



SweSum - An automatic text summarizer for Swedish news text

Hercules Dalianis
NADA-KTH
Royal Institute of Technology
100 44 Stockholm
ph: +46-8-790 91 05
mobile: +46 70 568 13 59
email: hercules@nada.kth.se



Hercules Dalianis page 1



Overview of talk

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



Hercules Dalianis page 2



Automatisk text summarization

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

Hercules Dalianis page 3





Summarization tools

<http://www.nada.kth.se/~hercules/HDbookmarks.html>



- Microsoft Word 97, 98 and Word 2000 have a summarizer for documents.
- ProSum . . . Text Summarisation For Word
- MuST Prototype
- Automated Text Summarization (SUMMARIST)

Hercules Dalianis page 4



- Intelligent Miner for Text -Summarization tool IBM
- Inxight (XEROX)
- Datahammer (Glucose Development Corporation)
- EchoSearch Search and Summarize
- Extractor: Text Summarization Software
- Infogist
- Autonomy

Hercules Dalianis page 5



- Mitre
- OracleContext
- Websumm
- WebCompass

Hercules Dalianis page 6

KUNGL. TEKNISKA HOGSKOLEN

Text summarization

- Extraction is much easier than abstraction
- Text summarization (extraction) uses statistic, linguistic and heuristic methods

Hercules Dalanis page 7

KUNGL. TEKNISKA HOGSKOLEN

- A text is divided into sentences
- Sentence positions (News/Reports)
- Title words
- Bold text, Numerical values, Citation, Names
- Keyword frequency and extraction (nouns, adverbs, adjectives)
- Use morphological information-lemma

Hercules Dalanis page 8

KUNGL. TEKNISKA HOGSKOLEN

- Word which are present in all other sentences.
- User adaptation
 - Use 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 page 9

KUNGL. TEKNISKA HOGSKOLEN

SweSum

- The first text summarizer for Swedish
- 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 page 10

KUNGL. TEKNISKA HOGSKOLEN

Key word lexicon

- Key words in news domain
- Also called "open class word lexicon"
- Key words can be noun, adjectives or adverbs

Hercules Dalanis page 11

KUNGL. TEKNISKA HOGSKOLEN

The keyword lexicon

700 000 words	40 000 words
<u>Inflected version</u>	<u>Lemma</u>
statsminister	statsminister
statsministern	statsminister
statsministerns	statsminister
statsministernas	statsminister
statsministrarna	statsminister
statsministrarnas	statsminister
..	..
regeringen	regeringen
regeringens	regeringen
regeringarna	regeringen
regeringarnas	regeringen
...	...

Hercules Dalanis page 12

• Swesum uses a Perl-CGI script, there is also a standalone version for plain text/html.

• <http://www.nada.kth.se/~xmartin/swesum/>

Hercules Dalanis page 13

Herc page 14

28/3 2002

ORIGINALTEXT

Minst 19 döda vid bombdåd i Israel
På onsdagskvällen dödades minst 19 personer vid ett palestinskt självmordsattentat i den israeliska staden Netanya.

En svensk 79-årig kvinna från Stockholm avled i samband med attentatet. Kvinnan var på besök i Israel och bodde på Park Hotell där dådet utfördes.

- Hon dog i samband med explosionen men jag är inte säker på att det var av skadorna. Hon kan ha dött av chocken, säger Åsa Arvidsson på ULD.

Kvinnan fös hem så snart kroppen är identifierad.

- Israeliska myndigheter är mycket noga med det, säger Åsa Arvidsson.

Den dödade kvinnan hade rest till Mellanöstern i sällskap med två andra svenskar. De båda undkom utan fysiska skador, enligt Åsa Arvidsson.

Förutom sällskapet befam sig en svensk barnfamilj tillfälligt på hotellet vid tidpunkten för explosionen. De hade gått till Park Hotel för att äta middag och fick enligt Åsa Arvidsson "bara lättare skador" vid attentatet.

En polis står beordrad de bröppar som läggs utanför ett strandhotell i kuststaden Netanya där en självmordbombare utförde sin laddning i en restaurang och dödade minst 16 människor. Foto Reuters

Sjlvsmordsattentatet har krävt omkring 20 människoliv och flera hundra har skadats.
Den palestinska gruppen Hamas har tagit på sig skulden för dådet som riktades mot Park Hotel vid strandpromenaden i Netanya strax efter söndagmorgnen då den judiska påskhelgen inleddes. Explosionen ägde rum i hotellets välbesatta restaurang.

I en första kommentar lade den israeliska regeringen skulden för dådet på Yasser Arafat.

- **Arafat måste åläggas skulden** för det våld som kommer från territorier som ligger under hans kontroll, sade regeringstalesmannen David Baker.

- **Det är uppenbart** att palestinierna är inställda på att utnyttja allt som står dem till buds för att döda och lemnläsa så många israrer som möjligt var som helst, när som helst, sade han.

En företrädare för den palestinska säkerhetsjätten fördömde dådet, som den radikala Hamasrörelsen sade sig ligga bakom, men påpekade att Israels gränser "föder terrorism".

TT-AFP
tillhika

© 2002 SVD Kontaktta SvD

Hercules Dalanis page 15

28/3 2002

Minst 19 döda vid bombdåd i Israel
På onsdagskvällen dödades minst 19 personer vid ett palestinskt självmordsattentat i den israeliska staden Netanya.

En svensk 79-årig kvinna från Stockholm avled i samband med attentatet. Kvinnan var på besök i Israel och bodde på Park Hotell där var av skadorna. Arvidsson.

Den dödade kvinnan hade rest till Mellanöstern i sällskap med två andra svenskar. Arvidsson. Förutom sällskapet befam sig en svensk barnfamilj tillfälligt på hotellet vid tidpunkten för explosionen, attentatet.

En polis står beordrad de bröppar som läggs utanför ett dödade minst 16 människor.
Sjlvsmordsattentatet har krävt omkring 20 människoliv och flera hundra har skadats.

- **Arafat måste åläggas skulden** för det våld som kommer från

- **Det är uppenbart** att palestinierna är inställda på att utnyttja allt

En företrädare för den palestinska säkerhetsjätten fördömde

30 % summarization

Hercules Dalanis page 16

28/3 2002

Minst 19 döda vid bombdåd i Israel

En svensk 79-årig kvinna från Stockholm avled i samband med

Textens språk: Svenska
Ord före 450
Ord efter 27
Sammanfattningsgrad: 6%
Typ av text: tidningstext
Nyckelord: israel arvidsson svensk minst skadad palestinsk hotell kvinna attentat

Hercules Dalanis page 17

Evaluation

- We had 100 news articles which were manually annotated with questions and answers
- Each of ten students within the framework of 2D1418 Språkteknologi (Human Language Technology course), summarized 10 texts each = totally 100 texts
- Increasing compression rates 20, 30 and 40 percent.
- If failed then the users could select their own key words to direct the summarizer-slanted summaries.
- Comparing the found answers with the manually annotated.

Hercules Dalanis page 18

Evaluation of the text summarizer SweSum

Summary/ Compression rate	20%	30%	40%	Keywords (20%)	Total correct answers
Number of texts	97	97	97	97	85
Given and correct answers	50	16	15	4	
Percent accumulated correct answer	52%	68%	84%	88%	

Hercules Dalianis page 19

- Accuracy 84% at 40% summary of 181 word length news texts.
- State of the art text summarizer
- The text summarizer is also available for English, Spanish, French and German texts.
- And soon in Danish and Norwegian !!

Hercules Dalianis page 20

Future improvements of SweSum

- Named Entity recognition
—Person names, Organization, Location, Time
- Tagging-clause level summarization
- Better pronominal resolution
- Lexical chains

Hercules Dalianis page 21

Applications of text summarizers

- Summary of news text which has been downloaded from Internet to a WAP phone
- An Internet browser for a PDA / Palmpilot
- Keyword directed subscription of news which are pushed and presented summarized to the user "Business Intelligence"
- To let a computer read summarized news in the telephone.

Hercules Dalianis page 22

- For presentation of pages in search engines
- In search engines to extract keywords
- MUST - Searches in foreign languages, summarizes and translates.

Hercules Dalianis page 23