knappen *Execute query* som leder til en

treffliste med ti poster. Andre valg kan stå urørt. Det indekserte materialet fungerer også som underlagsmateriale for diskusjonen i kapittel 4.

I kapittel 6 blir det forsøkt å samle trådene fra teori- og analysekapittelet. Resultatene blir kort oppsummert etterfulgt av en diskusjon for det i kapittel 7 gis noen forslag til videre undersøkelser.

2 Teoretisk forankring og tidligere forskning

2.1 Humør og kontekst

Lesere er ikke en homogen gruppe med like preferanser og skulle to lesere ha like preferanser vil sjelden konteksten være den samme. En god bok for leser A er ikke nødvendigvis en god bok for leser B. To undersøkelser (Weaver, 2007) (Ross, 2000) forsøker å avdekke hva som ligger bak en lezers valg av bok. Ross oppdager at lesere har vanskelig for å gi en god forklaring på hvordan de velger bøker og at de sjelden har en klar strategi (2000, s. 11), men begge undersøkelsene kommer likevel frem til et tilnærmet likt resultat. Weaver har bedt respondentene liste opp viktige kriterier og 50 prosent vurderte humør til å være det viktigste kriteriet (2007, s. 583). Ross har i motsetning ikke bedt respondentene liste opp viktige kriterier, men kommer frem til at grunnlaget for et bokvalg er humør (2000, s. 13). Humør går igjen i begge undersøkelser som enten grunnleggende for valg av bok eller som et viktig kriterium. Weaver definerer humør til å være: en manifestasjon av leserens kontekst. Definisjonen er kort og hva som er leserens kontekst er uklart. Med kontekst mener han alle de omliggende faktorene som influerer humøret, og bokvalget. Det kan for eksempel være leserens livssituasjon eller om bokvalget har et bestemt formål; eksempelvis en bok til lesesirkelen eller til et barn (2007, s. 583). Ut i fra Weavers beskrivelse er en mulig tolkning at det ikke nødvendigvis er det spesifikke humøret på tidspunktet for bokvalget som er styrende, men det mer generelle humøret og omstendighetene i leserens liv. Generelt er det vanskelig å finne enighet om hva følelser og tilhørende aspekter, slik som humør, innebærer. Allerede i 1981 var det nesten hundre registrerte definisjoner av følelse og i dag anvendes termer relatert til følelser innen mangfoldige fagområder med ulike oppfatninger av innholdet (Picard et al., 2004, s. 254). For å kunne arbeide med humør er det nødvendig med en presis forståelse av begreper og meningsinnhold i forhold til aktuelt domene det jobbes med samt se på tilnærmingen til tidligere undersøkelser.

2.2 Bøker og egenskaper

En bok har ulike bibliografiske data knyttet til seg. Forfatter og tittel er et minimum, men det kan også være undertittel, emneord, sjanger, språk, tidsperiode, oversetter, redaktør og forlag. Listen er ikke uttømmende og det er mulig å benytte flere bibliografiske data enn det er hensiktsmessig å liste opp. Bibliografiske data er det Pia Syversen i sin masteroppgave om anbefalingssystemer for litteratur (Syversen, 2011) kaller uttrykk for konkrete egenskaper. Konkrete egenskaper er oversiktlige, målbare, forholdsvis enkle å analysere og bruke med metoder fra informasjonsvitenskapen. Konkrete egenskaper alene er derimot ikke nok til å

beskrive en bok til tross for at det ofte, og spesielt i bibliotek, brukes emneord. En bok innehar også subtile egenskaper som ifølge Syversen kan si mer om boken som en gjenstand for brukeren, ment som egenskapene ved en bok som gjorde at en bestemt leser likte den. Subtile egenskaper er, i motsetning til konkrete egenskaper, vanskelig å måle, analysere og oppdage. Subtile egenskaper skjuler seg i teksten og kan oppdages ved å sette seg inn i boken. Syversen skriver ikke mer om hvordan disse egenskapene kan finnes utover å se på tagger som et alternativ til emneord for å formidle følelser, og at brukeranmeldelser kan være en kilde for nevnte egenskaper. Hun er derimot tydelig på at innsikt i hvorfor en bok appellerer til en leser, de subtile egenskapene som gjorde at leseren likte den, er viktig i et anbefalingssystem for litteratur (s. 50 - 53).

Koolen, Kamps, & Kazai (2013) ser, i likhet med Syversen, på tagger og brukeranmeldelser som en mulig kilde til det forfatterne av artikkelen kaller affektive aspekter ved en bok. Med affektive egenskaper menes det i artikkelen at lesere vurderer en bok ut fra hvor interessant, velskrevet, morsom eller populær den er, og ikke emnet alene (s. 38). Undersøkelsen forsøker å avdekke om brukeranmeldelser av bøker kan være nyttige i et gjenfinningsperspektiv, både ved å gjennomføre søkeeksperimenter og ved å be brukere vurdere data om en bok er relevant i forhold til et gitt emne og om boken eventuelt kan anbefales til en leser på jakt etter bøker om det gitte emnet. Resultatene viser at brukeranmeldelser både øker gjenfinningseffektiviteten og at brukere i større grad er i stand til å avgjøre om et dokument er relevant når minst en brukeranmeldelse er tilgjengelig.

I artikkelen *Using Collaborative Filtering in Social Book Search* (Hurdeman, Kamps, Koolen, & van Wees, 2012) sammenliknes innvirkningen av brukeruavhengige popularitetsmål i litteratursøk med anbefalinger basert på brukerprofiler. I artikkelen skilles det mellom tradisjonelle ordinære bokdata og brukergenererte data. Det distingveres mellom typer av brukergenererte data som i følge forfatterne er henholdsvis avhengige og uavhengige av den individuelle brukeren. Brukeruavhengige data er den generelle meningen til en gruppe lesere om en bok og kan finnes i anmeldelsene, karakterene, samt taggene som er tilknyttet en bok. Brukeravhengige data er derimot en brukers personlige smak og finnes i brukerprofiler, navigerings- og kjøpshistorikk og kan fortelle om det forfatterne kaller *andre aspekter ved relevans*. I undersøkelsen blir det gjort forsøk på finne ut om popularitet og karakterer er viktige kriterier når brukere skal velge en bok. Det antas at bøker med flere brukeruavhengige data og høyere karakter er mer populære enn bøker med færre tilsvarende data. Fordi

Huurderman et al. gjennomfører et eksperiment hvor målet er å undersøke hvor stor betydning popularitet har for bokanbefalinger, benyttes brukergenererte data kun som et mål for popularitet og brukeruavhengighet. Det betyr ikke at forfatterne utelukker brukergenererte data som relevant i andre sammenhenger. Det presiseres i artikkelen at tilgangen på brukeranmeldelser, karakterer og tagger lar lesere finne ut mer om en bok enn ordinære bokdata. Resultatene fra eksperimentene tyder på at popularitet alene ikke er effektivt for litteratursøk i sosiale sammenhenger.

The Book Genome Project ("Book Genome," udatert) analyserer innholdet i bøker for å avdekke underliggende strukturer slik at subtile likheter kan avsløres (thetextbookguru, n.d.). For å demonstrere konseptet har prosjektet opprettet siden booklamp.org som genererer bokforslag basert på inntastingen av forfatternavn eller boktittel. I følge grunnleggeren, Aaron Stanton, er prosjektets styrke at det ikke brukes noen form for menneskelig analyse og at analysen dermed blir objektiv (Herther, 2012)("BookLamp's FAQ," udatert). Stanton forklarer at fordelene med en objektiv tilnærming fremfor brukergenererte data er økt serendipitet. Booklamp.org beskriver han som en venn av bøkene på som ikke er på forlagenes topplister, målet er å analysere og gjøre flest mulig bøker oppdagbare. Sosiale nettsamfunn lider av det Stanton kaller "det sosiale hullet" og mener at en innholdsanalyse av alle bøker tilgjengelig kan være en del av løsningen (Herther, 2012, s. 46). Problemet han beskriver er velkjent; noen få populære bøker har tilknyttet mye data mens et stort antall bøker ikke har data. Koolen et al. peker på det samme i arbeidet med sitt datasett (2013, s. 38). Utfordringen er generell for anbefalingssystemer og er kjent som *The data sparsity problem* og *The Curse of Dimensionality* (Jannach, Zanker, Felfernig, & Friedrich, 2011, s. 24) (Ricci, Rokach, Shapira, & Kantor, 2011, s. 44). På prosjektets faq-side kommer det likevel frem at bare en liten del av bøkernes appell kan avdekkes gjennom objektiv analyse. Målet er først og fremst å hente ut mest mulig om innholdet i bøkene som stilles til rådighet for lesere ("BookLamp's FAQ," udatert). For å markere forskjellen mellom ordinære metadata og andre egenskaper ved bøker introduserer Stanton begrepet *multidimensjonale metadata*: "Any metadata that is a generational leap in DEPTH and SCOPE beyond the capabilities of a publisher to assign manually, or the crowd to describe effectively" ("Understanding the Book Genome Project," 2012). Begrepet har tidligere blitt benyttet i andre sammenhenger, blant annet i forbindelse med databaseteknologier, men kan ikke sammenlignes til tross for noen paralleller. Uavhengig av begrepet er Stantons viktige poeng at en bok har egenskaper utover det som gjenspeiles i tradisjonelle metadata. The Book

Genome Project sammenligner bøkens oppbygning med DNA og understreker dermed at bøker brytes ned i enkelte bestanddeler i analyseprosessen. Hvordan analysen gjennomføres i detalj forklarer ikke prosjektet nærmere enn at det er komplisert, teknisk, omfattende og resultatet av en utvikling gjennom flere år (“BookLamp’s FAQ,” udatert). På grunn av sparsommelig informasjon blir forklaringen vider også udefinert. Prosjektet har definert tre hovedelementer bøker kan bestå av; språkDNA, historieDNA og karakterDNA. Til hvert av de tre elementene har det blitt laget en genstruktur med målinger av spesifikke underelementer. SpråkDNA, som er det eneste delvis forklarte elementet, består av tempo, perspektiv, beskrivelse, tetthet, bevegelse og dialog som inneholder kombinasjoner av *gener* som fanger ulike aspekter ved språket. Hva kombinasjonen av gener innebærer er ikke klart, men beskrivelsen antyder at teksten brytes ned i mindre bestanddeler og kategoriseres. *Gene* i hver bok blir deretter sammenliknet med bøkene i korpuset (“Book Genome,” udatert).

2.2.1 Andre domener

Innenfor film- og bildegjenfinning har det blitt gjort lignende undersøkelser og forsøk på å utnytte andre egenskaper ved et verk. Hanjalic & Li-Qun Xu (2005) undersøker hvordan det forfatterne kaller *affective video content* kan brukes til å finne det underliggende humøret i en video slik at videotjenester kan tilpasses individuelle seere. I artikkelen differensieres oppfattelsen av video i *cognitive level* og *affective level*. Den førstnevnte er fakta om innholdet i en video. Eksempler kan være scener merket med lokasjon, skuespillere, vær, dialog og emne. *Affective level* defineres av Hanjalic et al. som mengden og type affekt, altså følelser en video rommer og som seeren er forventet å oppleve. I en artikkel om anbefalingssystemer for bilder brukes benevnelsene *generic metadata* og *affective metadata* for å distingvere mellom metadata fra innholdsprodusentene og metadata relatert til brukernes personlige og emosjonelle reaksjoner (Tkalčič, Burnik, & Košir, 2010). Bruken av ordet *affective metadata* videreføres i en avhandling (Kindt, 2013) om humørbaserte anbefalingssystemer for filmer. I Kindts prosjekt får hver film tilknyttet en *affective profile* basert på en kombinasjon av humor-tagger fra Jinni.com og sjanger-tagger fra IMDB.com. Jinni.com er en webside for filmanbefalinger og bruker noe de kaller *The Entertainment Genome Project*. Jinni deler ikke informasjon om fremgangsmåten utover at egenskapene ved en film grovt deles inn i to kategorier; opplevelse og historie. Videre analyseres anmeldelser, brukeranmeldelser og metadata ved prosessering av naturlig språk (Glick, 2014).

2.3 Forskjellen på egenskaper (kort drøfting)

Affektive aspekter, affektive metadata, affektivt innhold, subtile egenskaper, bokDNA, multidimensjonale metadata og beslektede uttrykk kan betraktes som forskjellige termer for egenskaper ved et verk som ikke innbefattes av ordinære bibliografiske data. I artiklene, oppgavene og nettstedene beskrevet over skilles det skarpt mellom bibliografiske data og data knyttet til andre egenskaper ved et verk. Felles er forståelsen av at det som kan kalles ordinære metadata, ikke kan dekke brukernes reelle informasjonsbehov. Verk, enten det er bøker, filmer eller bilder, innenfor samme emne kan være svært forskjellige og dermed appellere til ulike brukere og humør. Forståelsen av hva andre data knyttet til andre egenskaper er derimot ikke like entydig. Syversen (2011) har en bred definisjon av subtile egenskaper som hun mener har betydning for relasjonen mellom bok og leser. De subtile egenskapene kan gjenfinnes i bokteksten eventuelt i omtale av boken. Koolen et al. (2013) har muligens et enda bredere men lignende perspektiv når de forklarer affektive egenskaper til å være hvor interessant, velskrevet, morsom eller populært et verk er for leseren. Stanton ("Understanding the Book Genome Project," 2012) har en annerledes tilnærming når han hevder at multidimensjonale metadata er objektive data fra teksten. Tkalčič et al. (2010) definerer *affektive metadata* noe smalere og knytter begrepet i større grad til følelsesmessige reaksjoner hos brukeren. Også Hanjalic et al. (2005) knytter sitt begrep *affektive level* sterkt til følelser, men i motsetning til Tkalčič et al. er fokuset på hvilke følelser verket rommer.

En klar, tydelig og enhetlig definisjon av hva andre metadata egentlig innebærer mangler, men muligens kan noen av trådene samles. Det store skillet mellom definisjonene er først og fremst hvor fokuset settes, enten på brukerens følelsesmessig opplevelse eller på egenskaper i selve verket. Det er stor forskjell hvilke følelser det er forventet at brukeren skal oppleve og hvilke egenskaper i verket som gjør at det blir likt av brukeren. Spesielt er dette tydelig hos Stanton som kun forholder seg til det han kaller objektive egenskaper ved selve verket. Hvilke følelser verket vekker i brukeren er ikke like fremtredende som for eksempel hos Tkalčič et al. Et forenklet eksempel kan kanskje være oppklarende; Stantons tilnærming fortelle mye om innholdet i et verk men ingenting om hvordan det vil kunne oppleves av brukeren. Det er ikke sikkert to verk med lik mengde vampyrer, skog og kjærlighet vil appellere til samme bruker. Motsatt vil løsningen til Tkalčič et al. kunne avklare mye om brukerens reaksjoner men lite om hvorfor innholdet har den egenskapen. To verk kan være merket som morsomme og likevel ha ulik appell. Syversens subtile egenskaper forener ytterpunktene ved å legge vekt på egenskaper i verket som har innvirkninger for relasjonen med brukeren.

2.4 Brukergenererte data

Sosial innflytelse spiller en stor rolle for mange av valgene vi tar og muligens er betydning større i dag enn den har vært tidligere. I motsetning til tidligere hvor tjenestene på web var innholdssentrerte er nå de fleste brukersentrerte. Vi er alle påvirket av de sosiale forbindelsene vi har. Et mye brukt eksempel er valget av kinofilm. Om du og en venn skal finne en film velges det normalt noen kandidater ut fra personlig preferanser som drama, action, skuespiller, regissør osv. Valgene kan og påvirkes av anmeldelser i forskjellige medier. Det er større sannsynlighet for at valget står mellom filmer som har fått gode anmeldelser enn filmer som har fått utelukkende dårlige anmeldelser. Om vennen du skal på kino med har en kollega som har sett og anbefalt den ene filmkandidaten sannsynligheten stor for at nettopp den velges. Valget av film har i eksempelet blitt tatt med bakgrunn i tre viktige faktorer: egne preferanser, offentlige/offisielle anmeldelser/omtale og venner. I eksempelet er det og tydelig at venneinnflytelse ikke nødvendigvis trenger å være begrenset til nære venner. De vi har som venner har igjen venner og sosiale nettverk som indirekte påvirker våre egne valg. I denne konteksten vil nære venner være de som er på brukerens venneliste. Brukere av et sosialt nettsamfunn vil normalt kun gi vurdering til et lite antall av de ressursene som faktisk er tilgjengelig. I de fleste tilfeller innebærer dette at dine nære venner ikke har vurdert den aktuelle ressursen. Så lenge nære venner ikke har vurdert ressursen kan heller ikke påvirkningen deres utnyttes. Derimot kan en venn av en venn ha vurdert ressursen. Forsøk med utnyttelse av fjerne venner har vist at dekningsgraden kan øke med hundre prosent samtidig som nøyaktigheten kun har sunket med 5 prosent (He & Chu, 2010, s. 67).

Tagger, brukeranmeldelser og vurderinger er eksempler på det som kalles brukergenererte data. Andre benevnelser inkluderer blant andre brukerskapte data, sosiale data, sosialt skapte data og brukeravhengige data som alle er eller kan være varianter av brukergenererte data. Utrykkene brukes ikke konsekvent og både i lærebøker og i forskningsartikler er anvendelsen ulik. Kanskje skyldes det siste at konteksten er viktigst for forståelsen av meningsinnholdet i anvendt begrep. I materiale på engelsk ser forkortelsen *UGC* (User Generated Content) ut til å ha festet seg som et begrep og omfatter tagger, spørringer, vurderinger etc. (Ricci et al., 2011, s. 4). I denne oppgaven benyttes brukergenererte data fordi det ikke er noen tvil om betydningen; data generert av brukere. I de neste avsnittene er en gjennomgang av brukergenererte data, brukeranmeldelser, tagger og brukervurderinger.

2.4.1 Brukeranmeldelser

Brukeranmeldelser blir av Syversen sett på som en kilde til oppdagelse av subtile egenskaper og Koolen et al. har funnet at brukeranmeldelser kan hjelpe brukere med både gjenfinning og relevansvurdering. Sistnevnte påpeker også at det ikke trengs mer enn en brukeranmeldelse for å hjelpe brukere med å avgjøre om et dokument er relevant. Brukeranmeldelser blir i denne oppgaven brukt synonymt med bokanmeldelser av lesere, altså ikke profesjonelle anmeldelser. Antallet publiserte verker fører i dag til at profesjonelle anmeldere og ordinære mediekkanaler prosentvis kun har kapasitet til å dekke en brøkdel. Chandler fra Goodreads.com hevder at brukeranmeldelser allerede er med på å løse problemet. Han poengterer samtidig at brukeranmeldelser ikke brukes på samme måte som profesjonelle bokanmeldelser og må heller sammenlignes med å få personlige boktips av venner å bekjente eller iakttagelse av hva andre leser (Herther, 2012, s. 18). Implisitt i uttalelsen kan det leses at Chandler mener lesere stoler på omgangskretsen når det kommer til bokvalg. En lignende kan finnes hos Hu ved Denver University som sier at profesjonelle anmeldelser ofte blir ansett å være av høyere kvalitet og innehar en større autoritet enn brukeranmeldelser, men brukeranmeldelser har likevel en viktig, men annerledes, rolle fordi et lite antall anmeldere ikke kan ivareta alle preferanser og smak (s. 21). Stanton fra BookLamp.org har en annen utsikt med påstanden om at det sosiale hullet fører til mindre serendipitet når bøker uten anmeldelser havner i skyggen av synligheten til populære bøker med mye data og fremhever analyse av bokinnholdet som en bedre kilde (Herther, 2012, s. 46). Det er ikke Stantons hensikt å si at brukeranmeldelser er negativt i seg selv men heller å poengtere at det finnes andre kilder kan være bedre. Trieus tanker støtter Stantsons syn og mener at måten sosiale nettsteder kategoriserer og måler likhet mellom brukere på fører til en bobleeffekt som hindrer kreativitet og fri tekning (Trieu, 2011). Trieu adresserer primært algoritmer i anbefalingstjenester hvor brukere med like preferanser blir anbefalt det samme materialet og er således data skapt ut av sosiale relasjoner, ikke av brukerne direkte. Boblemetaforen kan likevel ha gyldighet også for mangelen på brukeranmeldelser om en velger å se disse som en form for anbefalinger slik som Chandler fra Goodreads.com. Lesere med et bokønske vil med stor sannsynlighet velge bøker innenfor sine preferanser, eksempelvis en viss sjanger, og med mange anmeldelser. Bøker uten anmeldelser har få muligheter til å bli valgt.

Katrine Judit Urke fant i sin masteroppgave fra 2012 om blogganmeldelser at det hun kaller bokomtaler er svært varierte både i lengde og innhold. De fleste er korte, under 200 ord, men fant også omtaler på mer enn 700 ord. Om innholdet sier Urke at enkelte omtaler bærer preg av å være litterære analyser og ikke en formidlingstekst. Andre beskriver handlingen eller

minner forlagsomtaler. Hun finner også at omtalene kan inneholde svært personlige tekster iblandet tekst uten tilknytning til det aktuelle verket (s. 28).

2.4.2 Tagger

Tagger er populært på mange sosiale nettsteder og kan i følge Syversen brukes som alternativer til emneord for å formidle følelser (2011, s. 51). Tagger er i praksis enkeltstående emneord som brukes til å merke en ressurs. En treffende sammenligning er blitt gjort med gule post-it lapper (Peters, 2009, s. 153). Tagging skiller seg fra klassifisering og kategorisering ved at det kreves mindre kognitive ressurser. Det er således en enkel inngang til klassifisering noe som kan ha hatt betydning for utbredelsen. Ved tagging vurderes ressursen kun mot termer i hukommelsen og termene brukes på ressursen (Peters 2009 s. 161-162). Det er en viss uenighet om om hva tagger faktisk er. Berendt og Hansers påstand er at tagger ikke kan regnes som metadata. Dette begrunnes med en undersøkelse hvor de fant at tagger i liten grad oppsummerer ressursen slik som metadata skal. Isteden fungerer tagger som ekstra innhold ved å komplimentere innholdet i ressursen. Tagger fungerer ikke som en beskrivelse av ressursen, men er en del av den. Hammond et al. argumenter for at tagger er en type metadata og kan fungere som supplement til formelle klassifikasjonssystemer (Peters, 2009, s. 156). For brukerne av de sosiale samfunnene derimot er ikke riktig og gal bruk, eller nøyaktighet ved tagging viktig. (s. 157). Motivasjonen for å tagge kan variere fra bruker til bruker men mye tyder på at selvorganisering er grunnleggende. I en undersøkelse av Ames og Naaman fra 2007 om bruken av tagger på billedelingstjenesten Flickr.com ble det identifisert fire typer motivasjon for å tagge en ressurs; selvorganisering, selvkommunikasjon, sosialorganisering og sosialkommunikasjon (Phuong, 2011, s. 19). Et skille er fremtredende; organisasjon for egen del og formidling av til andre brukere. En kvalitativ undersøkelse fra 2008 av Nov, Naaman og Ye konkluderes det med at selvorganisering og deling med andre er hovedmotivene for å tagge (s. 19). I en tilsvarende studie fra samme år argumenteres det for at brukere anvender tagger til å sette bilder inn i kontekst, enten for selvorganisering eller for deling (Peters, 2009, s. 80). Det finnes flere forskningsprosjekter som har undersøkt hvilke termer brukere benytter seg av når de skal tagge en informasjonsressurs. Spitieri fant i 2007 at tagger ofte består av slang, sjargong akronymer eller er feilstavet (Phuong, 2011, s. 22). I 2009 fant en undersøkelse av Vuorikari, et al. at mange tagger består av flere termer satt sammen uten separasjon. I tillegg brukes det ofte flere språk (s. 21). Samme året, 2009, finner Wu & Zhou at brukere ikke er konsekvente når de tagger. Over tid vil brukere glemme hvilken term de har brukt ved en tidligere anledning og benytte seg av en annen term. Det blir

derfor stor spredning av termer, også for den enkelte bruker (s. 21). I en undersøkelse fra 2008 hvor tagger tilordnet av brukere og søk gjort av brukere ble sammenlignet oppdages det at taggene var generelle (63%), mens søkene (51%) var spesifikke (Stvilia, Jørgensen, & Wu, 2012, s. 101). En bekreftelse på dette finnes i en undersøkelse, også fra 2008, hvor det konkluderes med at brukere foretrekker søk fremfor navigering med tagger når informasjonsbehovet er spesifikt. Tagger foretrekkes når brukeren ikke har et spesielt behov, det vil si browsing (Phuong, 2011, s. 14).

2.4.3 Brukervurderinger

Brukervurderinger kan være enten eksplisitte, hvor en bruker blitt bedt om å vurdere en ressurs eller implisitte hvor vurderingen blir avledet av brukers interaksjon med systemet, for eksempel ved produktklikk (Ricci et al., 2011, s. 9 - 10). For denne oppgaven er det den første varianten som er interessant og videre benyttes brukervurderinger synonymt med eksplisitt vurdering. Brukervurderinger kommer i ulike former hvor de vanligste er; numerisk, ordinal, binær og ensartet. Numerisk er den vanligste formen og blir brukt av for eksempel Amazon.com for å tilby andre bøker basert på brukerens vurderinger. Amazon benytter en skala fra 1-5 men det eksisterer mange varianter med ulik granularitet og kan for eksempel være på en skala fra 0 – 100 eller -10 - +10 eller gi brukeren mulighet til å gi halve vurderinger. Systemer hvor det benyttes stjerner eller andre symboler er i praksis det samme og symbolene er bare en representasjon som blir gjort om til tall av det bakenforliggende systemet (Cosley, Lam, Albert, Konstan, & Riedl, 2003, s. 585). Det er også vanlig med en ordinalvurdering hvor det velges et alternativ som samsvarer med brukerens mening. Typisk blir dette benyttet i spørreskjemaer med alternativer som ”svært enig”, ”nøytral” og ”svært uenig”. Ved anvendelse av en binær vurderingsform blir brukeren ganske enkelt bedt å vurdere en ressurs eller en gjenstand til å være god eller dårlig. Ensartet vurdering innebærer at det enten er gitt en vurdering eller ikke gitt en vurdering og kan eksempelvis bety at bruker har kjøpt et produkt og dermed gitt en positiv vurdering. Alle andre, ikke kjøpte produkter, har ingen vurdering (Ricci et al., 2011, s. 9). Den siste formen, ensartet, minner om og blir ofte regnet som en implisitt metode slik som hos Jannach et al. (Jannach et al., 2011, s. 23).

Den kanskje vanligste og mest anvendte formen er en skala fra 1 til 5, enten med symboler eller tall. Denne formen finnes også i datasettet brukt i denne oppgaven og vil derfor bli viet ekstra oppmerksomhet. Generelt kan det være vanskelig å overføre komplekse meninger til en enkelt verdi. Spesielt kan dette være vanskelig for litteratur hvor kvalitetskriterier, aspekter og

humør har betydning (Cosley et al., 2003, s. 586). Spalding bak boksamfunnet og anbefalingssiden LibraryThing.com sier i et intervju at de fleste lesere vurderer bøker til å være omtrent midt på skalaen, LibraryThings egen skala går fra 1 til 5 og gjennomsnittlig vurdering er 3.8. Spalding forklarer dette med at lesere velger bøker de selv tror de vil like og han gir uttrykk for at vurderingene derfor har liten verdi (Syversen, 2011, s. 54). Fordi LibraryThing er en enorm aktør med millioner av bøker og brukere må Spaldings meninger tillegges vekt, men om de stemmer for alle tilfeller er usikkert. Nesse fra bokelskere.no, det største boksamfunnet i Norge, sier at gjennomsnittlig terningkast på deres side er noe over fem og forklarer det med at ”vi stort sett legger inn bøker vi liker.” (Nesse, 2011). I likhet med tagger er ikke alltid brukere konsekvente i sine vurderinger. I en studie ble 212 brukere bedt om å vurdere filmer de allerede hadde gitt vurdering til på nytt og resultatet viste et samsvar mellom tidligere og ny vurdering i kun 60% av tilfellene (Cosley et al., 2003, s. 587).

2.5 Et rammeverk for appell

Joyce G. Saricks *Readers Advisory Service in the Public Library* (Saricks, 2005) er ifølge Moyers gjennomgang av litteratur om leseveiledning banebrytende og en av få forskningsbaserte utgivelser innen feltet (Moyer, 2005, s. 229). Verket er en praktisk veiledning for bibliotekarer i referansearbeidet og Saricks har skapt et rammeverk og en terminologi som kan brukes til å beskrive, snakke om og gjenkjenne elementene som utgjør en bok. Det er umulig å lese alle bøker og like umulig er det å like alle bøker, sjangere eller forfattere. For rådgivere er det derfor viktig å kunne snakke om bøker som en kun har kjennskap til gjennom anmeldelser, samtaler, omtale eller rask skumming av enkelte sider i en bok (Saricks, 2005, s. 62). Rammeverket fungerer således som et praktisk anvendelig verktøy for rådgivere og er ikke bare en teoretisk tilnærming. Saricks bruker begrepet rådgiver om personer som gir veiledning og anbefalinger i forbindelse med valg av litteratur og inkluderer dermed også personer utenfor bibliotekene og bokbransjen.

Ross og Weaver (Ross, 2000) (Weaver, 2007) peker, omtalt tidligere, på humør som essensielt, men gir ingen klar definisjon utover at det er leserens kontekst. Saricks benytter ikke humør i sitt arbeid men har funnet at lesere sjelden ser etter bøker om eller med et spesielt emne, men etter litteratur med en spesifikk stemning og følelse (2005, s. 40). Vi utvikler ofte en sjette sans for hvilke bøker som naturlig hører sammen og kan være av interesse fordi bøkene fremkaller lik respons og følelse. Det betyr ikke nødvendigvis at bøkene er i samme sjanger eller omhandler samme emne. Saricks avskriver ikke emneord,

mer er klar på at de alene gir lite informasjon om stemningen i en bok. Emneord kan formidle enkelte aspekter ved en boks appell men mangler meningsdybde og kan ikke fortelle hvordan emnene behandles av forfatteren. For å lettere kunne avgjøre om en bok er interessant trenger lesere en klar og gjennomtenkt oppsummering av handlingen inkludert setninger med viktige elementer som kan avsløre bokens appell. I følge Saricks er nøkkelen til å identifisere hva en leser liker ved en bok appell, ikke emneord (s. 41 - 42).

2.5.1 Appell

Omrisset av hva appell innebærer i denne sammenhengen kan skimtes i det foregående avsnittet; formidlingen av stemninger og følelser slik at leseren lettere kan avgjøre om boken vil falle i smak og passer humøret. Saricks egne undersøkelser med lesergrupper indikerer at lesere finner det enklere å avgjøre om en bok er av interesse når appellelementer brukes i samtale (Saricks, 2005, s. 64). Saricks definerer kort appell til å være elementene ved bøker som for leseren utgjør appelleringssevnen (s. 42). Til tross for den korte og tilsynelatende enkle definisjonen kan appell være vanskelig å sette ord på, selv for erfarne lesere og rådgivere. Alle som snakker om bøker gjenkjenner ofte en boks appell lenge før den kan defineres og bruker naturlig appellelementer i samtaler for å øke og komme til en felles forståelse. Ved å bruke termer for appell i samtaler om bøker kan boken, fordi stemningen og følelsen beskrives med større grad av nøyaktighet, åpnes opp for en eventuell leser og kombinert med emneord kan appell gjøre det mulig komme nærmere en boks essens slik at leseren kan avgjøre om den passer humøret. Den grunnleggende tanken bak er å gi rådgivere evnen til å oversette en enkelt lesers kommentarer og språk til appell slik at det blir mulig å anbefale litteratur uten nærgående kjennskap til leseren. Ved å benytte appell og et rammeverk rundt er det mulig for lesere å krysse sjangergrensene og bli introdusert for verk og forfattere de ellers kunne gått glipp av. Saricks har funnet at appell kan være med på å frigjøre rådgivere og lesere fra en plotorientert tankegang programmert inn i oss gjennom formelle samtaler om litteratur. Hun mener ikke å hevde at plot er betydningsløst, plotdetaljer blir for eksempel inkorporert i appellelementene, men understreker at kvaliteten på rådgivningssamtaler kan heves ved å tenke appell fremfor oppsummering av plot (s. 64). I siste instans kan bedre rådgivning gjennom appell være med på å aktivere større deler av bibliotekskatalogen (s. 63).

Appell må ikke forveksles med kvalitet eller kritikervurderinger, det er leserens oppfatning som er viktig. Saricks peker på at lesere er villige til å akseptere et bredt kvalitetsspekter så

lenge appellelementene stemmer (s. 42 - 44). For å forstå bøkernes og forfatterens appell må informasjonen som hentes fra bøker, anmeldelser, omtaler eller andre kilder legges inn i et rammeverk (s. 72).

2.5.2 Appellelementer

Appellelementene er kjernen rammeverket er strukturert rundt og som gjør det mulig å formalisere en boks appelleringssevne. I denne oppgaven brukes appelleringssevne som en betegnelse for hvordan en bok appellerer til en enkelt leser og ikke en boks generelle evne til appell. Det brukes ikke for å beskrive kvalitet eller popularitet. Saricks deler inn appell i fire grunnleggende elementer: tempo (pacing), karakterisering (characterization), ramme (frame) og handling (story line) (Saricks, 2005, s. 43). I tillegg beskriver hun et femte element, skrivestil (style). Elementene kan enkeltvis eller kombinert adressere hvordan en bok påvirker lesere, men i realiteten vil en bok bestå av en kombinasjon av appellelementer og et riktig element alene er ofte ikke nok til å overbevise leseren. At skrivestil ikke er et grunnleggende element begrunnes med at så lenge de andre elementene stemmer for leseren vil ikke skrivestil ha avgjørende betydning. Skrivestil blir likevel omfattet av rammeverket fordi enkelte kun ønsker å lese bøker med en bestemt skrivestil. Spesielt kan skrivestil ha en betydning for sakprosa hvor en leser ønsker at emnet skal presenteres på en bestemt måte, for eksempel ønsker noen at skrivestilen ikke skal være for komplisert (s. 61 - 62). En observasjon gjort i forbindelse med denne oppgaven er at skrivestil ofte og med få unntak inngår rammeverkets fire grunnleggende appellelementer og vurdere skrivestil kan i mange tilfeller være redundant. Skrivestil blir ikke benyttet videre i oppgaven. I påfølgende avsnitt blir Saricks fire grunnleggende elementer behandlet og må leses som en gjennomgang. Hensikten er å gi en oppsummering for å vise hvordan elementene har blitt forstått, hvilke aspekter som er sentrale og er ikke ment å være en gjennomgripende analyse.

2.5.2.1 Tempo

Tempo er ofte det første lesere oppdager, ofte ubevisst. Elementet er muligens det viktigste fordi bøker med forskjellig tempo ikke appellerer til den samme leseren og kan dermed brukes til å effektivt inkludere eller ekskludere aktuelle bøker. Saricks har funnet at lesere vurderer tempoet i en bok til å være enten raskt eller saktegående og det er disse kategoriene som brukes i elementet (Saricks, 2005, s. 44). Tempo handler primært ikke om hvor raskt en bok kan leses, selv om dette også er en faktor, men om hvordan leseren oppfatter boken. Det avgjørende for tempo er hvordan historien er bygget med tanke på hvilken måte leseren ser karakteren, beskrivelsen av omgivelsene, dialogen og hvor fort man dras inn i handlingen (s.

49). En bok kan altså ha et raskt eller saktegående tempo. At tempo er en enkel inngang til å inkludere eller ekskludere bøker som kan være aktuelle for leseren gjør også at elementet kan være vanskelig å bruke i en rådgivningssituasjon. Lesere har vanskeligheter med å identifisere hva de liker og stilt ovenfor valget mellom bøker med raskt eller saktegående tempo vil avgjørelsen nesten alltid falle ned på en med raskt tempo (s. 45). Saktegående assosieres ofte med tregt og har dermed en negativ klang. Rammeverket er som tidligere skrevet en metode for å formalisere språket vi bruker om bøker slik at det kan benyttes strukturert. Ved gjennomlesning av en bok er det forholdsvis enkelt å avdekke tempoet, verre blir det når man kun kan basere seg på noen korte utdrag, anmeldelser eller omtale. Det er derfor viktig å finne egenskaper som kan indikere hvilket tempo boken innehar. Saricks har forsøkt å finne frem til hva som kjennetegner tempo, henholdsvis raskt og saktegående. Kjennetegnene er uavhengige av hverandre og som regel er det ikke mulig å se etter alle i en og samme bok. Saricks bruker benevnelsen ledetråder og at disse kan hjelpe rådgivere med å identifisere tempo (s. 44). Flere ledetråder fører til en sikrere vurdering av bokens tempo og kjennetegnene kan med fordel brukes sammen. I de fleste tilfeller vil det være nærmest umulig å skille kjennetegnene, som tvinnes sammen og overlapper, fra hverandre og medfører at det er den samlede informasjonen som setter rådgiveren i stand til å ta en avgjørelse (s 49).

Den kanskje mest åpenbare indikatoren er hvor fort historien beveges fremover. Kastes karakteren ut i halsbrekkende sekvenser hvor leseren er vitne til hvordan karakteren agerer på hendelser nærmest umiddelbart antyder det et raskt tempo også i boken. Til motsetning vil bøker med et mer avmålt tempo gjerne avsløre karakteren gradvis ettersom historien utfoldes. Karakteren som reagerer på ytre hendelser og er fysisk deltakende, typisk i fare, indikerer gjerne et høyt tempo mens i fortellinger hvor karakteren er psykisk involvert og leseren får et intimt innblikk i karakterens tanker markere en roligere rytme. Likeledes vil bøker hvor plottet avsløres tidlig, gjerne med tvister, gi hint om et høyt tempo mens bøker hvor plottet avsløres over tid henter om et roligere tempo. Det betyr ikke at bøker med et raskere tempo ikke kan ha intrikate plot. Som regel er det slik at typiske bøker med stor fart, for eksempel mysterier eller spenningsbøker, har flere plotlinjer som må nøstes opp og beveger historien fremover. Flere plotlinjer kan være med å redusere hastigheten fordi leseren må konsentrere seg mer, men ofte er det motsatte tilfelle. Forfattere benytter seg av spennende avslutninger på kapitler, såkalte «cliffhangers» for å holde på leseren og gir en følelse av høy takt. Avslutninger kan i seg selv være et uttrykk for tempo. Bøker orientert mot en avslutning er

fokus på hvordan plottet vil løses opp fortrinnsvis særpreget av høyere fart enn bøker med åpne avslutninger og uten en definitiv konklusjon.

Forholdet mellom mengden dialog og beskrivelse kan fortelle mye om tempoet i en bok. Tekster med mye dialog vil naturlig ha et raskere tempo enn tekster med mye beskrivelse. En konsekvens av dialog er mye hvit plass på sidene og en kort titt gjennom sidene i en bok kan gi rådgiveren en anelse om bokens tempo. Enkelte sjangere, for eksempel thrillere, krever et høyt tempo og teksten blir følgelig mindre tettskrevet med mye dialog, korte setninger, avsnitt og kapitler som er med på å trekke leseren raskt gjennom boken. En effekt av en mindre tettskrevet tekst blir ofte en komprimert tidslinje hvor historien eller hendelsene utspiller seg over en kortere tidsperiode. I bøker med komprimerte tidslinjer er det ikke sjelden at historien utfolder seg over timer eller dager, og ikke over måneder eller år som er vanligere i tettere tekster. En tettskrevet tekst hentyder at skrivestilen er mer elegant med detaljer og beskrivelse som fører til et roligere tempo. Samtidig har tettskrevne tekster en tendens til å fremheve språket og være lyriske eller poetiske i formen hvor hvert ord teller. Slike bøker velges gjerne av lesere som ønsker å la seg underholde av teksten, språket eller poesien fremfor hendelsesforløpet og resulterer i en appelleringssevne som skiller seg fra bøker med mer luft. (Saricks, 2005, s. 46).

2.5.2.2 Karakterisering

Karakterer kan være avgjørende for leseropplevelsen. Å avgjøre hva som gjør en karakter minneverdig for den enkelte leser kan være svært vanskelig; enkelte lesere kan foretrekke introverte karakterer, andre liker utadvendte skikkelser med mange interaksjoner og vittige kommentarer. I motsetning til tempo kan ikke karakterer deles inn to enkle grupper og muligheten til å avskrive aktuelle bøker er ikke like ukomplisert. Fordelen med er at lesere har lettere for å huske og beskrive karakterer de har likt, ofte lettere enn plottet og historien (Saricks, 2005, s. 50 - 55). Noen forbehold må likevel tas. En lesers oppfatning av hva en velutviklet karakter innebærer varierer fra leser til leser og kan bety et mangfold av karakterer, alt fra å være lett gjenkjennbar til å ha mangfoldige personlighetslag. Saricks har forsøkt å gi en oversikt over hva som er sentralt når en karaktertype skal identifiseres.

Enkelte karakterer kan være lett gjenkjennbare, nesten stereotypiske, og bidrar til at leseren blir dratt direkte inn i handlingen. Selv om karakteren utvikles eller endres i løpet av boken er det likevel noe umiskjennelig som ikke avviker i stor grad fra leserens forventninger. Det

motsatte er karakterer som utvikles og avslører sin natur over tid hvor leserens oppfatning endres etter hvert som karakteren trer tydeligere frem med overraskende og unike trekk. Lesere som foretrekker en type kan etter Saricks erfaringer mislike å bli tilbudt bøker med karakter som ikke svarer til forventningene, spesielt gjelder dette for de som liker utvikling (s. 51). I moderne litteratur er det fortrinnsvis en hovedkarakter som dominerer men innen enkelte sjangre, primært gjelder det kvinnelige relasjoner, kan det være flere hovedkarakterer med sammenblandede liv. Det muliggjør en dypere karakterisering med flere nyanser. Selv om ikke kjennetegnet er fremtredende kan det være en indikator for om leseren ønsker flere hovedkarakterer eller er interessert i bøker hvor samspillet mellom fremtredende karakterer er viktig. Hvilken fortellerstemme som benyttes kan utgjøre en stor forskjell for hvordan leseren opplever historien. Saricks fremstiller synsvinkel på en ukomplisert og muligens for enkel måte. Hun understreker at det er viktig å være klar over forskjellen 1. Person og en aural fortellerstemme kan utgjøre, spesielt innen enkelte sjangere, men oppfatter hvilken karakters historie som fortelles som mer avgjørende. I spenningsromaner med seriemordere er det viktig at både etterforskerens og morderens historie blir fortalt for at leseren skal kunne få innblikk i mulige hendelsesforløp. For eksempel vil den samme historien med kun etterforskerens synsvinkel være i en annen sjanger; skrekk. Om fortellerstemmen er 1. Eller 3. Person er mindre viktig (s. 52).

I visse bøker kan karakteren(e) være så altoverskyggende at plot, omgivelser og atmosfære blir uvesentlig. Saricks betegner slike bøker som karaktersentrerte og mener med det at karakterenes indre liv, tanker, reaksjoner og ageringer er det sentrale i fortellingen (s. 52). Leseren vet mer om hovedpersonen enn om historien som foregår rundt. Karakterutvikling og tempoet i en bok henger nøye sammen, karakterer som utvikles over tid har ordinært et saktegående tempo, og spesielt gjelder dette for karaktersentrerte bøker som har lite ytre handling. Hva en karaktersentrert bok betyr kan skape forvirring fordi mange lesere følger serier på grunn av interessante karakterer og vil som oftest anta at de finner glede i karaktersentrerte bøker. Til tross for at det i utgangspunktet ikke er noen motsetninger mellom det å like seriekarakterer og karaktersentrerte romaner er det likevel store ulikheter. Lesere som liker og følger seriekarakterer er primært interessert i hva som skjer med karakteren og ikke indre tanker eller reaksjoner slik det er tilfelle i karaktersentrerte romaner (s. 53). Saricks benytter betegnelsen seriekarakter, muligens for å marke et ytterpunkt, men personene hun forsøker illustrere er karakterer som er lett gjenkjennelige uavhengige av serieformatet. Karakterene trenger ikke være den samme fra bok til bok eller fra forfatter til forfatter for å

være av typen seriekarakter, det gjenkjenneligheten som er viktig. Poenget er å vise at personskildringen i karaktersentrerte bøker er forskjellig fra bøker hvor det som skjer med karakteren er sentralt og medfører ulik appelleringsevne.

Fordi lett identifiserbare karakterer er ment å være sympatiske følges som regel handlingen gjennom hovedpersonens øyne og bildet leseren får er mer intimt enn i bøker hvor hovedpersonen følges med større distanse. Karakterer med egenskaper leseren enkelt kan identifisere seg med er med på å trekke leseren inn i hovedpersonens liv slik at historien kan komme i gang fortore uten en omstendelig dyptgående introduksjon. For eksempel kan heltene eller heltinnen i romantiske spenningsbøker være umiddelbart identifiserbare og sympatiske slik at oppløsningen av mysteriet kan starte. I fortellinger hvor karakteren betraktes på avstand kan sympatiske trekk settes under tvil og det er ikke nødvendigvis mulig for leseren å identifisere seg med eller like hovedpersonen. Avstanden gjør at leseren ikke trenger å følge handlingen på nært hold og forholdet til hovedpersonen føles tryggere. Det siste kan føre til karakterer med mer kompleksitet og leseren må i større grad ta stilling til flere interessante sider ved hovedpersonen, også de usympatiske (Saricks, 2005, s. 53 - 54).

2.5.2.3 Ramme

Lesere på jakt etter ny litteratur bruker først og fremst tempo som første sorteringskriteriet, etter humør samt personlig smak, og er kanskje det viktigste element fordi det setter anslaget. Ramme har en annen viktig rolle i motsatt ende av prosessen og har vist seg utslagsgivende når en bok tilslutt skal velges (Saricks, 2005, s. 58). Generelt kan indikatorene være til hjelp når leseren allerede har en utvalgt sjanger og må gjøre en ytterligere distingvering mellom aktuelle bøker. I følge Saricks er består ramme i hovedsak av grunntone, bakteppe og innramming (s. 59). Av de tre legger hun størst vekt på grunntonen som kan være vanskelig å definere, men paradoksalt nok er det nettopp grunntonen de fleste lesere forsøker å formidle når en bok eller forfatter beskrives. En bok kan gi leseren en viss følelse eller utgi en atmosfære som blir lagt merke til og er med på å bestemme hvordan boken oppfattes. Tonen kan for eksempel beskrives som dyster eller positiv, følelsen kan være morsom eller romantisk og atmosfæren kan være illevarslende eller oppløftende (s. 59).

Grunntonen kan være et viktig aspekt for enkelte sjangere og kan være et dominerende element innenfor eksempelvis skrekk, romantikk eller thrillere. Grunntonen kan også være med på arrangere historien ved å effektivt sette stemningen og følelsen slik at leseren begriper

boken på en bestemt måte; mordmysterier vil ha vidt forskjellig egenart om stemningen er mørk og dyster enn om stemning er lys og koselig. I likhet med grunntone kan bakteppe være avgjørende for hvordan leseren oppfatter boken. Med bakteppe mener Saricks informasjonen i omgivelsen som kan være svært spesifikke eller sparsommelige. Kunne historien forløpt seg på tilsvarende måte på et annet sted i tid og rom eller er historien bundet til et bestemt sted innenfor et fast tidsrom. Særlig for lesere av historisk fiksjon og tilgrensende sjangere kan et bakteppe med tidsriktige detaljer fra et spesielt geografisk område være betydningsfullt (Saricks, 2005, s. 60). Enkelte forfattere velger å gå enda lenger ved å skape en spesiell innramming rundt historien som kan være viktigere for leseren enn hvordan fortellingen forløper. Typiske eksempler er å finne innen fantastisk og fremtidsrettet litteratur hvor det gjerne skapes hele, ofte enorme, verdener med mange detaljer i omgivelsene. Lesere som liker en spesifikk type innramming, hvor karakterer og historien utfoldes innenfor, ønsker normalt å lese alt uavhengig av andre appellelementer så lenge rammen rundt stemmer (s. 61).

2.5.2.4 Handling

Når handlingen skal omtales i boksamtaler handler det om mer enn bare en ren oppsummering av bokens plott. Det essensielle er hvordan leseren oppfatter følelsen og stemningen i boken og fremstilles gjerne med ladete vendinger som for eksempel inspirerende eller tragisk for å gi uttrykk for en følelsesmessig innvirkning (Saricks, 2005, s. 56). Handlingen går utover emnet ved å gi et bilde av hvordan det behandles av forfatteren. For eksempel kan forfatteren ha gitt et tema en humoristisk vri eller ha skrevet det som en eventyrfortelling (s. 57).

Forfattere kan velge å fokusere ulikt i sin historie. Det kan være å fremheve personer fremfor omgivelser eller omvendt. Bøker kan inneholde sterke personligheter hvor fokuset likevel er på omgivelsene og historien. Saricks trekker frem klassiske westernfortellinger som eksempel. Motsatt kan bøker med detaljerte omgivelser fortsatt fokusere på karakterer (s. 56). Valg av fokus begrenses ikke til karakter eller omgivelse og kan også innbefatte ytre eller indre handling, hvor det første er orientert mot spenningsfylt handling og det siste mot det psykologiske på innsiden av karakteren. I nevnte tilfeller er det forholdet mellom fokus på omgivelse, karakterer, det ytre og det indre som er avgjørende, ikke mengden av hver indikator i seg selv. Handlingen i bøker kan velge å foregå over et eller flere plan med enten et enkelt plott eller med flere parallelle handlinger og lag som kommer til syne etter hvert som handlingen fremskrider. Handlinger over flere plan fordrer større konsentrasjon fra leseren og passer ikke alle typer humor og kan være viktig å legge merke til for en rådgiver (s. 57).

Appellelementet handling er ikke blitt gjort rede for like inngående som forutgående elementer fordi handling til dels blir tatt opp, inkludert, i tempo, karakterisering og ramme og til dels fordi elementet er vanskelig å oppdage i samtaler eller brukeranmeldelser (s. 58) og dermed også bruke i oppgaven. Noen eksempler kan muligens illustrere det første poenget; bøker hvor handlingen utspiller seg over flere plan vil inneha et roligere tempo enn bøker med et enkelt plot. Likeledes vil handlinger med fokus på karakterer fremfor historier som drives frem av omgivelse og direkte ageringer omfattes av appellelementet karakterisering. En bok hvor fokuset er på en karakters indre vil omfattes av både elementene karakterisering og tempo (s. 58). Selv om handling ikke brukes videre i oppgaven ble det funnet riktig å ta med en kort oppsummering av elementet for å begrunne hvorfor.

2.5.2.5 Oppsummering av appellelementer

Av Saricks fire grunnleggende elementer står tempo i en særstilling. Karakterisering, ramme og handling peker alle på hver sine måter tilbake til tempo. Stereotype personschildringer, en lys atmosfære og mye ytre handling kan være betegnende for en bok med et raskt tempo. Omvendt har mest sannsynlig en bok med et avmålt tempo mest sannsynlig en gradvis utfolding av karakteren og mer indre handling. Spesielt vil handling med få unntak peke tilbake til tempo og er derfor blitt valgt bort til tross for at et aspekt går tapt fordi det ikke er mulig å avgjøre hvordan et emne har blitt behandlet av forfatteren. Nærmere eksempel på hvordan handling omfattes av tempo vil bli gitt senere i oppgaven. Det er noen problemer med Saricks appellelementer. Enkelte av indikatorene kan være vanskelig å bruke uten å fysisk se i boken, for eksempel å få en kjenning med tempo ved å se på forholdet mellom dialog og beskrivelse. Et problem som kan vise seg større er at det elementene ikke har en naturlig rangering utover rådgiverens følelse. Problemene beskrevet over vil bli behandlet i forbindelse med metodebeskrivelse og i den praktiske realiseringen av oppgaven.

2.5.3 Vokabular, termer og eksempler

Saricks kaller ordene som brukes når lesere omtaler egne reaksjoner på bøker og litteraturer for termer, mer presist appelltermer. Sagt på en annen måte er appelltermene de ordene som beskriver appellelementene (Saricks, 2005, s. 65). I motsetning til emneord, som er substantiv, vil termer for appell nesten alltid være adjektiv og er i så måte bedre egnet til å avdekke innholdet bøker. Ulempen med appelltermer er mangelen på konsistent meningsinnhold og Saricks har derfor forsøkt å skape en tesaurus slik at appelltermene får samme betydning for alle lesere og bibliotekarer. Det har vist seg vanskelig fordi det er få ord

som fungerer på en så entydig måte og tesaurusen kan aldri bli utfyllende nok for formålet(s. 65). Saricks begrunnelse er kort og kan hende kryptisk men en mulig tolkning kan være at ordene og vendingene som brukes når bøker skal omtales, muntlig eller skriftlig, er for mangfoldig og i tillegg tvetydige fremfor konkrete. Inntrykket fra arbeidet med denne oppgaven bekrefter Saricks erfaringer. Lesere uttrykker seg ofte i antydende ordelag og sjelden finnes det anmeldelser som klart definerer en bok til å være rask med et lett tempo og stereotype karakterer. Å få lesere til å ta i bruk og skjønne appellelementene er problematisk og et kontrollert vokabular er ikke nødvendig, det viktigste er å gjenkjenne termer som kan gi hint om en boks appelleringssevne (s. 65 - 66). Saricks har altså ikke skapt et tesaurus men har laget en oversikt over aktuelle ord for hvert av appellelementene som vises i figuren under (Figur 1). Saricks skriver at listen kun er et utgangspunkt, må betraktes som foreløpig og at lesere ikke bestandig vil benytte termene (s. 66).

Figur 1: The Vocabulary of Appeal (Saricks, 2005, s. 66)

Pacing

breakneck, compelling, deliberate, densely written, easy, engrossing, fast paced, leisurely paced, measured, relaxed, stately, 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

Story Line

action oriented, character centered, complex, domestic, episodic, explicit violence, family centered, folksy, gentle, inspirational, issue oriented, layered, literary references, multiple plotlines, mystical, mythic, open-ended, plot centered, plot twists, racy, resolved ending, rich and famous, romp, sexually explicit, steamy, strong language, thought-provoking, tragic

Frame and Tone

bittersweet, bleak, contemporary, darker (tone), detailed setting, details of [insert an area of specialized knowledge or skill], edgy, evocative, exotic, foreboding, gritty, hard edged, heartwarming, historical details, humorous, lush, magical, melodramatic, menacing, mystical, nightmare (tone), nostalgic, philosophical, political, psychological, romantic, rural, sensual, small town, stark, suspenseful, timeless, upbeat, urban

Style

austere, candid, classic, colorful, complex, concise, conversational, direct, dramatic, elaborate, elegant, extravagant, flamboyant, frank, graceful, homespun, jargon, metaphorical, natural, ornate, poetic, polished, prosaic, restrained, seemly, showy, simple, sophisticated, stark, thoughtful, unaffected, unembellished, unpretentious, unusual

Det er som nevnt tidligere umulig å lese alle bøker og det er derfor viktig å kunne identifisere appelleringsveien ved å bruke andre kilder. En slik kilde er ifølge Saricks gode bokanmeldelser. For å belyse hvordan kjennetegn gjenkjennes gir hun eksempler på anmeldelser inkludert setninger med termer som ikke står i Saricks egen oversikt, men som likevel er talende for bokens appelleringsveie (s. 68). To av Saricks eksempler er plukket ut for å illustrere. Merk at de valgte setningene er plukket ut av Saricks selv. Det første eksempelet er hentet fra en spenningsbok med høyt tempo med følgende setninger fra anmeldelsen:

- epic action tale of intrigue, suspense, and adventure
- fights, schemes, and plots
- attempting to assassinate

- fast-moving action
- exciting plot

Selv om eksemplensetningene ikke inneholder noen termer som eksplisitt forteller at dette er en bok med et raskt tempo er kjennetegnene ikke til å ta feil av. Den første linjen gjengir at boken har en fortelling med action, intriger, spenning og eventyr og tilhører i utgangspunktet det bortvalgte appellelementet handling, men de indirekte kjennetegnene for bokens tempo tydelig og eksemplifiserer hvordan handling omfattes av tempo. Det andre eksempelet er fra anmeldelsen av en historisk roman med et lavt tempo og følgende setninger er plukket ut:

- meticulously researched and compelling historical
- witness
- observe
- flesh-and-blood viability
- major investment in reading time

Allerede ved den første utvalgte setningen gis det hint om en lavere takt enn i boken fra forrige eksempel. At det er gjort omhyggelige undersøkelser og at leseren er vitne til og observerer begivenhetene antyder detaljer og dermed et avmålt tempo. Den vanskelig oversettable setningen *flesh-and-blood viability* gir indikasjoner om en velutviklet karakter og viser samtidig hvordan karakterutviklingen og tempoet i en bok følger hverandre. Den siste setningen bekrefter tidligere hint om et saktegående tempo.

2.6 Tekstklassifisering

Automatisk klassifisering eller maskin klassifisering er det generiske uttrykket som ofte brukes om en klassifiseringsprosess hvor maskiner er involvert. Sean Owen definerer det slik: ”Computer classification systems are a form of machine learning that uses learning algorithms to provide a way for computers to make decisions based on experience and, in the process, emulate certain forms of human decision making” (Owen, 2012, s. 230). Klassifisering er i utgangspunktet en oppgave mennesker utfører med enkelhet hver eneste dag. I sin mest elementære form innebærer klassifisering, for både mennesker og for maskiner, besvarelsen av et lukket spørsmål ved å velge blant et begrenset antall muligheter (Owen, 2012, s. 227). Det som for mennesker er en enkel hverdagslig hendelse er vanskelig for maskiner som ikke kan trekke på erfaringer om hva som karakteriserer et bestemt objekt. Et illustrerende eksempel av Grant Ingersoll tydeliggjør forskjellen mellom menneskelig og maskinell klassifisering: hva skiller et fly og et helikopter? Mennesker vet forskjellen for begge deler

har blitt observert flere ganger og selv om vi ikke har tenkt eksplisitt over hvilke egenskaper som faktisk skiller de to så har det blitt gjort en klassifikasjon når det skilles mellom dem. Maskiner har ikke de samme mulighetene til å verken se et fly eller trekke på tidligere erfaring og må bli fortalt nøyaktig hvilke egenskaper som definerer og skiller et fly fra et helikopter. Begge kan fly, men det skiller dem ikke fra hverandre, i tillegg kan også mange andre objekter fly. Faktumet at begge frakter mennesker skiller dem fra andre ting som flyr. At et fly har faste vinger og et helikopter har rotor er derimot unike egenskap som kan brukes til å skille mellom de to objektene (Ingersoll, 2013, s. 178). En maskin må altså lære hvilke egenskaper som tilhører objektene i en bestemt klasse eller kategori. Det flere egenskaper maskinen kjenner desto mer presis blir klassifiseringen. Maskinklassifikasjon kan grovt deles inn i uadministrert og administrert (Owen, 2012, s. 230). For denne oppgaven er det administrert som er aktuelt og innebærer at algoritmen blir opplært til å skille mellom objekter gjennom eksempler. Uadministrerte algoritmer avgjør selv hvilke egenskaper som best kan skille mellom objekter men er foreløpig ikke gode nok og har et iboende problem i at de automatiske kategoriene ikke stemmer overens med menneskelig logikk (Baeza-Yates, 2011, s. 286).

2.6.1 Klassifikasjon og kategorisering

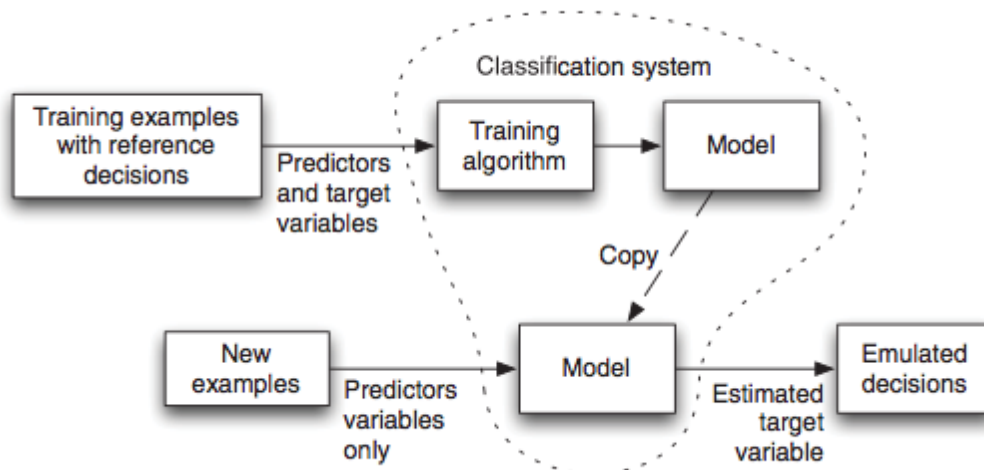
Klassifikasjon og kategorisering brukes omvekslende i bøker og artikler som omhandler maskinklassifikasjon. Enkelte skiller mellom systemer for binær- og multiklassifikasjon hvor binærklassifikasjon kun kan svare ja eller nei på et gitt spørsmål mens multiklassifikasjon kan tilegne objektet til en bestemt klasse ut av flere mulige. For eksempel kan et system som benytter binærklassifikasjon svare ja eller nei på om en kredittkorttransaksjon er svindel eller ikke. Multiklassifikasjon kan for eksempel avgjøre hvilken sjanger en film tilhører basert på gitte kjente kriterier. Multiklassifikasjon kan også deles inn ytterligere i en flat, som er kun en liste med merkelapper, eller hierarkisk struktur (Marmanis, 2009, s. 169). Andre igjen er mindre presise og sier at kategorisering er en spesialisering av klassifikasjon (Ingersoll, 2013, s. 177). Kategorisering er det samme som multiklassifikasjon og begge er en spesialisering av det som overordnet kalles maskinklassifikasjon. Kategorisering er den termen som benyttes oftest, vanligvis sammen med klassifikasjon som en term for maskinklassifikasjon. I denne oppgaven vil klassifikasjon benyttes synonymt med den generelle prosessen administrert maskinklassifikasjon og kategorisering benyttes for spesialiseringen multiklassifikasjon med flat struktur.

2.6.1.1 Trening

Innen maskinklassifikasjon er det mulig å benytte ulike algoritmer. Forskjellige algoritmer har ulike styrker og hvilken algoritme som er hensiktsmessig å bruke avhenger av bruksområdet. I denne delen av oppgaven blir det dog ikke valgt eller gjort en vurdering av hvilken algoritme som skal brukes. All utvikling av klassifikasjonssystemer følger den samme fremgangsmåten og består i hovedsak av to prosesser; trening og bruk. I tillegg er vanlig å teste systemet mellom trening og bruk.

I starten av kapittelet ble det lagt vekt på at algoritmen må lære hvilke egenskaper som karakteriserer et objekt. Det oppnås ved å trene en algoritme ved å gi eksempler på hvilke egenskaper som tilhører en kategori. Algoritmen produserer på bakgrunn av eksemplene en modell som benyttes av det ferdige klassifikasjonssystemet. Ved trening av algoritmen må det oppgis både en prediktor og en målvariabel. Prediktor er egenskapene som blir brukt i eksempelet, altså ordene, og målet er kategorien egenskapene tilhører. I det følgende eksempelet er det første ordet, *fast-paced*, målvariabelen og den påfølgende setningen prediktorene: "Fast-paced Even so, this book was an easy page turner". I det ferdige klassifikasjonssystemet får algoritmen kun oppgitt prediktorene og må avgjøre hvilken kategori som skal tilegnes. Figur 2 gir en oversikt over de ulike stadiene i prosessen og forskjell på trening og et ferdig system.

Figur 2: Stadiene i utviklingen av et klassifikasjonssystem (Owen, 2012, s. 233)



2.6.1.2 Modellbygging

Hvor godt klassifiseringssystemet presterer avhenger av kvaliteten på modellen, er det nok eksempler og er eksemplene relevante. Uten nok eksempler kan algoritmens feilantakelser i

treningssprosessen føre til at relasjonene mellom egenskaper og målvariabelen blir og føre til feil også i klassifikasjonsprosessen (Ingersoll, 2013, s. 184). Hvor mange eksempler som trengs avhenger av størrelsen og kompleksiteten på datasettet, men et betydelig antall er å foretrekke. Normalt blir det trent og testet med omtrent 10 prosent av det totale datamaterialet, men dette varierer (s. 192).

2.6.1.3 Overtilpasning

En vanlig utfordring i forbindelse ved trening av modeller er overtilpasning, kjent som overfitting og target leak (O'Neil, 2012)(Nigam, Lafferty, & McCallum, 1999)(Owen, 2012, s. 241). Det er noen subtile forskjeller mellom overfitting og target leak, men resultatet er det samme. Med en overfitting menes det at treningsdataene følger et mønster og algoritmen lærer seg mønsteret fremfor trendene i treningssettet. Overfitting kan oppstå når eksemplene er så tilpasset hver enkelt kategori at modellen ikke vil fungere for ukjente data som ikke er like de i treningssettet eller når treningsdataene ikke inneholder nok eksempler i forhold til virkelig data. Target leak betyr at prediktoren inneholder for sterke signaler om hva som er målvariabel, for eksempel ved å inkludere navnet på målvariabelen i navnet. Det kan også bety å gjennomføre treningen med feil prediktorer ved å inkludere eksempler som ikke er relevante for prediksjon i det virkelige datasettet, for eksempel id-nummer, slik at algoritmen lærer mønsteret på nummerrekkefølgen fremfor andre mer relevante egenskaper (O'Neil, 2012). Begge utfordringene er først og fremst relatert til testing av modeller, men har konsekvenser for det endelige klassifikasjonssystemet. Det er ønskelig å få et så godt testresultat som mulig og det er lett å fokusere for mye på testresultatene fremfor prestasjonen med det virkelige datasettet. Feilene kan være vanskelig å oppdage, men et for godt eller tilnærmet perfekt resultat indikerer som regel at noe er feil (Owen, 2012, s. 282).

3 Metodisk tilnærming

3.1 Datasett

Datasettet brukt i oppgaven består av 2.8 millioner, nærmere bestemt 2 780 299 millioner, representasjoner av bøker med metadata fra Amazon.com og LibraryThing.com hentet fra INEX 2014 Social Book Search Track. The INitiative for the Evaluation of XML retrieval (INEX) startet i 2002 og er et internasjonalt prosjekt for å skape evalueringsprosedyrer for innholdsbasert gjenfinning av xml-dokumenter. INEX har som hovedoppgave å stille til veie testkorpuser og evalueringsmetoder (Fuhr, Gövert, Kazai, & Lalmas, 2002). Gjenfinning har flere aspekter og er i stadig utvikling og INEX er derfor delt inn i flere spor for å reflektere de problemstillingene som til en hver tid er gjeldene. Et slik spor er Social Book Search Track som har til formål å utforske teknikker som kan hjelpe brukere med søk og navigering i både profesjonelle og brukergenererte metadata (“INEX,” 2014).

Alle bøkene i korpuset har et unikt ISBN-nummer som identifiserer hver enkelt bok og fordi et verk kan være publisert i flere utgaver, hver med sin unike ISBN, kan et verk forekomme mer enn en gang (Koolen, Kazai, Preminger, & Doucet, udatert). Hver representasjon er et eget xml-dokument og består av en blanding av formelle og sosiale metadata (vedlegg 1). Formelle metadata er det som tidligere i oppgaven har blitt omtalt som konkrete egenskaper, altså bibliografiske data. Sosiale data er ulike typer brukergenererte data. Fra Amazon.com er det bibliografiske data med blant annet forfatter, tittel, forlag, publiseringsdato, klassifiseringskoder samt brukergenererte anmeldelser og vurderinger (karakterer). Fra LibraryThing.com er det brukergenerert data i form av tagger, pristildelinger, personkarakterer og stedsangivelser. Det er ikke hensiktsmessig å liste opp alle tilgjengelige elementer fra datasettet her og en full oversikt finnes i figur 2. Antallet elementer i oversikten kan virke overveldende, men i praksis er kun et mindretall brukt regelmessig gjennom hele datasettet. De bibliografiske dataene som kan kalles vanlige er alltid med slik som ISBN, tittel, forfatter, forlag og publiseringsdato, andre elementer slik som dedikasjoner (dedications) og sitater (quotations) er sjelden brukt. Anvendelsen av brukergenererte data er i hovedsak begrenset til tre elementer: tagger, brukervurderinger og brukeranmeldelser. En analyse av korpuset etter indeksering avdekker at tagger finnes i 2.3 millioner dokumenter. Brukervurderinger og brukeranmeldelser finnes i 1.3 millioner dokumenter og tilsvarer nesten halvparten av korpus.

Figur 3: Xml-elementer i datasett fra INEX (Koolen, Kazai, Preminger, & Doucet, udatert)

tag name			
book	similarproducts	title	imagecategory
dimensions	tags	edition	name
reviews	isbn	dewey	role
editorialreviews	ean	creator	blurber
images	binding	review	dedication
creators	label	rating	epigraph
blurbers	listprice	authorid	firstwordsitem
dedications	manufacturer	totalvotes	lastwordsitem
epigraphs	numberofpages	helpfulvotes	quotation
firstwords	publisher	date	seriesitem
lastwords	height	summary	award
quotations	width	editorialreview	browseNode
series	length	content	character
awards	weight	source	place
browseNodes	readinglevel	image	subject
characters	releasedate	imageCategories	similarproduct
places	publicationdate	url	tag
subjects	studio	data	

Korpuset fra Social Book Search Track ligger på en MySQL-database ved Høgskolen i Oslo og Akershus og dokumentene brukt i denne oppgaven ligger som tekst i hver sin rad under en kolonne som vist i figur 3. For å gjøre arbeidet med indeksering og testing i forbindelse med gjennomføring av oppgaven har det blitt skrevet et enkelt skript i programmeringsspråket PHP som henter ut teksten fra radene i MySQL-databasen og lagrer teksten som xml-filer på en lokal datamaskin (vedlegg 2).

Figur 4: Dokumenter i MySQL-database

The screenshot shows the phpMyAdmin interface for a MySQL database. The table 'a_it' is selected, and the columns 'ind', 'isbn', 'XML', 'lgt', and 'lt_work_id' are visible. The table contains three rows of data:

	ind	isbn	XML	lgt	lt_work_id
<input type="checkbox"/>	974432	0316517283	<?xml version="1.0" encoding="UTF-8" standalone="n...	2974	5028813
<input type="checkbox"/>	951215	0340910283	<?xml version="1.0" encoding="UTF-8" standalone="n...	2918	2476050
<input type="checkbox"/>	1449624	1424169283	<?xml version="1.0" encoding="UTF-8" standalone="n...	4237	7208285

3.2 Verktøy

I gjennomføringen av underspørsmålet om hvordan avdekkingen av subtile egenskaper kan realiseres må det benyttes noen verktøy. Dokumentene i datasettet må indekseres og gjøres søkbare samt at det må gjennomføres en klassifikasjonsprosess. Verktøyene som i hovedsak blir benyttet i denne oppgaven er Solr og OpenNLP til henholdsvis indeksering og klassifisering. Av de to er Solr det mest omfattende og selv om det denne oppgaven ikke har til hensikt å gjøre rede for alle detaljer, spesielt med tanke på at størrelsen til Solr, må noen grunnleggende konsepter forklares. Det har blitt forsøkt å holde redegjørelsen, for både Solr og OpenNLP, så kortfattelig som mulig og kun gå nærmere inn i de elementene som er nødvendig for oppgaven. Enkelte elementer vil bli forklart overfladisk og bruken er ikke spesifikt forklart, men vil utgreies i større detalj i forbindelse med gjennomføringen. Begge verktøyene er skrevet på engelsk og for de ordene som lar seg oversette brukes norske uttrykk. For andre, enten fordi termene er vanskelig oversettbare eller fordi termene anvendes direkte av verktøyene, benyttes det opprinnelige engelske uttrykket.

3.2.1 Solr

Solr beskrives som en åpen kildekode søkeplattform bygget på toppen av biblioteket Lucene (Rafalovitch, 2013, s. 7). Solr brukes i dag av en rekke store aktører på web deriblant CNet.com og Netflix.com (Smiley, Pugh, & Eteve, 2009, s. 7) Lucene, kjerneteknologien i Solr, er et åpent kildekode bibliotek for indeksering og søk skrevet i java og er et prosjekt under Apache Software Foundation. I følge forfatterne av Lucene in Action er Lucene nå det biblioteket som oftest brukes i gjenfinningsoppgaver (McCandless, 2010, s. 3). Lucene er ikke et program som kan settes opp og brukes umiddelbart, men et verktøy som kan anvendes til å gi andre programmer indekserings- og søkemuligheter (2010, s. 7). Solr kan best forklares som en serverutgave av Lucene og kan i sin enkleste utgave benyttes som et program uten særlig konfigurering, i praksis er det likevel langt mer komplisert. Solr har i tillegg til indeksering og søkemuligheter også egne funksjoner som ikke kommer fra Lucene, for eksempel fasetter. Smiley et al. skriver i *Solr 1.4 Enterprise Search Server* at det ofte vanskelig å skille mellom hva som er Solr og hva som er Lucene (2009, s. 8). Ekstra vanskelig er det å skille når både Lucene og Solr kan utvides og det har blitt laget en rekke tillegg som blander funksjoner fra både Lucene og Solr. I oppgaven er det Solr som benyttes, men Lucene er viktig for å forstå hvordan Solr fungerer og spesielt viktig når egne funksjoner og metoder skal integreres. Fremover vil både Lucene og Solr normalt bli referert til som Solr

selv om enkelte av funksjonene tilhører Lucene. I motsetning til Lucene, som kun er et bibliotek, har Solr mulighet til å kommunisere over http samt et administrativt webgrensesnitt. Det må likevel presiseres at Solr er en server, ikke en søkemotor for eksempelvis webinnhold og eventuelle tjenester må bygges utenom.

Selve konfigureringen foregår i xml-filer, men webgrensesnittet er viktig for utprøving og forståelse av hvordan data indekseres og hentes ut (Smiley et al., 2009, s. 8 - 9). Solr bruker en invertert indeks for å lagre data og benytter det som omtales snaut ”A good scoring algorithm based on sound Information Retrieval (IR) principles to produce the more likely candidates first” (s. 8). En invertert indeks er en metode for å redusere dokumentene til en enkel liste for å øke effektiviteten og hastigheten ved søk. En enkel invertert indeks består av to deler; vokabular og forekomst. Vokabular er alle ordene i alle dokumenter og forekomst er i hvilke dokumenter det aktuelle ordet forekommer (Baeza-Yates, 2011, s. 340). Tabell # viser prinsippet med en invertert indeks med vokabularet som termer i venstre kolonne og hvilke dokumenter de forekommer i til høyre.

Tabell 1: Invertert indeks (Grainger & Potter, 2014, s. 54)

Term	Forekommer i dokument
a	1,3,4,5,6,7,8
becoming	8
beginners	6
buy	9
buying	4,5,6
car	4
...	...

En invertert indeks er i praksis mer komplisert enn den enkle forklaringen og eksempelet gitt ovenfor og vil normalt inneholde posisjonen til ordet eller de enkelte bokstavene i dokumentet. For eksempel kan innførselen for ordet ”tekst” se slik ut: [1 [1,7,19], 4 [9,12,17]]. Tallene i den innerste klammeparentesen indikerer hvor i dokumentet, ytterste klammeparentes, ordet tekst befinner seg. Den utvidete utgaven av invertert indeks kalles også full invertert indeks (Baeza-Yates, 2011, s. 341). I praksis brukes kun termen invertert indeks. Det Smiley et al. i sitatet over omtaler som ”A good scoring algorithm” innebærer i utgangspunktet en vektorbasert modell for å sammenligne en spørring og dokumentene i systemet med tf-idf vektning (2011, s. 77 - 78). Det er fullt mulig å velge andre modeller i Solr

, for eksempel probabilistiske metoder eller språkmodeller, men vektorbasert er forhåndsvalgt (Grainger & Potter, 2014, s. 567). For gjennomføringen av oppgaven har det liten betydning hvilken modell som benyttes og det vil ikke bli gitt en helhetlig gjennomgang av vektorbasert gjenfinning her.

Solr-konfigurasjonen består av minimum tre deler, solr.xml, schema.xml (Vedlegg 4) og solrconfig.xml (Vedlegg 5). I tillegg kan det benyttes egne konfigurasjoner for importering av data. solr.xml er et kort dokument som forteller Solr hvilke konfigurasjonsfiler og indekser som skal brukes. Schema har større grad av kompleksitet og definerer hvordan Solr skal behandle inndata, både ved indeksering og ved spørring

3.2.1.1 Skjema

Schema forteller Solr om innholdet i dokumentet som skal indekseres og formen på utdata; hvordan en Solr-post skal se ut. Dokumentet består av felter (<field>), definisjon av hvordan feltene skal behandles (<fieldType>) og angivelse av hvilket felt som skal brukes som en unik nøkkel for hver post (<uniqueKey>). Solr feltene får data fra et dokument som indekseres, for eksempel kan tittel hentes fra et xml-dokument med en tittel tagg. I et felt er det blant annet mulig å angi om hvilken felttype som brukes, om det skal indekseres, om det skal lagres eller om det skal kunne opptre mer enn en gang. I eksempelet under er det angitt navn på feltet (name), hvordan feltet skal behandles (type), om feltet skal indekseres (indexed), om det skal lagres (stored) og om det kan forekomme flere ganger i samme dokument (multiValued):

```
<field name="title" type="text_general" indexed="true" stored="true" multiValued="false"/>.
```

At et felt lagres innebærer at alt, eller deler av, innholdet i feltet lagres av Solr og kan fremvises i et søk. I eksempelet over vil ikke tittel bare være søkbart, men vil også kunne vises frem i treffet. Det har normalt størst betydning for den videre behandlingen av en Solr-post, for eksempel i forbindelse med programmering av en side eller utheving av søkeordet i en treffliste. Hvert felt har en type som må defineres og blir gjort i *fieldTypes*; er feltet en eller annen form for tall, ren tekst, en frase, dato, valuta etc.:

```
<fieldType name="nametext" class="solr.TextField">
```

I *fieldType* angis det også hvordan innholdet i et felt skal analyseres. En analyse av et felt kan bestå av tre komponenter, i tillegg til typedefinisjon: *analyzer*, *tokenizer* og *filter*. Det er viktig å presisere at Solr ikke indekserer tekst, men leksikalske enheter, det som i Solr kalles *tokenz*. Det betyr at teksten brytes ned til sine enkelte bestanddeler og noe som kan gjøres på flere måter. For eksempel kan et oppsett spesifisere at alt mellom to mellomrom skal gjøres

om til en token. Setningen: *jeg , har ikke, en stol* bli følgende tokens: "*jeg*" ", " "*har*" "*ikke,*" "*en*" "*stol*". Om komma blir en del av foregående token eller blir en egen avhenger av mellomrommet.

3.2.1.1.1 Analyse, enheter og filtre

Analyzers kan bestå enten en enkelt forhåndsdefinert oppsett, en klasse, eller være en kjede *tokenizere* og *filtre*. I *analyzers* med bare en klasse må den aktuelle klassen være fra Lucene. Solr kommer med en rekke *tokenizere* og *filtre* det er mulig å benyttes og som kan kombineres i en og samme kjede. I eksempelet under er det benyttet en standard *tokenizer*, og tre filter som behandler token-strømmen. Først gjøre alle store bokstaver om til små, deretter fjernes alle stoppord før alle enhetene fra token-strømmen stemmes:

```
<fieldType name="nametext" class="solr.TextField">
  <analyzer>
    <tokenizer class="solr.StandardTokenizerFactory"/>
    <filter class="solr.LowerCaseFilterFactory"/>
    <filter class="solr.StopFilterFactory"/>
    <filter class="solr.EnglishPorterFilterFactory"/>
  </analyzer>
</fieldType>
```

Tokenizer tar imot tekst og generer en token-strøm, teksten brytes som omtalt tidligere ned i enkeltkomponenter og gis en plass i den inverterte indeksen. Det er kun mulig å bruke en *tokenizer* i hver kjede, altså i hvert felt. *Filters* har ingen slik begrensning og kan kjedes ubegrenset innenfor samme *analyzer*. *Filters* tar i mot en token-strøm og behandler hver enkelt token etter reglene i filteret. For eksempel kan en token-strøm renses for stoppord eller store bokstaver. Det finnes og et lite antall filtre, såkalte *CharFilters*, som kan benyttes før en *tokenizer* slik at teksten som går inn i en *tokenizer* er forskjellig fra teksten i dokumentet som indekseres. For eksempel kan alle html eller xml koder stripres fra en tekst slik at de ikke blir *tokenz*. Inndata for Solr kan være ulike kilder og format: ren tekst, kommaseparert liste, pdf, xml for å nevne noen. I denne oppgaven er det brukt xml som inndata. Det er også flere måter å hente inn data til Solr, men de vanligste metodene er å enten sende et dokument til Solr eller ved å indekser innholdet fra en plassering. For denne oppgaven er det sistnevnte som er aktuelt. Når dokumenter skal indekseres fra en plassering, på disk eller server, må det benyttes en *importhandler* som defineres i et egen xml-fil (Vedlegg 6). I konfigurasjonsfilen for importering av data kan et felt i Schema linkes nøyaktig til en del av et dokument. I det vedlagte dokumentet er det brukt en *xpath-processor*, men andre prosessorer er tilgjengelig. Det er og mulig å gjøre forskjellige transformasjoner av dataene som kommer inn slik at dato får et bestemt format eller at enkelte tegn byttes ut er noen eksempler.

3.2.1.2 Konfigurasjonsfil

I Solrconfig defineres oppførselen til Solr samt angivelse av hvilke eksterne bibliotek, med plassering, Solr skal benytte. Det er mulig å spesifisere med stor detalj hvordan Solr skal behandle ulike aspekter ved indeksering og søk. For eksempel kan det defineres hvilke felter det skal søkes i, hvilke felter det skal legges ekstra vekt på eller mer komplekse operasjoner som hvordan resultatlistene skal behandles i hurtigbuffer. Heldigvis kommer Solr med forhåndsvalgt oppførsel og man trenger kun å forholde seg til de delene man ønsker å forandre. Defineres ikke oppførselen velger Solr å bruke standard. I Solr in Action uttrykkes det slik: "[...]solrconfig.xml has a number of complex sections. The good news is that you don't have to worry about these until you encounter a specific need" (Grainger & Potter, 2014, s. 86). Hvordan Solr konfigureres i denne oppgaven vil bli forklart nærmere i kapittelet som omhandler gjennomføring av underspørsmålet.

3.2.2 OpenNLP

OpenNLP et relativt nytt prosjekt under Apache Software Foundation og fikk først status som et offisielt Apache prosjekt så sent som i 2012 ("News," 2014). Dokumentasjon, bøker og veiledninger er derfor svært sparsommelig og preger gjennomgangen i denne oppgaven. OpenNLP er i likhet med Solr et åpent kildekodebibliotek skrevet i java og må betraktes som et verktøy fremfor et enkeltstående program. På prosjektets webside beskrives OpenNLP som en verktøykasse for prosessering av naturlig språk (NLP) basert på maskin læring og kan brukes til vanlige oppgaver innen NLP ("Welcome to Apache OpenNLP," 2014). Det prosjektet betegner som vanlige NLP-oppgaver innbefatter blant annet tokenisering, setningsidentifikasjon, identifisere ulike deler i en tekst (POS), syntaksanalyse og entitetsidentifisering samt tekstkategorisering. For denne oppgaven er det spesielt to av verktøyene som er interessante; setningsidentifikasjon og tekstkategorisering.

3.2.2.1 Setningsidentifikasjon

Setningsidentifikasjon, ofte kalt setningssegmentering, innebærer å finne ut hvor en setning begynner og slutter og hvor formålet er å identifisere meningsfulle hele setninger. Normalt vil en setning være alle bokstaver mellom to punktum hvor den første bokstaven er stor, men det mange unntak. Et mye brukt eksempel er setningen: Mr. Smith went to Washington. Problemet er tydelig og setningen vil bli delt opp i to mindre setninger med lite mening: 1. Mr., 2. Smith went to Washington (Ingersoll, 2013, s. 27). Et annet eksempel er datoer som kan skrives for eksempel slik 1. Januar. OpenNLP definerer ikke setninger annerledes men bruker i tillegg modeller for å identifisere dem. Modellene trenes ved å gi eksempler på hva

en som er en setning, på samme måte som for trening av modeller for tekstklassifisering (“Apache OpenNLP Developer Documentation,” udatert.) Ferdig trente modeller følger med OpenNLP for en rekke språk, men det er mulig å lage egne om ønskelig.

3.2.2.2 Tekstkategorisering med OpenNLP

OpenNLP har som nevnt en mulighet for å automatisk kategorisere tekst og gjør det ved å benytte algoritmen Maximum Entropy (MaxEntropy). Det overordnede prinsippet for MaxEntropy er at når ingenting er kjent skal fordelingen være så uniform som mulig, altså ha maksimum entropi (Kocik, 2004, s. 2). Det er ikke umiddelbart klart hva dette innebærer og et eksempel kan være oppklarende. I en situasjon hvor kategoriseringen kan få fire mulige utfall vil noen ord, kjent fra modellen, ha en viss prosentvis sjans for å havne i en av de fire kategoriene. For eksempelet skyld kan ordet *halsbrekkende* ha 40 prosent sjans for å havne i kategorien *raskt tempo*, og 20 prosent sjans for å havne i en av de tre andre kategoriene. Er derimot ordet ikke kjent vil fordelingen være uniform og det er 25 prosent sjans for hver kategori (Nigam et al., 1999, s. 62). Ved å sette avgrensninger for den uniforme fordelingen i et treningssett kan programmet lære betingelsene for kategorien (s. 61). I praksis betyr det å gi nok eksempler til en modell som kan vise programmet hvilke avvik som karakteriserer en definert kategori. Fordelen med å bruke MaxEntropy fremfor andre algoritmer er at den ikke antar at egenskapene er uavhengige

3.3 Tilpassning av Saricks rammeverk

For at Saricks rammeverk skal kunne brukes i denne oppgaven har det vært nødvendig å gjøre noen tilpasninger. Gjennomføringen av underspørsmål to er ment å være utforskende og konseptuell og rammeverkets størrelse er således for omfattende. Modellene i denne oppgaven blir relativt små i forhold til hva som er optimalt i en virkelig situasjon og det blir derfor vanskelig for klassifiseringsalgoritmen å skille mellom hva som tilhører de ulike kategoriene. Problemet eskalerer proporsjonelt med antallet kategorier fordi det blir flere kategorier som vil passe objektet som klassifiseres tilnærmet likt (Marmanis, 2009, s. 167). For at algoritmen skal få så enkle arbeidsvilkår som mulig har det blitt besluttet å dele inn i to distinkte kategorier med tydelige vilkår. Det har derfor blitt tatt noen valg i forhold til hvordan de ulike elementene kan forenkles. Første steg er å eliminere elementet handling. Som tidligere beskrevet er handling i følge Saricks vanskelig å oppdage i samtaler eller brukeranmeldelser og er i stor grad inkorporert i de tre andre elementene. Etter elimineringen gjenstår det nå tre elementer, tempo, karakterisering og ramme. Neste steg er å dele inn hvert av de tre elementene inn i kun to kategorier. Det vil si at hvert kategori kan få to utfall når bøkene i

datasettet senere skal klassifiseres. En slik forenkling er ikke uproblematisk når alle aspektene ved for eksempel elementet karakter det tas hensyn til i Saricks rammeverk når appell skal avdekkes. Til hvert element er det laget en tabell som vil bli brukt videre i forbindelse med gjennomføringen av klassifikasjonen og viser hva som kjennetegner hver kategori. Det er kun tempo som har to definerte kategorier hvor indikatorene med enkelhet kan plasseres. Karakterisering har blitt bearbeidet til å passe inn i to kategorier og med elementet ramme har det blitt gjort større avgrensninger.

3.3.1 Tilpassning av elementet tempo

Tempo er det elementet hvor tilpasningen er lettest å gjennomføre fordi det naturlig deles inn i kun to kategorier, raskt eller sakte tempo. Bøker med heseblesende takt eller intensitet fra første avsnitt og tettskrevne saktegående poetiske bøker må betraktes som ytterpunkter, og satt opp mot hverandre er det lett å henfalle til en kategorisering av det første som dårlig litteratur og det siste som god. Hensikten med inndelingen er derimot å forstå hva en leser ønsker eller i humør for, ikke å definere en bok til å være god eller mindre god. Tabellen under gir en oversikt over hva som kjennetegner enten høyt eller saktegående tempo.

Tabell 2: Indikatorer for tempo

Høyt tempo	Lavt tempo
Karakterene avdekkes raskt	Karakteren avsløres over tid
Karakteren reagerer på ytre hendelser	Karakteren er psykisk involvert i hendelser
Plottet avsløres tidlig	Plottet avsløres over tid
Orientert mot en avslutning	Åpen avslutning
Mye dialog	Detaljerte beskrivelser
Korte setninger, avsnitt og kapitler	Tettskrevet med fremhevet språk
Komprimert tidslinje	

3.3.2 Tilpassning av elementet karakterisering

Indikatorene for hvordan karakteriseringen i en bok opptrer sjelden alene og vil normalt gli over i hverandre. For eksempel kan en karakter betraktes fra flere synsvinkler og hovedpersonen er ikke enten overdrevent stereotyp eller bare ugjenkjennelig. Det er ikke meningen å definere en karakter presist men å gi rådgivere et verktøy for å øke forståelsen av hva slags karakter leseren ønsker og er i humør for. Trass i at karakterer ikke kan deles like enkelt i definerte kategorier slik som tempo er det tabellen nedenfor blitt gjort et forsøk på å dele inn indikatorene inn i to grupper. For at den skal kunne brukes videre i oppgaven er listen

outrert og merkelappene benyttet på gruppene representerer ytterpunktene av appellelementet karakterisering. Det betyr ikke at karakterer tilhørende den ene gruppen regnes som bedre eller mer litterært overlegen og må kun betraktes som en inndeling.

Tabell 3: Indikatorer for karakterisering

Overfladisk karakter	Kompleks karakter
Lett gjenkjennbar stereotyp karakter	Karakteren utvikles gradvis
Karakteren endres i liten grad	Karakteren fremstår med overraskende og unike trekk
En enkelt synsvinkel (person)	Flere synsvinkler (personer)
Ytre handling (hva skjer med karakteren)	Indre tanker og følelsesmessige reaksjoner
Leseren kan identifisere med karakteren	Leseren kan ikke identifisere med karakteren
Historien følges gjennom karakterens øyne	Historien betraktes på avstand

3.3.3 Tilpassning av elementet ramme

For å kunne lage en oversikt over kjennetegn slik at elementet ramme kan benyttes videre i oppgaven har det vært nødvendig å gjøre noen avgrensninger. Indikatorene er så forskjellige at de vanskelig samlet kan være definerende for en kategori og den indikatoren flest lesere forholder seg til, grunntone, har derfor blitt valgt. Bakteppe og innramming har ikke den samme graden av anvendelighet for lesere når litteratur skal beskrives og kan muligens sies å være knyttet opp mot spesifikke sjangere hvor tidskoloritt eller fantastiske verdener er sentralt. De bortvalgte indikatorene er ikke uvesentlige men kan lettere fanges opp av andre appellelementer. I tabellen nedenfor er grunntone, som representant for hele elementet, blitt delt inn i henholdsvis mørk og lys.

Tabell 4: Indikatorer for ramme

Mørk grunntone	Lys grunntone
Mørk eller dyster atmosfære og stemning	Lys eller lett atmosfære og stemning
Leseren opplever følelsen som mørk eller dyster	Leseren opplever følelsen som lys eller lett
Mørk eller dyster tone	Lys eller lett tone

3.4 Utarbeidelse av en modell

Klassifiseringssystemet bør som nevnt tidligere kjenne til så mange eksempler som mulig for å kunne prestere optimalt. Hva som er optimalt avhenger av formålet og kostnader forbundet med å utarbeide en modell fordi å lage en god modell kan være tidkrevende og krever ikke

bare nok eksempler, men også evaluering og finjustering. For denne oppgaven sin del kravene satt lavt og ment å vise et konsept og ikke være et fungerende system ment for produksjon. Det er laget tre modeller, en for tempo, en for karakterisering og en for ramme. Modellene blir ikke vist her men er vedlagt oppgaven i sin helhet. Tidligere har det blitt avklart at tempo er det styrende elementet og er det elementet det har blitt arbeidet mest med. Fremgangsmåten vil bli beskrevet nærmere i påfølgende avsnitt. For å forenkle prosessen med å lage en modell er det vanlig å bruke en teknikk kalt *bootstrapping* og betyr at egenskapene som trengs for trening blir forsøkt funnet automatisk. For eksempel kan nøkkelordssøk brukes til finne dokumenter med ønskelige egenskaper for deretter å hente ut disse (Ingersoll, 2013, s. 184). Et annen fremgangsmåte kan være å bruke en uadministrert klassifikasjonsalgoritme til å finne egenskaper som kan brukes i en administrert klassifikasjon (Owen, 2012, s. 238).

I arbeidet med oppgaven har det blitt gjort ulike forsøk på å forenkle utarbeidelsen av en modell. En initiell tanke var å kun bruke adjektivene i brukeranmeldelsene eventuelt utvidet med synonymer, men det viste seg fort at vanskelighetsgraden var for stor og at adjektiver alene ikke kunne dekke det utbroderende språket som ofte brukes i beskrivelsen av bøker slik Saricks pekte på. Spesielt problematisk er figurative uttrykk hvor meningen ikke kan forstås ved kun å undersøke de enkelte ordene (*Handbook of natural language processing*, 2010, s. 95). Det ble derfor besluttet å lage modellen manuelt og som en følge er modellene svært små og er mest sannsynlig overtilpasset. Det er brukt tabellene fra kapitlet om tilpasning av Saricks rammeverk for å lete etter setninger i datasettet som svarer til indikatorene. Eksemplene brukt til trening er overtydelige og vil for nye eksempler fra datamaterialet som er mer eller mindre like treningseksemplene fungere godt. Det betyr også at nye eksempler som ikke er like tydelige som treningseksemplene vil fungere særdeles dårlig.

3.4.1 Modell for tempo

Modellen for tempo (Vedlegg 7) var i utgangspunktet på 301 linjer fordelt omtrent likt mellom Fast-paced og Slow-paced. Det brukes engelske kategorier i modellen for å bruke samme språk som i datasettet. Fordi karakterisering henger nøye sammen med tempo ble eksemplene funnet til karakteriseringsmodellen lagt til modellen for tempo. Modellene er uavhengige av hverandre og klassifiseringen foregår i tre omganger slik at karakterisering ikke vil bli påvirket kopieringen av eksemplene til tempo. Eksempler som tydet på en enkel karakterutvikling ble lagt til som Fast-paced og eksempler med indikasjoner på en kompleks

karakter ble lagt som Slow-paced. Den utvidete modellen for tempo er nå på totalt 471 eksempler.

3.4.2 Modell for karakterisering

Modellen for karakteriseringer (Vedlegg 8) inneholder totalt 170 eksempler hvor av 92 målvariabelen Complex characters og 78 har målvariabelen Shallow_characters.

3.4.3 Modell for ramme

Modellen for ramme (Vedlegg 9) består av 135 eksempler til sammen og er fordelt med 64 i målvariabelen Light tone og 71 i målvariabelen Dark Tone.

4 Analyse av underspørsmål 1

4.1 Hvilke typer brukergenererte data er hensiktsmessig å benytte for å avdekke underliggende subtile egenskaper i et bokverk?

Et delmål for denne oppgaven er å undersøke hvilke brukergenererte data som kan anvendes for å avsløre de underliggende egenskapene til et verk. Underspørsmålet som skal besvares lyder: Hvilke typer brukergenererte data er hensiktsmessig å benytte for å avdekke underliggende subtile egenskaper i et bokverk? Det er viktig å understreke at resultatene må sees i sammenheng med datasettet og er ikke nødvendigvis generaliserbare. Det har ikke blitt gjennomført en brukerundersøkelse og det er ikke undersøkt hvordan dataene er brukt i forhold til innholdet i de enkelte dokumentene. Analysen er kun gjennomført ved fortolkning av datasettet som helhet slik det fremstår etter indeksering.ding

4.1.1 Tagger

Både Syversen (2011) og Koolen et al. (2013) nevner tagger som en mulig kilde til det de kaller henholdsvis subtile egenskaper og affektive aspekter. Om tagger kan regnes som metadata er det en viss uenighet om. Berendt og Hansers argumenterer for at tagger ikke oppsummerer en ressurs, men fungerer som ekstra innhold og dermed ikke kan regnes som metadata (Peters, 2009, s. 156). Hammond et al. har et motstridende syn og mener kun er en type ny type metadata og kan brukes som et supplement i formelle klassifikasjonssystemer. Etter arbeidet med datasettet fra INEX er inntrykket at taggene verken fungerer som ekstra innhold eller er anvendelig i klassifikasjonssystemer. En analyse av korpuset etter indeksering bekrefter langt på vei inntrykket. Figur 4 er en oversikt over de ti mest brukte taggene i datasettet og viser tydelig at selvorganisering er fremste anvendelsesområde (Tabell 4). Format, om det er skjønnlitteratur eller sakprosa, og om boken er lest eller ikke går igjen. En

oversikt over de 50 mest brukte taggene forandrer bildet til en viss grad fordi tagger med sjangerangivelse begynner å gjøre seg gjeldene (vedlegg 3). Ser man nærmere på de 500 hundre mest brukte taggene er det fortsatt mer av det samme, men nå utgjør også akronymer, uforståelige uttrykk og spesialtegn en del av listen. For eksempel består mange taggene kun av en bokstav og tegnet * er brukt mer enn 13500 ganger. Noen av taggene kan riktignok brukes i et klassifikasjonssystem, men svært mange vil ha liten verdi fordi taggene har som formål å organisere brukernes egen samling. Brukernes fokus på gjenfinning av egne bøker fører også til at innholdet ikke oppsummeres og reflekteres blant annet ved å benytte termer som ”ulest” og ”lest” eller ”bibliotek”. Det viktigste for brukerne er å huske.

Tabell 5: Ti mest brukte tagger i datasett fra INEX

526,777	Fiction
518,138	non-fiction
354,931	read
349,106	unread
291,153	History
280,810	own
213,657	novel
211,458	tbr
209,792	Paperback
162,231	Reference

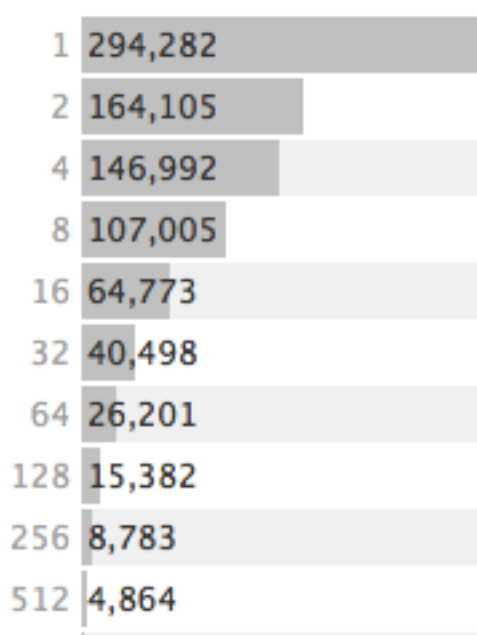
Lenke til server: <http://bibin.hioa.no:8090/solr-s162452/#/nlpCat2/schema-browser?field=tag>

Listen over topp 500 tagger er ikke vedlagt oppgaven, men kan enkelt finnes på [indekseringsserveren](#) med lenken under tabell 4. Etter topp 500 er antallet ganger hver tagg er brukt så lav sett i forhold til den totale mengden at nytteverdien av en videre analyse ikke rettferdiggjør påkrevd arbeidsmengde. En erkjennelse etter arbeidet med datasettet er at brukernes tagger er avhengig av nettsamfunnet, av konteksten. Ames og Naaman undersøkte hva som motiverer brukere til å tagge og fant at selvorganisering var et av fire hovedmotiver. Nov et al. kommer frem til et lignende resultat når de konkluderer med at selvorganisering og deling med andre er hovedmotivene for tagging (Phuong, 2011, s. 19). Undersøkelsene det henvises til utforsket motivasjonen for tagging på Flickr.com som er en billedelingstjeneste hvor formålet er å vise frem bilder til andre. En kvalitativ studie fra 2008 av Cox, Clough & Marlow viser et motstridende resultatet. De aller fleste bruker Flickr.com for å dele bilder og

motivasjonen er å få reaksjoner fra andre brukere. Organiseringen av egen fotosamling gjør brukerne på egen maskin. Kun en av i alt 100 intervjupersoner oppga å bruke Flickr som organiseringsverktøy (Peters, 2009, s. 80). I datasettet fra INEX bærer taggene preg av å være ment for selvorganisering i mye større grad enn det gis uttrykk for i undersøkelsene av Flickr.com. Årsaken er muligens å finne i konteksten. LibraryThing.com er en tjeneste hvor brukere legger til bøker i egen samling og er en form for selvkatalogisering. Bruken er dermed først og fremst ment å være til hjelp for navigering i egen samling og ikke for å kommunisere med andre brukere.

Av de 500 mest brukte taggene er det kun tre som er satt sammen av termer uten separasjon og det er kun en tagg på et annet språk en engelsk. Dette stemmer ikke med funnene til Vuorikari et al. som fant tagger ofte består av useparerte termer og på andre språk (Phuong, 2011, s. 21). Spiteri fant at tagger ofte består av slang, sjargong akronymer eller er feilstavet. Dette er ikke tilfelle blant de 500 mest brukte taggene i datasettet fra INEX. Akronymer er brukt kun 18 ganger og slang, sjargong og akronymer forekommer ikke. En antakelse er at forekomsten av problemene Spiteri peker på ville vært høyere om også taggene utenfor topp 500 listen hadde blitt undersøkt. Wu og Zhou fant i sin studie fra 2009 at tagger brukes inkonsekvent og at spredningen av termer derfor blir stor (Phuong, 2011, s. 21). Om dette er tilfelle for brukerne av LibraryThing.com er vanskelig å avdekke uten en brukerstudie som men ved å se på distribusjonen av taggene det kan likevel være mulig å se noen antydninger. Histogramet i tabell 5 viser at det totalt er det brukt 878 697 unike tagger. Av disse er 294 282 tagger kun brukt i et dokument, 164 105 tagger er kun brukt i 2 eller 3 dokumenter . Til sammen utgjør taggene brukt i 3 eller færre dokumenter mer enn halvparten alle taggene i datasettet. Å konkludere med at brukerne ikke er konsekvente, særlig for den enkelte, vil være forhastet. Like fullt er det tydelig at et større antall av taggene brukes sjelden. En årsak kan være at taggene ikke brukes for å oppsummere innholdet, men som personlige huskelapper.

Tabell 6: Fordeling av tagger



Lenker til server: <http://bibin.hioa.no:8090/solr-s162452/#/nlpCat2/schema-browser?field=tag>

Etter analysen spesielt et punkt fremtredende. Brukerne på LibraryThing.com bruker først og fremst tagger for selvorganisering for eksempel ved å angi om boken skal leses, sjanger eller bare en enkelt bokstav. Det kan betraktes som en bekreftelse på funnene til Stvilia et al. som fant at brukere primært benytter tagger ved navigering (2012). I denne konteksten innebærer det navigering i egen samling. Taggene formidler i liten grad lesernes følelser og er således dårlig egnet til å bruke i et system som forsøker å avdekke subtile egenskaper i et verk.

4.1.2 Brukervurderinger

Ingen av studiene under tidligere forskning nevner brukervurderinger som et alternativ for å avdekke underliggende elementer ved bøker. Det er likevel tatt med i oppgaven fordi det er en av tre typer brukergenererte data i datasettet med nok bruk til at de er aktuelle og for å vise hvorfor brukervurderinger ikke er tatt med inn i gjennomføringen. Å finne motivet bak en brukervurdering er ikke mulig uten en brukerundersøkelse eller en større sammenligning av gitt vurdering og beskrivelsen av hvorfor. At det er vanskelig å gi uttrykk for komplekse meninger gjennom en enkelt verdi er en åpenbar svakhet med brukervurderinger og Cosley et al. poengterer også det i sin studie av brukervurderinger (2003). Det finnes ingen enighet om anvendelsen av brukervurderinger og hva som faktisk blir vurdert varierer fra leser til leser. Det som er et viktig aspekt for en bruker er ikke nødvendigvis viktig for en annen bruker. En kan være fornøyd med prisen i forhold hva han eller hun får igjen, mens en annen bruker kan like at en bok var lettlest. Utdraget under, om Dan Browns *Iskaldt Bedrag*, kan være

illustrerende for problemet. Boken er ikke likt og anbefales ikke. Likevel har brukeren gitt boken vurderingen fem hvor fem er høyeste karakter.

<summary>nice value !! but disappointed story</summary>

<content>I have read all of his book but this one was't like the others. It is nice value (dollar amount wise when you compare with other off-line book store price) but I don't recomand this book.</content> <rating>5</rating>

Utdraget viser også at i likhet med tagger kan nettsamfunnet og konteksten ha en avgjørende betydning. Brukeranmeldelsene og –vurderingene i datasettet er hentet fra Amazon.com som er en kommersiell side for kjøp. Det reflekteres når kriteriet som leder til den gode vurderingen er pris. Boken blir nevnt, men ser ikke ut til å være avgjørende for brukeren når det skal gis en vurdering. Gjennomsnittlig brukervurdering hos LibraryThing.com 3.8 (Syversen, 2011, s. 54) og hos bokelskere.no er gjennomsnittlig terningkast over 5 (Nesse, 2011). Forklaringene er tilnærmet like. Spalding mener at lesere velger bøker de tror de vil like mens Nesse sier at det er bøkene lesere liker som blir lagt inn. Mulig er det samme tilfellet for brukervurderingene fra Amazon.com, hvor det kan tenkes at brukere kjøper og anmelder bøker de tror de vil like. Det er ikke hentet ut gjennomsnittsvurdering fra hvert dokument, men tallene fra hele datasettet etter indeksering er likevel avslørende. Vurderingen er 1 er gitt 308 321 ganger, 2 er gitt 310 302 ganger og 3 er gitt 435 976 ganger. Til sammen er vurderinger mellom 1 og 3 gitt 1 054 599 ganger. Til sammenligning er vurderingen 5 alene brukt 1 095 596 ganger, altså flere ganger enn 1, 2 og 3 samlet. Tallene bekrefter Spaldings bedømmelse av brukervurderinger som lite verdifulle.

Tabell 7: Brukervurderinger

1,095,596	5
718,178	4
435,976	3
310,302	1
308,321	2

Lenke til server: http://bibin.hioa.no:8090/solr-s162452/#/nlpCat2/schema-browser?field=user_rating

Hvilke aspekter og kriterier den enkelte bruker legger til grunn for gitt brukervurdering er ikke mulig å avdekke uten en nærmere forklaring om hvorfor og brukervurderingene blir derfor vanskelig å bruke alene. For å skjønne hvilke kriterier brukeren legger vekt på er det

nødvendig å se brukervurderingen i sammenheng med en brukeranmeldelsen. I eksempelet over, hvor det har blitt gitt en høy karakter til tross for at boken ikke kan anbefales, er dette tydelig. Bakgrunnen for denne oppgaven er å forsøke å finne de underliggende subtile elementene som utgjør en bok og å benytte brukervurderinger er lite formålstjenelig.

4.1.3 Brukeranmeldelser

Til tross for at enkelte ser på brukeranmeldelser som en mindre god kilde å finne de underliggende egenskapene ved en bok er det liten tvil om at brukeranmeldelser kan være et supplement til profesjonelle bokanmeldelser og samtidig ivareta lesernes personlige preferanser og smak. At bokbransjen og forfattere har begynt å ta brukeranmeldelser på alvor forsterker inntrykket (Springen, 2012). Goodreads tall kan også bekrefte populariteten hos brukerne med over 20 millioner registrerte brukere i juli 2013 (Chandler, 2013). En ytterligere bekreftelse på populariteten finnes i datasettet fra INEX. Brukeranmeldelser finnes i nær halvparten av dokumentene, i 1.3 millioner av totalt 2.8 millioner. Det vil på en annen side si at over halvparten ikke har anmeldelse og dermed har liten mulighet for å bli valgt slik at Stanton fra BookLamp.org muligens har et poeng i at det finnes bedre kilder (Herther, 2012, s. 46). Stantons metode, hvor bokens elementer trekkes ut av teksten fremfor fra brukeranmeldelser, er likevel ikke egnet for denne oppgaven fordi den ikke kan fortelle noe om boken som gjenstand for leseren. Tallene fra Goodreads kan også besvare bekymringene vedrørende manglende serendipitet; i januar 2012 representerte de 40 mest leste titlene kun 5 prosent og topp 5000 representerte 56 prosent av bøkene lagt til samme måned (Herther, 2012, s. 18). Det er ikke mulig å finne et eksakte tall vedrørende fordelingen av anmeldelser. Heller ikke hvor mange unike titler det finnes totalt i samlingen er mulig å oppdrive, men at det i august 2012 var katalogisert mer en 360 millioner bøker av brukerne taler for at antallet er betydelig (Chandler, 2012). Spørsmålet om serendipitet er i utgangspunktet ikke av betydning for oppgaven, men er tatt med for å vise at det finnes andre tilnærminger og at brukeranmeldelser ikke alltid blir ansett som en foretrukket kilde. Det er noen utfordringer med brukeranmeldelser. Urke fant i sin masteroppgave om blogganmeldelser at både lengde og innhold kan variere i svært stor grad (2012). Det er også tilfelle i brukeranmeldelsene fra Amazon.com. Urkes undersøkelse hadde et annet formål hvor bokomtalen ble grundig analysert og antallet omtaler begrenset seg derfor til 167. I denne oppgaven er det ikke anledning til en lignende gjennomgang av brukeranmeldelsene, men noen observasjoner har blitt gjort. I likhet med omtalene hos Urke er brukeranmeldelsene i datasettet varierende i både lengde og innhold. Svært mange av de observerte brukeranmeldelsene er i realiteten en

oppsummering av handlingen og ikke en beskrivelse av hvordan de oppfattet boken. I tillegg fører mange av brukeranmeldelsene ofte til utledninger eller diskusjoner om noe helt, for eksempel samfunnsforhold eller historie. Brukere vil alltid sammenligne boken som anmeldes med tidligere leseerfaringer og kan for eksempel skrive at det aktuelle verket har et tregt tempo i forhold til andre utgivelser av samme forfatter eller en annen forfatter i samme sjanger til tross for at den kan være rask i forhold til annen litteratur. Saricks rammeverk er ikke ment å være objektive kriterier og i en referansesamtale i et bibliotek er det ikke problematisk at brukeren relaterer til tidligere erfaringer, det kan være en fordel, men ikke nødvendigvis i et system som forsøker å sammenfatte og sammenligne opplevelsene til svært mange.

Brukeranmeldelser åpner opp for en formidling følelser og opplevelse som ikke er mulig med tagger og brukervurderinger. At svært mange bøker i datasettet mangler brukeranmeldelser er et problem og innebærer at kun halvparten av dokumentene kan brukes. utfordringen er imidlertid generell for alle typer brukergenererte data og som det arbeides med innen flere fagfelt, spesielt innenfor anbefalingssystemer. Hvordan mangelen på data, tidligere omtalt som *The data sparsity problem*, vil ikke bli behandlet i denne oppgaven. At lesere ikke har den samme referanserammen og ulike leseerfaringer som reflekteres i brukeranmeldelsene blir i avsnittet over beskrevet som en utfordring. Om det virkelig er et problem avhenger av målet og løsningene i gjennomføringen. Oppgaven har ikke til hensikt å avdekke en individuell lesers erfaringen, det interessante er den generelle opplevelsen til alle brukerne som har lest og anmeldt boken. Brukeranmeldelser har flere utfordringer, men er likevel den typen tilgjengelig brukergenerert data som er best egnet til å fange de subtile egenskapene.

5 Gjennomføring

I gjennomføringen har alle 2 780 299 millioner dokumenter blitt indeksert ved hjelp av Solr. To av dokumentene lot seg ikke enkelt indeksere og det totale antallet indekserte filer er nå 2 780 297. Det blir ikke sett på som viktig at to dokumenter mangler og det ikke blitt gjort noe forsøk på å løse problemet. Det har ikke blitt gjort noen bearbeidelse av dokumentene på forhånd og alle er blitt indeksert slik som de er. For å holde indekseringstid og størrelsen på indekseringsfilene nede har kun dataene som er nødvendig for gjennomføringen av oppgaven blitt indeksert. Alle ekstra felt og innholdsbehandling koster i form av tid og plass. Det må også bemerkes at en endring i Schema medfører at alle dokumenter bør, i noen tilfeller er det

også nødvendig, indekseres på nytt og det har vært ønskelig å unngå unødvendig behandling som kan skape problemer. En full indeksering av alle dokumentene i datasettet slik Solr er satt opp i denne oppgaven kan ta opp mot 20 sammenhengende timer. Schema angir både felt og behandling av innholdet i dokumentene og finnes leses i sin helhet i vedlegg 4.

5.1 Felter

Av ordinære bokdata har isbn, ean, tittel, forfatter, utgave, utgiver, publikasjonsdato, dewey-nummer, antall sider og emne blitt indeksert. Av brukergenererte data har tagger, brukervurderinger og brukeranmeldelser blitt indeksert. At et felt blir indeksert betyr ikke alltid at det har innhold og dokumentet vil bli lagt til i indeksen uavhengig av om feltet er tomt eller har innhold. Det er imidlertid mulig å spesifisere at et eller flere felt krever innhold for at dokumentet skal kunne indekseres. I dette tilfellet er det kun påkrevd med innhold i for isbn fordi feltet også brukes som en unik identifikator for det enkelte dokument. I tillegg til feltene fra dokumentene i datasettet er det lagt inn fem ekstra felter i Schema. To av feltene, resultat og test, er brukt i forbindelse med utvikling og testing og har ingen funksjon i det ferdige produktet. De tre andre tilleggsfeltene, category_pace, category_character og category_tone, brukes for å legge til ekstra felter til dokumentet i forbindelse med klassifikasjonsprosessen og blir beskrevet nærmere senere i kapittelet. Behandlingen av innholdet har som nevnt blitt holdt til et minimum og det er kun enkelte felt som behandles mer enn hva som kreves av Solr. Isbn, ean, utgave, dewey-nummer, antall sider og tagger behandles som tekststrenger og blir holdt sammen som en term. For eksempel vil en utgiver med navnet *New York Times publishing house* holdes sammen slik at det kan brukes i søk med fraser og kan brukes til organisering i kategorier, såkalt fasetering. Dette kan også oppnås med andre metoder, men det er ikke nødvendig for denne oppgaven. Feltene publikasjonsdato og brukervurderinger behandles som tall, henholdsvis som dato og heltall og kan dermed sorteres. Forfatter og emne blir behandlet som nøkkelord for å kunne brukes spesielt i oppsett av eventuelle automatiske søk, men har ingen funksjon slik de står nå utover å være indeksert som et tekst. Resterende felter er behandlet som tekst med noe prosessering hvilket betyr at alle html-karakterer blir strippet før ordene tokeniseringsprosessen. Etter tokenisering blir token-strømmen strippet for alle spesialtegn og alle store bokstaver blir transformert til små slik at ordene er lett søkbare uten at brukere trenger å tenke på verken spesialtegn eller store og små bokstaver. Alle feltene er kopieres inn i spesialfeltet text slik at eventuelle søk mot feltet text i praksis er et søk mot indeksert innhold i alle feltene. Innholdet i alle feltene, utenom tagger og spesialfeltet text, lagres i indeksfilene og kan fremvises ved

søk. Tagger har blitt utelatt fordi noen dokumenter inneholder så mange tagger at resultatlistene blir uoversiktlig. Taggene er likevel indeksert og er søkbare.

5.2 Konfigurasjon

Som nevnt i oversikten foregår i metodekapittelet kommer Solr med et standardoppsett og det er kun nødvendig å konfigurere det som skal avvike fra utgangspunktet. Konfigureringer foregår i dokumentet Solrconfig (Vedlegg 5) og det er også her Solr blir informert om hvilke eksterne java-klasser eller såkalte krukker(jars) som skal importeres ved oppstart. I motsetning til Schema medfører ikke endringer i Solrconfig at datasettet må indekseres på nytt, kun omstartes, og det er ikke kritisk om ikke alt går som planlagt. Enkelte av elementene i konfigurasjonen slik det står nå er strengt tatt ikke nødvendig, men det har blitt besluttet å ikke gjøre flere endringer nå som Solr er lagt ut på skolens server og fungerer etter hensikten. Elementene i konfigurasjonsfilen som ikke er nødvendig eller henger igjen fra utvikling og testing vil ikke bli omtalt i gjennomgangen, kun de elementene som er interessante og som faktisk har en funksjon. I Solrconfig er det kun lagt til noen få endringer. Hvilke ekstra java-klasser som skal lastes ved oppstart er angitt med henvisninger til plasseringen og er alle relatert til enten dokumentimport eller klassifisering i forbindelse med dokumentimport. Den henvisningen som utmerker seg spesielt er den siste henvisningen til plasseringen "lib" og er mappen med java-klassene som er laget i forbindelse med gjennomføring av oppgaven. Neste interessante konfigurasjon finnes i det som kalles en *requesthandler* med navnet */select* og er en standardpakke for søk. Det er utenfor oppgaven å forklare i detalj hvordan et søk blir behandlet men kortfattet består en søkeprosess i Solr av flere komponenter med funksjoner som kan inkluderes, ekskludere eller omrokkeres (Grainger & Potter, 2014, s. 94). Det er ikke nødvendig for denne oppgaven og standardoppsettet benyttes med kun noen mindre endringer i forhold til noen forhåndsvalg for søket resultatvisningen og innebærer at alle søk, med mindre noe annet eksplisitt spesifiseres vil bli utført mot feltet spesialfeltet text og alle resultat blir vist i formatet xml med innrykk.

I neste felt av interesse, elementet *UpdateRequestProcessor* med navnet *MaxEntCat*, foregår konfigureringen av den eksterne koden som utfører klassifiseringen. Å la den eksterne koden bli konfigurert fra Solr fremfor å hardkode verdiene inn i java-klassene er fordelaktiv fordi eventuelle endringer i plassering av modeller og feltdefinisjoner kan gjøres direkte uten omprogrammering og alle de ulempene det medfører i form av tid og kognitive ressurser. Først er det noe generell informasjon om at prosessen skal benytte en bestemt java-klasse.

Selve konfigurasjonen består av et *inputField* som forteller den eksterne koden hvilket felt i Schema informasjonen skal hentes og flere *outputField* som angir hvilke felt som skal benyttes for resultatet av klassifiseringen. Deretter følger henvisninger til hvor modellene som benyttes i klassifikasjonsprosessen.

Det siste feltet av interesse er en *requesthandler* med navnet */dataimport* som forteller Solr hvor konfigurasjonen for dataimport (Vedlegg 6) finnes og at det i importprosessen skal benyttes en *UpdateRequestProcessor* med navn *MaxEntCat*. Det siste er viktig å legge merke til. At klassifikasjonen blir gjort i forbindelse med dataimport medfører at alle dokumentene blir klassifisert på indekseringstidspunktet og når indekseringen er fullført er alle dokumentene klassifisert og kan brukes umiddelbart.

5.3 Utvidelse av Solr

I metodekapittelet ble det skrevet at Solr kan utvides. Ekstra funksjonalitet kan legges til Solr gjennom et såkalt *plugin* og betyr i praksis at det kan skrives kode for en prosess det er ønskelig at Solr skal utføre. Koden må skrives i java og består av minimum to java-klasser, en klasse (factory class) som initierer koden og en klasse med det som ønskes utført. Klassen som initierer koden har også ansvaret for formidlingen mellom den eksterne koden og Solr (Grainger & Potter, 2014, s. 190 - 195).

5.3.1 Integrering av OpenNlp i Solr

For at Solr skal kunne utnytte mulighetene som finnes i OpenNLP må det først skrives en såkalt *factory class* som beskrevet over initierer og formidler den eksterne koden fra OpenNLP til Solr og eventuelt tilbake om det er nødvendig (Vedlegg 10). Deretter må koden som utfører klassifiseringen skrives (Vedlegg 10). I tillegg er det skrevet en ekstra metode som ligger en egen klasse navngitt *FinnFlest* (Vedlegg 11) Alle detaljer ved programmeringen vil ikke bli beskrevet, men alle steg i klassifikasjonsprosessen vil bli gjennomgått.

Factory class har få oppgaver og utfører initiering og formidlingen som forklart i foregående avsnitt samt sørger for at indeksfilene blir riktig oppdatert. Klassen utvider en av Solrs egne java-klasser, i dette tilfellet klassen *UpdateRequestProcessorFactory* og overstyrer metoden *getInstance*. I java-klassen, navngitt *CategoriseDokument*, ligger koden som utfører klassifiseringen. Klassen utvider Solrs *UpdateRequestProcessor* og overstyrer metoden *processAdd*. Det første steget i klassifikasjonsprosessen henter, fra dokumentet som skal

klassifiseres, alle brukeranmeldelser og slår de sammen til en stor tekst. OpenNLP identifiserer deretter alle setningene og sender en og en setning videre til MaxEntropy algoritmen som klassifiserer hver setning basert på informasjonen i den tidligere produserte modellen. Klassifiseringen foregår i tre omganger, en gang for tempo, en gang for karakterisering og engang for ramme. Alle setninger klassifiseres altså tre ganger. For eksempel vil alle setninger som klassifiseres i henhold til modellen for tempo bli kategorisert til å være enten Fast-paced eller Slow-Paced. Alle setningene hentet fra brukeranmeldelsene i et dokument har etter klassifiseringen fått en kategori. Forholdet mellom antallet setninger i hver kategori avgjør klassifiseringen for hele dokumentet og blir utført av java-klassen *FinnFlest*. Er det flest setninger med kategorien *fast-paced* blir hele dokumentet kategorisert til *fast-paced*, og omvendt. Etter endt klassifisering er dokumentet merket med en kategori for tempo, en kategori for karakterisering og en for ramme. Det forutsetter at dokumentet som klassifiseres inneholder brukeranmeldelser. Mangler dokumentet brukeranmeldelser gjennomføres ikke klassifikasjonen. Resultatet av klassifikasjonen vil føre til følgende ekstra elementer i Solr-dokumentet:

```
<str name="category_pace">Slow-paced</str>
<str name="category_character">Complex_characters</str>
<str name="category_tone">Dark_Tone</str>
```

En full treffliste fines også på serveren og kan finnes ved å benytte følgende adresse: http://bibin.hioa.no:8090/solr-s162452/nlpCat2/select?q=%3A*. Det er også vedlagt oppgaven et forkortet Solr-dokument (Vedlegg 12).

6 Oppsummering

Oppgaven startet med å spørre: Kan brukergenerert data brukes til å automatisk avdekke underliggende elementer ved en bok og brukes i søk etter litteratur? Den overordnede problemstillingen ledet igjen til to underspørsmål som skulle besvares:

- Hvilke typer brukergenererte data er hensiktsmessig å benytte for å avdekke underliggende egenskaper i bøker?
- Hvordan kan automatisk klassifisering av underliggende egenskaper knyttet til underliggende egenskaper i bøker realiseres?

Underspørsmål 1 er besvart ved å undersøke tilgjengelige brukergenererte data. Tre typer finnes i et tilstrekkelig antall og brukeranmeldelser ble etter en analyse det foretrukne valget. Underspørsmål 2 er i utgangspunktet ikke et spørsmål som kan besvares, men et konsept har blitt laget og prosessen har blitt gjort rede for både forhold til teori og de valgene som har blitt fattet.

6.1 Resultater og diskusjon

6.1.1 Hvilke typer brukergenererte data er hensiktsmessig å benytte for å avdekke underliggende egenskaper i bøker?

Etter analysen i forbindelse med underspørsmål 1 i kapittel fire var det klart brukeranmeldelser er den typen som best formidler leseres følelser og reaksjoner til litteratur. Det må nevnes at resultatet var forventet og at analysen kan være farget av det forventede resultatet. Det ble også oppdaget at både brukervurderinger og tagger spesielt er avhengig av konteksten til brukeren. Flere tidligere undersøkelser har konkludert med at tagger brukes til selvorganisering og formidling. I datasettet brukt i oppgaven er det få tagger som har til hensikt å formidle, så godt som alle brukes til selvorganisering og leder til en konklusjon om at formålet med nettsamfunnet påvirker hvordan brukerne tagger. Det har heller ikke blitt funnet mange tagger med feilstavinger, akronymer eller slang og motstrider funnet til Vuorikari, et al (Phuong, 2011, s. 21). Det ble i analysendelen lansert to mulig årsaksforklaringer. I analysen er det kun sett på de 500 mest brukte taggene antakelsen er at feilstavinger og slang vil gjøre gjeldene lengre ned på listen over mest brukte tagger. Den andre årsaksforklaringen relaterer igjen til nettsamfunnet og brukerens kontekst. Antakelsen er at det er forskjell på nettsamfunn hvor formålet er å katalogisere bøker for egen samling og for eksempelvis et nettsamfunn for bilder hvor formålet formidling er en viktig element. Ved organisering av egen boksamling er det først og fremst å huske som er viktig og bruk av slang og feilstavinger tjener ikke formålet. Samtidig viser tallene fra analysen at en stor andel av taggene kun brukes svært gå ganger og tyder på en viss spredning av termene. Hvordan resultatet skal tolkes er usikkert. Muligens er metoden med å kun undersøke de 500 mest brukte termene ikke formålstjenelig. Et alternativ kunne vært å se på sekvenser av tagger med jevne mellomrom, men med godt over åtte hundre tusen tagger er oppgaven nærmest endeløs.

Brukervurderinger ble etter den teoretiske forankringen antatt å være utilstrekkelig til oppgaven med å formidle følelser. Det er vanskelig å si noe sikkert om motivene for karakteren som blir gitt av brukere. Å overføre komplekse meninger til en enkel tallverdi er

ikke enkelt og spesielt om en skal forsøke å finne motivasjonen for en gitt vurdering. Eksempelen i analysen illustrerer hvorfor. Brukernevurderingen var ujevnt fordelt på skalaen 1-5 slik det også var forventet etter den teoretiske gjennomgangen med et overveiende flertall av karakteren 5. Igjen er antakelsen av konteksten spiller en rolle, i tillegg til at lesere velger bøker de tror de vil like slik som det ble pekt på i analysen, er vurderingene hentet fra Amazon.com og det er rimelig å formode at lesere som kjøper en bok på nettet har gjort undersøkelser i forkant. En brukerundersøkelse hvor flere ulike nettsamfunn kan avklare til hvilken grad kontekst spiller en rolle. Underspørsmålet er likevel besvart og brukeranmeldelser er det alternativet som best kan fange de subtile egenskapene ved en bok.

6.1.2 Hvordan kan automatisk klassifisering av underliggende egenskaper knyttet til underliggende egenskaper i bøker realiseres?

Det har i forbindelse med underspørsmål to blitt valgt et rammeverk og det kan rettes et kritisk søkelys ved både valg av rammeverk og ved tilpassningen. Oppgaven er ikke ment å være en litteraturanalyse og rammeverket, det viktigste har vært å ha et rammeverk for utviklingen av et konseptuelt system Underspørsmål to har blitt besvart og det har blitt utviklet et konsept som viser hvordan subtile egenskaper kan brukes. Det er likevel mange problemer knyttet til både valg og gjennomføring. Det lyktes ikke å lage gode nok modeller fordi modellbyggingen etter ulike forsøk ble en manuell operasjon. Det er også problemer i forhold til gjennomføringen og det er ikke tatt hensyn til gradering innenfor kategoriene. Alle dokumenter er alltid kun en av to resultat men valget ble gjort for å forenkle utviklingsprosessen. Det har heller ikke blitt gjort en evaluering av søkeresultatene med og uten bruk av kategoriene produsert.

7 Forslag til videre undersøkelse

I forbindelse med tidligere diskusjon av underspørsmål en har det blitt nevnt at en brukerundersøkelse muligens kan avdekke om konteksten har like stor innvirkning som det antas i denne oppgaven. Det kan også være hensiktsmessig å ikke avskrive tagger og brukervurderinger men undersøke hvordan disse kan brukes sammen med andre brukergenererte data som brukeranmeldelser. Underspørsmål to består av mangfoldige løse tråder som kan nøstet opp i fremtidige undersøkelser. Det mest presserende er å finne bedre metoder for å lage modeller for klassifikasjonsprosessen. Det kan også bli gjennomført en ny vurdering av rammeverket som er benyttet i oppgaven for en bedre tilpassning, eventuelt utforske muligheten for å benytte andre rammeverk som er bedre egnet.

8 Litteraturliste

Andrejevic, M. (2013). *Infoglut: how too much information is changing the way we think and know*. New York: Routledge.

Apache OpenNLP Developer Documentation. (n.d.). *Sentence Detection*. Retrieved from <https://opennlp.apache.org/documentation/1.5.3/manual/opennlp.html#tools.sentdetect>

Baez, M., Birukou, A., Casati, F., & Marchese, M. (2010). Addressing Information Overload in the Scientific Community. *IEEE Internet Computing*, 14(6), 31–38.

doi:<http://doi.ieeecomputersociety.org/10.1109/MIC.2010.107>

Baeza-Yates, R. (2011). *Modern information retrieval: the concepts and technology behind search* (Second edition.). New York: Addison Wesley.

Book Genome. (n.d.). Retrieved March 10, 2014, from <http://bookgenome.com/>

BookLamp's FAQ. (n.d.). *StoryTime by BookLamp*. Retrieved March 10, 2014, from <http://storytime.booklamp.org/booklamps-faq/>

Chandler, O. (2012, August 13). Ten Million Thank Yous! *Goodreads Blog*. Retrieved from <http://www.goodreads.com/blog/show/380-ten-million-thank-yous>

Chandler, O. (2013, July 23). Goodreads Grows to 20 Million Readers. *Goodreads Blog*. Retrieved from <http://www.goodreads.com/blog/show/425-goodreads-grows-to-20-million-readers>

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). Is seeing believing?: how recommender system interfaces affect users' opinions (p. 585). ACM Press. doi:10.1145/642611.642713

Fuhr, N., Gövert, N., Kazai, G., & Lalmas, M. (Eds.). (2002). Proceedings of the First Workshop of the INitiative for the Evaluation of XML Retrieval (INEX). In *Proceedings of the First Workshop of the INitiative for the Evaluation of XML Retrieval (INEX)*. International Conference and Research Center for Computer

- Science: INEX. Retrieved from <https://inex.mmci.uni-saarland.de/static/proceedings/INEX2002-proceedings.pdf>
- Glick, Y. (2014). The Entertainment Genome. *jinni.com*. Retrieved from <http://www.jinni.com/info/entertainment-genome.html>
- Grainger, T., & Potter, T. (2014). *Solr in Action*. Greenwich, CT, USA: Manning Publications.
- Handbook of natural language processing*. (2010). Boca Raton, FL: Chapman & Hall/CRC.
- Hanjalic, A., & Li-Qun Xu. (2005). Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1), 143–154. doi:10.1109/TMM.2004.840618
- He, J., & Chu, W. W. (2010). A Social Network-Based Recommender System (SNRS). In M. Nasrulla, J. J. Xu, D. L. Hicks, & H. Chen (Eds.), *Data mining for social network data* (Vol. 12). New York ; London: Springer.
- Herther, N. K. (2012). From Book Reviewing to Crowdsourcing. *Searcher*, 20(5), 16–47.
- Hurdeman, H., Kamps, J., Koolen, M., & van Wees, J. (2012). Using Collaborative Filtering in Social Book Search. Amsterdam: Inex 2012. Retrieved from <https://inex.mmci.uni-saarland.de/static/proceedings/INEX2012-preproceedings.pdf>
- INEX 2014 Social Book Search Track. (2014). *INEX 2014 Social Book Search Track*. Retrieved May 23, 2014, from <https://inex.mmci.uni-saarland.de/tracks/books/>
- Ingersoll, G. S. (2013). *Taming text: how to find, organise, and manipulate it*. Shelter Island, NY : London: Manning ; Pearson Education [distributor].
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2011). *Recommender systems: an introduction*. New York: Cambridge University Press.
- Kindt, T. (2013). *Affective metadata from movies helps predict a viewer's mood for mood-based recommenders* (Master's Thesis for the Media Technology programme). Leiden University, The Netherlands. Retrieved from

<http://mediatechnology.leiden.edu/images/uploads/docs/kindt-affective-meta-data-and-mood-based-recommenders-thesis.pdf>

Kocik, K. (2004). *Question classification using maximum entropy models*.

Koolen, M., Kamps, J., & Kazai, G. (2013). Social Book Search: The Impact of Professional and User-Generated Content on Book Suggestions (pp. 38–39). Presented at the DIR.

Koolen, M., Kazai, G., Preminger, M., & Doucet, A. (n.d.). Overview of INEX 2013 social book search track. Retrieved from <http://www.clef-initiative.eu/documents/71612/08425031-a583-4692-b9eb-dc14f913f682>

Marmanis, H. (2009). *Algorithms of the intelligent web*. Greenwich, CT: Manning.

McCandless, M. (2010). *Lucene in action* (2nd ed.). Greenwich: Manning.

Moyer, J. E. (2005). Adult Fiction Reading. *Reference & User Services Quarterly*, 44(3), 220–231.

Nesse, A. (2011, May). Svar på Ankis melding datert 29. mai 2011. *Nye bokanbefalinger*.

News. (2014). *Apache OpenNlp*. Retrieved from <https://opennlp.apache.org/news.html>

Nigam, K., Lafferty, J., & McCallum, A. (1999). Using maximum entropy for text classification. In *IJCAI-99 Workshop on Machine Learning for Information Filtering*.

O’Neil, C. (2012). Columbia Data Science course, week 12: Predictive modeling, data leakage, model evaluation. *mathbabe*. Retrieved from <http://mathbabe.org/2012/11/20/columbia-data-science-course-week-12-predictive-modeling-data-leakage-model-evaluation/>

Owen, S. (Ed.). (2012). *Mahout in action*. Shelter Island, New York: Manning.

Peters, I. (2009). *Folksonomies: indexing and retrieval in Web 2.0*. Berlin: De Gruyter/Saur.

Phuong, L. T. T. (2011). *Processes of tagging behaviors : case studies of CiteULike, LibraryThing, YouTube and Flickr*. Høgskolen i Oslo. Avdeling for journalistikk,

- bibliotek- og informasjonsvitenskap, Oslo. Retrieved from
<http://hdl.handle.net/10642/994>
- Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., ... Strohecker, C. (2004). Affective Learning — A Manifesto. *BT Technology Journal*, 22(4), 253–269. doi:10.1023/B:BTTJ.0000047603.37042.33
- Rafalovitch, A. (2013). *Instant Apache Solr for Indexing Data How-to*. Packt Publishing.
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender systems handbook*. New York: Springer.
- Ross, C. S. (2000). Making Choices. *The Acquisitions Librarian*, 13(25), 5–21.
doi:10.1300/J101v13n25_02
- Saricks, J. G. (2005). *Readers' advisory service in the public library* (3rd ed.). Chicago: American Library Association.
- Smiley, D., Pugh, E., Brady, James, & Eteve, J. (2009). *Solr 1.4 Enterprise Search Server enhance your search with faceted navigation, result highlighting, fuzzy queries, ranked scoring, and more*. Birmingham, UK: Packt Pub. Retrieved from
<http://www.books24x7.com/marc.asp?bookid=32515>
- Springen, K. (2012, January 19). Should Authors and Agents Weigh In on Citizen Reviews? *PublishersWeekly.com*. Retrieved May 31, 2014, from
<http://www.publishersweekly.com/pw/by-topic/childrens/childrens-industry-news/article/50268-should-authors-and-agents-weigh-in-on-citizen-reviews.html>
- Stvilia, B., Jørgensen, C., & Wu, S. (2012). Establishing the value of socially-created metadata to image indexing. *Library & Information Science Research*, 34(2), 99–109.
doi:10.1016/j.lisr.2011.07.011
- Syversen, P. C. B. (2011). *Anbefalingssystemer for litteratur i den digitale hverdagen*. Master thesis. Retrieved October 17, 2013, from

https://oda.hio.no/jspui/handle/10642/941?mode=full&submit_simple=Show+full+item+record

thetextbookguru. (n.d.). The Textbook Guru | Book Genome Project. Retrieved from

<http://www.thetextbookguru.com/tag/book-genome-project/>

Tkalčič, M., Burnik, U., & Košir, A. (2010). Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction*, 20(4), 279–311. doi:10.1007/s11257-010-9079-z

Trieu, R. (2011, September 23). Social Cataloging May Be Trapping Users In A Bubble |

Neon Tommy. *Neon Tommy Annenberg Digital News*. Retrieved May 31, 2014, from

<http://www.neontommy.com/news/2011/09/social-cataloging-may-be-trapping-users-bubble>

Understanding the Book Genome Project. (2012, January 24). *StoryTime by BookLamp*.

Retrieved from <http://storytime.booklamp.org/2012/01/24/understanding-the-book-genome-project/>

Urke, K. J. (2012). «*Dersom du liker spenning ispedd historie, mystikk og kjærlighet, er dette boken for deg.*»: norske folkebiblioteks bokomtaler på nettet (Master thesis).

Retrieved from

https://oda.hio.no/jspui/handle/10642/1248?mode=full&submit_simple=Show+full+item+record

Weaver, M. (2007). Contextual metadata: faceted schemas in virtual library communities.

Library Hi Tech, 25(4), 579–594. doi:10.1108/07378830710840527

Welcome to Apache OpenNLP. (2014). *Apache OpenNLP*. Retrieved from

<https://opennlp.apache.org/index.html>

9 Vedlegg

9.1 Vedlegg 1

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!-- version 1.0 / 2009-11-05T01:55:19+01:00 -->
<!DOCTYPE book SYSTEM "books.dtd">
<book>
  <isbn>0373112580</isbn>
  <title>Elusive As The Unicorn (Harlequin Presents, 1258)</title>
  <ean>9780373112586</ean>
  <binding>Paperback</binding>
  <label>Harlequin</label>
  <listprice>$2.50</listprice>
  <manufacturer>Harlequin</manufacturer>
  <publisher>Harlequin</publisher>
  <readinglevel/>
  <releasedate/>
  <publicationdate>1990-03-01</publicationdate>
  <studio>Harlequin</studio>
  <edition/>
  <dewey>813</dewey>
  <numberofpages>187</numberofpages>
  <dimensions>
    <height/>
    <width/>
    <length/>
    <weight/>
  </dimensions>
  <reviews>
    <review>
      <authorid>AIBDTXX751CCP</authorid>
      <date>2007-02-20</date>
      <summary>Elusive as the Unicorn</summary>
      <content>Back Cover:
        &lt;br /&gt;When Eve Eden discovered that Adam Gardener, successful art
entrepreneur, wa searching fo
        the legendary English artist, The Unicorn, she Nervously shied away. The Unicorn's
true identity hit too
        close to home...
        &lt;br /&gt;Besides, Eve was rattled by Adam's mesmerizing prescence, especially
in the light of the
        ridiculous coincidence of their names-and his determination to take advantage of it!
But Eve was already
        engaged to marry her longtime friend, Paul.
        &lt;br /&gt;Yet Eve found herself troubled by the different choices Adam and Paul
presented. If only the
        answer to her dilemma didn't keep eluding her...
      </content>
      <rating>3</rating>
      <totalvotes>0</totalvotes>
```



```
<helpfulvotes>0</helpfulvotes>
</review>
</reviews>
<editorialreviews/>
<images/>
<creators>
  <creator>
    <name>Carole Mortimer</name>
    <role>Author</role>
  </creator>
</creators>
<blurbers/>
<dedications/>
<epigraphs/>
<firstwords/>
<lastwords/>
<quotations/>
<series/>
<awards/>
<characters/>
<places/>
<subjects>
  <subject>1958-1969</subject>
</subjects>
<tags>
  <tag count="1">unread</tag>
  <tag count="1">owned</tag>
  <tag count="1">tbr</tag>
  <tag count="2">Romance</tag>
  <tag count="1">hp</tag>
  <tag count="2">Harlequin Presents</tag>
  <tag count="1">box-w076</tag>
  <tag count="1">1258</tag>
</tags>
<similarproducts>
  <similarproduct>0373107803</similarproduct>
  <similarproduct>0373108524</similarproduct>
</similarproducts>
<browseNodes>
  <browseNode id="17">Literature & Fiction</browseNode>
  <browseNode id="23">Romance</browseNode>
  <browseNode id="1000">Subjects</browseNode>
  <browseNode id="10125">General</browseNode>
  <browseNode id="10129">Contemporary</browseNode>
  <browseNode id="12831">Authors, A-Z</browseNode>
  <browseNode id="13153">( M )</browseNode>
  <browseNode id="13361">General</browseNode>
  <browseNode id="13391">Series</browseNode>
  <browseNode id="13412">Harlequin Presents</browseNode>
  <browseNode id="283155">Books</browseNode>
```

```
<browseNode id="466458">Mortimer, Carole</browseNode>
<browseNode id="746228">Regency</browseNode>
<browseNode id="388186011">Refinements</browseNode>
<browseNode id="394174011">Binding (binding)</browseNode>
<browseNode id="400272011">Paperback</browseNode>
<browseNode id="618072011">Format (feature_browse-bin)</browseNode>
<browseNode id="618083011">Printed Books</browseNode>
<browseNode id="713365011">General AAS</browseNode>
<browseNode id="713392011">General AAS</browseNode>
<browseNode id="713660011">General AAS</browseNode>
</browseNodes>
</book>
```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>

<!-- version 1.0 / 2009-11-05T01:55:19+01:00 -->

<!DOCTYPE book SYSTEM "books.dtd">

<book>

<isbn>0373112580</isbn>

<title>Elusive As The Unicorn (Harlequin Presents, 1258)</title>

<ean>9780373112586</ean>

<binding>Paperback</binding>

<label>Harlequin</label>

<listprice>\$2.50</listprice>

<manufacturer>Harlequin</manufacturer>

<publisher>Harlequin</publisher>

<readinglevel/>

<releasedate/>

<publicationdate>1990-03-01</publicationdate>

<studio>Harlequin</studio>

<edition/>

<dewey>813</dewey>

<numberofpages>187</numberofpages>

<dimensions>

<height/>

<width/>

<length/>

<weight/>

</dimensions>

<reviews>

<review>

<authorid>AIBDTXX751CCP</authorid>

<date>2007-02-20</date>

<summary>Elusive as the Unicorn</summary>

<content>Back Cover:

When Eve Eden discovered that Adam Gardener, successful art entrepreneur, wa searching fo

the legendary English artist, The Unicorn, she Nervously shied away. The Unicorn's true identity hit too

close to home...

Besides, Eve was rattled by Adam's mesmerizing prescence, especially in the light of the

ridiculous coincidence of their names-and his determination to take advantage of it! But Eve was already

engaged to marry her longtime friend, Paul.

Yet Eve found herself troubled by the different choices Adam and Paul presented. If only the

answer to her dilemma didn't keep eluding her...

</content>

<rating>3</rating>

<totalvotes>0</totalvotes>

<helpfulvotes>0</helpfulvotes>

</review>

```
</reviews>
<editorialreviews/>
<images/>
<creators>
  <creator>
    <name>Carole Mortimer</name>
    <role>Author</role>
  </creator>
</creators>
<blurbers/>
<dedications/>
<epigraphs/>
<firstwords/>
<lastwords/>
<quotations/>
<series/>
<awards/>
<characters/>
<places/>
<subjects>
  <subject>1958-1969</subject>
</subjects>
<tags>
  <tag count="1">unread</tag>
  <tag count="1">owned</tag>
  <tag count="1">tbr</tag>
  <tag count="2">Romance</tag>
  <tag count="1">hp</tag>
  <tag count="2">Harlequin Presents</tag>
  <tag count="1">box-w076</tag>
  <tag count="1">1258</tag>
</tags>
<similarproducts>
  <similarproduct>0373107803</similarproduct>
  <similarproduct>0373108524</similarproduct>
</similarproducts>
<browseNodes>
  <browseNode id="17">Literature & Fiction</browseNode>
  <browseNode id="23">Romance</browseNode>
  <browseNode id="1000">Subjects</browseNode>
  <browseNode id="10125">General</browseNode>
  <browseNode id="10129">Contemporary</browseNode>
  <browseNode id="12831">Authors, A-Z</browseNode>
  <browseNode id="13153">( M )</browseNode>
  <browseNode id="13361">General</browseNode>
  <browseNode id="13391">Series</browseNode>
  <browseNode id="13412">Harlequin Presents</browseNode>
  <browseNode id="283155">Books</browseNode>
  <browseNode id="466458">Mortimer, Carole</browseNode>
  <browseNode id="746228">Regency</browseNode>
```

```
<browseNode id="388186011">Refinements</browseNode>
<browseNode id="394174011">Binding (binding)</browseNode>
<browseNode id="400272011">Paperback</browseNode>
<browseNode id="618072011">Format (feature_browse-bin)</browseNode>
<browseNode id="618083011">Printed Books</browseNode>
<browseNode id="713365011">General AAS</browseNode>
<browseNode id="713392011">General AAS</browseNode>
<browseNode id="713660011">General AAS</browseNode>
</browseNodes>
</book>
```

9.2 Vedlegg 2

```
<?php
try {
    $tilkobling = new PDO('mysql:host=ark1.hioa.no; dbname=inex', 'bruker', 'passord');
} catch (PDOException $sunntak){
    echo "Tilkoblingsfeil: " . $sunntak->getMessage();
}
$status =
"Server: " . ($tilkobling->getAttribute(PDO::ATTR_CONNECTION_STATUS)). "<br>".
"Version: " . ($tilkobling->getAttribute(PDO::ATTR_SERVER_VERSION)). "<br>".
"Client: " . ($tilkobling->getAttribute(PDO::ATTR_CLIENT_VERSION));
echo $status;

$tilkobling->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION);

<?php
include 'tilkobling_ark1_pdo.php';
header('Content-type: text/plain');
date_default_timezone_set('Europe/Oslo');
//totalt 2,780,299
echo "Startet". ": ".date('d-m-Y H:i:s'). "\n";
$start = 0;
$rader = 2780299;
$query = "SELECT `XML` FROM `a_it` LIMIT $start, $rader";

foreach ($tilkobling->query($query) AS $row){
    $xml = $row['XML'];
    //printf("%s", $xml);

    $dom = new DOMDocument();
    $dom->loadXML($xml, LIBXML_NOBLANKS);
    /*$root = $dom->documentElement;

    $test = $root->getElementsByTagName("title");
    //print $test->item(0)->nodeValue. "\n";
    $tittel = $test->item(0)->nodeValue;*/

    if ($counter < $start){
        $counter = $start+1;
    }else{

        $counter++;
    }
}
//print $tittel. "\n";
$dom->save("/Volumes/skole/masteroppgave/solr/nlpCat/bookfiles_import/$counter.xml");
```

```
if($counter == $start+$rader/4){
    echo($counter). "---- ".date('d-m-Y H:i:s'). "\n";
}
if ($counter == $start+$rader/2){
    echo($counter). "---- ".date('d-m-Y H:i:s'). "\n";
}

}

echo "Ferdig". ": ".date('d-m-Y H:i:s'). "\n";
echo($start+$rader). "\n";
```

9.3 Vedlegg 3

526,777	Fiction
518,138	non-fiction
354,931	read
349,106	unread
291,153	History
280,810	own
213,657	novel
211,458	tbr
209,792	Paperback
162,231	Reference
161,724	literature
145,400	20th Century
136,732	children's
129,054	Biography
115,724	mystery
112,570	hardcover
111,687	owned
111,221	wishlist
108,240	Children
105,816	fantasy
104,438	series
99,556	religion
94,699	english
89,789	american
89,153	Romance
86,680	British
84,651	Art
84,352	England
83,760	Science
78,790	classic
76,671	philosophy
73,137	historical fiction
72,496	Poetry
71,821	classics
71,647	picture book
71,524	science fiction

9.4

9.5 Vedlegg 4

```
1 <schema name="min_test" version="1.5">
2 <fields>
3 <field name="isbn" type="string" indexed="true" stored="true" required="true"
4 <field name="ean" type="string" indexed="true" stored="true" />
5 <field name="title" type="text_general" indexed="true" stored="true" />
6 <field name="publisher" type="text_general" indexed="true" stored="true" />
7 <field name="publicationdate" type="date" indexed="true" stored="true" />
8 <field name="edition" type="string" indexed="true" stored="true" />
9 <field name="dewey" type="string" indexed="true" stored="true" multiValued="true"
10 <field name="numberofpages" type="string" indexed="true" stored="true" multiValued
11
12 <field name="user_review_content" type="text_general" indexed="true" stored="true"
13 <field name="creator" type="keyword" indexed="true" stored="true" multiValued
14 <field name="subject" type="keyword" indexed="true" stored="true" multiValued
15 <field name="user_rating" type="int" indexed="true" stored="true" multiValued
16 <field name="tag" type="string" indexed="true" stored="false" multiValued="true"
17
18 <field name="text" type="text_general" indexed="true" stored="false" multiValued
19 <copyField source="*" dest="text"/>
20
21 <field name="category_pace" type="string" indexed="true" stored="true"/>
22 <field name="category_character" type="string" indexed="true" stored="true" />
23 <field name="category_tone" type="string" indexed="true" stored="true" />
24 <field name="resultat" type="string" indexed="true" stored="true" />
25 <field name="test" type="string" indexed="true" stored="true" />
26 </fields>
27
28
29 <uniqueKey>isbn</uniqueKey>
30
31
32 <types>
33 <fieldType name="int" class="solr.TrieIntField" precisionStep="0"
positionIncrementGap
34 <fieldType name="date" class="solr.TrieDateField" precisionStep="6"
positionIncrementGap
35 <fieldType name="string" class="solr.StrField" />
36
37 <fieldType name="text_general" class="solr.TextField" positionIncrementGap="100"
38 <analyzer>
39 <charFilter class="solr.HTMLStripCharFilterFactory"/>
40 <tokenizer class="solr.StandardTokenizerFactory" />
41 <filter class="solr.ICUFoldingFilterFactory"/>
42 <filter class="solr.LowerCaseFilterFactory"/>
43 </analyzer>
44 </fieldType>
45
46 <fieldType name="keyword" class="solr.TextField" positionIncrementGap="100">
47 <analyzer>
48 <tokenizer class="solr.KeywordTokenizerFactory"/>
49 </analyzer>
50 </fieldType>
51 </types>
52 </schema>
```

9.6 Vedlegg 5

```
1 <?xml version="1.0" encoding="UTF-8" ?>
2 <config>
3
4 <!--<abortOnConfigurationError>${solr.abortOnConfigurationError:true}
</abortOnConfigurationError-->
5
6
7 <lib dir="/usr/share/solr/s162452/solr-4.6.0/contrib/analysis-extras/lucene-libs/" />
8 <lib dir="/usr/share/solr/s162452/solr-4.6.0/contrib/analysis-extras/lib/" />
9 <lib dir="/usr/share/solr/s162452/solr-4.6.0/dist/" regex="solr-cell-\.d.*\.jar" />
10 <lib dir="/usr/share/solr/s162452/solr-4.6.0/contrib/extraction/lib" regex=".*\.jar" />
11 <lib dir="/usr/share/solr/s162452/solr-4.6.0/dist" regex="solr-dataimporthandler-.*\.jar" />
12 <lib dir="/usr/share/solr/s162452/solr-4.6.0/dist" regex="solr-solrj-.*\.jar" />
13
14 <lib dir="/usr/share/solr/s162452/nlpCat2/lib" />
15
16
17
18
19 <.luceneMatchVersion>LUCENE_46</luceneMatchVersion>
20 <requestDispatcher handleSelect="false">
21 <requestParsers enableRemoteStreaming="true" multipartUploadLimitInKB="204800" />
22 <httpCaching never304="true" />
23 </requestDispatcher>
24
25 <requestHandler name="/admin/luke" class="org.apache.solr.handler.admin.LukeRequestHandler" />
26 <requestHandler name="/select" class="solr.SearchHandler">
27 <lst name="defaults">
28 <str name="wt">xml</str>
29 <str name="indent">>true</str>
30 <str name="df">text</str>
31 <!--<str name="defType">edismax</str>
32 <str name="qf">title^10 isbn^5 publisher^3 publicationdate numberofpages user_review_summary^5
33 user_review_content^5 editorialreviews_content creator_name^9 series characters places
text</str-->
34 </lst>
35 </requestHandler>
36
37 <requestHandler name="/update" class="solr.UpdateRequestHandler" />
38
39
40 <requestHandler name="/admin" class="solr.admin.AdminHandlers" />
41 <requestHandler name="/analysis
/field" class="solr.FieldAnalysisRequestHandler" startup="lazy" />
42
43 <updateRequestProcessorChain name="MaxEntCat" default="true">
44 <processor class="com.nlpsolr.cat.tre.CategorizeDocumentFactory">
45 <str name="inputField">user_review_content</str>
46 <str name="outputFieldPace">category_pace</str>
47 <str name="outputFieldCharacter">category_character</str>
48 <str name="outputFieldTone">category_tone</str>
49 <str name="resultField">resultat</str>
50 <str name="testField">test</str>
51 <str name="modelPace">/usr/share/solr/s162452/nlpCat2/egne_opennlp-modeller
/model_pace</str>
52 <str name="modelCharacter">/usr/share/solr/s162452/nlpCat2/egne_opennlp-modeller
/model_character</str>
53 <str name="modelTone">/usr/share/solr/s162452/nlpCat2/egne_opennlp-modeller
/model_tone</str>
54 <str name="openNlpSentenceModel">/usr/share/solr/s162452/nlpCat2/opennlp-models
/en-sent.bin</str>
55 </processor>
56 <processor class="solr.RunUpdateProcessorFactory"/>
57 <processor class="solr.LogUpdateProcessorFactory"/>
58 </updateRequestProcessorChain>
59
60 <requestHandler name="/dataimport" class="org.apache.solr.handler.dataimport.DataImportHandler">
61 <lst name="defaults">
62 <str name="config">dih-def_mappe.xml</str>
63 <str name="update.chain">MaxEntCat</str>
64 </lst>
65 </requestHandler>
66
67
68 </config>
69
```

9.7 Vedlegg 6

```
1 <dataConfig>
2 <dataSource type="FileDataSource" encoding="UTF-8"/>
3 <document>
4
5 <!--lager liste over filene i mappen. Ikke nødvendig når
filer 6 <entity name="f" processor="FileListEntityProcessor"
7 fileName="*.xml"
8 recursive="true"
9 rootEntity="false"
10 dataSource="null"
11 baseDir="/Volumes/skole/masteroppgave
/solr/nlpCat/bookfiles_import"
12 >
13
14 <!--Denne konfigurasjonen bruker XPathEntityProcessor
og filepath-->
15 <entity name="bok"
16 processor="XPathEntityProcessor"
17 url="{f.fileAbsolutePath}"
18 forEach="/book"
19 transformer="DateFormatTransformer"
20 >
21
22 <field column="isbn" xpath="/book/isbn"/>
23 <field column="ean" xpath="/book/ean"/>
24 <field column="title" xpath="/book/title"/>
25 <field column="publisher" xpath="/book
/publisher"/>
26 <field column="publicationdate" xpath="/book
/publicationdate" dateTimeFormat="yyyy"/>
27 <field column="edition" xpath="/book/edition"/>
28 <field column="dewey" xpath="/book/dewey" />
29 <field column="numberofpages" xpath="/book
/numberofpages"/>
30
31 <field column="user_review_content" xpath="/book
/reviews/review/content"/>
32 <field column="creator" xpath="/book/creators
/creator/name"/>
33 <field column="subject" xpath="/book/subjects
/subject"/>
1 of 2
34 <field column="user_rating" xpath="/book
/reviews/review/rating"/>
35 <field column="tag" xpath="/book/tags/tag" />
36
37 </entity>
38 </entity>
```

```
39 </document>  
40 </dataConfig>
```

9.8 Vedlegg 7

Fast-paced Easy to read.

Fast-paced This encyclopedia provides simple explanations for technical terms, offers a guide to the cause and treatment of common conditions, and supplies background information allowing the understanding of medical matters.

Fast-paced It left me wanting more.

Fast-paced Even so, this book was an easy page turner.

Fast-paced Lots of fast-paced action keeps the heroine thinking on her feet throughout the book.

Fast-paced Mostly she's trying not to get killed as things spiral out of control.

Fast-paced She's undercover and on the run. It is exciting and fun to read.

Fast-paced The book is full of violent chases.

Fast-paced It is exciting and fun to read.

Fast-paced What follows is a book-long wild ride into the underworld in the company of a young hustler.

Fast-paced It's an exciting, fast read with plenty of descriptions of barrio life, morals and codes, which holds the reader's interest, so much so that I'll be looking out for some more.

Fast-paced A cliff-hanging suspense mystery novel.

Fast-paced She slightly prolongs things too much.

Fast-paced This plot just wasn't chilling, mysterious, or really in any way a page-turner.

Fast-paced This is a "cliffhanger," and I couldn't wait to read "I" and find out what would happen next.

Fast-paced If you want high pace excitement and thrills of high rolling stakes, this is the book for you.

Fast-paced This book is as usual a fast paced thriller.

Fast-paced It makes for light and entertaining reading.

Fast-paced This is a rollicking roller coaster of a rapturous read.

Fast-paced I shall not abandon hope, especially since the book's blend of profound spirituality and nail-biting suspense guarantees.

Fast-paced I started reading the book late last night with only the idea of reading one chapter and ended up reading until the sun started rising outside my window.

Fast-paced I got a little teary eyed especially the treatment of Zoe by her stepmother and also Zoe's father's heartbreak over losing the woman he so truly loved because he was not strong enough to do what his heart was telling him but he loved his daughter Zoe, the product of that great love.

Fast-paced I began reading it at 7pm this evening and did not put it down until I completed it at 12:30am.

Fast-paced I truly enjoyed this book from beginning to end. I could not put it down.

Fast-paced This book is simply scrumptious! Couldn't put it down!

Fast-paced Filled with a lovable damsel in distress, a handsome hero, an evil step-mother and a host of evil doers

Fast-paced I got this book 2 days ago and am already finished with it. Not because it was a short book, but because I couldn't put it down.

Fast-paced It is an epic tale of love, betrayal, and prejudice which will take you on a journey you won't soon forget.

Fast-paced I was immediately engaged in the story of instant and

physical attraction between the main characters.

Fast-paced It is an amazing book that you can't put down. It was a definite page-turner, and I loved it.

Fast-paced Zoé takes the reader beyond color and shares the true nature of humanity and passion in a page-turning tale that will keep you reading.

Fast-paced Rattling paced novel.

Fast-paced An exciting fast-paced thriller that will keep you turning the pages.

Fast-paced It's a quick read.

Fast-paced It was a fast paced read that I couldn't put down.

Fast-paced Time and space collide in this science-fiction thriller, and you won't be able to stop turning the pages.

Fast-paced The story takes place over a short space of time

Fast-paced Author Graham Marks writes an exciting, fast-paced thriller.

Fast-paced It is, however, a rather difficult book to read, especially if you're expecting a plot-driven story, and requires patience.

Fast-paced You'll find a great adventure ahead of you and many additional books to carry you through.

Fast-paced This book is exciting and adventuresome.

Fast-paced It is a good versus evil kind of book.

Fast-paced IMagination of the author is just stunning and the suspense and the thrill of this book is gripping

Fast-paced The book keeps you at the edge.

Fast-paced I started and just could not put it down.

Fast-paced I think it is an excellent book, with a full range of emotions and lots of action.

Fast-paced I needed to know what was happening, and I devoured this book very quickly.

Fast-paced This book is full of suspense and is definitely a page turner.

Fast-paced This read is a good one and you will be hooked.

Fast-paced This action packed thriller that will keep you on the edge of your seat.

Fast-paced If anyone out there loves suspense thriller novels as I do, really needs to read this book.

Fast-paced Lots of fast-paced action keeps the reader throughout the book.

Fast-paced The plot twists kept me reading.

Fast-paced She befriends the woman she is investigating, and of course there are plot twists.

Fast-paced You have to re-read this after you've finished because a final three-word sentence stands the whole plot on its head.

Fast-paced The third star is for the last sentence which was the best plot twist for the storyline.

Fast-paced The two big plot twists can be seen coming a mile away.

Fast-paced This is a cliff-hanging suspense mystery novel.

Fast-paced This book is as usual a fast paced thriller.

Fast-paced Halfway through it picked up and moved very fast to the end, which was exciting, surprising, and had a lot of action.

Fast-paced Lots of page-turning action and Patterson's trademark gruesome, misogynistic murders.

Fast-paced Patterson hits one right out of the park with this book.

1st to Die is an extremely fast-paced, mystery novel full of

wonderful characters and a plot with twists upon twists.

Fast-paced The setting is described magnificently, each scene exploding in the reader's mind like a Hollywood blockbuster.

Fast-paced The plot is amazing, keeping you glued to the book until the very last page.

Fast-paced There are several unexpected turns and twists to this plot and just when you think you've rounded the last turn, Patterson throws another one right in your face.

Fast-paced This plot is similar to most of the contemporary mystery/suspense/thriller/ good for Lifetime stories.

Fast-paced The plot was predictable you can see the "twists coming 200 pages ahead.

Fast-paced The chapters are short and the story moves at a quick pace.

Fast-paced This is a wonderful mystery thriller with many twists and turns to keep the reader guessing.

Fast-paced They have short chapters, are written in a fast pace, have more than one layer of plot going on.

Fast-paced The murders are twisted and the plot moves along quickly.

Fast-paced The novel works and the story builds suspense to the end with--of course--a last minute twist.

Fast-paced Overall it was a riveting, edge of your seat, intense murder mystery/thriller.

Fast-paced This is a wonderful mystery thriller with many twists and turns to keep the reader guessing.

Fast-paced Once you start you will find it difficult to put this one down.

Fast-paced This is the beginning of an exciting series in the crime/thriller genre, and it had me turning pages right up to the end.

Fast-paced There's lots of action as the Cahills seek to thwart each other.

Fast-paced Young readers will get a smattering of history and the landscape as they barrel along through the furious action and blistering dialogue.

Fast-paced I felt there was plenty of action in the "minor skirmish" sense.

Fast-paced It had lots of action and suspense.

Fast-paced A world was built here where there is a lot of space for many action-packed battle stories.

Fast-paced It was a fun read, full of action, excitement and characters who seemed to act true to form.

Fast-paced Sure it bogged down every now and then, but over all the story and action was great.

Fast-paced And yes, there is some splendid action.

Fast-paced I found myself skipping pages when the plot strayed from what I considered to be the major storyline.

Fast-paced For a person who likes tightly plotted novels, which move along at a fast pace to a dramatic climax.

Fast-paced The dialog is still a bit forced at times, and people spend long periods of time explaining things to each other.

Fast-paced It basically drags both authors down into a long-winded mire of dialogue.

Fast-paced The dialogue and prose snap, and though you can see where the story is headed a mile away.

Fast-paced From America's leading mystery authority, this informative guide is packed with plot twists.

Fast-paced It has plenty of action, both gunfights and fistfights.

Fast-paced But whether you read it for the plot twists and turns, the hard, unadorned prose served cold, or simply as a nostalgic walk down fiction's memory lane, "361" is prime pulp fiction, a quick thrill to savor and enjoy.

Fast-paced Westlake's seamless plotline, deft dialogue, and high pacing.

Fast-paced The plot jumps around very quickly, and as a reader I found myself confused several times over just exactly what was going on.

Fast-paced It is full of suspense, violence and clean prose you expect in the genre.

Fast-paced The authors developing mastery of dialogue is on display here.

Fast-paced Yet crisp dialogue and very readable prose make it an enjoyable read.

Fast-paced The book is full of action; something is always going on.

Fast-paced Its so action packed that it's really hard to put this novel down.

Fast-paced I would recommend this novel to anyone who is looking for something that is full of jam packed action and is ready for the many twists that are thrown out at you by the author.

Fast-paced It may sound sick and weird, but I enjoy reading books with plenty of action and some violence.

Fast-paced Interesting and hard to put down.

Fast-paced I would recommend this book to anyone who likes mysterious books and suspenseful books.

Fast-paced A strength of this book is the suspense that the author describes.

Fast-paced The story moves at a quick pace and the action never stops.

Fast-paced As a connoisseur of techno-thrillers, suspense and espionage novels, I think it ranks up there with the greatest.

Fast-paced Characterizations is on-target and makes for a fun, can't-put-it-down reading.

Fast-paced Several plot changes make for a good ending.

Fast-paced Also, the way the characters talk and the dialogue is so fitting to the character.

Fast-paced I think that the book is great. It is educational and action packed.

Fast-paced It is full of action and doesn't bore the reader with so many details he or she is lost.

Fast-paced Lots of suspenseful situations.

Fast-paced Great read, fun and lots of action.

Fast-paced The action won't stop, and never lets up.

Fast-paced If you like action its the book for you.

Fast-paced The action is fast and furious.

Fast-paced Overall it was a fun read and one I could not put down.

Fast-paced The book has top notch action.

Fast-paced The entire series is action packed.

Fast-paced Bottom line, if you want hardcore action with no holds barred, this series is tough to beat!

Fast-paced It is a fast moving thriller that never slows down from the first page.

Fast-paced It's fast paced, exciting, and uses both the elements of a thriller.

Fast-paced It is a fast moving thriller that never slows down from the first page.

Fast-paced She has given readers exciting characters and heartstopping suspense.

Fast-paced Everyone converses in a stilted manner.

Fast-paced There is no lack of action.

Fast-paced The best part of the book is that it is written in short chapters.

Fast-paced Easy to understand.

Fast-paced This is a real page-turner.

Fast-paced The author gives the reader hints of uncoming events while still leaving the reader in suspense.

Fast-paced The plot has many fine twists and surprises.

Fast-paced Once again King has provided a page-turning, non-stop read.

Fast-paced It's a gritty urban "noir thriller" that is hard hitting and exciting.

Fast-paced I think it is an excellent book, with a full range of emotions and lots of action.

Fast-paced The plot twists and turns enough to prevent the reader from predicting the events completely.

Fast-paced Several twists are in this story that will keep you up all night.

Fast-paced There are many twists throughout the novel.

Fast-paced It was a roller coaster ride of emotions and twists and turns and I couldn't wait to know what would happen next.

Fast-paced The information delivered about the 16th century was interesting and easy to read.

Fast-paced Very wonderful page turner.

Fast-paced I couldnt put it down.

Fast-paced It's filled with action, romance and a great plot.

Fast-paced I highly recommend this fast-paced book to anyone who loves medieval romances.

Fast-paced This is the kind of reading you want to take on vacation with you - to the beach, on the airplane, alongside the pool while the kids try to dunk one another.

Fast-paced This novel is a quick and easy read.

Fast-paced The writing gave the feeling of a suspenseful page turner despite knowing what the tragic outcome would be.

Fast-paced Suspenseful, though we all know the ending.

Fast-paced Intensely suspenseful despite the known outcome.

Fast-paced It is a beautiful love story and a suspenseful mystery

Fast-paced Easy to read, I read it in 4 nights, when usually it takes me 3 months to read a book.

Fast-paced We're left with a generally static narrative of fairly pointless minor incidents among uninteresting, vaguely threatening people.

Fast-paced She comes across as a bit stupid here.

Fast-paced When she causes her first accident she feels a "thrill."

Fast-paced Who is this person?

Fast-paced As always, her characters are, well for lack of a better description, characters.

Fast-paced You just can't believe that she would put herself in these ridiculous positions

Fast-paced The dialogue is weak, the sexual situations seem

gratuitous, and there is a fair amount of stereotyping.

Fast-paced However, Callard doesn't present a full enough picture of the woman's life and accomplishments.

Fast-paced The love between the two characters was never really explained, and the fact that the book happened in the space of about a week was wholly unbelievable.

Fast-paced Implausible plot twists

Fast-paced I found the characters in this book simplistic and obvious.

Fast-paced The characters in this book is childish and ridiculous.

Fast-paced I just never understood why the main character was making the decisions-radical and unorthodox-that he was making.

Fast-paced The rest of the characters also have the problem of conforming to stereotypes.

Fast-paced There's at times something more deeper to them, such as Asriel's true colors coming through and his discussion as to the nature of man, but it's really just a spark in a pool of shallow character development.

Fast-paced Also, I don't think the characters were very exciting and interesting because they would never really act like a real person would act during a similar situation.

Fast-paced The story of the narrating character and his little brother had generated no tension in me.

Fast-paced This book is filled with cool people and zany good times.

Fast-paced Rather than delving very deeply into emotional issues, the emphasis is mostly on how beautiful, albeit tortured, these women are.

Fast-paced The author doesn't delve very deeply into that line of the story.

Fast-paced This probably makes character development of the "other" woman difficult, and this book suffers for it.

Fast-paced Top it off with well-developed characters who are attractive, interesting, and complex and this author has delivered a real winner.

Fast-paced Most of the characters are likable.

Fast-paced I was hoping for funny characters with an edge.

Fast-paced I knew it would be a light read, and I wasn't sure if I would like it, but I loved it.

Fast-paced Little was revealed about Maggie's relationship with her soon to be estranged husband.

Fast-paced Maggie is not too believable as a good girl gone crazy, but you want to believe.

Fast-paced Some of the characters is predictable

Fast-paced The main character as well as incidentals weren't drawn very well and you never felt that attachment.

Fast-paced The Walsh characters are just that, stereotype.

Fast-paced I couldn't have cared less about Maggie.

Fast-paced I found the character very uninteresting.

Fast-paced The main character seems boring.

Fast-paced I thought the characters in this book were especially unrealistic.

Fast-paced I am very pleased to find out that the characters in this book are charming.

Fast-paced I thought the characters were one-dimensional and had very little depth.

Fast-paced His characters seem simple - a collection of traits,

perhaps.

Fast-paced The characters are very genuine.

Fast-paced It was a story which had compassion for all of the characters.

Fast-paced What impress me the most was how the main character is being really true to her heart and follows the way she feels for her one truly love.

Fast-paced The author brings out the divinity of wonderful relationships between all the book characters.

Fast-paced My problem was that I didn't know the characters, and I think there was an assumption by the writers that I already did.

Fast-paced The characters are so lifeless they might as well be made out of cardboard.

Fast-paced You identify with all these people and believe they are who they say they are.

Fast-paced The reader can identify similar situations and emotions in her or his own life.

Fast-paced Bianca is a character any teen, or adult for that matter, can identify with.

Fast-paced The author did a great job in depicting kids that don't seem to "fit in" to the general flow of high school, which is easily identifiable by many kids.

Fast-paced Unbelievable character, and this is one of the major problems of this book.

Fast-paced And even if the characters aren't completely likeable or easy to identify with, that doesn't make it a bad book.

Fast-paced She is very one dimensional and this makes the story less interesting.

Fast-paced The main character, Eve Duncan, was not someone that I even cared what happened to.

Fast-paced Fast and involved reading. Full of action.

Fast-paced Boring characters (even their names are dull), plodding story line, tacky and padded dialogue.

Fast-paced I have read all of Iris Johanson's books and enjoy the romance, suspense and interaction between the characters.

Fast-paced The characters are very likable and believable.

Fast-paced In fact, all of the characters are very likeable.

Fast-paced The characters were stereotyped and paper-thin.

Fast-paced Very little character development and unnecessarily complicated plot.

Fast-paced The characters don't develop or grow and have as much depth as the paper they're printed on.

Fast-paced The characters are very real and likeable.

Fast-paced Its a story that women of today can identify with.

Fast-paced She relates wonderful stories of her own experiences.

Fast-paced The author goes beyond, showing compassion for her characters and is painting realistic people.

Fast-paced I identified with Poly more than any other character I'd ever read.

Fast-paced The character development was also a tad on the thin side.

Fast-paced And she can't make her reader *feel* anything, because her characters have no depth.

Fast-paced I loved how she wrote of the main characters' emotions; you really get into the love story.

Fast-paced Thats good writing in my book, if you can feel like you

know the characters.

Fast-paced I felt every individual character's emotions and I fell in love.

Fast-paced They have no depth, and their psychology is unbelievable.

Fast-paced I was immediately captivated by its simple language and great metaphores.

Fast-paced The characters are pretty realistic and very typical of the new Miami.

Fast-paced This novel is an entertaining and easy read.

Fast-paced Every character - major and minor - was alive.

Slow-paced This volume collects some of the most significant works on the novel from a wide variety of theoretical perspectives, including Marxist, psychoanalytical, historicist and feminist, giving a collection which engages with questions about the psychological and social significance of this transgressive and popular text.

Slow-paced This thing goes on and on and not once was I absorbed.

Slow-paced But does Grafton ever deliver with a little excitement in her plots?

Slow-paced All she seemed to do was go to the bar with them or hang out in their apartment.

Slow-paced The plot doesn't seem to advance very quickly in these segments.

Slow-paced It rambles aimlessly before it comes to a frustrating and disappointing end.

Slow-paced Refreshing and scholarly take on an area.

Slow-paced writer Evelyn Scott succeeds in interesting the reader with sordid details of Scott's escapades.

Slow-paced It gives you very detailed information.

Slow-paced I really think that this book is very informative.

Slow-paced This is great book with very detailed information.

Slow-paced Incredibly detailed piece of work.

Slow-paced The result is a well-centered volume that is disciplined and restrained in its presentation of research, but which is layered and multivocal in presentation, and which ends with some provocative conclusions.

Slow-paced A rewarding and challenging read.

Slow-paced This enchanting book will make you want to sit down with a soothing piece of needlework and relax.

Slow-paced Kaleidescopic in structure and movement, daringly personal yet intensely political, his verse encompasses narrative, satire, meditation, and high-spirited hilarity.

Slow-paced Line by line we turn from zany wisecracks about globalization to solemn invocations of the moon, from intensely felt lyricism to splintered, televisual, attention-fractured deferrals of feeling.

Slow-paced However, as I continue to ponder the implications of the unquestioned support.

Slow-paced These dazzling lyrics and sequences create one of the most compelling portraits we have of a mind, a sensibility, a language.

Slow-paced He is one of those gifted poets whose words need only be read once to reverberate in memory for a lifetime.

Slow-paced It is an amazing, entertain and beautiful read that you should definitely check out

Slow-paced The descriptions of the scenery as well the character's

dresses and actions made it seem like a period movie played out in my head as I read.

Slow-paced All of this gives the reader a richly complex kaleidoscopic view.

Slow-paced I am currently struggling to read through this.

Slow-paced His mythical yet factual life is accentuated by stories of a currandera who "sees" his life slowly unfolding.

Slow-paced His story is rich with historical detail.

Slow-paced Presented in a historical and literary context that makes this worth reading.

Slow-paced Can be a bit slow going.

Slow-paced The writing is clear and has a wonderful swaying quality to it.

Slow-paced I read the L-O-N-G run on sentences that clogged up reading flow in just the first two pages.

Slow-paced Reading it's long, winding, meandering sentences was like walking through molasses.

Slow-paced It took me nearly forever to finish it.

Slow-paced This book is not for people who need lots of explicit flash-bang style action.

Slow-paced This is a book for those that like a challenge in reading.

Slow-paced It is not passive reading.

Slow-paced It needs your whole attention throughout the novel, and if you are lucky and get inside you will have a wonderful experience, you will be there with her and the characters, reading with a dim light.

Slow-paced The simplest and tiniest of actions are examined in microscopic detail so as to capture the finer feeling of the moment.

Slow-paced It's a rich and complex work, and totally unlike anything I've ever read.

Slow-paced But in the end, what makes this most remarkable is the rich detailing, the marvelous nuances, the lush delineations.

Slow-paced This is a book to be read slowly and savored, much as one would read a long poem.

Slow-paced But keep an open mind for it takes patience to fall into this masterpiece.

Slow-paced I found myself re-reading passages over and over again, just to get the meaning.

Slow-paced This book will take time and patience. But if you keep working at it and savour every word, then you will be in for a rewarding literary experience.

Slow-paced He uses very poignant expressions constantly throughout the book that can get the reader off track because they tend to cause him to attempt to frequently stop and process what has just been said.

Slow-paced This isn't a hard concept, but Wilde's writing is very challenging.

Slow-paced This was a rather difficult book to get through.

Slow-paced In the end, it was worth the time and effort but does take some patience to get through.

Slow-paced In fact, perhaps those slow zones serve the purpose of allowing more pondering.

Slow-paced Also, the language was hard to understand.

Slow-paced I would only recommend this to someone who has a lot of time on their hands or to someone who likes slow paced books.

Slow-paced Once I got past the enormously long sentences, this book is a charmer.

Slow-paced I stayed up late into the night reading what would happen.

Slow-paced This book droned on and on.

Slow-paced People say the plot was confusing but if you stop and reread a few pages everything will come together.

Slow-paced The main character, Eve, is annoying, pompous, sanctimonious and rude.

Slow-paced You'll want to read this one slowly.

Slow-paced Slow moving at first, but interesting just the same.

Slow-paced I very much look forward to that plot development as the portions of history I've read on the man tend to impress me.

Slow-paced Weber write an engrossing story with convincingly real characters.

Slow-paced On the other hand, there are fewer battles and a generally slower pace.

Slow-paced The necessary downside in the writing is more exposition and slower pace.

Slow-paced I have not minded the slower pace because it has given me a better opportunity to get to know the characters and there are a lots of interesting characters in this book.

Slow-paced The material was dense, probably historically accurate.

Slow-paced I found myself laboring through entire chapters of the book forced to read what I considered some rather dull, pages long descriptions of constitutional banter.

Slow-paced Naipaul's passages of description and reflection postpones the book's impact as a true novel.

Slow-paced His writing style is compelling and elegant, while the plot and characterization are superb.

Slow-paced I'm pleased to report that although it's not quite as genius as his debut, it is still exceedingly funny, and has a bit more depth to it.

Slow-paced I thought the historical and geographical descriptions to be adequate in and of themselves.

Slow-paced It is written in slang and is difficult to pick up on at first, is told from the first person.

Slow-paced Never before in my entire life have I ever read a plot so vivid and detailed.

Slow-paced The smattering of scintillating imaginary slang that dance around the story doesn't take away from the often ironic and thought provoking plot.

Slow-paced In a compact plot and a short novel, the author elucidates the issues of moral determinism vs moral choice.

Slow-paced It also touches on the psychological and social aspects that are prevalent in today's society.

Slow-paced I had to read this book more slowly than others.

Slow-paced The author does an excellent job of presenting this issue, developing both sides of the argument without bias.

Slow-paced An amazingly creative, engrossing novel.

Slow-paced The language keeps the reader's attention on the events of the story.

Slow-paced I have never seen language used in such a way.

Slow-paced It is without a doubt going to be a slight challenge for the reader, because most of it is written in a completely different language.

Slow-paced Great book, a little hard to read a first because of the language.

Slow-paced Overall, the book excellently demonstrates an intellectual grasp on society and also his commanding use of language.

Slow-paced Let me tell ya... the language takes a bit to sink in.

Slow-paced It takes a while for your brain to process this language.

Slow-paced Some will be turned off by the difficulty of reading the made up language.

Slow-paced Don't let the language stop you from reading this powerful novel.

Slow-paced One should not let the language stop them from enjoying a truly classic piece of literature after all.

Slow-paced The author puts it down so intelligent and repeats certain things to give you this poetic feeling.

Slow-paced This book is extremely long and rarely has any action to keep you awake.

Slow-paced I found this book very boring because it didn't have much action to keep my attention.

Slow-paced There was really no action in it.

Slow-paced His wit and caustic style is just one of a kind, and his novels, with the in depth plot and yet witty and sardonic characters.

Slow-paced The author provides a philosophical proposition within an interesting plot.

Slow-paced This book has a real lack of depth of character development and lots of stilted dialogue.

Slow-paced Even though the book starts out slow the fast chapter breaks make it a book that needs no bookmarks.

Slow-paced The description is campy; the authors apparently decided to focus on the clothing of women. There are attractive wimples and red low cut blouses galore

Slow-paced The plot builds up nicely, although a little sadly.

Slow-paced You'll like this book if you enjoy the impassive style of writing and if you're not particularly looking for a book with a dozen heart-hammering twists.

Slow-paced The absence of a central plot as well as the length makes this a harder read than other novels.

Slow-paced This is a book rich in layers of meaning.

Slow-paced The imagery... wow, although nothing is ever just imagery with her, it's life, every sense, as always, is brought to life in her descriptions.

Slow-paced The impact generated is the result of a slow but steady accretion of little things done extremely well.

Slow-paced There is NO carefully observed detail of characterization OR place.

Slow-paced The book starts out slow, so you have to stay with it.

Slow-paced What makes this an extraordinary novel is the vibrant characters, well-detailed descriptions and just the sheer poignancy of the story that is guaranteed to enthrall.

Slow-paced In a story that is more about characters than about action, motivations often start out hazy, and only gradually come into focus.

Slow-paced Beautifully descriptive of a bygone era and laconic in its pace, this is most certainly a novel to be savored.

Slow-paced Compelling in its description of time and place.

Slow-paced It has the elegance of a waltz but the depth and richness of a symphony.

Slow-paced This book was a little slow to start, but about halfway through I couldn't put it down.

Slow-paced Written in a language poetic, it is nothing short of a gift of an experience if read.

Slow-paced I call this writing pure poetry.

Slow-paced It can be hard to read at times, and hard to follow.

Slow-paced The concise nature of this story does not reflect upon the depth of the prose and, in fact, the author imbues every line with description and dialogue so wonderfully rich that the length of the work is irrelevant.

Slow-paced This is a delightful book -- an offbeat story line, illuminated by beautiful writing, excellent characterizations, and an intriguing denouement.

Slow-paced He brings us along a story that unfolds like flowers slowly opening to the sun, offering many precious moments which linger in you mind and tug at your heart.

Slow-paced Things slow down at times and we do wonder where it is all going, but patience wins the day and we are brought to the brink of both pathos and joy.

Slow-paced The author is excellent in his description of intricate details, as well as in his unraveling of an exceptionally original plot, to reaching one's deepest feelings of sadness about experiences of unrequited and unfulfilled love, and finally to the revealing of the mural behind the old paint and the secret of the original artist himself.

Slow-paced But the crowning glory of the book is just its sparkling, gorgeous prose throughout.

Slow-paced Lyrical, poetic escape.

Slow-paced A unique book which gathers in its 135 pages a wonderful, well told story full of poetry.

Slow-paced The writing is excellent; the characters are complex, and the detail (I presume) is correctly rendered.

Slow-paced If you like antiques, quirky experts (nobody can do quirky like the British!).

Slow-paced The line between right and wrong begins to shift along with Hannah's personality and involvement.

Slow-paced Raymond and Bibiana are particularly memorable supporting characters as well.

Slow-paced We read her for the richness of her writing, her psychological insights, dry wit.

Slow-paced Herbert Brokering is a master with language, and this book is no exception.

Slow-paced The parables in this book are simple, but have several different layers of meaning. The neatest thing is that there is no "right" meaning for each parable; the meaning might be different for each person.

Slow-paced Ruud is able to present the value and beauty of Chaucer's lyric poetry in an accessible, but not patronizing, style.

Slow-paced A cheap book full of dirty gossips.

Slow-paced Zoe is definitely a strong independent female character forced to grow up wiser than her years due to the inhumanity she grew up around as a child and that she was reminded of by her stepmother everyday.

Slow-paced But overall the book feels like it could have taken place

in a very real time not too long ago.

Slow-paced In contrast to Zoe, she encounters the Counte , a strong, assertive, but troubled character who drags her into his own web of issues.

Slow-paced The power of this novel is that as a reader I could feel Zoe's pain.

Slow-paced One minute I was feeling the loves happiness and the next I was feeling their pain

Slow-paced The author creates characters that are easy to love despite their flaws.

Slow-paced Moving, funny, and at times tragic, the layered dimensions of their world are mesmerizing.

Slow-paced His life story really gives you an incite.

Slow-paced Napoleon has crafted a tale blending humor, sex, tragedy and social consciousness.

Slow-paced He has created interesting and memorable characters to support the story of the unique and brilliant.

Slow-paced The story is bright and crisp, and Meisel does an excellent job of showing both close-ups of Sigmund and his daughter Zara, and scene-setting long perspectives as well.

Slow-paced She explores as well Eliot's keen sense of the important distinctions between specific schools of thought.

Slow-paced A story that reveals once again that the extraordinary lies just beneath the surface in even the most "ordinary" of people.

Slow-paced Enough suspense to keep the reader turning the pages.

Slow-paced As the characters try to make sense of their own circumstances and find a way to move forward with their lives, dangerous collisions become inevitable and choices must be made.

Slow-paced The little details in the relationships between the characters, their idiosyncrocies and habits, are intriguing.

Slow-paced She also knows how to weave a story and her characters, even those that aren't as well fleshed out, stay with you long after you've read the last page.

Slow-paced The ebbs and flows of relationships, the frailty of life....it's all in this amazing piece of literature.

Slow-paced She also, quite eloquently and precisely, somehow manages to describe "the dance" known as marriage.

Slow-paced "To a Lighthouse" is a masterpiece; it succeeds as a group and character portrait While a difficult "read", the story, the plot, the character definition, and the passage of time...largely accomplished through internal dialogue.

Slow-paced It is a book that deals with the inner workings of the mind.

Slow-paced The book provides food for thought on the author's strange and sometimes twisted views on society as played out by the characters.

Slow-paced In spite of being a hero, he's shown as a man with selfdoubt

and sometimes (especially on land) not knowing how to behave.

Slow-paced The character development (within the book, but especially over the series) is engrossing.

Slow-paced It is also, however, a deeply psychological novel.

Slow-paced Particularly intriguing are the characters.

Slow-paced He describes her awful character with such meticulous detail.

Slow-paced This is no doubt a book that makes you think about the

psychological aspect of life.

Slow-paced the characters are portrayed in a realistic and poignant way that helps the reader to understand who they are and why they are being slammed by forces outside of their control.

Slow-paced I cannot even begin to imagine putting together so many characters and situations and ideas into such beautiful words and images.

Slow-paced I've never read a book that has taken such a sentimental turn for the worse than this.

Slow-paced I'd recommend this book for deep thinkers and people not afraid of a disappointing ending.

Slow-paced It is so real, and the feelings are so real and strong that you suffer with the characters.

Slow-paced I have never felt such strong emotions towards a literary character.

Slow-paced His descriptions of the world and the characters he's created are vivid

Slow-paced The author does a great job of putting a picture of what of what the main character is seeing and experiencing.

Slow-paced The details are so well put that you can almost believe you were there.

Slow-paced There are so many aspects of it to enjoy, and it is just an all around entertaining book.

Slow-paced Colorful and sublimely written characters.

Slow-paced What an incredible account of the author's pain, hope, love, fears, hatred.

Slow-paced I became so entwined in the storyline that it felt like I was a part of this disjointed and hilarious family.

Slow-paced There are reoccurring instances where the main character precariously falls into the most atrocious situations, of which he has no control.

Slow-paced The characters are drawn with care and whose attraction to one another is utterly convincing.

Slow-paced There are three main characters and they are simply a treat to get to know.

Slow-paced The leading ladies and the supporting cast of this debut novel by newcomer K.E Lane will charm you, entertain you, and leave you with a smile on your face, eager for Ms. Lane's next offering.

Slow-paced What makes a story stand out is the depth and breadth of characterization

Slow-paced The characters has a great conversational, casual style loaded with great details and nuances.

Slow-paced Great character development and just a fun book to read.

Slow-paced Both the leading ladies and all the secondary characters are well-developed, smart, funny, tender and incredibly likable.

Slow-paced These are the masters of odd but endearing characters.

Slow-paced She is at her best when she juxtaposes humor with dark, psychological element to expose her characters' nature.

Slow-paced I enjoyed them for the wonderful characters: flawed, but funny and decent people.

Slow-paced As with all the other sisters, she is a multi-dimensional woman who is coming to grips with some major developments in her life.

Slow-paced Both characters were well developed.

Slow-paced There's much more plot depth and characters that were overly stereotypical in the first novel (ie; Simpson) are given

actual personalities, and on this latter account the story gains much.

Slow-paced I have not minded the slower pace because it has given me a better opportunity to get to know the characters and there are a lots of interesting characters in this book.

Slow-paced There is much more character development, which fleshes out some of the characters.

Slow-paced There is way too much character development.

Slow-paced The plot and character development are good and the time period is interesting, to say the least.

Slow-paced The character are interesting and with psychological depths not seen before.

Slow-paced Not only a good story, but depth to characters, as well.

Slow-paced It is a good book with well developed characters, all together a good novel.

Slow-paced Although the story is somewhat slow at times, and is occasionally idealistic, there is a wonderful depth and development of the main character as well as a poignant theme about life and love.

Slow-paced Just proved to me once again that a good story with profoundly good character development is ageless and timeless.

Slow-paced Manchester has grown into an extremely strong secondary character, to the extent that one cannot help but wish to see him used more frequently.

Slow-paced His character development is subtle but has a way of making me feel invested in his characters

Slow-paced Another commendable aspect is the growth in character that happens to the main characters.

Slow-paced It were worth the read because of the depth of the characters she has the reader follow.

Slow-paced The relationships were as varied and complicated as the characters, and that's what makes A Lost Lady so great for analyzing.

Slow-paced A very complex character, I was amazed by this novel because of the ambiguity of its protagonist.

Slow-paced It brings otherwise unremarkable characters to life with very concisely-worded descriptions.

9.9 Vedlegg 8

Complex_characters The writing is excellent; the characters are complex, and the detail (I presume) is correctly rendered.

Complex_characters If you like antiques, quirky experts (nobody can do quirky like the British!).

Complex_characters The line between right and wrong begins to shift along with Hannah's personality and involvement.

Complex_characters Raymond and Bibiana are particularly memorable supporting characters as well.

Complex_characters We read her for the richness of her writing, her psychological insights, dry wit.

Complex_characters We are introduced to a cast of historic characters, both heroes and villains

Complex_characters Herbert Brokering is a master with language, and this book is no exception.

Complex_characters The parables in this book are simple, but have several different layers of meaning. The neatest thing is that there is no "right" meaning for each parable; the meaning might be different for each person.

Complex_characters Ruud is able to present the value and beauty of Chaucer's lyric poetry in an accessible, but not patronizing, style.

Complex_characters A cheap book full of dirty gossips.

Complex_characters Zoe is definitely a strong independent female character forced to grow up wiser than her years due to the inhumanity she grew up around as a child and that she was reminded of by her stepmother everyday.

Complex_characters But overall the book feels like it could have taken place in a very real time not too long ago.

Complex_characters In contrast to Zoe, she encounters the Counte , a strong, assertive, but troubled character who drags her into his own web of issues.

Complex_characters The power of this novel is that as a reader I could feel Zoe's pain.

Complex_characters One minute I was feeling the loves happiness and the next I was feeling their pain

Complex_characters The author creates characters that are easy to love despite their flaws.

Complex_characters Moving, funny, and at times tragic, the layered dimensions of their world are mesmerizing.

Complex_characters His life story really gives you an incite.

Complex_characters Napoleon has crafted a tale blending humor, sex, tragedy and social consciousness.

Complex_characters He has created interesting and memorable characters to support the story of the unique and brilliant.

Complex_characters The story is bright and crisp, and Meisel does an excellent job of showing both close-ups of Sigmund and his daughter Zara, and scene-setting long perspectives as well.

Complex_characters She explores as well Eliot's keen sense of the important distinctions between specific schools of thought.

Complex_characters Her characters have flaws, but often great nobility.

Complex_characters She is an intelligent and gifted writer with a knack for making characters come to life.

Complex_characters A story that reveals once again that the extraordinary lies just beneath the surface in even the most

"ordinary" of people.

Complex_characters Enough suspense to keep the reader turning the pages.

Complex_characters As the characters try to make sense of their own circumstances and find a way to move forward with their lives, dangerous collisions become inevitable and choices must be made.

Complex_characters The little details in the relationships between the characters, their idiosyncrocies and habits, are intriguing.

Complex_characters She also knows how to weave a story and her characters, even those that aren't as well fleshed out, stay with you long after you've read the last page.

Complex_characters To the Lighthouse is a dazzling work of art that illustrates with great clarity the dynamic nature of human thought and perception.

Complex_characters Without ever being a mother herself, she captured a mothers fears, various feelings, and love for her children.

Complex_characters The ebbs and flows of relationships, the frailty of life...it's all in this amazing piece of literature.

Complex_characters She also, quite eloquently and precisely, somehow manages to describe "the dance" known as marriage.

Complex_characters "To a Lighthouse" is a masterpiece; it succeeds as a group and character portrait While a difficult "read", the story, the plot, the character definition, and the passage of time...largely accomplished through internal dialogue.

Complex_characters It is a book that deals with the inner workings of the mind.

Complex_characters The book provides food for thought on the author's strange and sometimes twisted views on society as played out by the characters.

Complex_characters In spite of being a hero, he's shown as a man with self-doubt and sometimes (especially on land) not knowing how to behave.

Complex_characters The character development (within the book, but especially over the series) is engrossing.

Complex_characters It is also, however, a deeply psychological novel.

Complex_characters Particularly intriguing are the characters.

Complex_characters He describes her awful character with such meticulous detail.

Complex_characters This is no doubt a book that makes you think about the psychological aspect of life.

Complex_characters the characters are portrayed in a realistic and poignant way that helps the reader to understand who they are and why they are being slammed by forces outside of their control.

Complex_characters I cannot even begin to imagine putting together so many characters and situations and ideas into such beautiful words and images.

Complex_characters I've never read a book that has taken such a sentimental turn for the worse than this.

Complex_characters I'd recommend this book for deep thinkers and people not afraid of a disappointing ending.

Complex_characters It is so real, and the feelings are so real and strong that you suffer with the characters.

Complex_characters I have never felt such strong emotions towards a literary character.

Complex_characters His descriptions of the world and the characters

he's created are vivid

Complex_characters The author does a great job of putting a picture of what of what the main character is seeing and experiencing.

Complex_characters The details are so well put that you can almost believe you were there.

Complex_characters There are so many aspects of it to enjoy, and it is just an all around entertaining book.

Complex_characters Colorful and sublimely written characters.

Complex_characters What an incredible account of the author's pain, hope, love, fears, hatred.

Complex_characters I became so entwined in the storyline that it felt like I was a part of this disjointed and hilarious family.

Complex_characters There are reoccurring instances where the main character precariously falls into the most atrocious situations, of which he has no control.

Complex_characters The characters are drawn with care and whose attraction to one another is utterly convincing.

Complex_characters There are three main characters and they are simply a treat to get to know.

Complex_characters I think the most delightful part of the novel is the character's view of themselves.

Complex_characters The leading ladies and the supporting cast of this debut novel by newcomer K.E Lane will charm you, entertain you, and leave you with a smile on your face, eager for Ms. Lane's next offering.

Complex_characters What makes a story stand out is the depth and breadth of characterization

Complex_characters The characters has a great conversational, casual style loaded with great details and nuances.

Complex_characters Great character development and just a fun book to read.

Complex_characters The extended length of the book gives the characters a chance to really develop their personalities

Complex_characters Both the leading ladies and all the secondary characters are well-developed, smart, funny, tender and incredibly likable.

Complex_characters These are the masters of odd but endearing characters.

Complex_characters She is at her best when she juxtaposes humor with dark, psychological element to expose her characters' nature.

Complex_characters I enjoyed them for the wonderful characters: flawed, but funny and decent people.

Complex_characters As with all the other sisters, she is a multidimensional

woman who is coming to grips with some major developments in her life.

Complex_characters This book captured the complex creature known as man.

Complex_characters Both characters were well developed.

Complex_characters I did enjoy the extra pages spent on developing the characters

Complex_characters There's much more plot depth and characters that were overly stereotypical in the first novel (ie; Simpson) are given actual personalities, and on this latter account the story gains much.

Complex_characters I have not minded the slower pace because it has

given me a better opportunity to get to know the characters and there are a lots of interesting characters in this book.

Complex_characters There is much more character development, which fleshes out some of the characters.

Complex_characters There is way too much character development.

Complex_characters It is highly detailed and intricately woven, with characters developing nicely.

Complex_characters The plot and character development are good and the time period is interesting, to say the least.

Complex_characters The character are interesting and with psychological depths not seen before.

Complex_characters Not only a good story, but depth to characters, as well.

Complex_characters It is a good book with well developed characters, all together a good novel.

Complex_characters Although the story is somewhat slow at times, and is occasionally idealistic, there is a wonderful depth and development of the main character as well as a poignant theme about life and love.

Complex_characters Just proved to me once again that a good story with profoundly good character development is ageless and timeless.

Complex_characters Her main character faces some problems that bring about a lot of mature growth.

Complex_characters Manchester has grown into an extremely strong secondary character, to the extent that one cannot help but wish to see him used more frequently.

Complex_characters His character development is subtle but has a way of making me feel invested in his characters

Complex_characters Another commendable aspect is the growth in character that happens to the main characters.

Complex_characters It were worth the read because of the depth of the characters she has the reader follow.

Complex_characters The relationships were as varied and complicated as the characters, and that's what makes A Lost Lady so great for analyzing.

Complex_characters A very complex character, I was amazed by this novel because of the ambiguity of its protagonist.

Complex_characters The storyline was unique and the characters intricate and amusing.

Complex_characters It brings otherwise unremarkable characters to life with very concisely-worded descriptions.

Shallow_characters The hero infiltrates an insurance scam ring that turns out to be run by LA gangsters with Touretts syndrome. She puts on some trashy clothes and bad makeup and they just invite her in for a beer.

Shallow_characters We're left with a generally static narrative of fairly pointless minor incidents among uninteresting, vaguely threatening people.

Shallow_characters She comes across as a bit stupid here.

Shallow_characters When she causes her first accident she feels a "thrill."

Shallow_characters Who is this person?

Shallow_characters As always, her characters are, well for lack of a better description, characters.

Shallow_characters You just can't believe that she would put herself in these ridiculous positions

Shallow_characters The dialogue is weak, the sexual situations seem gratuitous, and there is a fair amount of stereotyping.

Shallow_characters However, Callard doesn't present a full enough picture of the woman's life and accomplishments.

Shallow_characters The love between the two characters was never really explained, and the fact that the book happened in the space of about a week was wholly unbelievable.

Shallow_characters Implausible plot twists

Shallow_characters I found the characters in this book simplistic and obvious.

Shallow_characters The characters in this book is childish and ridiculous.

Shallow_characters I just never understood why the main character was making the decisions—radical and unorthodox—that he was making.

Shallow_characters The rest of the characters also have the problem of conforming to stereotypes.

Shallow_characters There's at times something more deeper to them, such as Asriel's true colors coming through and his discussion as to the nature of man, but it's really just a spark in a pool of shallow character development.

Shallow_characters Also, I don't think the characters were very exciting and interesting because they would never really act like a real person would act during a similar situation.

Shallow_characters The story of the narrating character and his little brother had generated no tension in me.

Shallow_characters This book is filled with cool people and zany good times.

Shallow_characters The main character was something of a pinball, bouncing from one place and situation to another as a consequence of the seemingly random decisions of those around him.

Shallow_characters Rather than delving very deeply into emotional issues, the emphasis is mostly on how beautiful, albeit tortured, these women are.

Shallow_characters The author doesn't delve very deeply into that line of the story.

Shallow_characters This probably makes character development of the "other" woman difficult, and this book suffers for it.

Shallow_characters Top it off with well-developed characters who are attractive, interesting, and complex and this author has delivered a real winner.

Shallow_characters Most of the characters are likable.

Shallow_characters I was hoping for funny characters with an edge.

Shallow_characters One of the reasons why I love the characters in her books is because they are shamelessly witty and fun

Shallow_characters I knew it would be a light read, and I wasn't sure if I would like it, but I loved it.

Shallow_characters Little was revealed about Maggie's relationship with her soon to be estranged husband.

Shallow_characters Maggie is not too believable as a good girl gone crazy, but you want to believe.

Shallow_characters I loved the character being compared to plain yogurt at room temperature.

Shallow_characters Some of the characters is predictable

Shallow_characters The main character as well as incidentals weren't drawn very well and you never felt that attachment.

Shallow_characters The Walsh characters are just that, stereotype.

Shallow_characters I couldn't have cared less about Maggie.

Shallow_characters I found the character very uninteresting.

Shallow_characters The main character seems boring.

Shallow_characters I thought the characters in this book were especially unrealistic.

Shallow_characters I am very pleased to find out that the characters in this book are charming.

Shallow_characters I thought the characters were one-dimensional and had very little depth.

Shallow_characters His characters seem simple - a collection of traits, perhaps.

Shallow_characters The characters are very genuine.

Shallow_characters It was a story which had compassion for all of the characters.

Shallow_characters What impress me the most was how the main character is being really true to her heart and follows the way she feels for her one truly love.

Shallow_characters The author brings out the divinity of wonderful relationships between all the book characters.

Shallow_characters My problem was that I didn't know the characters, and I think there was an assumption by the writers that I already did.

Shallow_characters The characters are so lifeless they might as well be made out of cardboard.

Shallow_characters You identify with all these people and believe they are who they say they are.

Shallow_characters The reader can identify similar situations and emotions in her or his own life.

Shallow_characters Bianca is a character any teen, or adult for that matter, can identify with.

Shallow_characters The author did a great job in depicting kids that don't seem to "fit in" to the general flow of high school, which is easily identifiable by many kids.

Shallow_characters Unbelievable character, and this is one of the major problems of this book.

Shallow_characters And even if the characters aren't completely likeable or easy to identify with, that doesn't make it a bad book.

Shallow_characters She is very one dimensional and this makes the story less interesting.

Shallow_characters The main character, Eve Duncan, was not someone that I even cared what happened to.

Shallow_characters Fast and involved reading. Full of action.

Shallow_characters Boring characters (even their names are dull), plodding story line, tacky and padded dialogue.

Shallow_characters I have read all of Iris Johanson's books and enjoy the romance, suspense and interaction between the characters.

Shallow_characters The characters are very likable and believable.

Shallow_characters In fact, all of the characters are very likeable.

Shallow_characters The characters were stereotyped and paper-thin.

Shallow_characters Very little character development and unnecessarily complicated plot.

Shallow_characters The characters don't develop or grow and have as much depth as the paper they're printed on.

Shallow_characters The characters are very real and likeable.

Shallow_characters Its a story that women of today can identify with.

Shallow_characters She relates wonderful stories of her own experiences.

Shallow_characters The author goes beyond, showing compassion for her characters and is painting realistic people.

Shallow_characters I identified with Poly more than any other character I'd ever read.

Shallow_characters The character development was also a tad on the thin side.

Shallow_characters And she can't make her reader *feel* anything, because her characters have no depth.

Shallow_characters I loved how she wrote of the main characters' emotions; you really get into the love story.

Shallow_characters Thats good writing in my book, if you can feel like you know the characters.

Shallow_characters I felt every individual character's emotions and I fell in love.

Shallow_characters They have no depth, and their psychology is unbelievable.

Shallow_characters I was immediately captivated by its simple language and great metaphores.

Shallow_characters The characters are pretty realistic and very typical of the new Miami.

Shallow_characters This novel is an entertaining and easy read.

Shallow_characters Every character - major and minor - was alive.

9.10 Vedlegg 9

Light_Tone Scuse Me While I Kiss This Guy, and its followup, He's Got The Whole World in His Pants (both by Gavin Edwards) had me, as well as everyone in my office, laughing so hard we couldn't speak.

Light_Tone This book contains the most outraged, hilarious, and humorous misunderstood lyrics.

Light_Tone Her books show her very funny sense of humor and nothing is immune to her wit.

Light_Tone Who else could come up with a singles bar named "The Meat Locker" and get away with the reader chuckling about it?

Light_Tone The typography is excellent and fun to read.

Light_Tone I can't remember when a book last had me cheering out loud.

Light_Tone It was filled with romance, jealousy, betrayal, surprises, and so much more.

Light_Tone Wise's playful poetry.

Light_Tone Hilarious watercolor illustrations highlighting each creature's distinct personality.

Light_Tone I laughed until I cried.

Light_Tone The most fun required reading I ever did.

Light_Tone The author has created a heartwarming story about a loving family

Light_Tone It had a lot of humorous quips and paradigms as presented by Harry that I laughed out loud at.

Light_Tone It is a great adventure and mystery book.

Light_Tone The book starts in an upbeat comical mood.

Light_Tone Jokes are being thrown, and people are having fun times.

Light_Tone The novel is filled with humor and at times takes an optimistic stance on technology and modern age versus middle age, but there're also some touching moments and more than meets the eye at a first stance.

Light_Tone I would recommend this to anyone who has an imagination and loves to read adventures with humor as well.

Light_Tone The story is captivating and realistic (except, of course, for how he got there), punctuated by witty humor and hilarious circumstances.

Light_Tone In conclusion I did like this book because it was witty and slightly far-fetched.

Light_Tone The idea of a telephone system, a printing press or the King riding a bicycle is simply prepsterous and therefore humorous.

Light_Tone There is no doubt that this is thoroughly enjoyable and humorous work.

Light_Tone It is not so much the dark and serious underlying tone that has made this work a classic but Twains brilliant humour, creativeness, and story-telling.

Light_Tone Along the way, the book brought both tears and laughter to this reader.

Light_Tone This a good story, very well written, intelligent, amusing and every so often something comes up in it that makes you laugh out loud.

Light_Tone Do yourself a favor and read this book because its highly entertaining, very humorous, and definitely an excellent read.

Light_Tone I recommend this book for people who want something light and less serious, because this book will give you a good laugh.

Light_Tone I think that some of his funniest moments are in this

novel.

Light_Tone I am especially fond of humorous and witty mysteries.

Light_Tone Eggers has great language, a wonderful sense of humor, and a vivid imagination.

Light_Tone This book is brilliant. Its humor is perfectly balanced with its raw life lessons.

Light_Tone It is wonderful in that Dave Eggers CAN really turn a phrase and be hilarious, poignant, amazing.

Light_Tone I laughed out loud on every page.

Light_Tone Wit, wisdom, a sense of humour, vision, style, flair, and the passion that enables him to masterfully craft such a truly genius work.

Light_Tone His extraordinary sense of humour also is commendable

Light_Tone Trust me, you'll be hitting yourself over the head, this is a fairly comical book

Light_Tone Its a comic adventure alright and its also a book I would recommend you read.

Light_Tone Laugh out loud funny.

Light_Tone What a funny novel!

Light_Tone The humor, the gotchas and the satire are just marvelous.

Light_Tone Marian didn't try too hard to satirise it but instead portrayed its madness with a nice, light touch.

Light_Tone Her adventures were at times funny.

Light_Tone Did I mention it makes for a great summer read.

Light_Tone I could not stop laughing aloud in public.

Light_Tone The novel are filled with laugh out loud, wipe your eyes moments that are absolutely priceless.

Light_Tone The book is great for a holiday or summer (vacation) read.

Light_Tone There was so much wit and humor to keep the book upbeat.

Light_Tone It adds humor to real life and has similar situations that occur in your life.

Light_Tone Cabot adds humor and wit to an ageless story.

Light_Tone I highly recommend this novel for anyone looking for an enjoyable quick read.

Light_Tone A cute love story and a suspenseful mystery.

Light_Tone The funniest humor piece I've read in the last few years

Light_Tone His detailing of life's small moments, rife as they are with the potential for embarrassment, will leave readers laughing aloud.

Light_Tone If you like a razor sharp British wit, then this is the book for you.

Light_Tone The poems are so bright and cheerful.

Light_Tone The world of outdoor humor without McManus would be in a sad, sad state.

Light_Tone The moody, lyric writing about the landscape of Louisiana in the books rings true, delicious and delirious.

Light_Tone This book sets a tense and uneasy tone throughout.

Light_Tone She takes an advantage to work under cover in a plot that is both fascinating and hilarious

Light_Tone Her hilarious dialogue, and various "paranormal" touches mesh together to make one of the most hilarious YA books of the year.

Light_Tone Odgers has invented a funny situation, and wonderful characters to act it all out.

Light_Tone I found it ridiculously funny and overly exaggerated.

Light_Tone I could not contain my laughter.

Dark_Tone A handful of merely apocalyptic scenarios with practically zero chance of success

Dark_Tone This book is likely to frighten younger kids, with its depictions of frightened- and unhappy looking children being chased by snarling animals.

Dark_Tone they discover some surprising and frightening secrets

Dark_Tone This book provides a deep pessimism.

Dark_Tone Cam escapes his worst nightmare, only to find a more bizarre reality waiting for him.

Dark_Tone I recommend this book only if you are in the mood to read about three depressing and unhappy lives.

Dark_Tone The climax of the book is a horrifying murder-suicide.

Dark_Tone The novel begins with high hopes for it's main character, but soon turns into a never-ending landslide of tragedy

Dark_Tone It is just depressing and painful. Not to say you can't learn a lot from books that aren't pleasant.

Dark_Tone Hardy's ideas about the negatives of society and the harsh realities of the human condition are apparent throughout the novel.

Dark_Tone It is a book I would recommend reading once, but definitely not for amusement and only if you can handle 400 pages of gloom.

Dark_Tone He really pours on the tragedy.

Dark_Tone Lyra's world seems to be very dark, mysterious and secretive.

Dark_Tone This world Pullman has creative is a fantasy and mystical adventure.

Dark_Tone If there's one word to summarize the atmosphere around the book, it would be Dark.

Dark_Tone From start to end there's hardly a sliver of light that seeps through, understandable as there is an apocalyptic outcome that everyone is expecting.

Dark_Tone What I can say is that it's a brilliant dark fantasy book which combines our world with another.

Dark_Tone Still, as i said, the book goes into very dark, deep themes.

Dark_Tone People who enjoy a mysterious plot would enjoy this book because the plot is very exciting and stimulating.

Dark_Tone It gets even darker when Hank murders 10 knights with a pistol, then later when he and a handful of other men kill off 50,000 knights with high-voltage electric fences, land mines, and trenches with sharp spears sticking out on the bottom, only to almost die from the stench of the dead.

Dark_Tone This masterpiece is a stinging dystopian satire.

Dark_Tone But then, the whole story is Mark Twain's gloomy statement on Mankind's uneasy place in a dysfunctional world

Dark_Tone I don't think I have ever read anything so honest and stark in its emotional content.

Dark_Tone Dave does not hold back from writing his own manicdepressive

thoughts and goes into great detail on how he feels about certain events, even if it usually doesn't fit in the stereotype of the teenager on television, but one readers will be able to relate too even if they haven't gone through the same hardships Dave has gone through.

Dark_Tone I'm sure there are people with tragedies just as

heartbreaking, but with a less entitled outlook, out there for our sympathy and support.

Dark_Tone For anyone who isn't afraid to delve into the realities of dysfunctional, struggling

Dark_Tone It is dealing with some of life's heartaches.

Dark_Tone A Heartbreaking Work.

Dark_Tone If you read the book, look for the claustrophobic motif.

Dark_Tone It is funny and sad, gentle and harsh, full of hope and despair.

Dark_Tone A symbol of war and violence.

Dark_Tone It's just too disturbing.

Dark_Tone Karl is a victim throughout in a series of improbable adventures, and constantly struggles through a confused labyrinth of streets and buildings and random acts of cruelty and compassion.

Dark_Tone Rossman then 'disappears' into the poor working class landscape of America, where he encounters many less than scrupulous characters.

Dark_Tone The atmosphere of nightmare and difficulty that pervades Kafka's work was felt by me then as in absolute contradiction to the spirit.

Dark_Tone A KILLING NIGHT is a textured story told with a grace and nuance that belies the gruesome nature of the crimes with which it is concerned and the rather desperate circumstances of the main characters involved.

Dark_Tone But this one had me up all night long, crying in the dark after I turned out the light.

Dark_Tone How plausible are our nightmares.

Dark_Tone I think this book is atmospheric and frightening.

Dark_Tone I didn't realize that it would be the kind of book that would continue to disturb my thoughts even after I shelved it unfinished.

Dark_Tone This book produces nightmares.

Dark_Tone The ever-present aura of impending disaster.

Dark_Tone Tantalizing, horrifying and compelling book about two naives who've fallen into very perverse company.

Dark_Tone They're drawn into a web of psychological control, with disastrous results.

Dark_Tone It is extremely frightening.

Dark_Tone It includes some really twisted, frightening stuff.

Dark_Tone The "normal" lovers are drawn into the lives of a pair of perverts: a woman who loves to have extreme physical pain inflicted on her and a man who loves to inflict it.

Dark_Tone Things things become downright revolting.

Dark_Tone The initial "comfort" soon turns to uneasiness, and later to horror

Dark_Tone Astonishing stories from acclaimed author Ursula K. Le Guin carry us to worlds of wonder and horror, desire and destiny, enchantment and doom.

Dark_Tone While it is grim, it leaves the reader with a sense of hope, that beauty can come from tragedy.

Dark_Tone This story is the most grim and depressing, but don't let that discourage you.

Dark_Tone I found it quite depressing.

Dark_Tone Do not read it if your in a happy mood, because it will put you right off!!!

Dark_Tone Depressing but gripping reading.

Dark_Tone I read this book with a deepening sadness.

Dark_Tone This book is part of a series, but stands on its own with its own characters and dark atmosphere.

Dark_Tone Yet the story exudes passion and eroticism, in Aschebach's homosexual attraction for a young adolescent, the dank gondolas of Venice, the fetid epidemic that plagues the city, and the atmosphere of death and destruction that Mann captures in his work

Dark_Tone It was so beautifully laid out, the tone, the story, the dark deep themes - as a short novel, it stands with the best of literature.

Dark_Tone The author skillfully traces how this caused a violent pendulum swing in the story.

Dark_Tone Brides and grooms are being killed on their wedding nights. More than killed, the brides are being defiled by the murderer.

Dark_Tone The plot is a good deal deeper and darker.

Dark_Tone Future sequels will even waste more pages on illness, misery and suffering.

Dark_Tone It is both at once horrifying and irresistible.

Dark_Tone And there is plenty of shock: Burgess spares no detail while dictating the horrible reality of urban violence.

Dark_Tone The book is full of action; something is always going on whether it is a rape, a beating, or even burning down some one's home.

Dark_Tone It may sound sick and weird, but I enjoy reading books with plenty of action and some violence.

Dark_Tone The narrator gleefully partakes in some terrible actions.

Dark_Tone This plot will open up the minds of those with the emptiest imagination and show them a world of darkness.

Dark_Tone If you have a weak stomach or cannot tolerate descriptions of brutal violence or rape, I would like to suggest that you don't pick up this title.

Dark_Tone The plot involves murder, incest, and familial cruelty.

Dark_Tone An extremely sad narrative.

9.11 Vedlegg 10

CategorizeDocumentFactory.java

```
1 package com.nlpsolr.cat.tre;
23
import java.io.FileInputStream;
4 import java.io.IOException;
5 import java.io.InputStream;
6 import java.util.ArrayList;
7 import java.util.Collection;
89
import opennlp.tools.doccat.DoccatModel;
10 import opennlp.tools.doccat.DocumentCategorizerME;
11 import opennlp.tools.sentdetect.SentenceDetectorME;
12 import opennlp.tools.sentdetect.SentenceModel;
13
14 import org.apache.solr.common.SolrInputDocument;
15 import org.apache.solr.common.params.SolrParams;
16 import org.apache.solr.common.util.NamedList;
17 import org.apache.solr.request.SolrQueryRequest;
18 import org.apache.solr.response.SolrQueryResponse;
19 import org.apache.solr.update.AddUpdateCommand;
20 import org.apache.solr.update.processor.UpdateRequestProcessor;
21 import
org.apache.solr.update.processor.UpdateRequestProcessorFactory;
22 import org.slf4j.Logger;
23
24
25 public class CategorizeDocumentFactory extends
UpdateRequestProcessorFactory {
26 public static final Logger log =
org.slf4j.LoggerFactory.getLogger(CategorizeDocumentFactory.class);
27 SolrParams params;
28
29 public void init( NamedList args ) {
30 params = SolrParams.toSolrParams((NamedList) args);
31 }
32
33 @Override
34 public UpdateRequestProcessor getInstance(SolrQueryRequest
req, SolrQueryResponse rsp, UpdateRequestProcessor next)
35 {
36 return new CategorizeDocument(next);
37 }
38
39
Page 1
CategorizeDocumentFactory.java
40 public class CategorizeDocument extends
UpdateRequestProcessor {
41
42 public CategorizeDocument( UpdateRequestProcessor next)
{
43 super( next );
44 }
```



```

45
46 @Override
47 public void processAdd(AddUpdateCommand cmd) throws
IOException {
48
49 try{
50 SolrInputDocument doc =
cmd.getSolrInputDocument();
51
52 String inputField = params.get("inputField");
53 String outputFieldPace =
params.get("outputFieldPace");
54 String outputFieldCharacter =
params.get("outputFieldCharacter");
55 String outputFieldTone =
params.get("outputFieldTone");
56 String resultField = params.get("resultField");
57 String testField = params.get("testField");
58
59 String modelPace = params.get("modelPace");
60 String modelCharacter =
params.get("modelCharacter");
61 String modelTone = params.get("modelTone");
62 String nlpSentencePath =
params.get("openNlpSentenceModel");
63
64 Collection<Object> input =
doc.getFieldValues(inputField);
65
66 ArrayList<String> inputList = new
ArrayList<String>();
67
68 for (Object token : input){
69 inputList.add(token.toString());
70 }
71
72 String helSetning = inputList.toString();
Page 2
CategorizeDocumentFactory.java
73 String nySetning = helSetning.replaceAll("\\\[",
"").replaceAll("\\\]", "");
74
75 InputStream modelIn = new
FileInputStream(nlpSentencePath);
76 SentenceModel model = new
SentenceModel(modelIn);
77 SentenceDetectorME sentenceDetector = new
SentenceDetectorME(model);
78 String setninger[] =
sentenceDetector.sentDetect(nySetning);
79
80 InputStream isPace = new
FileInputStream(modelPace);
81 DoccatModel mPace = new DoccatModel(isPace);
82 DocumentCategorizerME catPace = new
DocumentCategorizerME(mPace);

```

```

83
84 InputStream isCharacter = new
FileInputStream(modelCharacter);
85 DoccatModel mCharacter = new
DoccatModel(isCharacter);
86 DocumentCategorizerME catCharacter = new
DocumentCategorizerME(mCharacter);
87
88 InputStream isTone = new
FileInputStream(modelTone);
89 DoccatModel mTone = new DoccatModel(isTone);
90 DocumentCategorizerME catTone = new
DocumentCategorizerME(mTone);
91
92 ArrayList<String> resultListPace = new
ArrayList<String>();
93 ArrayList<String> resultListCharacter = new
ArrayList<String>();
94 ArrayList<String> resultListTone = new
ArrayList<String>();
95 /* ArrayList<String> resultListAll = new
ArrayList<String>();*/
96
97 for (int i = 0; i < setninger.length; i++) {
98 double[] resultPace =
catPace.categorize(setninger[i]);
99 double[] resultCharacter =
catCharacter.categorize(setninger[i]);
Page 3
CategorizeDocumentFactory.java
99 double[] resultCharacter =
catCharacter.categorize(setninger[i]);
100 double[] resultTone =
catTone.categorize(setninger[i]);
101
resultListPace.add(catPace.getBestCategory(resultPace));
102
resultListCharacter.add(catCharacter.getBestCategory(resultChar
acter));
103
resultListTone.add(catTone.getBestCategory(resultTone));
104 /
*resultListAll.add(catPace.getAllResults(resultPace));
105
resultListAll.add(catCharacter.getAllResults(resultCharacter));
106
resultListAll.add(catTone.getAllResults(resultTone));*/
107 }
108
109 FinnFlest bestResult = new FinnFlest();
110 String paceToSolr =
bestResult.HentBesteKategori(resultListPace);
111 String characterToSolr =
bestResult.HentBesteKategori(resultListCharacter);
112 String toneToSolr =
bestResult.HentBesteKategori(resultListTone);

```

```
113
114
115
116
117 if (paceToSolr != null &&
bestResult.HentBesteKategori(resultListPace) != ""){
118 doc.addField(outputFieldPace, paceToSolr);
119 doc.addField(outputFieldCharacter,
characterToSolr);
120 doc.addField(outputFieldTone, toneToSolr);
121 /* doc.addField(resultField,
resultListAll.toString());
122 doc.addField(testField, nySetning);*/
123 //log.warn("kategorisering gikk bra");
124 }
125 else {
126 log.warn("ingen resultat");
127 }
128
```

Page 4

CategorizeDocumentFactory.java

```
129 }
130 catch(IOException e1){
131 e1.printStackTrace();
132 }
133 catch(Exception e){
134 e.printStackTrace();
135 }
136 super.processAdd(cmd);
137 }
138 }
139
140
141
142 }
143
```

Page 5

9.12 Vedlegg 11

FinnFlest.java

```
1 package com.nlpsolr.cat.tre;
23
import java.util.List;
45
public class FinnFlest {
67
    public String hentBesteKategori(List<?> liste) {
8     int size = liste.size();
9     if(size == 0)
10    return null;
11
12    int teller = 0;
13    int maksTeller = 0;
14    String kategori = (String) liste.get(0);
15    String besteKategori = kategori;
16
17    for(int i = 0; i<size; i++) {
18        if(liste.get(i).equals(kategori)) {
19            teller++;
20            if(teller > maksTeller) {
21                maksTeller = teller;
22                besteKategori = kategori;
23            }
24        } else {
25            teller = 1;
26        }
27        kategori = (String) liste.get(i);
28    }
29    return besteKategori;
30 }
31
32 }
33
```

9.13 Vedlegg 12

```
<doc>
<str name="edition">2</str>
<arr name="subject">
<str>Psychology - Dictionaries.</str>
</arr>
<arr name="numberofpages">
<str>320</str>
</arr>
<date name="publicationdate">2001-12-31T23:00:00Z</date>
<arr name="dewey">
<str>300</str>
</arr>
<str name="publisher">Routledge</str>
<arr name="user_rating">
<int>5</int>
[... ]
</arr>
<arr name="creator">
<str>Konrad Lorenz</str>
</arr>
<str name="ean">9780415283205</str>
<str name="title">On Aggression (Routledge Classics)</str>
<arr name="user_review_content">
[... ]
<str>
Is TV to blame for human violence? Some people believe so.
But in the countries with censored violence-free media the
crime rate is even worse. According to the author
aggression is innate, it evolved because environment made
it necessary for the survival of the species. Aggression
is also responsible for the development of personal
relationships: one has to know whom not to kill. If there
was no aggression there would be no biological need for
friendship. The author proves his explanation of human
behavior using numerous observations on various animal
species, including the species that exhibit aggression and
the ones that don't.
</str>
[... ]
</arr>
<str name="isbn">0415283205</str>
<str name="category_pace">Slow-paced</str>
<str name="category_character">Complex_characters</str>
<str name="category_tone">Dark_Tone</str>
</doc>
<doc>
```