

# Slutrapport för projektet Vidareutveckling av Folkets lexikon

Viggo Kann  
viggo@algoritmica.se

Joachim Hollman  
joachim@algoritmica.se

20 september 2011

## 1 Inledning

Internet har gjort det möjligt att skapa stora verk (som lexikon) genom att massor av människor drar var sitt strå till stacken. Ett verk som kräver en enorm arbetsinsats kan därmed skapas utan att varje enskilt bidrag behöver vara så stort. Ju fler personer som lämnar bidrag desto mindre behöver varje bidrag bli. På engelska kallas detta för crowdsourcing. Det mest kända verket som byggts upp på detta sätt är encyklopedin Wikipedia.

År 2008 startade projektet Folkets engelsk-svenska lexikon och i mars 2009 gjordes detta Folkets lexikon tillgängligt för allmänheten på <http://folkets-lexikon.csc.kth.se>

Detta stora svensk-engelska och engelsk-svenska lexikon kan utvidgas av folket genom att användarna bedömer översättningsförslag som presenteras automatiskt när man gör uppslagningar i lexikonet.

I detta fortsättningsprojekt, som startade i oktober 2009, har Folkets lexikon utvecklats på många sätt. Projektet blev klart sommaren 2011.

## 2 Mål och syfte

Projektet har fyra målgrupper:

1. Internetanvändare som behöver slå upp i ett svensk-engelskt eller engelsk-svenskt lexikon.

2. Språkintresserade Internetanvändare som vill hjälpa till att utveckla lexikonet.
3. Utvecklare av fri programvara som behöver ett fritt tillgängligt svensk-engelskt, engelsk-svenskt eller svenskt lexikon.
4. Forskare som studerar lexikonanvändning.

Syftet är att tillfredsställa alla målgrupperna genom förbättringar av Folkets lexikon.

Inledningsvis hade projektet följande mål:

- Utvidga Folkets lexikon med två andra fria lexikonresurser: Folkets synonymlexikon och Saldo, Svenskt associationslexikon.
- Utveckla stöd till forskare som vill studera hur Folkets lexikon används. Samla statistik för både uppslagningar och inställningar.
- Inför klickbara länkar inom lexikonet.
- Inför fritextsökning.
- Konvertera huvudorden till grundform (i Lexin används verbens presensform som uppslagsord).
- Skapa ett system där volontärer kan granska uppslagsord och säga okej till andra användares ändringar.
- Skapa ett forum där användare kan diskutera lexikonet.

Under projektets gång byttes de två sista målen, i samråd med .SE-stiftelsen, mot följande två nya mål:

- Utveckla möjligheter för användarna att utvidga och korrigera lexikonets övriga fält, inte bara översättningarna av huvudordet. Fyll i ordklass automatiskt, om möjligt.
- Utveckla en webbservice så att andra tillämpningar kan anropa Folkets lexikon.

### 3 Projektbeskrivning

Projektet har drivits av Viggo Kann och Joachim Hollman på Algoritmica Hollman & Kann HB och finansierats av IIS Internetfond. Det bygger på det ursprungliga projektet Folkets engelsk-svenska lexikon och har till stor del genomförts parallellt med det.

Patrik Lilja har utvecklat utvidgningsdelen inom ett examensarbete i datalogi för civilingenjörsexamen vid KTH.

Tiotusentals Internetanvändare har bidragit genom att rösta på översättningsförslag och själva bidragit med förslag.

### 4 Leverabler

1. Stöd till användningsstudier  
För att forskare ska kunna hur Folkets lexikon används har en statistikinsamling av både uppslagningar och inställningsändringar införts. Alla bedömningar och användarförslag har redan tidigare lagrats. Uppslagningsstatistiken visas månadsvis på Folkets lexikons webbsida (under *Om Folkets lexikon*).
2. Inkorporering av Saldo och Folkets synonymlexikon  
Alla synonymer i Folkets synonymlexikon

med en bedömning över 3 på den femgradiga synonymhetskalen har införts i det svensk-engelska lexikonet. Det är över 37000 synonymer. Över 25000 referenser till Svenskt associationslexikon Saldo har införts i det svensk-engelska lexikonet. I Saldo, som också finns fritt tillgängligt på webben, finns bland annat ordens alla böjningsformer och associationer. Eftersom det är så mycket information och eftersom Saldo uppdateras kontinuerligt har vi valt att lägga länkar till Saldo istället för att importera lexikonets information i Folkets lexikon.

3. Konverterade huvudord  
I Lexin används verbens presensform som uppslagsord. Eftersom detta inte är standard för lexikon utan snarare ett särdrag för Lexin så har vi konverterat alla huvudord till grundform. Den som söker på presensformen kommer ändå att få träff eftersom vi lagt in presensformen bland böjningsformerna i lexikonet.
4. Fritextsökning  
För att göra det möjligt att söka även på delar av lexikonet som inte är indexerade som sökord så har vi infört fritextsökning. Då användaren skriver in ett eller flera ord och klickar på knappen *Fritext* så visas alla poster i lexikonet som innehåller dessa ord eller uttryck.
5. Uppslagningstjänst för externa tillämpningar  
Vi har utvecklat en webbservice så att andra tillämpningar kan anropa Folkets lexikon, göra en uppslagning och få svaret tillbaka i HTML. Formatet för anropet beskrivs på webben under *Om Folkets lexikon*.
6. Möjlighet för användarna att utvidga och korrigera lexikonets övriga fält

Folket vill kunna påverka inte bara översättningarna av huvudordet utan även resten av uppslagsposterna. Detta är den största leverabeln. Den består av utveckling av ett användargränssnitt för utvidgningar och ändringar, en metod som sällar bort skräpinskiökningar, en ordklassgissare som automatiskt föreslår ordklass på både svenska och engelska ord, samt en funktion som delar upp lexikonändringarna i mindre delar och presenterar dessa för andra användare i bedömningsförslag.

## 5 Resultat

De sex leverablerna ovan har levererats. Resultaten beskrivs kort ovan för varje leverabel. För en mer detaljerad resultatbeskrivning hänvisas till Patrik Liljas exjobbssrapport, kapitel 6.

Folkets lexikon finns tillgängligt för alla på <http://folkets-lexikon.csc.kth.se>

## 6 Utvärdering och analys

### 6.1 Utvärdering av resultat

Folkets lexikon är mycket välanvänt och tycks uppskattat av användarna. Användningsstatistik finns på webbsidan (under Om Folkets lexikon). I maj 2011 gjordes cirka sju miljoner uppslagningar i lexikonet.

Målen med projektet har uppfyllts mycket väl, men det har tagit längre tid än beräknat att genomföra hela projektet. Efter ett år av projekttiden insåg vi att vi inte hade tid att genomföra återstående mål. Då lyckades vi engagera den mycket duktiga Datateknikstudenten Patrik Lilja, som gärna åtog sig att utveckla det stora utvidgnings- och ändringssystemet i form av ett examensarbete vid KTH. Patrik blev klar våren 2011, och hans fina resultat bevisade att en utomstående person kan sätta sig in i systemet och bygga vidare på det.

### 6.2 Förslag på förbättringar

Se Patrik Liljas exjobbssrapport, kapitel 7.

## 7 Framtida arbeten

Folkets lexikon är gjort så att det ska kunna fortleva med minimalt underhåll. Folkets egna förslag och bedömningar gör att innehållet utvecklas och förhoppningsvis blir både större och bättre. Vid årsskiftet 2011/2012 kommer även XML-koden till det svensk-engelska lexikonet att göras tillgänglig för nedladdning.

Det finns redan flera webbsidor och appar som använder Folkets lexikon, antingen webbservicen eller som har laddat ner lexikonet som XML.

Översättarna av ett svensk-franskt lexikon som inte längre trycks har hör av sig till oss och skulle vilja göra det tillgängligt som Folkets svensk-franska lexikon. Om vi kan få någon som sponsrar så skulle det vara roligt att göra ett sånt lexikon också.

## 8 Referenser

Lars Borin, Markus Forsberg & Lennart Lönngrén: SALDO 1.0, Språkbanken, Göteborgs universitet, 2008,

[http://spraakbanken.gu.se/personal/markus/publications/saldo\\_1.0.pdf](http://spraakbanken.gu.se/personal/markus/publications/saldo_1.0.pdf)

Viggo Kann & Magnus Rosell: Free construction of a Swedish dictionary of synonyms, NoDaLiDa 2005, Joensuu, 2005,

<http://www.nada.kth.se/theory/projects/infomat/rapporter/kannrosell105.pdf>

Patrik Lilja: Folket har makten över ordet – ett ordboksredigeringsystem för allmänheten, Examensarbete, KTH CSC, 2011, TRITA-CSC-E 2011:071,

[http://www.nada.kth.se/utbildning/grukth/exjobb/rapportlistor/2011/rapporter11/lilja\\_patrik\\_11071.pdf](http://www.nada.kth.se/utbildning/grukth/exjobb/rapportlistor/2011/rapporter11/lilja_patrik_11071.pdf)