




Språkteknologin i industrin

Hercules Dalianis
DSV-KTH-Stockholms universitet
email:hercules@kth.se
070-568 13 59 / 08-674 75 47

Hercules Dalianis
1




Översikt

- Maskinöversättning
- Naturligt språkgränssnitt
- Taligenkänning
- Stavningsstöd, lexikon
- Textsammanfattning
- Omvärldsbevakning
- Informationssökning

Hercules Dalianis
2




Maskinöversättning

- 1950-talet, ryska-engelska
- 1960-talet, SYSTRAN (1968)
- Systran-Frankrike, Babelfish,
- Wordlingo, Las Vegas, USA.
- Prompt, ryskt företag
- Esteam,
 - Islandsägd svenskt företag i Aten, Grekland

Hercules Dalianis
3




Naturligt språkgränssnitt

- Till databaser 1980-talet, NL-SQL->NL
 - Silvia (IBM Sverige)
 - HSQL (Nordiska ministerrådet)
- Generering av NL-rapporter från data.
 - Formella specifikationer Ericsson
 - Väderdata,
 - SMHI,
 - Börsinformation
- Cogentex, Kanada, <http://www.cogentex.com/>
- Data2Text, Skottland, <http://www.data2text.com/>

Hercules Dalianis
4




Taligenkänning

- Boka resor, Televerket, (Telia), SJ.
- Diktering
 - Philips FreeSpeech,
 - Dragons Dragon Dictate
 - VoiceExpress (Tidigare från Lernout & Hauspie)
- Telephone commander
 - Philips SpeechPearl, Nuance
 - Bilprovningen, Ringa och fråga om bilregistreringsuppgifter.

Hercules Dalianis
5




PipeBeach - HP

- PipeBeach (från Rörstrandsgatan)
- HP OpenCall Media Platform
- <http://h20208.www2.hp.com/opencall/products/media/ocmp/index.jsp>
- VoiceXML
- SiteSeeker Voice-Talstyrd sökmotor

Hercules Dalianis
6

Stockholms universitet

KTH Data och systemvetenskap

- Översättning tal till text
 - Virage.com (ett Autonomyägd företag) använder
 - <http://www.bbn.com/speech/>
- Några länkar
 - http://en.wikipedia.org/wiki/Speech_recognition

Hercules Dalianis 7

Stockholms universitet

KTH Data och systemvetenskap

Talsyntes

- Readspeaker läser upp webbplatsers innehåll
- Främst offentlig sektor

Hercules Dalianis 8

Stockholms universitet

KTH Data och systemvetenskap

Stavnings- och grammatiksystem

- Lingsoft, Microsoft Word, finska, svenska, danska och norska
- Viggo Kanns - Stava
- Stavningsstöd i Gula sidorna

Hercules Dalianis 9

Stockholms universitet

KTH Data och systemvetenskap

Textsammanfattning

- Nyheter i flera format, tidning, webb, wap, sms, tal.
- Omvärldsbevakning, Business Intelligence

Hercules Dalianis 10

Stockholms universitet

KTH Data och systemvetenskap

Namnigenkänning-lexikon

- Automatisk namnigenkänning
 - Named entity tagging/recognition
 - Svenska, engelska, kinesiska...
- Terminologilistor
 - Sjukvård
 - Tillverkningsindustri
 - Automatisk tillverkning av sådana för översättning och för domänkunskap

Hercules Dalianis 11

Stockholms universitet

KTH Data och systemvetenskap

Omvärldsbevakningsföretag

- Cision (Observer), Affärsdata, Retriever, Agent25, mm
- Ungefär 10 företag i Sverige
- 1 000 anställda?
- Omsätter 1 miljard?
- 50 000 kunder?

Hercules Dalianis 12




Text mining

- Text mining (Text brytning)
- Sentimentanalys
- Blogganalys
- Produktutvärdering
- Bevakning av bloggar

Hercules Dalanis
13




Informationssökning

- Globala sökmotorer
 - Google, Altavista, Vivisimo, Kartoo,
- Lokala sökmotorer för webbplatser och intranät
- SiteSeekersökmotorn från Euroling
 - Över 200 kunders webbplatser och intranät
- Picsearch-bildsök bildtolkning

Hercules Dalanis
14




Informationssökning

- Intelligent agenter
 - Hanna på Försäkringskassan
- Artificial Solution
- Humany
- Automatiskt e-postbesvarande
 - 30 procent av e-posten kan kapas

Hercules Dalanis
15




SiteSeekersökmotorn

- Innehåller språkteknologi
 - Stemming / Ordböjning
 - Stavningsstöd
 - Synonymer
 - Textextrakt
 - Tvärspråklig sökning
- Men också sökstatistik

Hercules Dalanis
16




Marknad

- Sverige- 10 miljoner invånare
- Tyskland - 80 miljoner invånare
- USA - 300 miljoner invånare
- Kina - 1 300 miljoner invånare

Hercules Dalanis
17




Marknad

- Få ut produkten på marknaden
- Marknaden mogen
- Lobba för teknologin
- Nyttä kontra kostnad
- Konkurrenter

Hercules Dalanis
18



Slutsatser

- Gör omvärldsbevakning
- Hitta nya användningsområden
- Förbättra dom nuvarande produkterna.
- Ställ krav på produkter.

Hercules Dalianis

19



Exjobb

- EMAIL
 - <http://syslab.dsv.su.se/group/imal/forum/topics/three-master-thesis-proposals>
- KEA
 - <http://dsv.su.se/forskning/ithalsa/projekt/kea/>

Hercules Dalianis

20