

LINKÖPINGS UNIVERSITET
Institutionen för Datavetenskap

1 augusti 2013

Användbarhetstestning av gränssnittsprototyp
för projektet Webbkustering

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

Innehåll

1 Inledning	1
2 Bakgrund	1
2.1 Läsbarhetsmått	1
2.2 Beskrivning av systemet	1
3 Genomförande	2
3.1 Analysmetod	2
4 Analys	3
4.1 Task Success	3
4.2 Tid	3
4.3 Problem	4
4.3.1 Förvirring vid tolkning	4
4.3.2 Att växla mellan dokument- och klusternivå	5
4.3.3 Använda dokumentlistan	5
4.3.4 Svårfunna läsbarhetsmått	6
4.4 SUS-poäng	6
5 Diskussion	6
5.1 Metoddiskussion	7
6 Slutsats	8
A Gränssnitt: Klusternivå	10
B Gränssnitt: Dokumentnivå	11
C Instruktioner och uppgifter	12

1 Inledning

Projektet Webbklustring som drivs på Linköpings universitet är ett webbaserat system som syftar till att effektivisera informationssökning. Ett övergripande mål att bistå ovana användare eller personer med läs- och skrivsvårigheter med ett enkelt verktyg för att underlätta informationssökning på internet.

Det system som utvecklas skall kunna jämföra hemsidor med samma innehåll och således utesluta dubletter, men också sammanfatta texter och avgöra hemsidors läsbarhet.

Denna undersökning har ämnat undersöka användarnas upplevelse av en första gränssnittsprototyp och ge förslag på eventuella förbättringar.

2 Bakgrund

Projektet är delvis baserat på resultat från tidigare projekt: Webblättläst[2] och FriendlyReader[3]. Webblättläst rangordnar texter efter läsbarhetsmått och FriendlyReader sammanfattar texter i syfte att göra dem mer lättlästa. Inom projektet Webbklustring[4] mäts hemsidors informationsinnehåll.

2.1 Läsbarhetsmått

De läsbarhetsmått som tillämpas i systemet är läsbarhetsindex (LIX) och ordvariationsindex (OVIX). LIX är ett i svenskan vedertaget mått som baseras på antalet långa ord (ord längre än sex bokstäver) och medelantalet ord per mening.

OVIX är ett läsbarhetsmått som 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 svårlästa än texter med liten ordvariation och till skillnad från LIX-värdet tas ingen hänsyn till ordens längd.

2.2 Beskrivning av systemet

Den prototyp som utvärderades finns illustrerad i Appendix A samt Appendix B. Gränssnittet bestod av ett rutnät av hexagonala celler där dokument placerades i kluster beroende på likhet. Y-axeln angav sökresultatens relevans och X-axeln angav läsbarhet.

Ett sökfält fanns lokaliserat högst upp på sidan och till höger fanns en lista på samtliga dokument som ingick i sökningen. Antalet sökresultat kunde regleras med ett reglage till höger om sökfältet. Både rutnätet och dokumentlistan gick att organisera enligt läsbarhetsmåten LIX och OVIX, och dokumentlistan kunde även sorteras enligt dokumentens relevans.

Genom ett musklick på ett kluster öppnades det specifika klustret upp och visade de unika dokumentens placering enligt samma rutnät som beskrivits ovan. Genom att klicka på ett dokument, antingen ur dokumentlistan eller inuti det specifika klustret, visades en *snippet* i nedre delen av sidan, samt det aktuella dokumentets LIX- och OVIX-värde. För att underlätta förståelsen av de värden som visades återgavs även läsbarheten i ord.

3 Genomförande

Fem personer, varav tre män och två kvinnor, genomförde testet. Medelåldern var 22,4 år ($SD = 0,5$) och ingen av deltagarna hade någon form av läs- eller skrivsvårighet. Samtliga deltagare var universitetsstuderande. Den egenskattade¹ datorvanan var genomgående hög ($M = 4,4$, $SD = 0,89$).

I användbarhetstestningen användes tänka-högt-protokoll samt ett efterföljande SUS-formulär. Det senare syftar till att få kvantitativa värden för framtida jämförelse av kommande versioner av systemet.

SUS-formuläret[1] översattes till svenska och ordet *system* ersattes av *hemsida* för att underlätta förståelsen för användarna.

Uppgifterna (Appendix C) utformades för att täcka systemets samtliga funktioner och deltagarna ombads att muntligt beskriva sin tankegång vid utförandet av uppgifterna. Data samlades in genom ljudupptagning samt skärminspelning.

3.1 Analysmetod

För att analysera resultatet av uppgifterna ur ett kvantitativt perspektiv mättes *task success*. Till vilken grad deltagaren lyckades genomföra uppgiften mättes genom att gradera varje uppgift deltagaren genomförde med 1 (*lyckades*), 0,5 (*vissa problem*) eller 0 (*misslyckades*). Medelvärdet av dessa gav ett värde mellan 0 och 1 som kan ses som riktlinjer för hur svåra uppgifterna var att utföra.

Två uppgifter plockades bort ur analysen då de liknade två redan utförda uppgifter. Syftet med dessa var endast att bidra till ett verklighetstroget arbetssätt, och utgjorde inte ett adekvat mått på de faktiska funktionerna.

¹På en skala mellan 1-5 där 5 var ”mycket hög”

4 Analys

I detta avsnitt presenteras resultat och analys från testets samtliga delar.

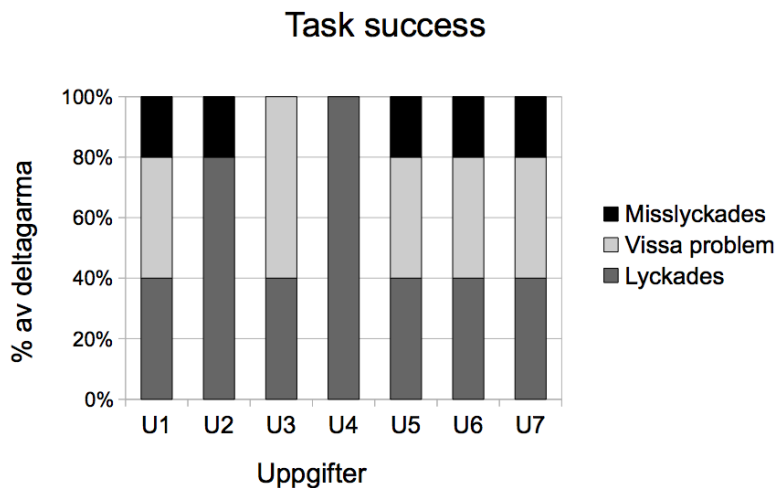
4.1 Task Success

Generellt presterade deltagarna väl. Uppgifternas natur och medelvärdet av den erhållna poängen presenteras i Tabell 1 Uppgiften rörande sökning och reglering av antal sökresultat (U4) klarade samtliga deltagare utan problem. Även U2 och U3 fick en relativt hög poäng (0,8 respektive 0,7). U1 och U5-U7 fick lägst poäng, samtliga 0,6.

Uppgift	Funktion	Poäng
U1	Sökfunktionen	0,6
U2	Informationshämtning	0,8
U3	Navigera	0,7
U4	Reglaget	1
U5	Läsbarhet	0,6
U6	Dokumentlistan	0,6
U7	Läsbarhet	0,6

Tabell 1: Uppgifternas huvudsakliga funktion samt medelvärdet av poängen.

Den procentuella fördelningen för varje uppgift presenteras i Figur 1.

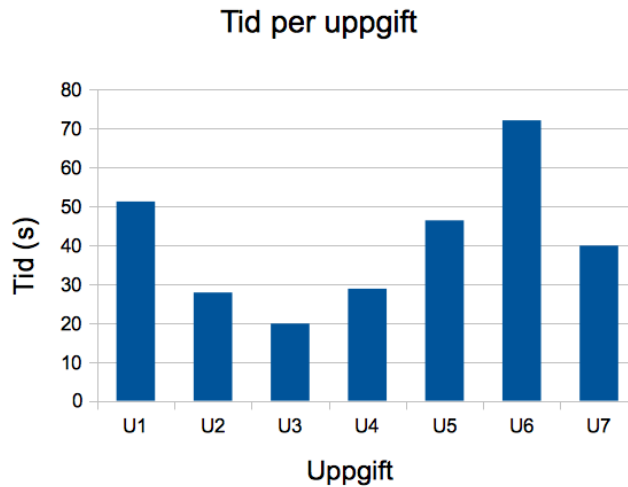


Figur 1: Task success.

4.2 Tid

Figur 2 visar den genomsnittliga tiden som deltagarna spenderade på varje uppgift. De uppgifter som krävde minst tid var U3, tätt följt av U2 och U4. Uppgift U6 var mest tidskrävande.

Tabell 2 visar mer detaljerad data för varje uppgift. Uppgift U6 hade högst



Figur 2: Time on task.

genomsnittlig tid men ett stort konfidensintervall där tiden varierar från 4,4 sekunder till 186,1 sekunder.

Uppgift	Medelvärde	Upper bound	Lower bound	Konfidensintervall
U1	51,25	74	31,7	17,1
U2	27,85	36,3	18,1	6,9
U3	19,86	29,2	8,7	7,4
U4	28,82	39,2	18,8	6,9
U5	46,40	70,8	9,7	25,5
U6	72,13	186,1	4,4	72,7
U7	39,90	54,2	26,1	12,2

Tabell 2: Tid i sekunder per uppgift.

4.3 Problem

Med utgångspunkt från de uppgifter som deltagarna misslyckades med identifierades de faktorer som orsakade dessa problem.

4.3.1 Förvirring vid tolkning

Många deltagare uttryckte förvirring vid introduktion för gränssnittet, och det tog tid att tolka gränssnittet. Exempel på detta presenteras nedan.

Okej... hur gör jag då då? Ska jag klicka på de här då eller? [För musen över rutnätet] Nä.. Jag fattar ingenting! [skratt] jag är inte van vid sånt här. [klickar runt lite över rutnätet, klickar i sökrutan]

En annan deltagare uppvisade liknande förvirring vid första uppgiften:

Och då är den någonstans här alltså [för muspekaren över rutnätet]. I dom här.. [skratt] okej... men det händer ju ingenting om jag klickar. Varför gör det inte det? [scrollar upp och ner på sidan].

Att inte förstå gränssnittets design skapar viss problematik vid utförandet av uppgifter, vilket illustreras tydligt nedan. Här har deltagaren troligen inte sett eller kunnat tolka x- och y-axeln och kan således inte förstå meningen bakom placeringen av klustren. Istället för att öppna det kluster som är längst till höger på läsbarhetsskalan öppnas det kluster som är störst.

Men jag öppnar... det är en som är grön som... som har flest kluster. Som jag öppnar [klickar på det kluster som har flest hexagoner]. Ah! Å då.. när jag klickade på den så.. delades dom upp i massa små.. delar.

Samma problem återkom sedan i nästföljande uppgift (uppgift U2), *I detta kluster: hitta det dokument som har högst läsbarhet och relevans.:*

Aa.. Det måste ju vara de här som är tre som sitter ihop.

och även i uppgift U5, *Hitta den hemsida som har högst läsbarhet enligt OVIX.:*

Men då måste ju den vara störst [håller muspekaren över det största klustret]. Då klickar vi på det klustret så får vi se.

4.3.2 Att växla mellan dokument- och klusternivå

Ett problem som uppstod var att bläddra mellan dokument- och klusternivå (uppgift U3). Inuti ett kluster fanns en vänsterriktad pil vid sidan om texten dokumentnivåsom tar användaren tillbaka till klusternivån. Detta visade sig innebära problem i vissa fall och användarna provade andra möjliga vägar innan de hittade rätt:

Det måste väl vara att bara trycka tillbaka [trycker tillbaka i webbläsaren, fungerar inte]. Aha där! [hittar dokumentnivån omgående, klickar]

En annan deltagare provade att radera det tidigare sökordet:

Okej... okej. Men om jag tar bort det där då [raderar texten i sökrutan] och söker på ingenting. Nä, det kanske inte går [söker, ingenting händer]. Nää... det gick inte. Dokumentnivå då.. [klickar på dokumentnivå]. Yes!

Samtliga deltagare klarade denna uppgift, men majoriteten upplevde vissa problem.

4.3.3 Använda dokumentlistan

Uppgift U6 vållade problem för flera användare. De som klarade uppgiften hade vid tidigare uppgifter använt sig av dokumentlistan och var bekanta med dess funktion, men de deltagare som endast hade använt sig av rutnätet uppvisade vissa svårigheter. Dessa var mer benägna att försöka använda själva rutnätet för att sortera dokumenten efter relevans, exempelvis genom att klicka på -axelns uppåtpil eller i origo. En deltagare antog att de två drop down-menyer var likadana och missade att det var möjligt att välja relevanssom alternativ:

Nä, händer ingenting.. OVIX.. Relevans- NU! Nu klickade jag på relevans [i dokumentlistan]. Jag trodde inte att det var ett alternativ.

4.3.4 Svårfunna läsbarhetsmått

Uppgift U7 bestod av att rapportera läsbarhetsmått, vilket fanns representerat underst på sidan. Denna text krävde att deltagaren scrollade ned. Eftersom LIX-värdet inte var direkt synligt och deltagarna inte hade behövt att scrolla vid tidigare uppgifter innebar detta vissa problem.

4.4 SUS-poäng

SUS-poängen varierade mellan 62,5 och 75 och gav ett medelvärde på 68,5.

5 Diskussion

De uppgifter som fick lägst poäng (0,6) var U1 och U5-U7. Detta var också de uppgifter som deltagarna spenderade mest tid på (>40 sekunder). Dessa fyra uppgifter behandlade sökfunktionen, sortering i dokumentlistan samt tolkning av läsbarhetsmått och återkom även som teman i tänka högt-protokollet.

Problemet med uppgift U1 var svårigheter i att tolka och förstå gränssnittet. Trots att deltagarna ansåg sig ha god datorvana var systemets funktion inte omedelbart intuitivt och förvirring uppstod. Förutom systemets faktiska funktion och syfte verkade den faktor som var mest svårförståelig vara klustrens placering på rutnätet. Detta skulle kunna tydliggöras genom att ha en bättre introducerande text där dokumentklustrarens funktion beskrivs mer ingående. En grundläggande förståelse för hur systemet funkar är grundläggande för att kunna använda det. Detta blev tydligt då en deltagare inte förstod grundsyftet och således inte klarade av att utföra de efterföljande uppgifterna på ett tillfredsställande sätt.

En förändring som kanske skulle göra skillnad är en tydligare x- och y-axel. Detta kunde möjligtvis fånga upp de användare som inte omedelbart förstår att klustrens placering fyller en funktion. Detta skulle kunna förstärkas genom att göra pilarna vid axlarnas ändar tydligare, alternativt förstora upp texten "Relevans" samt "Läsbarhet".

Ett ytterligare sätt att tydliggöra systemets funktion är att implementera ett hjälpavsnitt. Det är många funktioner och begrepp som kan verka krångliga för nya användare, och om denne kan slå upp dessa vid behov elimineras en förvirrande faktor.

Att växla mellan dokument- och klusternivå visade sig också vara problematiskt. Deltagarna tenderade att backa i webbläsaren och en deltagare försökte använda sökrutan för att gå tillbaka till klusternivån. Att den faktiska tillbakaknappen inte var ett förstahandsval hos deltagarna indikerar att den är något otydlig. I den utvärderade prototypen har den aktuella knappen formen av en vänsterriktad pil, placerad invid texten "Dokumentnivå". Det är möjligtvis denna placering och detta ordval som gör det otydligt. En eventuell förbättring kunde bestå i att ändra benämningen till "Gå tillbaka till klusternivå", för att förtydliga knappens funktion eller eventuellt möjliggöra användning av webbläsarens backfunktion.

Dokumentlistans funktion var inte självklar för alla deltagare då fokus tenderade att till största del ligga på rutnätet. Det är möjligt att de två dropdown-menyer kan ha inneburit problem. Om användaren har använt den meny som

endast sorterar efter LIX och OVIX verkar denna vara benägen att anta att även den andra dropdown-menyn endast innehåller dessa två alternativ. Detta skulle kunna lösas genom att ta bort den meny som endast sorterar efter LIX och OVIX och byta ut dessa mot radioknappar. Eventuellt skulle detta öka nyfikenheten mot dokumentlistans sorteringsfunktion och således förenkla användandet av denna. Detta kräver dock att endast ett fåtal läsbarhetsmått implementeras i systemet, då det riskerar att bli en rörig design.

Slutligen fanns det en svårighet i att finna dokumentens läsbarhetsmått. Detta berodde troligen på flertalet faktorer. I uppgifterna användes termen "LIX-värde", och i gränssnittet benämndes det som "läsbarhetsmått". Detta kan ha haft en förvirrande effekt och krävt att deltagarna förstod konceptet LIX för att kunna finna det. Detta är ett problem som eventuellt har sin lösning i en bättre introducerande beskrivning och hjälpavsnitt.

Den skärm som användes var 11", vilket gjorde att hela gränssnittet inte fick plats på skärmen. Detta krävde att deltagarna scrollade ned för att hitta läsbarhetsmåten, vilket troligen utgjorde en försvårande faktor. Detta bör dock inte vara några problem då en större skärm används.

5.1 Metoddiskussion

Antalet deltagare var fem, vilket kan tyckas vara ett lågt antal. Detta val gjordes utefter rekommendationer [1] motiverat med att nya problem sällan visar sig efter de fem första deltagarna. Då antalet funktioner som utvärderades var begränsade samt deltagargruppen var relativt homogen i ålder och utbildningsnivå ansågs fem vara ett tillräckligt antal för att uppfylla studiens syfte.

Systemets slutanvändare kommer framför allt att vara personer med lite vana av informationssökning eller personer med läs- och skrivsvårigheter, och det hade troligen varit fördelaktigt om systemet hade testats mot denna målgrupp. Detta kan möjligtvis ge annorlunda resultat än de som framkommit i denna studie, och är att rekommendera vid vidareutveckling av systemet.

I testet användes ett tänka-högt-protokoll, vilket kan ha påverkat tiden som deltagarna spenderade på varje uppgift. Det råder delade meningar om tid är ett lämpligt mått vid användning av denna metod, då det kan påverka tiden i både en positiv och negativ riktning. Att detta ändå användes motiveras med att de data som behandlade tid inte var bärande för slutsatserna utan endast fungerade som komplement till dessa.

I analysen av tidtagningen fanns en stor spridning i uppgift U6. Konfidensintervallet var 72,7 och värdena varierade från 4,4 sekunder till 186,1 sekunder. Dessa värden innebar problem då det inte fanns något värde som kunde pekats ut som outlier. Samtliga värden var vitt skilda från varandra och då det inte hade blivit mer korrekt vid eliminering av något värde, behölls alla värden.

6 Slutsats

Denna undersökning har ämnat undersöka användarnas upplevelse av en första gränssnittsprototyp och ge förslag på eventuella förbättringar. De ändringar som rekommenderas är följande:

- Förbättrad introducerande text
- Hjälpsnitt
- Tydligare pilar och text vid rutnätets axlar
- Annat ordval vid bläddring mellan nivåer
- Radioknappar för val av läsbarhetsmått under rutnätet

Referenser

- [1] Tullis, Tom och Albert, Bill. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Elsevier Science, 2008.
- [2] Projektet Webblättläst, IDA. <http://www.ida.liu.se/~arnjo/webblattlast/>. Hämtad 2013-08-01.
- [3] Projektet FriendlyReader, IDA. <http://www.ida.liu.se/projects/friendlyreader/>. Hämtad 2013-08-01.
- [4] Projektet Webbklustring, IDA. <http://www.ida.liu.se/projects/webbklustring/>. Hämtad 2013-08-01.

A Gränssnitt: Klusternivå

Webbklustring Sök Katter 40

Webbklustring är ett verktyg för att enkelt dela upp ett sökresultat i olika grupper (kluster) beroende på likhet mellan hemsidorna. Kluster placeras ut på rutnätet beroende på dess läsbarhet och relevans. 2

Klicka på ett kluster för att öppna det och se de enskilda hemsidorna.

Katter Läsbarhet LIX 3

Klusternivå

Dokument

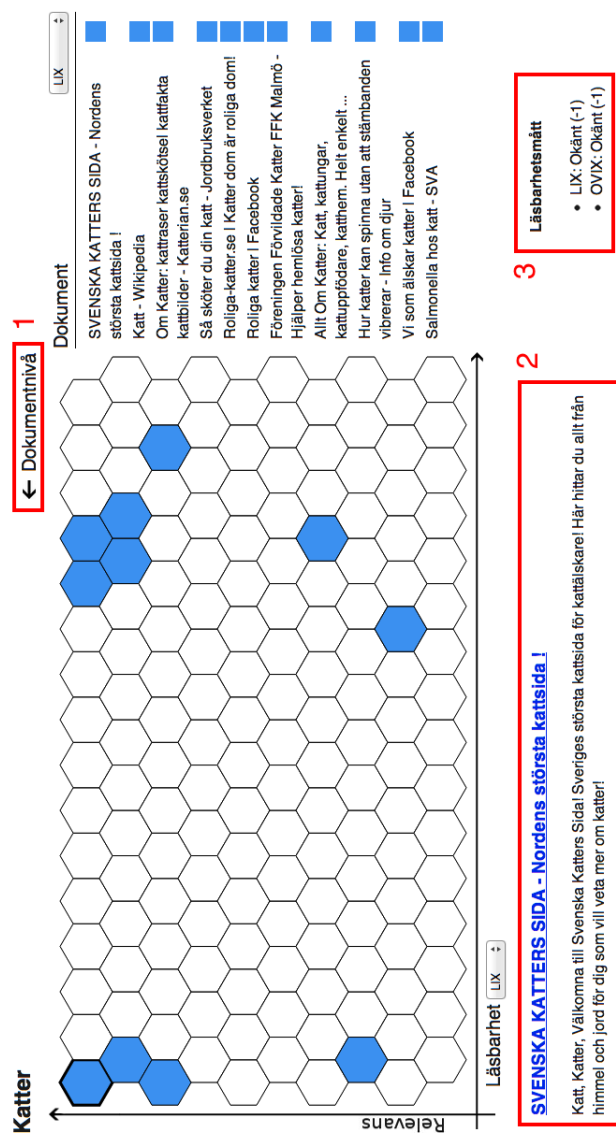
- SVENSKA KATTERS SIDA - Nordens största kattsida!
- Katter i Fokus - Artiklar
- Söta katter - YouTube
- DjurID.se - SKKs Ägarregister - Svenska Kennelklubben
- katter, kattprylar, kattartiklar på Katter.se
- Katter-spel - Spelo
- Föreningen för Höörs Hemlösa Katter
- Katt - Wikipedia
- Om Katter: kattrares kattsötsel kattraaktakattbilder - Katterian.se
- Så sköter du din katt - Jordbruksverket
- Roliga katter.se | Katter dom är roliga dom!
- Roliga katter | Facebook
- Föreningen Förvildade Katter FFK Malmö - Hjälper hemlösa katter!
- Allt Om Katter: Katt, kattungar,

Läsbarhet LIX 4

Figur 3: Det utvärderade gränssnittet på klusternivå efter en sökning på "katter".

1. Sökrutan, 2. Reglaget som reglerar antalet sökresultat, 3. Dropdown-menyn som sorterar dokumentlistan (innehåller värdena *Relevans*, *LIX* samt *OVIX*), 4. Dropdown-menyn som sorterar rutnätet (innehåller värdena *LIX* samt *OVIX*).

B Gränssnitt: Dokumentnivå



Figur 4: Det utvärderade gränssnittet på dokumentnivå efter en sökning på "katter". 1. Knappen som bläddrar tillbaka till klusternivå, 2. Snippet av valt dokument, 3. Läsbarhetsmätt.

C Instruktioner och uppgifter

Instruktioner:

Du kommer få ett antal uppgifter att utföra på en hemsida. Jag vill att du, under tiden du utför uppgifterna, pratar om vad du tittar på, tänker på och hur du resonerar. Tanken är att jag ska kunna följa din tankegång under utförandet av uppgifterna.

Du får avbryta när du vill och resultaten kommer inte att kunna spåras till dig.

Uppgifter:

U1. Sök efter "Linköping" och öppna det kluster som har högst läsbarhet enligt LIX.

U2. I detta kluster: hitta det dokument som har högst läsbarhet och relevans.

U3. Gå tillbaka till de ursprungliga klustren.

U4. Sök efter "Katter". Begränsa antalet sökresultat till 70.

U5. Hitta den hemsida som har högst läsbarhet enligt OVIX.

UX. *Gå tillbaka till de ursprungliga klustren*

UX. *Sök efter "Ibuprofen". Maximera antalet sökresultat.*

U6. Sortera alla sökresultat efter hur relevanta de är.

U7. Hitta den hemsida som har högst läsbarhet enligt LIX. Hitta LIX-värdet för denna hemsida.