KTH Data- och systemvetenskap

TvärSök - Tvärspråklig sökning på skandinaviska

Hercules Dalianis

Department for Computer
and System Sciences

Stockholms University and KTH

Email:hercules@kth.se

KTH Data- och systemvetenskap

TvärSök - Tvärspråklig sökning på skandinaviska

- Partners
- KTH - Stockholm University
- University of Bergen
- CST-Copenhagen University
- + kommande isländska och finska partners

KTH Data- och systemvetenskap

Bakgrund

- Internet har fjärmat de skandinaviska språken
- Sökning sker på modersmål samt engelska
- För att hitta information på de skandinaviska språken krävs aktiva kunskaper i grannspråken
- Skandinaver missar information på våra grannspråk

KTH Data- och systemvetenskap

Domäner

- Arbetssökande
- Studerande
- Nordiska företag inom banker, byggbranchen, turism m.m.

KTH Data- och systemvetenskap

Mål

- En tvärspråklig sökmotor för danska, norska och svenska
- Lärande av skandinaviska genom att använda språken
- 20 miljoner skandinavisktalande personer i Norden
- 1-2 miljoner skandinavisktalande personer utanför Norden

KTH Data- och systemvetenskap

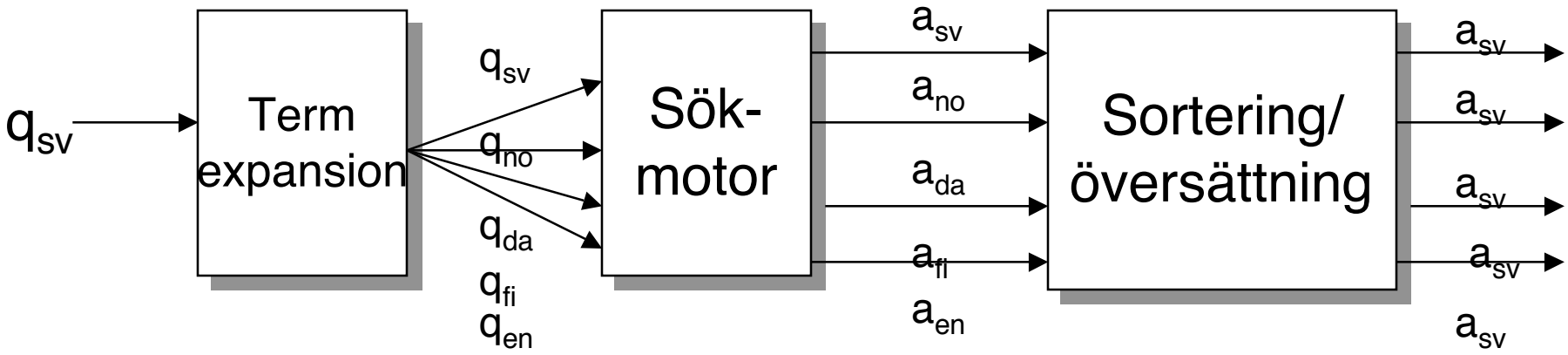
Tekniker för TvärSök

- Infrastruktur: SiteSeekersökmotorn och Eurolings servrar
- Språkigenkänning för 40 europeiska språk inkl danska, norska, finska, isländska, samiska, lettiska, litauiska, estniska, ryska och engelska.
- SiteSeeker har stemming för svenska, danska och engelska.
- SiteSeeker har dynamiskt stavningsstöd

KTH Data- och systemvetenskap

Lexikonuppslagning

- Tillgängliga nordiska översättningslexikon som Skandlexikon 2004 och NordTerm 2004.
- Domänlexikon skapas ”on the fly” genom termkoppling
- Indata texter från webbplatsen(rna)
- Jämföra parallella eller icke parallella texter


q_{sv}
Term expansion
q_{sv}
q_{no}
q_{da}
q_{fi}
q_{en}
Sök-motor
a_{sv}
a_{no}
a_{da}
a_{fi}
a_{en}
Sortering/ översättning
a_{sv}
a_{sv}
a_{sv}
a_{sv}
a_{sv}

KTH Data- och systemvetenskap

- Frågeexpansion eller termexpansion på flera språk?
- Sökträffar på flera språk
- Hur sorterar man träffarna?
- Per språk och relevans?
- Per relevans och språk?
- Flera fönster - ett språk per fönster

KTH Data- och systemvetenskap

Fuzzy matching

- Scandinaviska språken liknar varandra
- Fuzzy matching av liknande sökord
- Förslå ord på de andra språken


Sök efter:
maskin översättele
Hitta!
Avancerad sökning »
Resultat: Inga träffar på maskin och översättele inom Alla Nordoknets webbsi
Sökfrågan kan vara felstavad
Menade du maskinoversættelse eller maskinöversättning?
Jag vill söka efter maskinoversættelse.
Jag vill söka efter maskinöversättning.
POWERED BY SiteSeeker

KTH Data- och systemvetenskap

Random Indexing

- Random indexing skapar från parallella texter ett översättningslexikon.
- Random indexing mer effektiv än Latent semantic indexing

KTH Data- och systemvetenskap

- En bra fungerade tvärspråklig sökmotorsdemonstrator skall finnas tillgänglig på nätet senast januari 2006.