# Status for WP3.2-søgemaskinen. T30.

Selve søgemaskinen er færdig og kan leveres til DKClarin som en standalone applikation, evt. med flere databaser tilknyttet, hver baseret på et på forhånd udvalgt sæt TextGrid eller Chat-filer konverteret til TextGrid.

Siden sidst har den været igennem en gennemgribende refaktorering og er nu baseret på Spring MVC-frameworket.

Førstnævnte betyder, at WP3.2. kan lever søgemaskinen ’i eksportabelt format’ i ovenstående betydning. DGCSS har testet, at en sådan leverance er mulig ved at levere en sådan, standalone, tilpasset søgemaskine til DanPass-projektet.

Sidstnævnte betyder, at søgemaskinen nu har et antal nye funktioner:

1. En statistikside med overordnede oplysninger om det korpus, der søges i. Dvs. ting som antal ord og typer i korpus og den slags.
2. Ved søgning på et enkelt ord i ortografen, kan man vælge at få vist kollokationer for ordet med MI-score beregnet (på samme måde som KorpusDK har det). Resultatet kan downloades som csv-fil.
3. Ordfrekvenser og log-likelyhood kan vises for hele korpus, en enkelt fil eller et af brugeren defineret sæt af filer. Resultatet kan downloades som csv-fil.
4. En statuslinie med meddelelser om databassens status på visningstidspunktet, f.eks. om der er en opdatering af korpus i gang
5. En voldsomt forbedret visning af søgeresultater, afspilning af underliggende lyd o.a.
6. Mulighed for download af søgeresultatet, såvel som en csv-fil af alle slices den bagvedliggende TextGrid i resultatområdet som en csv-fil kun med de valgte tiers, men til gengæld med et valgfrit antal slices før og efter (konkordansvisning).

Der er ikke nyt at bemærke vedr. leverance af inputintegreret søgemaskineværktøj, idet WP5 ikke vendte tilbage på T18-rapporten eller på det tidligere fremsendte notat desangående. Deri blev det gjort klart at det var en forudsætning for at søgemaskinen kunne integreres med infrastrukturen at

1. Leveringsservice (2.8.4 i ”Clarin kravspec 051009.doc”)
2. Service til søgning i metadata (ibid. 2.8.2).

var færdige eller i al fald specificeret tilstrækkeligt præcist til at man kunne gå i gang med udviklingsarbejdet.

Dersom der ønskes en service/funktion, hvorved søgemaskiner baseret på DGCSS’ søgemaskine dynamisk kan genereres og gøres tilgængelige indenfor projektets deadline, skal WP5 øjeblikkeligt tage kontakt til WP. 3.2. v. undertegnede og aftale det videre forløb, herunder helt specifikt hvilke servicekontrakter servicen skal kunne indgå i og der skal etableres en (minimal) følgegruppe med en repræsentant for hhv. WP3.2. og WP5.2. til kommunikation og test af integrationen mellem de to platforme.

En sådan service kunne f.eks. fungere sådan, at infrastrukturen vil kunne få udleveret et link til en netop genereret søgemaskine, som kan søge i et sæt TextGrids fremfundet ved en nylig metadatasøgning, men så skal WP5 have indbygget at kunne tilspørge en sådan web-service vha. ovennævnte kontrakt og siden kunne omdirigere brugeren til det returnerede link (eller sende det til hende i en mail, hvis hun ikke gider vente foran skærmen mens den bliver genereret).

Et tool som vores søgemaskine, vil så kunne programmeres til at lytte efter sådanne web-service-kald fra infrastrukturen, bruge en anden webservice suppleret af WP5 til at hente TextGrid-filerne ud af infrastrukturen, generere en søgemaskine ud fra dem og returnere et link hertil, igen på eksakt den aftalte facon.

f. DGCSS,   
Michael Barner-Rasmussen  
IT Ansvarlig  
DGCSS, Nordisk Forskningsinstitut (NFI)  
Njalsgade 136, lok. 27.5.61  
2300 København S  
Tlf. Institut: 35 32 83 82  
Tlf. Direkte: 31 25 31 49