

LINKÖPINGS UNIVERSITET
Institutionen för Datavetenskap

Studie av gränssnittsprototyp i projektet
Webbklustring - Nyttoupplevelsen

Namn Evelina Rennes
E-mail `evere305@student.liu.se`

Innehåll

1 Inledning	1
1.1 Komplexitetssmått	1
1.2 Beskrivning av systemet	1
2 Genomförande	2
2.1 Tänka högt-protokoll och intervju	2
2.2 Task Success	2
2.3 SUS-utvärdering	2
2.4 Uppgifter	2
3 Resultat	3
3.1 Task Success	3
3.2 SUS-utvärdering	3
3.3 Intervju	4
3.3.1 Upplevd nytta	4
3.3.2 Jämförelse med andra sökmetoder	5
3.3.3 Otillfredsställande klustersammanfattningar	5
4 Diskussion	6
5 Slutsats	7
Referenser	7
A Gränssnitt: Startside	8
B Gränssnitt: Klusternivå	9
C Gränssnitt: Hovring över kluster	10
D Gränssnitt: Dokumentnivå	11

1 Inledning

Projektet Webbklustring drivs på Linköpings universitet och finansieras av Internetfonden och SICS East Swedish ICT AB. Projektet har som mål att utveckla ett webbaserat system som ska underlätta informationsökning, bland annat genom att jämföra hemsidor med samma innehåll och således utesluta dubletter, men också sammanfatta texter och avgöra hemsidors läsbarhet. Ett övergripande mål med projektet är att bistå ovana användare eller personer med läs- och skrivsvårigheter med ett enkelt verktyg för att underlätta sållande bland information. Projektet integrerar delvis även ramverk framtagna i tidigare projekt: Webblättnäst och FriendlyReader. Webblättnäst rangordnar texter efter komplexitetsmått och FriendlyReader sammanfattar texter i syfte att göra dem mer lättlästa. Inom projektet Webbklustring mäts hemsidors informationsinnehåll.

Studien har ämnat undersöka den upplevda nyttan av det system som presenteras i projektet Webbklustring, som komplement till befintliga användbarhetsstudier utförda på tidigare prototyper av gränssnittet.

1.1 Komplexitetsmått

De komplexitetsmått som tillämpas i systemet är läsbarhetsindex (LIX) och ordvariationsindex (OVIX). LIX är baserat på antalet långa ord (ord längre än sex bokstäver) och medelantalet ord per mening.

OVIX ser till hur många unika ord som finns i en text jämfört med totalt antal ord. En text som har stor ordvariation anses vara mer svårläst än en text med liten ordvariation, och till skillnad från LIX-värdet tas ingen hänsyn till ordens längd.

1.2 Beskrivning av systemet

Gränssnittet består av ett antal sektioner; en sökfunktion, ett rutnät som representerar sökresultaten visuellt samt en dokumentlista. Sökfunktionen fanns lokaliserat högst upp på sidan och består av en textruta och, till höger om rutan, ett reglage där användaren kontrollerar hur många sökresultat som ska visas (se Bilaga A). Vid sökning sorteras dokumenten i kluster baserat på likhet och sökresultaten representeras på två sätt: som en hexagon i rutnätet samt i en klassisk dokumentlista baserat på relevans. Dokumentets placering på rutnätet beror på vilken läsbarhet och relevans dokumentet har (se Bilaga B). Så väl rutnätet som dokumentlistan går att sortera efter det komplexitetsmått som önskas användas (LIX eller OVIX) och dokumentlistan kunde utöver det sorteras efter relevans. Defaultvärdet för både rutnätet och dokumentlistan vid sökning är LIX. När användaren för muspekaren över ett kluster öppnas en informationsruta som visar en multidokumentsammanfattning av det specifika klustret och dess komplexitetsmått, angivna i ord och siffror (se Bilaga C).

Ett musklick på ett kluster öppnar det och samma gränssnitt visas på dokumentnivå (se Bilaga D). Här återfinns de unika dokumentens placering i samma rutnät som beskrivits ovan. För användaren muspekaren över de enskilda dokumenten visas en informationsruta liknande den på klusternivå, med skillnaden att den istället innehåller den specifika hemsidans snippet. Ett musklick öppnar den specifika hemsidan.

2 Genomförande

Deltagarna bestod av 10 studenter utan läs- eller skrivsvårigheter från Linköpings universitet, varav 6 var män och 4 var kvinnor. Åldern på deltagarna varierade mellan 20 och 27 ($M = 23$, $SD = 2,6$) och samtliga deltagare ansåg sig ha god datorvana.

Innan testet blev deltagarna informerade om att de var fullständigt anonyma, att de hade möjlighet att avbryta testet när de ville samt att resultat inte skulle kunna spåras tillbaka till dem.

2.1 Tänka högt-protokoll och intervju

Ett tänka högt-protokoll användes för att få insikt i användarens upplevelse vid arbete med systemet. Försöksdeltagaren ombads att verbalt beskriva sina tankar och upplevelser vid interaktion med systemet och dessa dokumenterades med hjälp av en ljudupptagningenhet samt skärminspelning. Efter utfört test genomfördes en ostrukturerad intervju i syfte att fånga upp användarens tankar om systemet.

2.2 Task Success

För varje uppgift bedömdes deltagarens förmåga att fullfölja den, i syfte att se hur verktyget fungerar vid realistiska informationssökningssituationer. I detta ändamål skapades tydliga kriterier för vad som definierade en fullständigt utförd uppgift och uppgifterna utformades för att vara entydiga. Efter varje uppgifts slutförande ombads försöksdeltagaren uttala lösningen på uppgiften och i de fall svaren innehöll onödigt eller tvetydigt information ställdes kompletterande frågor av försöksledaren i syfte att undersöka huruvida deltagaren hade klarat uppgiften eller inte. Detta motiveras enligt riktlinjer föreslagna i Tullis och Albert (2008). Det fanns endast två alternativ för uppmärksningen av *task success*: 1 eller 0, där 1 representerade att deltagaren hade klarat uppgiften och 0 representerade ett misslyckande. Att använda sig av binära mätvärden innebär att en försöksdeltagare aldrig kan klara uppgiften delvis, vilket ställer ytterligare krav på entydigheten i uppgifternas design.

2.3 SUS-utvärdering

En efterföljande SUS-utvärdering utfördes för att få ett kvantitativt jämförande värde. En SUS-utvärdering täcker upp ett systems effektivitet, måluppfyllnad och tillfredsställelse. Det SUS-formulär som användes i vid detta tillfälle var baserat på Brookes originalformulär (Brooke, 1996) och identiskt med det som användes vid en tidigare användbarhetstestning av systemet. Det var sedan tidigare översatt till svenska och ordet *system* var ersatt av *hemsida* för att underlätta förståelsen för användarna.

2.4 Uppgifter

Uppgifterna utformades i syfte att motsvara systemets samtliga funktioner och togs fram av personer som arbetar med projektet Webbklustring.

De uppgifter som användes var:

- Hitta den mest läsbara hemsidan om Helan och Halvan enligt LIX.
- Hitta det kluster som är det mest relevanta.
- Hitta den hemsida är den mest relevanta.
- Hitta det kluster som är det mest lättlästa.
- Vilken sammanfattning tycker du är bäst och varför?
- Hitta en sida som beskriver hur du lagar en punktering.
- Hur många olika typer av svar på frågan hur du lagar en punktering hittar du?
- Hur många olika kluster hittar du?

De två sista frågorna var tänkta att vara relativa ekvivalenta och på så vis fånga användarens förståelse för hur sökresultaten sorteras i de olika klustren.

3 Resultat

I denna sektion presenteras de resultat som erhöles från testsessionerna.

3.1 Task Success

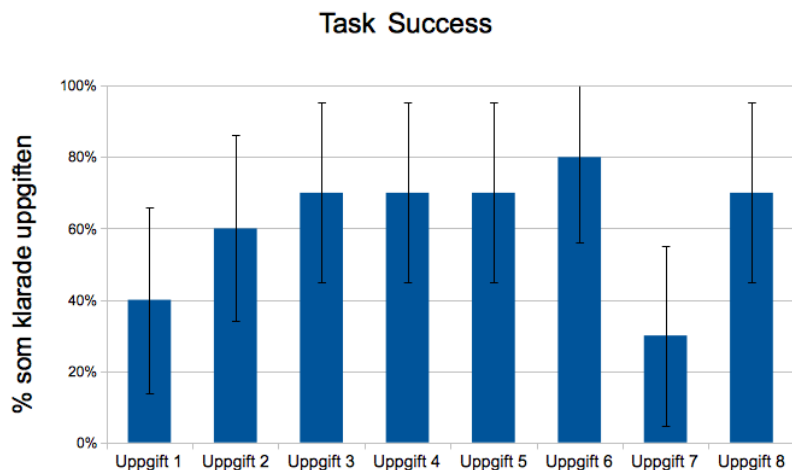
Det fanns inte någon uppgift som ingen deltagare klarade av. Den uppgift som flest (80% av deltagarna) klarade av var uppgift 6, *hitta en sida som beskriver hur du lagar en punktering*. Deltagarna presterade sämst på uppgift 7, *hur många olika typer av svar på frågan hur du lagar en punktering hittar du?*, där endast 30% av deltagarna fullföljde uppgiften tillfredsställande. Samtliga resultat med konfidensintervall återfinns i tabell 1.

Uppgift	Antal klarade	Konfidensintervall låg	Konfidensintervall hög
Uppgift 1	40%	17%	69%
Uppgift 2	60%	31%	83%
Uppgift 3	70%	39%	90%
Uppgift 4	70%	39%	90%
Uppgift 5	70%	39%	90%
Uppgift 6	80%	48%	95%
Uppgift 7	30%	10%	61%
Uppgift 8	70%	39%	90%

Tabell 1: Andel deltagare som genomförde uppgifterna på ett tillfredsställande sätt med konfidensintervall baserat på en binomial fördelning av data (*adjusted Wald interval*).

3.2 SUS-utvärdering

SUS-utvärderingen visade inte på någon förbättring från utvärdering av tidigare prototyper. Medelvärde var 57,25 (*SD*: 19,52), men spridningen var mycket stor.



Figur 1: Andel deltagare som utförde uppgifterna tillfredsställande. Felstaplarna representerar konfidensintervallet (95%) baserat på en binomial fördelning (*adjusted Wald interval*)

3.3 Intervju

I denna sektion presenteras de teman som framkom ur de efterföljande intervjuerna, och delvis även ur tänka högt-protokollet.

3.3.1 Upplevd nytta

Det tema som var genomgående i intervjuerna var att deltagarna ser en nytta med verktyget, men framhåller även att förbättringar måste genomföras för att det ska fungera som hjälpmedel. En del uttryckte att det finns en nytta för gemene man, och tror att informationssökning i allmänhet kan komma att gå fortare med hjälp av systemet. Detta speglas i följande uttalande:

P01: Det är hjälpsamt också och värdefullt på det sättet att det kan gå snabbare för dig att hitta vad du söker liksom om inte allt kommer i klump utan faktiskt visar vad den tror är relevant för ditt sökord tänker jag. Vilket blir hjälpsamt också.

En deltagare framhöll fördelarna med att få lättlästa texter för en snabb överblick av ett ämne och betonade även att det är värdefullt att ha ett system som underlättar kunskapsinhämtning:

P06: Jag ser verkligen hur det kan hjälpa. Inte bara inom skolan utan även utanför, att man kan ta fram en text som är om man väljer att man vill ha en text som kanske bara beskriver någonting lättförståeligt inte så djupgående så man väljer den texten istället för att ta en svår och väldigt ehh omfattande. Och värdefullt är lite samma spår att det är viktigt att människor ska kunna tillgodose sig kunskap på olika sätt och efter egen vilja och förmåga.

Det fanns även deltagare som inte ansåg att dokumentklustraren skulle vara till nytta för deras egen del, men som däremot ansåg att systemet med fördel kunde användas av andra målgrupper:

P02: Kanske om man är jag vet inte, dyslektiker och väldigt mån om att endast ha lättlästa träffar till exempel, då kanske det är bra. Men jag kommer ju inte byta ut det här mot google anytime soon.

3.3.2 Jämförelse med andra sökmetoder

Flertalet deltagare jämförde dokumentklustraren med sökmotorn google, med både positiva och negativa undertoner.:

P08: Okej det var ganska tydligt tycker jag och man kan väl ha användning av det här men det känns lite ofärdigt. Sen var det väldigt kul att söka på något annat sätt än bara googlerutan.

P03: Den gör ju ingenting bättre än google. Det är ju liksom den använder säkert samma sorts sortering som google så det blir liksom, va ska man ha den till?

Det uttrycktes även åsikter om att dokumentklustraren hade fördelar mot vanliga sökmotorer.

P07: Det kan till viss del vara hjälpsamt att se asså väldigt tydligt att de här resultatens finns och framför allt med relevans och läsbarhetsaxeln så är ju det på ett sett en fördel gentemot google som sökmotor. Jag antar att google som sökmotor använder sig av samma sak liksom men det var på ett sätt hjälpsamt att få se också de olika klustrena.

3.3.3 Otillfredsställande klustersammanfattningar

Någonting som framkom i intervjuerna och tänka högt-protokollet var att de klustersammanfattningar som ämnade skapa en överblick över klustrens innehåll var otillfredsställande. Det var multidokumentsammanfattningarna på klusternivån som ansågs röriga, vilket illustreras i följande uttalande från en deltagare:

P02: (...) typ den här tvåraderssummeringen för ett kluster på fem sidor, det är lite såhär här vet man inte riktigt vad man får om man går in i det här klustret. Klustersammanfattningarna var inte så bra tycker jag, eller dom är väl snarare för små för att kunna göra en bra sammanfattning på sju olika hemsidor tror jag. Det är väl svårt att sammanfatta sju sidor på tjugoåtta ord eller vad det nu är.

Flertalet deltagare uttryckte åsikten att sammanfattningarna var något svårförståeliga och att de inte beskrev vad klustrena innehöll:

P01: Här exempelvis när det står klassiker från 1943 ehh då ser jag förvisso att det står blueray och dvd inom parentes men det förklarar inte om det är en film eller om det är en klassikerbok eller så.

P06: Ehh den gula känns mer mer beskrivande om just helan och halvan att det är en svensk benämning på komikerparet Laurel and Hardy medan andra kan vara lite rörigt, man förstår inte riktigt vad det är. Exempelvis den gröna där det bara står att allt är kul och inget är kravfullt och sen avslutas med helan går, så då kan det lika gärna vara en sång som nånting annat.

Speciellt hemsidor som baseras mer på video än på text uppgavs ge svårförståeliga sammanfattningar:

P04: Den här är ju lite oklar liksom om meningarna är asså om det ens är hela meningar och sen kommer det nåt sånt här video subscription så.

P09: Ja där finns ju sammanfattningar. Jag tycker inte den gröna är så bra för det är massa blandat. Typ som det är en youtube-titel eller nånting. Så det verkar ju inte vara så bra sammanfattning för Helan och Halvan.

De sammanfattningar som beskrevs som bäst var de som var längst, mest sammanhängande och som hade innehåll som stämde överens med vad användaren hade förväntat sig:

P02: Den här sammanfattningen var bäst. Dels för att den upplevdes som längst och dels för (paus) ja det andra kändes så liksom fragmenterade på nåt vis men den här var för det mesta ett stycke och det blev väl nåt annat där på slutet men aa, den mest sammanhållna sammanfattningen.

P03: De andra var inte så himla relevanta överhuvudtaget. Den handlar om helan och halvan och det gör typ inte dom andra.

Att de längsta och mest sammanfattningarna bedömdes som bäst ledde till att de sammanfattningar som endast återgav ett enda dokument bedömdes som bäst:

P04: Då tycker jag den här röda är bäst. Inte för att den säger så mycket heller men det känns som att det är början, alltså det är början på en text som man gärna skulle läsa vidare.

P08: Jag tycker wikipedia är bäst. Eller den högsta här uppe på den här lila. Det är för att den är kort och koncis och man får ut mest information om man inte vet någonting om helan och halvan.

P10: Jag tycker den här mörkblåa är bäst. Dels så vet man att Wikipedia är bra på att sammanfatta men sen vet man också att dom andra hemsidorna inte stämmer överens med det man det jag är ute efter.

4 Diskussion

De resultat som erhöles vid mätning av *Task success* visar på att det är relativt många som lyckas med uppgifterna och att systemet går att använda i praktiken. Den uppgift som användarna lyckades sämst på var *Hur många olika typer av svar på frågan hur du lagar en punktering hittar du?* Denna fråga var tänkt att fånga användarnas förståelse för hur de olika dokumenten sorterades i kluster, och således på ett sätt vara ekvivalent med frågan *Hur många olika kluster hittar du?* Det blev tydligt att denna koppling inte gjordes, förmodligen på grund av att klustringen inte var så entydigt som hade önskats.

Anledningen till att resultat var förhållandevis dåligt på första frågan beror troligen på att uppgifternas ordning inte ändrades mellan testsessionerna, och användarna blev duktigare på att använda verktyget ju längre in i testet de kom. Detta hade kunnat korrigeras för genom att ändra uppgifternas ordning.

I de efterföljande intervjuerna uttryckte deltagarna en upplevd nytta av systemets koncept, men framhöll också att vidare utveckling bör ske i syfte att göra det användbart som hjälpmedel. Användarna trodde att verktyget kunde

komma att bli användbart för dyslektiker, men även för deras egen del för att underlätta informationssökning. Detta visar på att det finns ett behov för det grundkoncept som Webbklustring står för, även om vissa förändringar bör ske i utformningen.

Flertalet deltagare återkom till en jämförelse med andra sökmetoder. En del ansåg att det var en störande faktor i för informationssökningen, medan andra tyckte att det var motiverande och roligt att få sökresultaten representerade grafiskt. Dessa skilda meningar motiverar till en omdesign av systemet där den grafiska komponenten blir valbar och resterande delen av gränssnittet blir mer likt etablerade sökmotorer. Detta skulle leda till att användaren själv kan välja om denna vill representera sina sökresultat på ett annat vis, och därmed eliminera en störande faktor då användaren inte behöver se eller använda den grafiska klusterrepresentationen.

De multidokumentsammanfattningarna som ämnade beskriva klustrens innehåll på ett överskådligt vis ansågs vara otillfredsställande. De bedömdes vara svårförståeliga och ansågs i vissa fall inte heller beskriva innehållet. Speciellt problematiskt var det för kluster som innehöll dokument som inte var textbaserade, så som videoklipp, och en deltagare menade på att det inte är möjligt att beskriva flertalet hemsidor med så korta sammanfattningar. Detta motiverar en omdesign där multidokumentsammanfattningarna tas bort på klusternivån och ersätts av andra element, exempelvis nyckelordsmoln, eller möjligtvis en enkel lista av nyckelord.

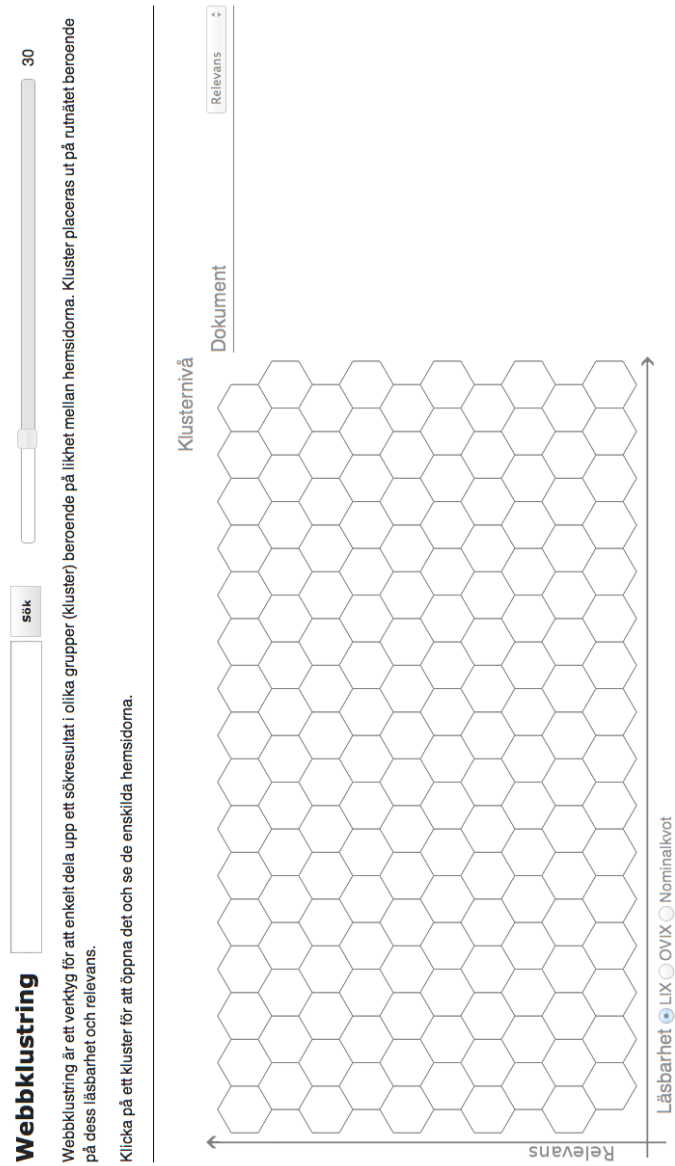
5 Slutsats

Denna studie ämnade undersöka nyttoupplevelsen av det webbaserade verktyget Webbklustring. Genom analys av intervjuer och tänka högt-protokoll identifierades tre teman: upplevd nytta, jämförelse med andra sökmetoder och otillfredsställande klustersammanfattningar. Systemet ansågs ha ett användbart koncept, även om det behöver viss vidareutveckling. Två förbättringsförslag togs fram: att göra den grafiska representationen till ett valbart alternativ och göra det övriga gränssnittet mer likt befintliga sökmotorer samt byta ut multidokumentsammanfattningarna av klusterinnehållet mot någon annan beskrivning av klustret, exempelvis ett nyckelordsmoln.

Referenser

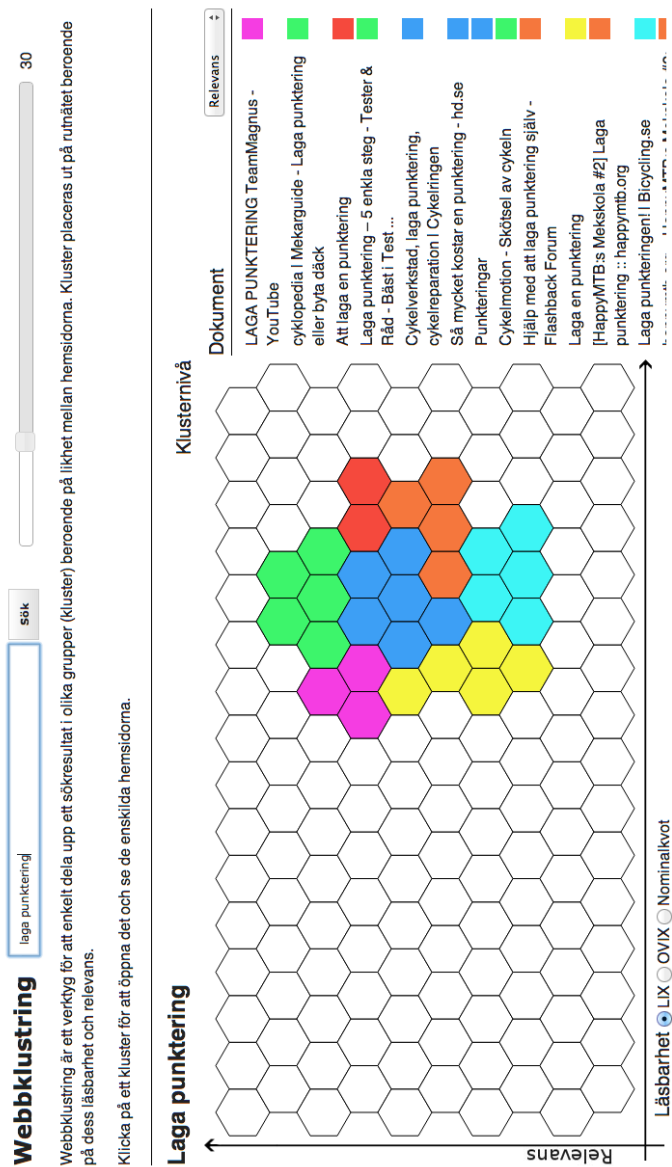
- Brooke, J. (1996). *Sus – a quick and dirty usability scale*. Hämtad 2014-01-20, från <http://hell.meiert.org/core/pdf/sus.pdf>
- Friendlyreader*. (u. å.). Hämtad 2013-12-29, från <http://www.ida.liu.se/projects/friendlyreader/>
- Tullis, T. & Albert, B. (2008). *Measuring the user experience: Collecting, analyzing and presenting usability metrics*. Denise E. M. Penrose.
- Webbklustring*. (u. å.). Hämtad 2013-12-29, från <http://www.ida.liu.se/projects/webbklustring/>
- Webblättnäst*. (u. å.). Hämtad 2013-12-29, från <http://www.ida.liu.se/arnjo/webblattlast/>

A Gränssnitt: Startside



Figur 2: Gränssnittet som det ser ut innan sökning.

B Gränssnitt: Klusternivå

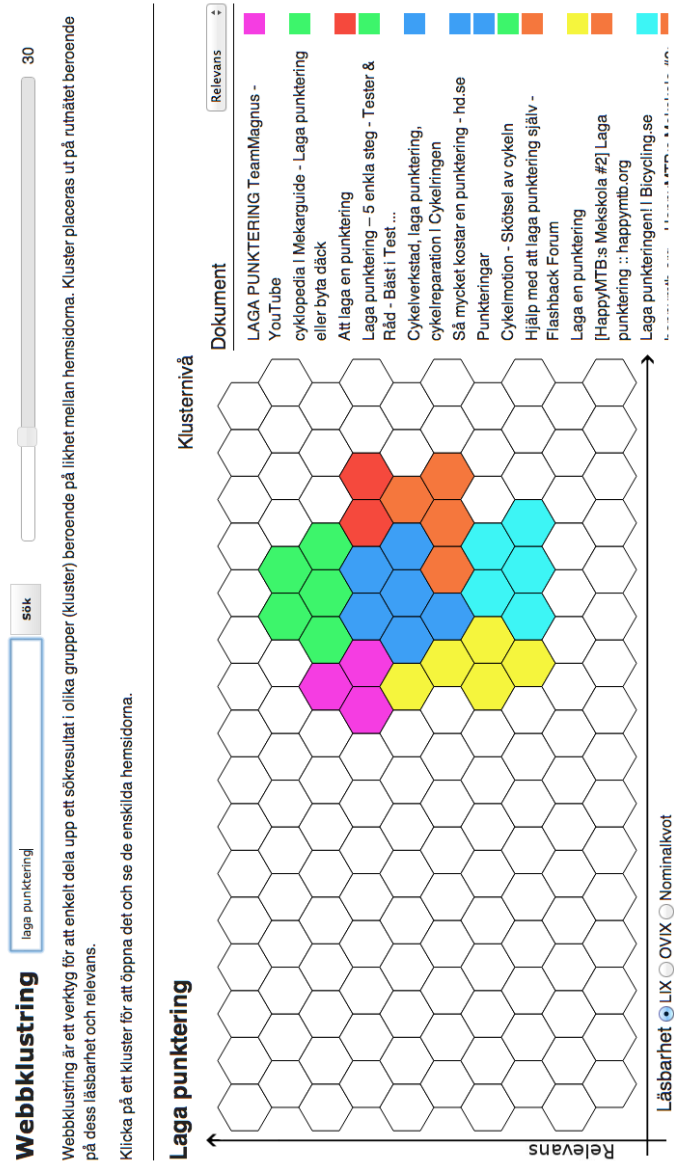
Figur 3: Gränssnittet efter en sökning på *laga punktering*. Klusternivå.

C Gränssnitt: Hovring över kluster



Figur 4: Den information som visas om ett visst kluster när användaren hovrar över det.

D Gränssnitt: Dokumentnivå

Figur 5: Gränssnittet efter en sökning på *laga punktering*. Dokumentnivå.