

---

# **INFO221 – Multimediahåndtering**

## **Obligatorisk oppgave 2, Våren 2009**

---

**Regine O. Sagstad, Ingrid Nese og Marianne Midbø (Gruppe: RIM)**

## Innhold

1.1 Innledning.....	3
1.2 Antagelser.....	3
1.3 Visjon og mål .....	4
1.4 Brukergrupper og tilknytninger .....	4
1.5 Case 1 .....	5
1.6 Case 2 .....	5
2.1 Systemoppbygging .....	7
2.5 Relevans.....	12
2.6 Søk .....	12
2.7 Tagger .....	14
2.8 Kvalitetssikring.....	15
2.9 Indeksering .....	16
3.1 Datavisualisering .....	16
3.2 Integrering med andre systemer.....	17
3.3 Åpen standard .....	17
4.1 Diskusjon .....	18
5.1 Konklusjon .....	18
6.1 Kildehenvisning.....	20
7.1 Inspirasjonskilder.....	20
8.1 Vedlegg.....	21

## **1.1 Innledning**

Vi skal lage et system som skal hjelpe Kings College i Storbritannia å knytte potensielle forskere og studenter sammen. Det skal være en nyskapende og kreativ informasjonsgjenfinningsteknologi, der målet er å samkjøre forskning og konstellasjoner av studenter og forskere. Dette skal vi løse med å lage modeller og illustrasjoner til en funksjonalitet som samler relevant informasjon. Gjennom systemet, som vil være en nettside, skal man som forskere og studenter kunne publisere og søke etter relevante dokumenter fra andre forfattere. Resultatene fra et søk skal ha likhetstrekk med egenproduserte dokumenter og baseres på fag- og interesserelatert informasjon. For at dataen som blir vist skal være relevant for brukeren og av god kvalitet, skal systemet ha funksjoner som gir mulighet for favorisering av dokumenter og forfattere, tagging og sammenlikning av dokumentinnhold og brukerinteresser. Informasjonen skal bli visualisert på en måte som gjør det enkelt for brukeren å finne det han leter etter.

I dag legger mange forskere og studenter selvskrevede dokumenter ut på internett. For å gjøre denne informasjonen best mulig tilgjengelig for en søker, kan vi bruke et elektronisk informasjonssystem (IR) som forenkler lokaliteten og viser dokumenter som er relevante for den som har søkt etter informasjonen. (Nordbotten, 2008) Vi tar i bruk et IR-system for å gjøre det enkelt for forskere og studenter å få tilgang til denne informasjonen på en visuell og oversiktlig måte. Siden dokumentene kan være ulike medieplattformer som for eksempel bilde, lyd, tekst og video, må informasjonssystemet vårt ha mulighet til å motta informasjon fra varierte medierepresentasjoner.

## **1.2 Antagelser**

Vi antar at dette skal være et verdensomspennende system og ikke bare et system for Kings College i Storbritannia. Dette mener vi er den beste løsningen for Kings College, ettersom de da vil kunne dele kunnskap på tvers av landegrensene. Dermed antar vi at det som blir lastet opp i systemet sin database er skrevet på engelsk. Systemet vil også gi mulighet for å søke innenfor stedet man tilhører. På den måten vil Kings College få et system med muligheter for kommunikasjon og søking både internt og internasjonalt.

Når det gjelder brukere, antar vi at en person ikke kan være både forsker og student samtidig. En student er kun en potensiell forsker.

### **1.3 Visjon og mål**

*Vi vil gi forskere og studenter muligheten til å dele selvlaget materiale, finne andres materiale og kommunisere på tvers av landegrensene. Resources Community vil organisere all verdens dokumentert forskning i et stort og åpent system.*

Som visjonen sier ønsker vi å hjelpe forskere og studenter til å kunne kommunisere med andre innenfor samme fagfelt, i tillegg til å kunne lese og skaffe seg kunnskap fra andre deler av verden. Noen universitet forsker gjerne mye på et fagområde, mens et annet universitet forsker innenfor andre områder. Ved å benytte seg av vårt system vil vi gjøre det enklere for disse å dele de erfaringene og den kunnskapen de ulike universitet og personer har. I tillegg til å lese artikler innenfor samme fagfelt ønsker vi å gi alle muligheten til å utforske andre fagfelt og tilegne seg kunnskap gjennom å lese artikler publisert av eksperter på området. Dette vil gjøre systemet lærerikt for forskere og studenter, både innenfor deres fagfelt, men også på andre plattformer. Målet vårt er at forskere og studenter skal få den kunnskapen de er interessert i, oppdage ny kunnskap de ikke visste de var interessert i og knytte nettverk med interessante personer fra hele verden.

### **1.4 Brukergrupper og tilknytninger**

Av data vil vi at forskerne skal kunne finne dokumenter fra ulike forfattere. De skal også få metadata fra artiklene i form av fra hvilke tema de omtaler, tittel, forfatter, relevans, størrelse og publikasjoner. Organiseringen av disse vil være visualisert i form av et interaktivt system der brukeren kan navigere seg gjennom et informasjonsunivers bestående av fargekoder, former og relevansstørrelser.

Systemet skal også være tilrettelagt for å knytte studenter sammen. Dette skal være studenter som har like interesser eller studerer samme fag. De vil ha bruk for systemet for å finne relevante artikler til egne oppgaver, eller kommunisere med andre om deres fagfelt. For å avgjøre om en student er lik en annen student vil en profil kreve at personens interesser blir lagt til som favoritttagger. Studenten vil da få opp forslag på relevante dokumenter som er

publisert av forfattere med samme interesser. Det skal også være mulig for studenten å finne samarbeidspartnere til



grupperarbeid dersom de er lokalisert på samme universitet. Lokalisering og studiested må dermed også registreres.

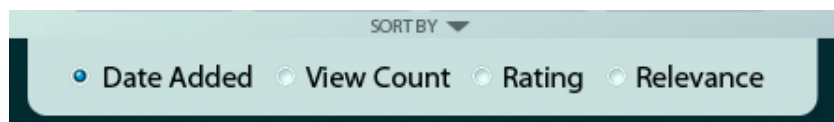
### **1.5 Case 1**

Mona er forsker. Hun driver med utvikling av en teknologi for "Wearables" også kalt "smarte klær". Hun ønsker å finne artikler som omtaler denne type teknologi, i tillegg til å få kjennskap til andre forskere som driver med det samme. Hun benytter seg av "Resources Community" for å finne den informasjonen hun er på jakt etter. Mona søker på "Wearables" og får da opp et verdenskart med ulike artikler som omtaler dette emnet. De artiklene som inneholder emneordet flest ganger, i form av tagger, kommer først ettersom det er de artiklene sannsynligheten er størst for at handler om nettopp dette. Mona finner noen interessante artikler som er skrevet i USA, men hun er litt usikker på hvilke artikler som er mest relevante. Hun går derfor inn på noen av artiklene og skummer raskt gjennom teksten. Det tar kort tid før Mona skjønner hvilke artikler som svarer til det hun var på jakt etter. Hun velger å laste ned de artiklene, samtidig finner hun to av forfatterne veldig interessante. De jobber med å utvikle ulike prototyper som Mona syns er svært interessante og som hun vil vite mer om når utviklingen fortsetter. Hun velger derfor å la disse to forfatterne bli hennes favoritter for å få oppdateringer når de publiserer nye artikler. Hun kan da følge med i utviklingen av prototypene, lære mer om denne type teknologi og lese mer om deres meninger og tanker. En av artiklene som omfattet et av Monas pågående problemområder, ønsket hun å kunne ta frem i senere anledninger. Mona velger derfor å la denne artikkelen også bli en del av hennes favoritter. Slik kan hun enkelt ta den frem når hun måtte ønske, uten å måtte søke artikkelen fram på ny.

### **1.6 Case 2**

John er medisinstudent og veldig interessert i faget og utdanningen sin. Han har nå studert i fem år og gleder seg til å begynne å praktisere utdanningen sin. I løpet av studieårene har han funnet ut at det blir forsket og skrevet mye interessant innenfor fagfeltet. Naturligvis er det ikke alt han får med seg gjennom forelesninger og pensum. Han prøver derfor å finne

interessante artikler andre plasser for å lære enda mer om sitt fremtidige yrke. For John er det viktig å være oppdatert på flere områder innenfor medisin ettersom han vil bli allmennpraktiserende lege. Allmennpraktiserende leger møter pasienter med ulike symptomer og behov hver eneste dag. Det er krevende å ha kunnskap innenfor alle medisinfagfeltene og utfordrende å stille helt forskjellige diagnoser i løpet av en dag. John mener derfor det er viktig å ha et bredt kunnskapsfelt innenfor all forskningen som blir gjort. Det er tidkrevende og vanskelig å oppdatere seg innenfor forskningstemaene. John vet verken hva han skal søke på, eller hva som er mest interessant og viktig å bruke tid å lese. Derfor benytter John seg av "Resources Community". Her kan han enkelt finne interessante artikler selv om han ikke vet hva han eksakt leter etter. John klikker seg inn på fagområdet medisin og deretter velger han temaet neurologi som han har lyst å lære mer om. John vet ikke mer om neurologi enn det han har hatt på studiet, men han vet at dette er noe han trenger kunnskap om som allmennlege. En ny side viser bokser med ulike tagger som omhandler neurologi. Størrelsen på boksene varierer etter hvor mange dokumenter som er publisert. De viktigste sykdomsgrupper hvor det er knyttet høy basal og klinisk forskningsaktivitet til og som det er publisert flest artikler om er multipel sklerose, mitokondriesykdommer, paraneoplastiske neurologiske sykdommer, hjerneslag og myasthenia gravis og Parkinson sykdom. John mener at han som allmennlege bør ha mer kunnskap om hjerneslag og velger derfor denne boksen. Da kommer han inn på verdenskartet som viser artiklene som omhandler temaet. I tillegg til plasseringen på verdenskartet har John mulighet til å se de siste publiserte artiklene, de som er mest favoriserte, de mest leste eller de som er mest relevante, i forhold til



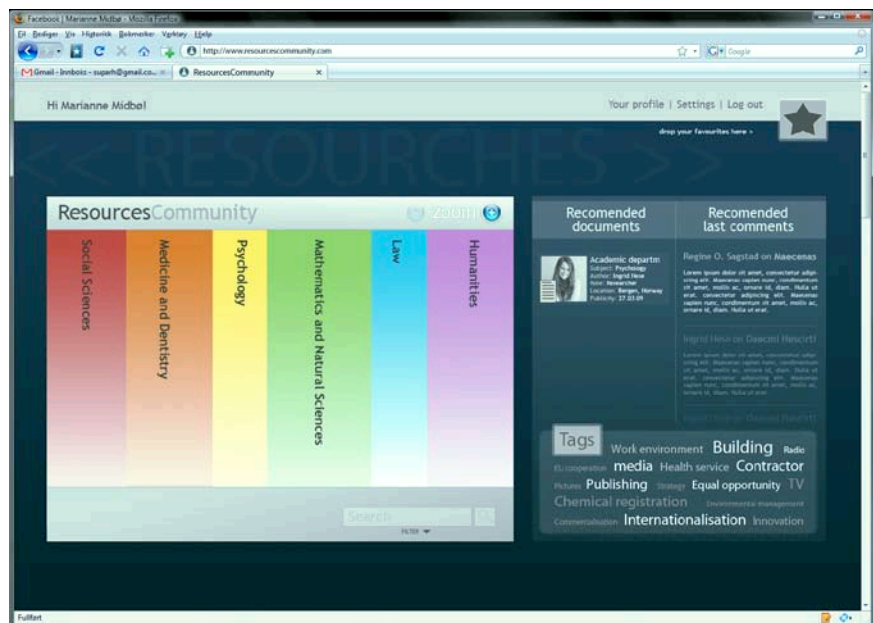
hvilke som inneholder ordet flest ganger. Ettersom John er interessert i å anskaffe seg generell kunnskap om temaet regner han med at han vil få det ved å gå inn på de artiklene som er mest favoriserte ("Rating"). Han antar at artiklene er gode, har mye nyttig informasjon og er relevante innenfor det gitte fagområdet. Det er ikke første gangen John har vært på kunnskapsleting i "Resources Community", han gjør det regelmessig. Det er en fin måte å skaffe seg mer kunnskap på. Derfor blir det mindre tidkrevende når han har mulighet til å lese de mest favoriserte artiklene, i stedet for å lete i en verden av artikler og informasjon.

## 2.1 Systemoppbygging

For lettere å kunne forklare hvordan systemet vårt fungerer vil vi nå beskrive, side for side, hvordan funksjonene og visualiseringen fungerer.

Inngangssiden er en innloggingstjeneste der brukeren første gang må registrere personlige opplysninger som navn, alder, e-postadresse og lokalisering. I tillegg må man oppgi hvorvidt man er forsker eller student og hvilken institusjon man holder til. Brukeren må også oppgi interesser og fagfelt, og har muligheten til å laste opp et profilbilde. Forskerne må også oppgi stillingstittel. Det er mulig å hoppe over registreringen og bare benytte seg av søkefunksjonene, men da vil mye av funksjonaliteten forsvinne. Systemet har vanskeligere for å hjelpe deg til å finne de artiklene som er mest relevant for deg, i tillegg til at man mister oppdateringstjenestene, favoriseringer og muligheten til å publisere egne artikler.

Når innloggingen er unnagjort kommer man inn på hovedsiden. Toppmargin inneholder muligheter for å endre innstillinger, gå inn på profilen, og legge til noe i dine favoritter. Denne vil være tilgjengelig til en hver tid. Det samme vil to spalter, til høyre, med oppdateringer være.

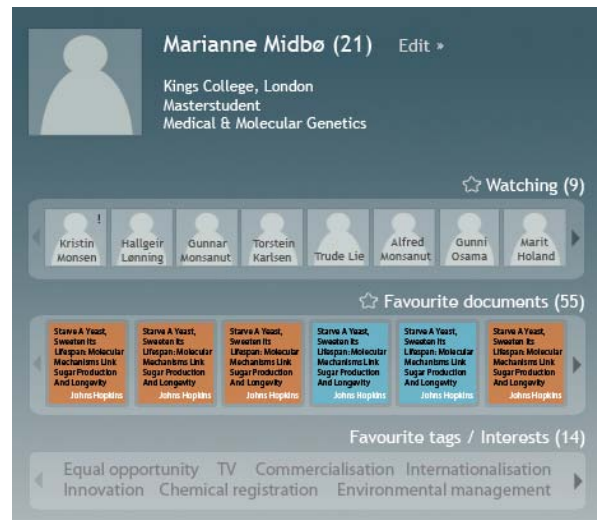


Systemet vil være lik for både studenter og forskere, utenom at forskerne i utgangspunktet ikke får opp studentarbeid når de bruker søkefunksjonen. På innstillinger kan forskerne endre på dette slik at søkefunksjonene også innebærer studentarbeid. Grunnen til at denne begrensingen finnes i søkefunksjonen, er at vi anser det som mest relevant for forskere å få opp publikasjoner gjort av andre forskere. Studentarbeid vil i sjelden grad tilføyne noe for

forskerne, men for studentene vil det absolutt være en fordel å kunne anskaffe seg kunnskap fra publikasjoner gjort av forskere.

## 2.2 Personlig side

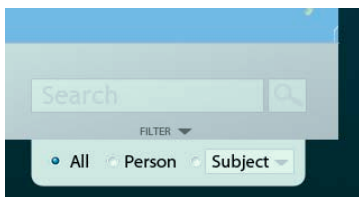
Profilen er brukerens personlige side som i tillegg til å vise personlig informasjon har en oversikt over brukerens publiserte artikler. Det er også mulighet for å legge til artikler, forfattere og tagger som favoritter. De vil være synlig på profilsiden. Favoritter kan til enhver tid slettes fra siden. På profilen har forskere og studenter også muligheten til å publisere nye



artikler og studentoppgaver. Man må da oppgi relevante tagger for det bestemte dokumentet før man publiserer. Disse taggene skal gjøre det enklere for andre interesserte forskere og studenter å finne denne publikasjonen. På profilen kan man altså kort fortalt publisere nye artikler, vise de tidligere publiserte artiklene og ha en oversikt over favoritter. Brukeren har ikke mulighet til å publisere et enkelt bilde, ettersom det allerede finnes gode bildeportaler til dette formålet, men tekst som inneholder bilder er mulig å publisere. Publikasjonene må være i formatene pdf, avi, mpg eller swf.

## 2.3 Søking i informasjonsuniverset

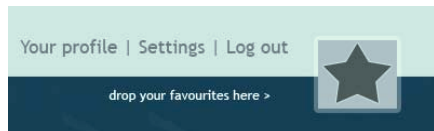
Brukeren har to valg når han vil finne artikler innenfor et spesielt område. Den ene muligheten er å søke i søkefeltet på det emne han ønsker å finne artikler om, søkingen vil da ha ord som er relatert til hverandre slik at et "dårlig" søk skal gi treff som søkeren sannsynligvis er interessert i. I tillegg til å skrive emneordet inn i søkefeltet kan brukeren velge et av fire alternativer for å spesifisere søket. Forhåndsinnstillingen er "All", da vil



brukeren få opp alt systemet finner av treff på søkeordet. Dersom brukeren ikke gjør noen endringer vil systemet søke gjennom alt av publikasjoner og informasjon for å finne likhetstreff med søkeordet. Brukeren kan også velge "Person", da vil søket kun treffe personer. For å få et begrenset søketreff fra lokaliseringen man tilhører, velges "lokalisering" (ikke skissert på bildet). Den siste muligheten er en combobox som viser

de ulike fagområdene slik at systemet bare søker gjennom det valgte fagområdet. En tagg kan ha ulike betydninger innenfor ulike fagfelt, ved å benytte denne søkemetoden vil man da unngå urelevante artikler.

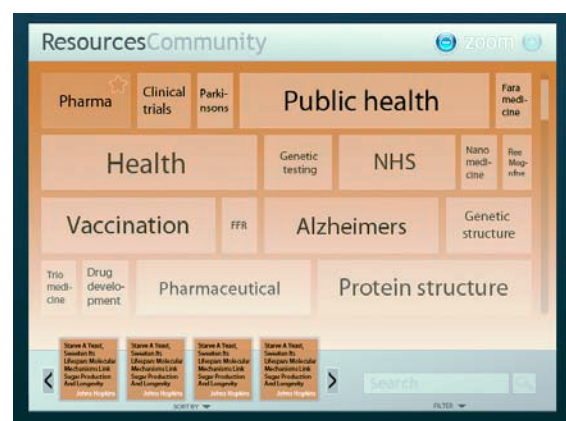
Søker man etter en person i søkefeltet vil man komme inn på personens personlige side. Her har man mulighet til å lese artiklene han har publisert og se hvilke favoritter han har i form av



forfattere, artikler og tagger. Både forfatteren, artiklene og favorittene kan gjøres om til dine favoritter. Dersom man ønsker å legge til forfatteren eller noen av artiklene som sine favoritter kan man enkelt dra disse opp til stjernen

som ligger i toppmargen, og det vil bli en del av dine favoritter. Har man valgt forfatteren som en av sine favoritter vil man få oppdateringer når forfatteren publiserer nye artikler. I tillegg vil man i senere søk få opp artikler skrevet av denne forfatteren som relevante.

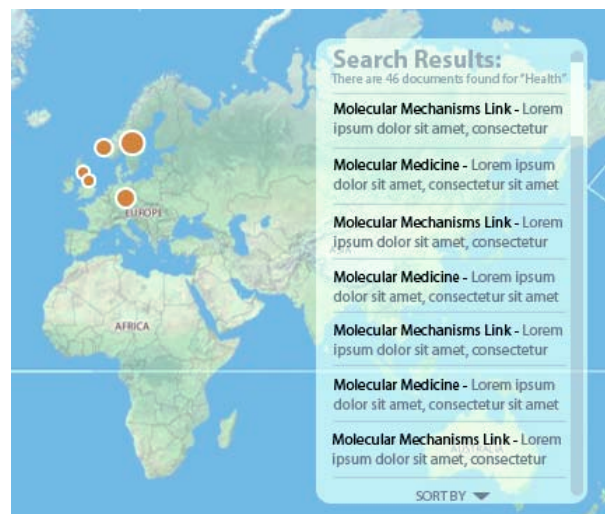
Det andre valget brukeren har er å bruke fagkolonnene og trykke seg inn på det fagområdet han er interessert i. Man vil da komme videre til en underside med ulike fagfelt (rangert A-Z) innenfor dette fagområdet. Denne siden vil ha en fargenyanse basert på fagområdet sin farge. Har man for eksempel gått inn på fagområdet medisin, som er oransje, vil fagfeltene være ulike nyanser av oransje. Ved bruk av farger visualiserer vi på en god og enkel måte for brukeren hvilket fagområde han befinner seg i. Etter å ha valgt et bestemt fagfelt vil man komme inn på en ny side. Siden vil bestå av bokser som inneholder tagger relatert til fagfeltet. Størrelsen på boksene markerer hvor mange dokumenter som finnes innenfor hver enkel tagg. Er det mange publikasjoner om en tagg vil denne boksen være større enn boksen til en tagg det er mindre publikasjoner om. Fargene på boksene vil være samme fargenyanse som fagområdet, men styrken på fargen fortelle hvor relevant de ulike taggene er, ut ifra brukerens registrerte interesser. Hva har



brukeren søkt på før? Hvilke artikler har han som favoritter? Hvilke tagger er registrert som favoritter og hvilke interesser har han? Dette vil være med å avgjøre hvilke bokser som er sterke i fargen og som da regnes som mer relevante enn andre bokser. De vil også være rangert først. Dersom man ønsker å legge til en av taggene som favoritter kan man som nevnt enkelt dra boksen opp til stjernen som ligger i toppmarginen, og taggen vil da bli lagt til i dine favoritter.

## 2.4 Emnevisualisering

For å komme videre fra taggesiden må man velge én, eller flere av taggeboksene. Ved å velge flere taggebokser begrenser man søket og det vil være større sjans for å finne de artiklene som er mest relevante. Fra taggesiden kommer man videre til et kart. Her er det en oversikt over publiserte artikler som inneholder de gitte taggene, over hele verden. Verdenskartet viser hvor i verden artiklene er publisert, og er rangert etter hvilke artikler som er mest relevant for deg. De artiklene som inneholder taggene flere ganger, vil komme før de artiklene som bare inneholder taggene et par ganger. På denne måten vil det være større mulighet for at artiklene nettopp handler om det man er interessert i å finne. Brukeren kan også selv velge hvordan artiklene skal rangeres, enten etter de sist publiserte artiklene, de mest leste artiklene, de som er favorisert mest, eller de av mest relevans (basert på hvilke som inneholder det definerte ordet flest ganger). Dersom mange artikler med de gitte taggene er publisert fra samme plass vil de mest relevante være synlige, for å se de andre må man trykke inn på det bestemte området. Ved søk kan det være vanskelig å få hele oversikten over søketreffene på verdenskartet, derfor vil det i tillegg komme opp enn informasjonsboks, som lister



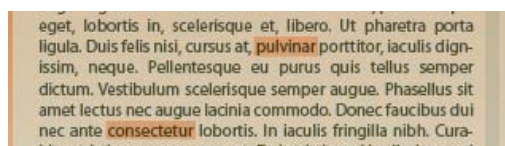
søketreffene opp i en liste og viser antall søketreff. Informasjonsboksen og verdenskarte vil gi samme informasjon, men ved å bruke begge måter å illustrere informasjon på håper vi at systemet kan nå ulike brukerønsker. Informasjonsboksen vil være enkel å fjerne.

Dersom man trykker på området for å se flere artikler fra samme sted, eller trykker på landet/byen en forfatter er fra, får man se et utsnitt av verdenskartet. Her vil forskningsinstitusjoner og universitet vise, med publikasjoner derfra.

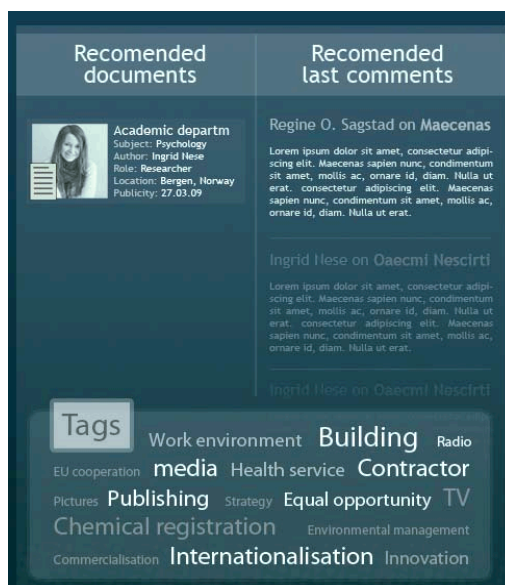
Når man trykker på et dokument vil man i første omgang få se dokumentet og informasjonen om dette i form av metadata. Dersom man mener dokumentet bør inneholde flere tagger enn det forfatteren har lagt inn, kan man foreslå andre relevante tagger for dokumentet. Dette må forfatteren godkjenne før taggen eventuelt blir lagt



til. I tillegg til selve dokumentet og informasjonen vil det også være mulig å se kommentarer andre har lagt inn. På den måten kan forskere og studenter kommunisere med andre som også er interessert i samme fagfelt. Her blir det mulighet for å komme med innspill og meninger, som et diskusjonsforum. Taggene vil være uthevet med farge i teksten. Finner man ut at teksten ikke er relevant, kan man gå ut av siden og komme tilbake til kartsiden. Er dokumentet interessant kan det lastes ned og/eller legges til i favoritter.



I tillegg til de ulike sidene vi har beskrevet nå, vil det, som nevnt, være to spalter som alltid er synlige. Disse spaltene skal gjøre det lettere å finne artikler og personer som er relevante i forhold til brukerens tidligere nedlastninger, favoritter og interesser. Slik vil vi at systemet ikke bare skal hjelpe studentene og forskerne til å oppsøke og finne relevante publikasjoner, men også få oppdateringer om dette kontinuerlig når det blir publisert nye interessante publikasjoner.



Den ene spalten skal oppdatere om nylig publiserte

dokumenter som brukeren kan være interessert i ut fra interesser/favoritt-tagger. Den andre spalten vil oppdatere med relevante kommentarer. Dersom man har lagt til kommentarer på et dokument, vil andre kommentarer fra samme dokument dukke opp i denne spalten, uavhengig av hvilket fagområde dokumentet tilhører. Her vil det også vises kommentarer fra personer man har favorisert.

## **2.5 Relevans**

Et dokument kobles til et annet dokument ved at en algoritme gjennomløper teksten, ord for ord, for å lete etter tekst som befinner seg begge steder. Her vil hyppig brukte ord som, ”og”, ”i”, ”en” osv unngås. Denne løsningen ville mest sannsynlig fungert dårlig med dagens CPUer, men vi ønsker likevel ikke å avgrense denne funksjonaliteten, da oppgaven tilsier at vi skal abstrahere bort overflødig teknisk informasjon.

For hvert ord i dokumentet

    Loop gjennom hvert ord i dokument X

    Loop gjennom hvert ord i dokument Y

        Hvis ord X == ord Y

            dokumentrelevans ++

totalsum = dokumentsum / dokumentlengde

## **2.6 Søk**

I vårt system bruker vi termvekting i søkefunksjonen vår. Systemet søker gjennom og ser hvor mange ganger taggene brukes i dokumentet. Slik får vi bedre kvalitet på svarsettet. Dokumenter som er lik forespørselen returneres og rangeringsformelen sorterer i forhold til søketerm. For å bruke vektning innad i dokument bruker vi metadata, ettersom metadata gjør det enkelt med gjenfinning. (Personlig kommunikasjon T. Gjørseter 11.3.2009) Metadata er data om data, forfatterne som bruker systemet vårt må oppgi informasjon om artikkelen som i seg selv er data, og blir definert som metadata. Nedenfor er pseudokoden for termvekting. Systemet vil gå gjennom hvert ord i dokument x. Dersom ordet i dokumentet er ekvivalent med ordet som er blitt søkt på (Y), skal taggtelleren for Y økes slik at man vet at det ordet ofte blir søkt på. Gjeldende tagg vil bli mer populær fordi den får et større antall.

For hvert ord i dokument X  
Hvis ordet er ekvivalent med Y  
Øk taggtelleren for Y

Det er brukerne som gir publikasjonene tagger som blir brukt ved søking, og vektingen skjer derfor manuelt. Vi mente det var den beste løsningen ettersom det er brukerne selv som skal søke, og publisere verk i systemet. Med en manuell vekting unngår man at en liten gruppe mennesker bestemmer definisjoner og hva som er hva når det er mennesker fra hele verden som skal bruke systemet. Det er en vanskelig problemløsning å finne ut hvordan man skal definere noe, alle har sin oppfatning av virkeligheten. Men ved å la alle få ta del i vektingen unngår man problemet i stor grad, og vi kan bruke mulighetene vi får med ontologi. Vi kan definere en ting i forhold til noe annet. I vårt system bruker vi det ved at brukeren får opp anbefalinger som er relatert til metadata som er registrert på profilen, interesser og favoritter.

Metadata, data og indekser henger sammen til en helhet slik at det blir enkelt for brukeren å søke. Hvert dokument inneholder informasjon som er gitt av forfatteren, og dokumentene med lignende metadata kan dermed kobles sammen ved å søke gjennom indeksen som består av all data. Når brukeren da søker på et emneord (i form av en tagg) vil systemet søke gjennom indeksen og finne alle publikasjoner som inneholder metadata. I søket registrerer systemet hvor relevant artiklene er i forhold til hvor mange ganger dokumentet inneholder taggen du søkte på. Dersom forskjellen på antall tagger i artikkelen ikke er så stor, vil systemet rangere dokumentet som er favorisert flest ganger først.

#### **Søk på et dokument som inneholder et ord**

Søk etter dokument med tagg X  
Dersom X er flere ganger i DokumentY enn i dokumentZ  
Gi ut Dokument Y

AND

Dersom taggforskjellen er mindre enn 5  
AND Dokument Z er favorisert flere ganger enn Dokument Y  
Gi ut Dokument Z

#### **Søk innenfor en lokalisering:**

Søk i lokasjon A etter dokument med tagg X  
Dersom X er flere ganger i DokumentY enn i dokumentZ  
Gi ut Dokument Y

AND

Dersom taggforskjellen er mindre enn 5  
AND DokumentZ er favorisert flere ganger enn Dokument Y  
Gi ut Dokument Z

#### Søk innenfor et fagområde:

Søk i fagområde A etter dokument med tagg X  
Dersom X er flere ganger i DokumentY enn i dokumentZ  
Gi ut Dokument Y

AND

Dersom taggforskjellen er mindre enn 5  
AND DokumentZ er favorisert flere ganger enn Dokument Y  
Gi ut Dokument Z

## 2.7 Tagger

En tagg er et ikke-hierarkisk system som blir brukt i form av nøkkelord eller terminologier. Forskere og studenter kan tagge ulike "items" og denne samlingen med taggs blir det vi kaller en folksonomi (Wikipedia 2009b). Tagger blir brukt i sammenheng med informasjonen som er lagt ut i forbindelse med dokumenter. Både studenter og forskere kan legge ut dokumenter, og når de gjør det må de tagge det. Eksempel på det er at hvis en forsker legger ut en artikkel om medisin, kommer det opp en liste med relevante ord som han kan velge å tagge artikkelen sin med. Flere tagger gir søkere større mulighet til å finne det de leter etter.

I oppgaven vår har vi brukt tagger hyppig for å få mer spesifiserte søk. Og for å gjøre det enklere å finne frem. Neste gang studenter, eller forskere skal søke etter et dokument med navn, men ikke finner det igjen, kan de søke på taggene (altså nøkkelordene), eller ord som finnes i dokumentet, og finne det raskere igjen. I tillegg til tagger og ordsøk, kan brukerne legge til sine favorittartikler. Hvis man trykker på profilen sin, vil man se hva man har lagt til.

På den første siden man kommer til etter å ha logget inn vil det være en dynamisk boks nederst i høyre spalte med de taggene som er klassifisert etter hvilke emner man er inne på. Et eksempel er at boksen med tagger vil forandre seg hvis



en bruker en inne på medisin, men klikker seg videre til nevrologi. Tagger med relasjoner til hverandre vil forekomme.

Et viktig aspekt av folksonomi er at det inkluderer termer i et flatt navnerom. Det vil si at det ikke er noe hierarki, som nevnt over. En form for direkte parent: child form er det heller ikke. Men en automatisk generert relasjonstagger. (Wikipedia 2009b). Dette baserer seg på hvilke ord som har blitt tagget av brukerne og i hvilken sammenheng. Ulempen med dette taggesystemet er at ulike brukere kan legge forskjellige meninger i det de tagger. De kan bruke samme ord, for eksempel, medisin, men ha helt ulike oppfatninger av hva medisin er, eller ulike sammenhenger. Likevel vil brukeren som søker på "medisin" få opp begge resultatene og det kan skape forvirring, og uønskede resultater på søket. Vi tillater mellomrom i taggene, på samme måte som Flickr. Forholdene mellom store og små bokstaver vil ikke bli tatt hensyn til, av den enkle grunn at det kompliserer taggingen.

I systemet med tagger er det ingen synonym – kontroll. Det vil si at "mac", "macintosh", og "apple" vil være tre ulike tagger og vil ofte dukke opp i lag. På samme måte vil det være med de grammatiske formene. Data, datamaskiner og lignende vil også kunne dukke opp samtidig, selv om det er unødvendig og tar opp plass til andre relevante tagger. Likevel er det vokabularet til de ulike brukerne som er i sentrum, og dette er positivt. Forhåpentligvis vil dette systemet appellere til forskere og studenter som bruker mye av de samme terminologiene, ordene og uttrykkene.

## ***2.8 Kvalitetssikring***

For at brukerne skal få opp de beste, og mest relevante dokumentene, er det ulike faktorer som spiller inn. Faktorene forekommer i form av ulike farger, størrelser og rangeringer. Alle som har en profil vil kunne dra innhold (både dokumenter, tagger og forfattere) opp til et favorittsymbol i form av en stjerne. Denne funksjonen blir brukt på blant annet nettsamfunnet Digg.com i dag. Brukere kan "digge" noe. Dette bidrar til å kvalitetssikre dokumentene, ut i fra brukernes egne meninger og synspunkter. Dersom et dokument blir favorisert av mange brukere, vil dokumentet fremstå som et dokument av god kvalitet.

## **2.9 Indeksering**

De forskjellige søkemotorene vi har på nettstedet er tekstbaserte. Innholdet som man da søker etter må være tekst for å kunne indekseres.

I praksis går dette ut på å lage en oversikt over alle ord i et dokument med referanse, enten bare til hvilket dokument de ligger i (invertert file index). Eller både hvilket dokument de ligger i samt hvor i dokumentet de ligger (full inverted index). Dette vil databasen som ligger bak vår nettside gjøre automatisk (Wikipedia, 2009d).

Indekseringen inneholder stemming, som går ut på å redusere ord til deres ordstamme. Et eksempel på det er at datamaskiner reduseres til datamaskin. Stemming utfører man for å gjøre søk mer fleksible, slik at man ikke må søke på nøyaktig den samme bøyningen av et ord for å finne treffe relatert til ordet.

## **3.1 Datavisualisering**

Hovedmålet med datavisualisering er å formidle informasjon klart og effektivt gjennom grafiske virkemidler. Det betyr at ideene må formidlest effektivt, både i estetisk form og funksjonalitet. Det er viktig å oppnå den rette balansen mellom design og funksjon. Det er ingen poeng i å ha flotte datavisualiseringer dersom den ikke klarer å tjene formålet; å formidle informasjon (Wikipedia, 2009a). Vi har prøvd å finne denne balansen, og håper at brukeren, bare ved å se på informasjonsuniverset skal skape en forståelse og innsikt i informasjonen. Ved å bruke farger og størrelser til å gi informasjon om mengde og relevant stoff, håper vi å oppnå denne effekten. I tillegg vil brukeren oppfatte hvor i verden de ulike dokumentene er publisert ved å se på kartet. Det vil gi brukeren en form for informasjon. Er det spesielle tagger det er skrevet mye om på en kant av verden, kan det fortelle noe om aktualiteten for taggen, hvordan samfunnet er der og lignende.

Det finnes mange måter og illustrere informasjon på. Data kan presenteres både elegant, enkelt og beskrivende. Den mest vanlige måten å illustrere informasjon på er ved hjelp av lister. Men det finnes mange ulike varianter og visualisere data på – diagram, tabeller, søylediagram, sirkeldiagram, eller tidslinje er noen eksempler (Wikipedia, 2009a).

### ***3.2 Integrering med andre systemer***

Vi vil at systemet vårt skal integrere med andre system som brukerne allerede benytter seg av. Derfor vil vi at det skal være mulig for brukeren, om man ønsker det, at publiserte artikler også blir publisert på andre nettsteder som for eksempel Facebook. Dette kan forfatteren gjøre i det han publiserer et nytt dokument på "Resources Community". Andre brukere kan også gjøre det når de leser en artikkel de synes er interessant og vil dele med venner de har på nettsamfunn som Facebook. Systemet vårt gir studenter og forskere en unik mulighet til å søke og finne interessante artikler fra hele verden. Mange av våre brukere vil være aktive brukere på ulike sosiale nettsteder og vi ser det som positivt at brukerne kan få muligheten til å dele artikler han finner med andre venner og bekjente. Akkurat som de i dag har muligheten til å dele nyhetsartikler, quizer og underholdning, vil de da også få mulighet til å dele interessante forsknings- og studentarbeid.

Vi har vurdert å bruke Open id, som er en meget nyttig og velbrukt ressurs. Å bruke en digital id, slik at du bare trenger å logge deg på én plass når du er på internett er en tidsparende funksjon. Problemet i vårt tilfelle er at man må kunne kvalitetsjekke at det enten er en forsker, eller en student som logger seg på nettstedet. Kontroll av dette er nødvendig for å opprettholde en seriøsitet og relevans innenfor fagplattformen. Derfor har vi valgt at brukerne må logge seg på med sine utdelte brukernavn fra organisasjonen de tilhører.

### ***3.3 Åpen standard***

Åpne standarder er spesifikasjoner på hvordan et fritt tilgjengelig format er oppbygd. Vi ønsker å bruke åpne standarder i vår løsning slik at vi kan utveksle informasjon med andre programmer. På denne måten sikrer man konkurranse ved at brukerne ikke blir låst til en bestemt programvare eller en bestemt leverandør (wikipedia 2009c). Standarden vår vil bli publisert og det vil være gratis å laste opp dokumenter og lignende. Det som kreves er en brukerkonto for å bevise at man er forsker, eller student. Fordelene ved at det er en åpen standard er at den er tilgjengelig utbredt over hele verden. Med en verdensomspennende utstrekning er det ideelt for oss at dette er en ressurs som kan være åpen og tillatt for alle som studerer eller forsker. Hadde dette vært en ressurs som man måtte betale for, er det vanskelig å se for seg at nettsiden hadde hatt potensial til å vokse seg stor nok til å bli

verdensomspennende. Hadde vi privatisert og lagt på en avgift ville vi avgrenset omfanget. Resources Community vil være velegnet til å være en åpen standard.

## 4.1 Diskusjon

Ved å bruke systemet vårt vil brukeren få mulighet til å finne relevante dokumenter og personer over hele verden. Systemet er gjennomtenkt, lagrer data på en fornuftig måte og gir ut informasjonen oversiktlig. Det er lett å navigere seg frem og rundt på nettsiden. Likevel kan man spørre seg hvorvidt systemet vil appellere til forskere i ulike aldersgrupper. Vil det bli for mye å sette seg inn i? Vi mener selv at systemet er logisk bygd opp. Virkemidler vi har brukt som størrelse, farger og tagger er velkjent for de fleste. Hvis det mot formodning ikke skulle være kjent, er det fullt mulig å få forklart hvordan systemet fungerer ved en hjelp-funksjon på nettsiden.

Vi har basert systemet vårt på at søkingen skal foregå på en ryddig og systematisk måte. En ulempe med systemet vårt er at man må beskrive godt det man velger å laste opp. Et eksempel på det er hvis en kunststudent skal laste opp en videofil, må kunststudenten skrive inn metadata, som tittel og lignende. I tillegg må studenten beskrive innholdet i filmen og tagge den deretter. Grunnen til at det må gjennomføres slik er for at andre enkelt skal kunne finne den igjen ved å søke. Det er mer arbeid på dem som laster opp, men alt i alt blir det mer effektivt. Hvis du ikke har noe konkret du vil søke etter er det mange steg du må gjennom. Da må du mer manuelt gå gjennom flere sider og trykke på emner.

## 5.1 Konklusjon

Formålet med oppgaven var å gi oss grunnlaget som er nødvendig for at vi skal kunne diskutere forskjeller med ulike IR-system. For at vi, på best mulig måte, skulle klare å utvikle et system som kunne fungere for Kings College måtte vi sette oss inn i forskjellene. Ved å bruke kjente fagterminologiske ord og uttrykk får vi frem det systemet vi har laget for å laste opp og dele informasjon. I tillegg til rapporten har vi laget SSM for å skildre viktige entiteter og relasjoner i systemet. Utover disse er det skisser laget i Adobe Photoshop som best representerer nettsiden vi har laget. Vårt system vil kunne fungere utmerket til sitt bruk med tanke på spesifikasjonene beskrevet i oppgaveteksten. Vi mener at de forskjellige typer søk

som var krevd at skulle være mulig, er laget på en god måte. Interaksjonen er formidlet med lettfattelighet og selv om ikke brukervennlighet er det vi har fokusert på her, er poenget at vi har laget tilsynelatende funksjoner som innfrir en brukers forventning til systemet.

## 6.1 Kildehenvisning

[Nordbotten](#), C, J. (2008) *Multimedia Information Retrieval Systems* [Internett]. Tilgjengelig fra: <[http://nordbotten.com/ADM/ADM\\_book/MIRS-frame.htm](http://nordbotten.com/ADM/ADM_book/MIRS-frame.htm)> [Nedlastet 31.mars 2009].

Wikipedia (2009a) *Data visualisering* [Internett].  
Tilgjengelig fra: <[http://en.wikipedia.org/wiki/Data\\_visualization](http://en.wikipedia.org/wiki/Data_visualization)> [Nedlastet 31.mars 2009].

Wikipedia (2009b) Tag (*metadata*) [Internett].  
Tilgjengelig fra: <[http://en.wikipedia.org/wiki/Tag\\_\(metadata\)](http://en.wikipedia.org/wiki/Tag_(metadata))> [Nedlastet 31.mars 2009].

Wikipedia (2009c) *Open standard*[Internett].  
Tilgjengelig fra: <[http://en.wikipedia.org/wiki/Open\\_standard](http://en.wikipedia.org/wiki/Open_standard)> [Nedlastet 31.mars 2009].

Wikipedia (2009d) *Inverted index* [Internett].  
Tilgjengelig fra: <[http://en.wikipedia.org/wiki/Inverted\\_index](http://en.wikipedia.org/wiki/Inverted_index)> [Nedlastet 31.mars 2009].

## 7.1 Inspirasjonskilder

<http://www.digg.com>

<http://www.facebook.com>

<http://www.last.fm>

<http://rssvoyage.com/#RSS=loaded>

<http://marumushi.com/apps/newsmap/>

<http://www.bbc.co.uk/history/interactive/timelines/british/index.shtml>

## **8.1 Vedlegg**

Vedlegg 1: SSM-modell

Vedlegg 2: Inngangsside med fagoppdeling (steg 1)

Vedlegg 3: Fagområder A-Z (steg 2)

Vedlegg 4: Tags (steg 3)

Vedlegg 5: Kart

Vedlegg 6: Artikkel

Vedlegg 7: Frisøk - resultat

Vedlegg 8: Profilside