




Textextraktion och automatisk textsammanfattning

Hercules Dalianis
DSV-SU-KTH
e-post:hercules@kth.se
070-568 13 59 / 08-674 75 47

Hercules Dalianis 1




Overview of talk

- Background
- Other summarizers
- Technique
- Future improvements
- Applications
- Evaluation

Hercules Dalianis 2




Automatic text summarization

- is the method where a computer summarizes a text.
- An extract from a longer original text.
 - A text is given to the computer and it returns a non-redundant shorter text
- The technique has it's roots in the 60's.
- With the Internet and the WWW it has been an awakening interest in summarization techniques.

Hercules Dalianis 3




Summarization tools

<http://www.dsv.su.se/~hercules/HDbookmarks.html>
http://www.ics.mq.edu.au/~swan/summarization/projects_full.htm

- Microsoft Word 97, 98 and Word 2000 have a summarizer for documents.
- Intelligent Miner for Text - Summarization tool IBM
- Inxight (XEROX)

Hercules Dalianis 4




- Corpporum Summarizer- Cognit AS (Norway)
- Pertinence (France)
- Copernic Summarizer
- Automated Text Summarization (SUMMARIST)
- Columbia Newsblaster
<http://www1.cs.columbia.edu/nlp/newsblaster>
- OracleContext
- Safari web reader for Macintosh under "Safari->Services"

Hercules Dalianis 5




What is Automatic summarization good for?

- Search engine - extracts in hit lists
- Business Intelligence- survey news flow
- Translation - make the text shorter before translating the text
- Summarize news for SMS, WAP, 3G-format
- News paper setting and printing
- Speech synthesis - summarize text before synthesize. Text-To-Speech

Hercules Dalianis 6

KAROLINSKA INSTITUTET
Sök

Sök på Karolinska Institutets externa webbplatser

Sök efter:

Avancerad sökning »

Resultat: 141 träffar på stamcell inom Karolinska Institutet

Sortera efter: Relevans Datum

- Regeringen backar om stamceller - Karolinska Institutet**
« Tillbaka - 2004-06-03 - Regeringen backar om stamceller - Källa: nys teknik.se, hämtad 20040603 - Helena Mayer Informationsavdelningen - « Tillbaka - Länkar - http://www.nys teknik.se/pub/psart.asp?art_id=35106 ...
Beskrivning: Regeringen backar om stamceller.
===== Info.ki.se/article, se.html?ID=1446
HTML-dokument - 2004-06-15 - Visa med sökorden markerade
- Särskilda stamceller bakom cancer - Karolinska Institutet**
« Tillbaka - 2004-06-04 - Särskilda stamceller bakom cancer - Källa: Vetenskapsraden, hämtad 20040604 - Helena Mayer Informationsavdelningen - « Tillbaka - Länkar - Läs mer på sr.se - Sidan ...
===== Info.ki.se/article, se.html?ID=1448
HTML-dokument - 2004-06-15 - Visa med sökorden markerade
- Stamceller - Karolinska Institutet**
« Tillbaka - 2004-06-04 - Särskilda stamceller bakom cancer - Källa: Vetenskapsraden, hämtad 20040604 - Helena Mayer Informationsavdelningen - « Tillbaka - Länkar - Läs mer på sr.se - Sidan ...
===== Info.ki.se/news/items/stemcell_lines_se.html
HTML-dokument - 2004-06-15 - Visa med sökorden markerade
- Center f Hearing and Communication Research - Karolinska Institutet**
to ki se - search ki se - search stuff at ki se - NEWS - RESEARCH - STAFF ... med Mats Ulfendahl om pågående forskning kring bl.a. stamceller. I framtiden kanske läkarna kan bota hörselskador? - Welcome to ...
Beskrivning: Center for Hearing and Communication, Karolinska Institutet, Stockholm, Sweden. Center för Hörsel- och

KAROLINSKA INSTITUTET
Sök

Sök på Karolinska Institutets externa webbplatser

Sök efter:

Avancerad sökning »

Resultat: 6 träffar på stamcell, lars, åhrund och richter inom Karolinska Institutet

Sortera efter: Relevans Datum

- Cell till Samhälle - Karolinska Institutet**
« Sökesresultat « Forskningsområde: Molekylärbiologi - Embryologi ... odling av embryonala stamceller. - Den preimplantationa embryologi ... mer om tidig utveckling. Lars Åhrund-Richters forskning inom embryologi har ... började Lars Åhrund-Richter bygga upp Sveriges första ... frysta musenbyon, menar Lars Åhrund-Richter, eftersom det minskar användningen ...
===== Info.ki.se/... /professors/article, se.html?ID=117&Area=Molekylärbiologi
HTML-dokument - 2004-06-15 - Visa med sökorden markerade
- Stamceller - Karolinska Institutet**
Svenska stamcellslinjer - Intresset i USA för svensk stamcells forskning är ... forskare, säger Lars Åhrund-Richter, professor vid kliniskt forskningscentrum ... med anslag från NIH, säger Lars Åhrund-Richter. Det är dock alls inte självklart ... universitetssjukhus, säger Lars Åhrund-Richter. I dagläget finns ingen reglering ... under utveckling. Våra stamceller är under utvärdering. Vi vet ...
===== Info.ki.se/news/items/stemcell_lines_se.html
HTML-dokument - 2004-06-15 - Visa med sökorden markerade
- Cell till Samhälle - Karolinska Institutet**
« Tillbaka till startsidan - Sortering på efternamn : - « Tillbaka till ... kirurgiska ingrepp - Lars I Eriksson - Professor i Anatomi ... i Patologi - Cancer - stamceller som tappat kontrollen? - Hans-Olov ... räddar liv - Lars Åhrund-Richter - Professor i Molekylär ...
===== Info.ki.se/... /professors/lista_se.html?Bokstav=...
HTML-dokument - 2004-06-15 - Visa med sökorden markerade

KTH
KTH Data och systemvetenskap

Summarization approaches

- Single-document summarization
 - One document summarized
- Multi-document summarization
 - Several documents summarized to one
- Multi-lingual summarization
 - Several documents in several languages to one doc.
- Multimodal summarization
- User adapted summary-Slanted summary
 - Summarize around some key-concepts

Hercules Dalanis 9

KTH
KTH Data och systemvetenskap

Extract not Abstract

- Extraction is much easier than abstraction
- Abstraction needs understanding and rewriting
- Most automatic summarization tools makes extracts not abstracts
- Uses original sentences or part of sentences to create "abstract"

Hercules Dalanis 10

KTH
KTH Data och systemvetenskap

Algorithm

- Find what the text is about
- Then decide what so say
- Then decide how to say it

Hercules Dalanis 11

KTH
KTH Data och systemvetenskap

SweSum

- SweSum is the first summarizer for Swedish news text
- Started in 1999
- Is available on the Internet

Hercules Dalanis 12




Techniques

- Remove all sentences except the first one and the title.
- Done!

Hercules Dalanis
13




Techniques

- Text summarization (extraction) uses statistic, linguistic and heuristic methods
- A text is divided into sentences
- Sentence positions (News/Reports)
- Title words
- Bold text, Numerical values, Citations
- Named Entities (Frequency based)

Hercules Dalanis
14



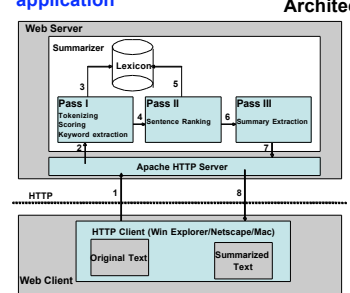

Key word lexicon

- Keyword extraction and frequency
- Language specific
- Domain specific
- Also called "open class word"
 - nouns, adverbs, adjectives
- Use morphological information-lemma
 - skatt, skatterna, skatten, skattens .. => skatt
 - öka, ökning, ökningens.. => öka
- Key words in news domain

Hercules Dalanis
15




A Client/Server web-based application SweSum Architecture



The diagram illustrates the SweSum architecture. It shows a Web Server and a Web Client connected via HTTP. The Web Server contains a Summarizer, a Lexicon, and an Apache HTTP Server. The Summarizer consists of three passes: Pass I (Tokenizing, Scoring, Keyword extraction), Pass II (Sentence Ranking), and Pass III (Summary Extraction). The Lexicon is connected to Pass I and Pass II. The Apache HTTP Server connects to the Web Client. The Web Client contains an HTTP Client (Win Explorer/Netscape/Mac) which sends the Original Text to the Web Server and receives the Summarized Text back.

Figure by Nima Mazdak

Hercules Dalanis
16




Techniques

- User adaptation
 - Utilize user keywords - Obtain slanted summaries
- Combination function of all rankings with different weights gives the rank of each sentence.
- Generate all high ranking sentences
- Voilà the summary !

Hercules Dalanis
17




SweSum

- Summarizes Swedish news paper text in HTML/text format on the WWW.
- Uses a Swedish key word lexicon that contains 40 000 words and their possible 700 000 inflections.
- During the text summarization are 5-10 key words produced which describes or categorizes the text
- Key words - A miniature summary.

Hercules Dalanis
18




The Swedish keyword lexicon

SwoSum - Automatic Text Summarizer with NE Recognition by [Martin Hassel](#) and [Hercules Dalianis](#)
Localization, Interfaces and Swedish Prominential Resolution by [Marta Hassel](#)

 [På svenska, tack!](#) [Lester options, please!](#) [\[URL\]](#)

Please type or paste a text of your own to summarize:

Alternatively, you can upload a text/HTML file from your own computer:

[Browse ...](#)

Keywords that may be important for the text Choose type of text Newspaper Choose language of the text Swedish

Summary of the original text: 30 percent

Print keywords and statistics ☒ Number of keywords: 10
 Use previous resolution (only for Swedish)

Set weights for discourse parameters:

First line	Bold	Numeric values	Keywords	User keywords	Named entities
1000	10	1.133	0.360	500	2

Choose which named entities should be tagged and weighted (named for Swedish):

☒ Personal names ☒ Places ☒ Companies/organizations ☐ Dates, weekdays, etc.

☒ Mark with colors ☐ List named entities

[Summarize](#)



Summarization/Compression

- 30 percent summarization rate
 - 70 percent compression rate (remove 70 percent of text)
- Gives 80 percent intact information
- ~70 percent is redundant information

The slide features a white background with a blue header bar at the top. On the left side of the header is the KTH logo, which includes a blue square with a white crown and the text 'KTH' and 'KTHENSKA TEKNISKA HOGSKOLEN' below it. On the right side of the header is the Stockholm University logo, a circular emblem with a blue border containing the text 'STOCKHOLM' and 'UNIVERSITY' and a central blue figure. The title 'SweSum' is centered in a large, bold, black font. Below the title is a bulleted list with two items: 'SweSum is available for 8 languages' and 'SweSum is available to summarize news texts in Swedish, Danish, Norwegian, English, Spanish, French, German and in Farsi (Iranian)'. At the bottom of the slide, there is a blue horizontal bar containing the text 'Hercules Dalanis' on the left and the number '22' on the right.

The slide features a blue header with the KTH logo on the left and a circular seal on the right. The main title is 'Users on the www'. Below it is a bulleted list of statistics.

- <http://swesum.nada.kth.se>
- SweSum has 2 200 visitors per month since March 2002
- 1 863 unique visitors (100 per month)
- A large amount of that are Spanish and American universities

The footer contains the text 'Hercules Dalianis' and the page number '23'.

[illegible]

Cellfynd revolution för blodsjuka	Konstgjort händer smitta (SvD)	
(Baltimore, SvD) Nu har en forskargrupp från Philadelphia i USA klarat det som många trodde var en omöjlighet. "Barnen" - miljontals andra blodceller - dör snabbt och försvinner. Stamcellerna bildar ständigt nya. Forskare har i åratal jagat metoder för att försöka hitta och identifiera dessa bennärrens ursprungsceller. De tänkbara applikationerna är svindlande: de kan användas för att utfylla stamcellerna oöda från enorma mängder av alla sorters blodceller i laboratoriet och därmed ersätta dagens blodtransfusioner där blodet tas direkt från människor, alltså främstallt konstgjort blod. Om stamcellerna kan identifieras blir hela denna procedur betydligt enklare och mer effektiv, eftersom egentligen enbart stamcellerna behövs för att återuppsbygga ett nytt fungerande immunförsvar. Generapi testas också för att försöka skapa blodceller som inte kan infekteras med eller som kan bekämpa hivus. Den stora stötestenen för generapi på blodceller har just varit att de "riktiga" stamcellerna inte kunnat nås och effekten därför blivit kortvarig. Den markör för identifiera stamcellerna som forskarna hittat är dessutom förvånande. Inger Attersten Original text Lexicon: Swedish Words before 630 Words after 179 Summary length: 28% Type of text: tidningsstext Keywords: stamcellerna blodceller generapi identifiera markör forskargrupp ursprung forskare leukemi metod		25
Hercules Dalanis		

Cellfynd revolution för blodsjuka	Konstgjort händer smitta (SvD)	
(Baltimore, SvD) Nu har en forskargrupp från Philadelphia i USA klarat det som många trodde var en omöjlighet. "Barnen" - miljontals andra blodceller - dör snabbt och försvinner. Stamcellerna bildar ständigt nya. Forskare har i åratal jagat metoder för att försöka hitta och identifiera dessa bennärrens ursprungsceller. De tänkbara applikationerna är svindlande: de kan användas för att utfylla stamcellerna oöda från enorma mängder av alla sorters blodceller i laboratoriet och därmed ersätta dagens blodtransfusioner där blodet tas direkt från människor, alltså främstallt konstgjort blod. Om stamcellerna kan identifieras blir hela denna procedur betydligt enklare och mer effektiv, eftersom egentligen enbart stamcellerna behövs för att återuppsbygga ett nytt fungerande immunförsvar. Generapi testas också för att försöka skapa blodceller som inte kan infekteras med eller som kan bekämpa hivus. Den stora stötestenen för generapi på blodceller har just varit att de "riktiga" stamcellerna inte kunnat nås och effekten därför blivit kortvarig. Den markör för identifiera stamcellerna som forskarna hittat är dessutom förvånande. Inger Attersten Original text Lexicon: Swedish Words before 630 Words after 179 Summary length: 28% Type of text: tidningsstext Keywords: stamcellerna blodceller generapi identifiera markör forskargrupp ursprung forskare leukemi metod		26
Hercules Dalanis		

Problems	
- Pronoun and other anafora referenser - Kalle sprang. Han sprang fort. - Pronoun resolution - Clauses can be too long or too short - Clause reductions- and clause combination rules - Aggregation	27
Hercules Dalanis	

SweSum without PRM	
Analysera mera! Regi: Harold Ramis Medv: Robert De Niro, Billy Crystal, Lisa Kudrow Längd: 1 tim, 45 min ... Ett av många skäl att glädjas åt Analysera mera är att Robert De Niro här verkligen utövar skådespelarkonst igen. Han accelererar emotionellt från 0 till 100 på ingen tid alls, för att sedan kattmjukt bromsa in och parkera, lugnt och behärskat. Och han är tämligen oemotståndlig. Här har han åstadkommit ännu en intelligent komedi för alla oss vänner av intelligens och komedi, gärna i kombination. SvD 99-10-08	28
Hercules Dalanis	

SweSum with PRM	
Analysera mera! Regi: Harold Ramis Medv: Robert De Niro, Billy Crystal, Lisa Kudrow Längd: 1 tim, 45 min ... Ett av många skäl att glädjas åt Analysera mera är att Robert De Niro här verkligen utövar skådespelarkonst igen. Robert accelererar emotionellt från 0 till 100 på ingen tid alls, för att sedan kattmjukt bromsa in och parkera, lugnt och behärskat. Och Robert är tämligen oemotståndlig. Här har Harold åstadkommit ännu en intelligent komedi för alla oss vänner av intelligens och komedi, gärna i kombination. SvD 99-10-08	29
Hercules Dalanis	

Evaluation	
- We found that if one summarizes the text to 30 percent of original length one will obtain around 70-80 percent accuracy on 3-4 pages news articles. .. but query based evaluations are based on subjective opinions - These evaluation need large human effort - Small overlap of opinions - We need man-made extracts to compare the machine made extracts automatically	30
Hercules Dalanis	

KTH

Berättelse och experimentell forskning

UNIVERSITETET I STOCKHOLM

KTH Extract tool

<http://www.nada.kth.se/ST4/NEP.html>
[Heracles Web Site](#)
[PDB Web Site](#)
[App](#)

KTH extraktkorpus

Den utgörs av en lista över de meningar som du tycker borde lyftas ut för att utgöra ett rättvisande extrakt av texten, dvs en sammanfattning.

Totalt antal valbara meningar i denna text är 33 och av dessa blir du inte valda mer än 20 - vilket blir viktiga att skapa en rättvisande sammanfattning på 20-30% av originaltexten (dvs att du inte bara tar de viktigaste, viktiga, nygripande och/eller mest intressanta och/eller i originaltexten).

Den ursprungliga och utvalda meningarna (Läs mer om pilen)

Denna text är ID är svenska > text - test001.htm

☒ Cellfynd revolution för blodsjuka

☐ (Baltimore, SvD)

☒ Nu har en forskargrupp från Philadelphia i USA klarat det som många trodde var en omöjlighet.

☐ De har lyckats finna en metod att hitta de försvagade blodcellerna i benmärgen som skapar alla andra celler i vårt blod, det vill säga vita och röda blodkroppar samt blodplättar.

☒ Dessa så kallade hematopoietiska stamceller är ytterst få och finns i benmärgen.

☐ De smälter del av sig till de vita blodcellerna som reglerar blodet och som fungerar som vårt kroppens immunförsvar, dels av att de har en unik förmåga att bilda blod. Berörda sig vidare.

☒ Stamcellerna är försvagade för att de inte längre fungerar som i ett friskt blod, transporteras inte till alla vävnader och blodkärl.

☐ "Stam" - mjukaste av alla blodceller - där snabbt och försvinner.

☒ Stamcellerna bildar ständigt nya.

☐ Stamcellerna bildar ständigt nya.

☒ Forskare har i årtal jagat metoder för att förstå, hitta och identifiera dessa benmärgens ursprungsceller.

☐ Hitta inget.

☒ De har länge sökt efter att göra utgående och pågående systematiska, säger Cesare Piccoli vid Thomas Jefferson-universitetet i Philadelphia när han presenterade resultaten vid den pågående konferensen arrangerad av Institutet för hematologiska sjukdomar.

☐ Förklaringen till att det inte fungerat är att dagens metoder är så svåra att de inte kan användas för att hitta de stamcellerna.

☒ Cesare Piccoli och hans forskargrupp har alltså lyckats identifiera ett specifikt stamceller som utgör kärnan för de vita blodcellerna.

☐ Det går det möjligt att använda dem och separera dem från alla andra blodceller.

☒ De två blodcellerna är svårare att separera från alla andra blodceller och från många andra celler av alla sorters blodceller i laboratoriet och därmed ersätta dagens blodtransfusioner där blodet tas direkt från människor, vilket

Heracles

31

KTH

Berättelse och experimentell forskning

UNIVERSITETET I STOCKHOLM

KTH extraktkorpus

Nedan visas 30% av originaltexten. Denna sammanfattning representerar det bästa extraktet i enlighet med majoritetens beslut baserat på 11 extrakt. Vardet inom hakparenteser för varje mening representerar antalet gånger denna mening blivit utvald till ett extrakt.

Idéer

1) Cellfynd revolution för blodsjuka

2) Nu har en forskargrupp från Philadelphia i USA klarat det som många trodde var en omöjlighet.

3) De har lyckats finna en metod att hitta de försvagade blodcellerna i benmärgen som skapar alla andra celler i vårt blod, det vill säga vita och röda blodkroppar samt blodplättar.

4) Dessa så kallade hematopoietiska stamceller är ytterst få och finns i benmärgen.

5) Forskare har i årtal jagat metoder för att förstå, hitta och identifiera dessa benmärgens ursprungsceller.

6) De smälter del av sig till de vita blodcellerna som reglerar blodet och som fungerar som vårt kroppens immunförsvar, dels av att de har en unik förmåga att bilda blod. Berörda sig vidare.

7) Stamcellerna är försvagade för att de inte längre fungerar som i ett friskt blod, transporteras inte till alla vävnader och blodkärl.

8) "Stam" - mjukaste av alla blodceller - där snabbt och försvinner.

9) Stamcellerna bildar ständigt nya.

10) Forskare har i årtal jagat metoder för att förstå, hitta och identifiera dessa benmärgens ursprungsceller.

11) Hitta inget.

12) De har länge sökt efter att göra utgående och pågående systematiska, säger Cesare Piccoli vid Thomas Jefferson-universitetet i Philadelphia när han presenterade resultaten vid den pågående konferensen arrangerad av Institutet för hematologiska sjukdomar.

13) Förklaringen till att det inte fungerat är att dagens metoder är så svåra att de inte kan användas för att hitta de stamcellerna.

14) Cesare Piccoli och hans forskargrupp har alltså lyckats identifiera ett specifikt stamceller som utgör kärnan för de vita blodcellerna.

15) Det går det möjligt att använda dem och separera dem från alla andra blodceller.

16) De två blodcellerna är svårare att separera från alla andra blodceller och från många andra celler av alla sorters blodceller i laboratoriet och därmed ersätta dagens blodtransfusioner där blodet tas direkt från människor, vilket

17) 1990 var de först med att hitta de celler som utgör ett mellanstadium mellan stamcellerna och de endokala blodcellerna.

Via sammanfattning på 30% procent

Sammanfattningen ovan är baserad på 11 extrakt

Kortaste extraktet representerar ovan är 13%, längsta är 43% och medelvärdet är 26%.

Täckningen för [tag sammanfattning, text, text] är 52% och precisionen för demamma är 46%.

Täckningen för [tag, tag, tag, tag, tag] är 52% och precisionen för demamma är 31%.

Täckningen för [tag, tag, tag, tag, tag] är 52% och precisionen för demamma är 46%.

Täckningen för [tag, tag, tag, tag, tag] är 52% och precisionen för demamma är 54%.

Via alla sammanfattningar för den här texten

Via sammanfattning, tag, tag, tag

Via sammanfattning, tag, tag, tag

Heracles

32

KTH

Berättelse och experimentell forskning

UNIVERSITETET I STOCKHOLM

Cellfynd revolution för blodsjuka

[Heracles Web Site](#)
[PDB Web Site](#)
[App](#)

KTH extraktkorpus

Den utgörs av en lista över de meningar som du tycker borde lyftas ut för att utgöra ett rättvisande extrakt av texten, dvs en sammanfattning.

Totalt antal valbara meningar i denna text är 33 och av dessa blir du inte valda mer än 20 - vilket blir viktiga att skapa en rättvisande sammanfattning på 20-30% av originaltexten (dvs att du inte bara tar de viktigaste, viktiga, nygripande och/eller mest intressanta och/eller i originaltexten).

Den ursprungliga och utvalda meningarna (Läs mer om pilen)

Denna text är ID är svenska > text - test001.htm

☒ Cellfynd revolution för blodsjuka

☐ (Baltimore, SvD)

☒ Nu har en forskargrupp från Philadelphia i USA klarat det som många trodde var en omöjlighet.

☐ De har lyckats finna en metod att hitta de försvagade blodcellerna i benmärgen som skapar alla andra celler i vårt blod, det vill säga vita och röda blodkroppar samt blodplättar.

☒ Dessa så kallade hematopoietiska stamceller är ytterst få och finns i benmärgen.

☐ De smälter del av sig till de vita blodcellerna som reglerar blodet och som fungerar som vårt kroppens immunförsvar, dels av att de har en unik förmåga att bilda blod. Berörda sig vidare.

☒ Stamcellerna är försvagade för att de inte längre fungerar som i ett friskt blod, transporteras inte till alla vävnader och blodkärl.

☐ "Stam" - mjukaste av alla blodceller - där snabbt och försvinner.

☒ Stamcellerna bildar ständigt nya.

☐ Stamcellerna bildar ständigt nya.

☒ Forskare har i årtal jagat metoder för att förstå, hitta och identifiera dessa benmärgens ursprungsceller.

☐ Hitta inget.

☒ De har länge sökt efter att göra utgående och pågående systematiska, säger Cesare Piccoli vid Thomas Jefferson-universitetet i Philadelphia när han presenterade resultaten vid den pågående konferensen arrangerad av Institutet för hematologiska sjukdomar.

☐ Förklaringen till att det inte fungerat är att dagens metoder är så svåra att de inte kan användas för att hitta de stamcellerna.

☒ Cesare Piccoli och hans forskargrupp har alltså lyckats identifiera ett specifikt stamceller som utgör kärnan för de vita blodcellerna.

☐ Det går det möjligt att använda dem och separera dem från alla andra blodceller.

☒ De två blodcellerna är svårare att separera från alla andra blodceller och från många andra celler av alla sorters blodceller i laboratoriet och därmed ersätta dagens blodtransfusioner där blodet tas direkt från människor, vilket

Heracles

33

KTH

Berättelse och experimentell forskning

UNIVERSITETET I STOCKHOLM

Future improvements of SweSum

- Tagging instead of static lexicons
- Clause level summarization
- Improved Named Entity recognition
- Improved Pronominal Resolution
- Lexical chains using SIMPLE and/or EuroWordNet
- Automatic evaluation method

Heracles

34

KTH

Berättelse och experimentell forskning

UNIVERSITETET I STOCKHOLM

Demonstrators

- SweSum – Standard version
 <http://swesum.nada.kth.se/index-eng.html>
- SweSum – Experimental NE version
 http://www.nada.kth.se/~xmartin/swesum_lab/index-eng.html

(SweSum uses a Perl-CGI script, there is also a standalone version for plain text/html)

Heracles

35

KTH

Berättelse och experimentell forskning

UNIVERSITETET I STOCKHOLM

Conclusions

- Automatic text summarization will be widely used
- Business Intelligence
- News publishing
 - News paper, web, WAP, 3G, TV, Radio
- Machine translation preprocessing
- Advanced Search engines,
 - Keyword and Named Entities extraction
 - Semantic Webb
- Human language technology will play key role

Heracles

36