

Referater

6. juni 2007

Referat - 11/maj-2007

Info

Tilstedeværende: Martin Dybdal, Troels Henriksen og Dina

Fraværende: Jesper Reenberg.

Ordstyrer: Martin Dybdal

Referent: Troels Henriksen.

Møde om endelig kravspecifikation

- Der skal tilføjes flere overskifter til afsnittene. Specielt analysen.
- Vi kan evt rende ind i problemer med hensyn til "code review" fra andre grupper grunet vores valg af programmerings sprog, men det blev aftalt at dette ikke skulle være en hindring for os.
- Dina påpegede at der mangler at blive defineret hvad "sidens tekst" er.
- Der blev diskuteret hvorvidt vores kravspec passede ind i "skabelonen" da eksempler var tynde og begrundelser ofte var gentagelser. Men Dina kommenterede : "Det spiller ingen rolle på så lille skala som det foregår på."

Heraf at hvert afsnit i analysen skal indeholde en problemstilling og en løsning.

- Krav 2 blev diskuteret og der skal beskrives entydigt at vi med en analyse af en underside, mener at foretage analyser med de nedenfor beskrevne algoritmer og knytte deres resultater sammen med siden som output (dog strippet for eventuel formatering / billeder som kunne forvirre det overordnede billede),
- Dina påpeger at vi generelt skal læse vore ting mere igennem.
- Sidesværhedsgraden skal defineres som værende en værdi baseret på resultatet af LIX- of FKRT-algoritmen, men da disse først beskrives senere er der behov for referencer.
- Afsnittet der beskriver krav til målgruppe og læsere skal pilles ud for sig og have en individuel overskrift.
- Dina påpeger at det indledende afsnit i analysen kunne undværes da dette var for meget "hyggesnak".

Aftaler

- lexer og parser skal laves færdig.
- HTTP modul skal tilpasse og fikses så det virker ordenligt.
- Util modul skal udbygges.
- Ordenlig brug af navngivning så som filnavne begynder med stort.

- Der skal laves et Robots modul som kan parse en “robots.txt” og ud fra dette angive hvorvidt et link må crawles eller ej.
- Der skal laves en funktion til at finde links i de parsede HTML dokumenter.

Tidsforbrug

Tidsforbrug er anslået til ca. 10 timer pr. mand.

Referat - 11/maj-2007

Info

Tilstedeværende: Martin Dybdal, Troels Henriksen og Jesper Reenberg

Fraværende: Dina.

Ordstyrer: Martin Dybdal

Referent: Jesper Reenberg.

Møde

- Dina kunne ikke komme til mødet grundet sygdom.

Fra Dina Friis
Til Hjemmesideanalyse-grupperne
Dato 18-05-2007 08:13
Emne jeg er desværre syg idag

- Grundet dette blev der ikke holdt noget officielt møde, men gruppen havde en hurtig snak om fordeling af opgaver og tidsforbrug.

Aftaler

- Der skal laves flere kommentarer til den skrevne kode.
- Lexer og parser skal optimeres og opdateres til at bruge funktioner fra Util modulet.
- getLinks skal kunne håndtere relative links og gøre dem absolutte inden de resulteres videre i systemet.
- Nogle funktioner skal simplificeres og eventuelt omskrives så de evt. kan blive hurtigere.
- Utils skal udbygges mere flere funktioner.
- HTTP modulet skal yderligere fikses så det ikke blot kaster "badURI" exception.
- HTTP modulet skal yderligere fikses så det også kan håndtere andre "content-type"'s.
- HTTP modulet skal også ydbygges så det kan håndtere http "response codes" ordenligt (specielt 404').
- Der skal tilføjes kommandolinje parametre til at kontrollere programmet.
- Der skal ikke crawles links uden for det angivet domæne.
- Der skal laves en tekst udtrækker og et analyse modul til videre håndtering af dokumenter.
- Ydermere skal der også laves et modul til at vise den analyserede txt.

Tidsforbrug

Tidsforbrug er anslået til ca. 10 timer pr. mand.

Code review - 25/maj-2007

Info

Tilstedeværende: Martin Dybdal, Troels Henriksen og Jesper Reenberg

Andre grupper: Gruppe #17

Ordstyrer: Martin Dybdal

Referent: Jesper Reenberg.

Code review med gruppe #17

- Gruppe 17 mente at vi havde udemærket/nærmest eksemplarisk kode, så de havde ikke rigtigt nogle kommentarer eller kretik.
- Dog kom de med en henvisning til at vi også kunne have brugt en lexer/parser der var lavet i forvejen som fx mosmlyacc eller mosmllex, men vi har valgt at det var nemmere selv at lave det da det er forholdsvis simpelt og at vi ellers skulle til at sættes os ind det andet.

Referat - 25/maj-2007

Info

Tilstedeværende: Martin Dybdal, Troels Henriksen, Jesper Reenberg og Dina

Ordstyrer: Martin Dybdal

Referent: Jesper Reenberg.

Møde efter code review

- Dina er utroligt imponeret over hvor langt gruppen egentlig er nået sammenlignet med mange af de andre, da gruppen på nuværende tidspunkt har implementeret 99% af baseline designet.
- Med dette ville Dina utroligt gerne vide hvad gruppen egentlig manglede:
 1. Der skulle gerne ske en form for speciel semantisk håndtering af html tags så som `strong`, `em` osv.
 2. Der skal findes en løsning på håndtering af små sætninger. Fx sætninger på 0-3 ord som giver det enkelte tekst afsnit en meget forvrænget score (Dette gør sig gældende på alle tekste under ca. 1.000 ord).
 3. Finde en anden formel for "sideslemhedsfaktoren", der på nuværende tidspunkt bare udregnes rent ud fra lix tallet
- Dina påpeger at vi skal huske at have farve prints af uddata fra analyzen til den udprintede rapport, så man har en chance for at se uddata nu da der bruges mange farver.
- Der var en lille diskussion om hvorvidt det endelige design var lavet færdigt, og det blev til at der manglede ganske lidt før det kunne afleveres. Der var blot nogle småting til Dataflow men ellers ikke det store
- Der blev snakket en smule om forskellige test, og vi kunne dog kun stå ved at vi var begyndt at lave unit tests og herigennem konstatere at det var utroligt få ting som ikke virkede efter hensigten.
- Gruppens mente at programmet så småt var på vej til at kunne "bruger testes" så i den anledning påpegede Dina at der *skulle* ligge en brugervejledning og en beskrivelse af hvordan det er planen at udfører disse bruger tests så der kunne være lidt sammenhæng i resultaterne.
- Til aller sidst blev gruppens code review diskuteret og selvom det var denne der skulle have været sankket mest om måtte gruppen detsvære sætte sig med takke at der ikke var kommet noget konstruktiv kritik af værdi. Eneste ting gruppen også havde fået nævnt et par gange tidligere er at: den farveskala som bruges til at farvelægge teksts afsnit og selve sætningerne går fra grøn til brun til rød, og dette kunne forvire med hvorledes brun intuitivt ligger i forhold til rød og grøn. Så med dette er det således også blevet en opgave for gruppen at ændre skalaen til at gå fra grøn til gul til rød hvilken er mere intuitivt for alle Med hensyn til code reviewet havde Dina heller ikke nogle synderlige kommentarer andet end vi blot skulle klø på.

Aftaler

- Farvekodningen af teksten skal ændre så den ikke går i grøn-brun-rød men derimod grøn-gul-rød
- Analyse modulet skal smukkeseres lidt.
- Uddata skal også gøres lidt pænere.
- HTTP modulet skal stadig fikses lidt.
- Index siden for en analyse af et websted skal også farvelægges
- Tilføjelse af flere commandolinje parametre
-

Tidsforbrug

Tidsforbrug er anslået til ca. 10 timer pr. mand.

Referat - 1/juni-2007

Info

Tilstedeværende: Martin Dybdal, Troels Henriksen, Jesper Reenberg og Dina

Ordstyrer: Martin Dybdal

Referent: Jesper Reenberg.

Møde om test specifikation

- Dina spørger til hvad gruppen selv syntes der mangler og gruppen kommer frem til at der mangler indhold på det, før mødet, sendte materiale til dina.
- Dina opfordre kraftigt gruppen til at tage følgende strategi i brug:
 1. Tag hver krav og find ekvivalens klasser
 2. Tag derefter hvert krav og find grænse tilfælde.
 3. Lav derefter test ud fra dette.
- Dina påpeger herunder også at vi skal sørge for at takle problematiske tests på en smart måde. Af problematiske tests har gruppen bl.a. tegnsæts håndtering og HTML dokument håndtering.
- Der blev specielt diskuteret omkring håndtering af tests for HTML dokument håndtering, da dette vil være utroligt omfattende at teste at al html håndteres efter hensigten og specielt også utroligt plads krævende i rapporten, sagde Dina at vi kunne lave "nok" tests og så herefter sige at "resten må man stole på også virker efter hensigten når det første gør, da der ikke er tid nok til at udpensle alle tests vedrørende HTML" eller noget i den stil
- Dina påpeger at i det udkast der har været udleveret før mødet ikke er dokumenterende nok mht. hvor og hvilke krav de enkelte tekst afsnit kommer fra.
- I det udleverede udkast bliver der bl.a. nævnt aspell men der er ikke umildbart før dette været noget andet forklarende tekst omkring dette så enten en fodnote eller en henvisning til hvor der kan læses nærmere om dette,
- Dina påpeger herved også at Lexeren på nuværende tidspunkt nok ikke er godt nok beskrevet. Selvom den formentlig bliver beskrevet i rapporten ses det gerne om muligt at der eventuelt lå et separat billag omkring dette og evt også parseren.
- Gænsetilfælde er mer eller mindre slet ikke beskrevet i det udleverede udkast.
- I det udleverede udkast står der at alt relevant tekst skal vises i analysen. Hvis der vælges at skrive sådan, skal det virkelig gøres 1000% klart hvad relevant tekst er. Ellers vil gruppen formentlig blive udspurgt om dette til eksamen og eventuelt "hængt op i en galge" på dette.

- Med hensyn til dokumentation af tests foreslog dina at der blev lavet en tabel som indeholdte:

1. Test nr
2. Grænse tilfælde (henvisning)
3. Ind data (henvisning)
4. Forventet uddata (kort beskrivelse og eventuel henvisning til billag med dybere beskrivelse eller specifik uddata)
5. Resultat

Da dette skal stå til sidst som konklusionen på vores tests skal henvisningerne selvfølgelig være til allerede defineret data. Fx kunne en html side være inddata og denne kunne være vedlagt (vist som HTML kode og vist i en browser) som billag for sig selv.

Test Nr	Grænse tilfælde	inddata	forventet uddata	Resultat
1	G1	I1	xxxx	ok
2	G2	I1	xxxx	ok
3	G1	I2	xxxx	ok

Aftaler

- Opdater sideslemhedsfaktoren så den ikke kun afhænger af lix men af alle analyserne (som oprindeligt ment i kravspec)
- Småændringer til uddata,
- Håndtering hvis stavekontrol (aspell) ikke er installeret på maskinen.
- Lav testspecifikationer færdig.

Tidsforbrug

Tidsforbrug er anslået til ca. 10 timer pr. mand.