



SweSum and the Problem of Evaluating Automatic Text Summarizers

Martin Hassel
NADA-KTH
Royal Institute of Technology
100 44 Stockholm
ph: +46-8-790 66 34
mail: xmartin@nada.kth.se



Martin Hassel page 1



Overview of talk

- Background
- Technique
- Past evaluations
- Building new corpora
- Suggested interface to the corpora
- (Discussion)



Martin Hassel page 2



Summarization approaches

- Extraction vs. Abstraction
- Generic vs. Query based
- Indicative vs. Informative
- Restricted vs. Unrestricted domain
- Background information vs. New information (TDT)
- Single-document vs. Multiple-document
- Monolingual vs. Multilingual
- Textual vs. Multimedia



Martin Hassel page 3



Text summarization



- Extraction is much easier than abstraction
 - Can be extended with rewriting rules: Jing & McKeown [NAACL'00], Jing [ANLP'00]
- Text summarization (extraction) uses statistic, linguistic and heuristic methods

Martin Hassel page 4



- A text is divided into sentences
- Each sentence is assigned a rank
 - Sentence positions (News/Reports)
 - Title words
 - Bold text, Numerical values, Citation
 - Named Entities
 - Keyword frequency (nouns, adverbs, adjectives)
 - Use morphological information – lemma

Martin Hassel page 5



- Words present in many sentences are central
- 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, we have the summary!

Martin Hassel page 6

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 5-10 key words are produced which describe or categorize the text

Martin Hasselpage 7

Keyword lexicon

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

Martin Hasselpage 8

The keyword lexicon

700 000 words 40 000 words

Inflected version	Lemma
statsminister	statsminister
statsministern	statsminister
statsministrarna	statsminister
statsministrarnas	statsminister
...	...
regeringen	regeringen
regeringens	regeringen
regeringarna	regeringen
regeringarnas	regeringen
...	...

Martin Hasselpage 9

Evaluating SweSum – 1st Run

- Fall 2000
 - 9 participants (students, HLT course)
 - 10 news texts and movie reviews
- Steps:
 - Read the text to be summarized
 - Summarize the text gradually lowering the resulting length of the summary
 - Note when important information is given and when coherence is broken

Martin Hasselpage 10

Informativity

Important information is missing at

Text/Person	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	37%	25%	40%	15%	35%	20%	30%	35%	48%
2	40%	25%	55%	45%	45%	20%	40%	25%	53%
3	30%	30%	45%	25%	50%	25%	20%	30%	55%
4	20%	15%	20%	15%	35%	35%	50%	25%	29%
5	50%	20%	10%	15%	20%	15%	20%	45%	15%
6	37%	30%	98%	35%	40%	40%	30%	25%	
7	44%	25%		30%	30%	25%	50%	50%	41%
8	30%	30%	30%	25%	35%	40%			
9	20%	15%	20%	25%	15%	30%			19%
10	15%	10%	10%	35%	10%	20%			

Martin Hasselpage 11

Coherence

Coherence is broken at

Text/Person	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	25%	20%	35%	30%	35%	25%	20%	20%	
2	20%		30%	55%	30%	50%	20%	15%	53%
3	27%	25%	25%	25%	70%	25%	15%	25%	55%
4	15%			15%	15%		10%	5%	
5	30%		20%	10%	20%	20%	15%	25%	
6	30%	10%	30%	35%	35%	10%	15%	15%	
7	20%	10%		25%	30%	40%	30%	35%	41%
8		20%	60%	25%	35%	35%	20%		
9		5%		20%	40%	10%	10%		44%
10		5%		10%	35%	10%	10%		

Martin Hasselpage 12

And the results were ... ?

	Information	Coherence
Total median:	30%	24%
Total average:	31%	26%

Satisfying, but not very reliable!?

* Note: These results are comparable to results achieved with English language systems

Martin Hassel page 13

Evaluating SweSum – 2nd Run

- Spring 2001
 - 10 participants (students, HLT course)
 - 97 short news texts (telegram style)
- Question and answering schema
 - 1 question central to the text for each text
 - 1 answer to each question

Martin Hassel page 14

And the results this time were ... ?

Summary	30%	30%	40%	Keywords (30%)	Total correct answer
Comparison rate	97	97	97	97	
Number of texts	50	18	15	4	85
Percent accumulated correct answer	52%	83%	84%	88%	

More reliable, and still quite satisfying

- But still too much manual labour ...
- And not repeatable with an acceptable effort and comparability

Martin Hassel page 15

Evaluating SweSum – 3rd Run

- Create corpora for automatic evaluation
 - Still a lot of manual labour ...
 - But hopefully only once!
 - Reusable for re-evaluation without use of test subjects

Martin Hassel page 16

Suggested Composition of the Corpus

- Full texts
 - These are the original texts made available for reference and for summarization purposes

Martin Hassel page 17



The Original Text

Självordsattentatet i helgen har gjort de amerikanska soldaterna nervösa. I synnerhet som den irakiska ledningen har hotat med många fler sådana attacker. Det har sagts att tusentals frivilliga har kommit till Irak från hela arabvärlden. Döds-skjutningarna väntas ytterligare öka den arabiska vreden mot invasionsstyrkorna. Soldaterna utanför Najaf sköt skarpt när bilen med 13 kvinnor och barn inte slannade trots varningskott. Soldaternas överordnade försvarade efteråt deras agerande. De gjorde helt rätt eftersom bilföraren uppträdde onormalt, hette det. Det var vid en vägsnäring i samma område som fyra amerikanska soldater dödades i lördags.

Brittiska företrädare medger dock att irakierna lätt kan komma att se de allierade soldaterna som skurkar i stället för en befrielsestyrka, rapporterar BBC. Från amerikansk sida uppges att händelserna ska utredas. Incidenten utanför Shatra inträffade några timmar efter gryning. Här var det en obehäpnad man som körde i sin pickupbil tvärs över det taggtrådshinder som låg över vägen. När marinkärsoldaterna öppnade eld dödades föraren och en medelålders man som åkte med i bilen skadades svårt. Marinkärens sjukvårdare försökte efteråt rädda mannens liv.

Bagdad utsattes under natten och morgonen för upprepade flyganfall. Runt midnatt träffades ett av president Saddam Husseins palats vid floden Tigris. Både Saddam och hans son Ousay uppges använda palatset. - Ett stort, stort, stort rökmoln bolmar upp från anläggningen. Kanske använder de nu större bomber än tidigare, rapporterade en Reuters-korrespondent. Det stora palatsområdet hade tidigare träffats flera gånger under måndagen. Söder om Bagdad uppgavs på formiddagen strider utspelas mellan amerikanska



Martin Hassel page 18

Suggested Composition of the Corpus

- Full texts
- Man-made indicative summaries
 - These abstracts are made by editors at the news papers or by the article authors themselves



Martin Hassel page 19

Man-made indicative summaries

Fredsaktivister uppges anfallna
 Enligt Iraks informationsminister har 18 personer dött sedan i måndags i flyganfallen över Bagdad. Han hävdar också att USA attackerat fredsaktivister. Flera dödsskjutningar av civila som inträffat i Irak det senaste dygnet har spätt på den arabiska vreden mot invasionsstyrkorna.



Martin Hassel page 20

Suggested Composition of the Corpus

- Full texts
- Man-made indicative summaries
- Man-made informative extracts
 - These are gathered by using test subjects and generated by majority vote on sentence level

Martin Hassel page 21






Man-made informative extracts

[4] Självordsattentatet i helgen har gjort de amerikanska soldaterna nervösa.
[3] Dödsskjutningarna väntas ytterligare öka den arabiska vreden mot invasionsstyrkorna.
[3] Soldaterna utanför Najaf sköt skarpt när bilen med 13 kvinnor och barn inte stannade trots varningsskott.
[4] Från amerikansk sida uppges att händelserna ska utredas.
[2] Incidenten utanför Shatra inträffade några timmar efter gryningen.
[3] När marinkårssoldaterna öppnade eld dödades föraren och en medelålders man som åkte med i bilen skadades svårt.
[4] Bagdad utsattes under natten och morgonen för upprepade flyganfall.
[4] Runt midnatt träffades ett av president Saddams Husseins palats vid floden Tigris.
[2] Det stora palatsområdet hade tidigare träffats flera gånger under måndagen.

Sammanfattningen ovan är baserad på 4 extrakt [...]



Martin Hassel page 22

Suggested Composition of the Corpus

- Original texts (full texts)
- Man-made indicative summaries
- Man-made informative extracts
- Automatically created extracts

Martin Hassel page 23

Automatically created extracts

- Can be created by aligning the sentences in the man-made indicative summaries with one or more sentences in the full text and extracting these to an extract
- Several well-known techniques from the creation of parallel corpora and information retrieval can be modified and applied for this
- Or perhaps Random Indexing?

Martin Hassel page 24

Possible caveats

- Man-made extracts may be worse than baseline (no consensus)
 - Test run on three text by two persons
 - Fairly good results on two texts and not so good on the third
 - New test run with 5 texts and 5 people coming week
- But then we will still have:
 - Baseline extracts easy available
 - The man-made abstracts in one central place
 - The automatically generated extracts
 - The man-made extracts (at least for reference)
 - And last but not least, one experience richer

Links

- SweSum – Standard version
<http://swesum.nada.kth.se/>
- SweSum NE – Experimental NE version
http://www.nada.kth.se/~xmartin/swesum_lab/
- SweNam (NE tagger used in SweSum NE)
<http://www.nada.kth.se/iplab/hlt/swenam/>
- SEE (Summarization Evaluation Environment by C-Y Lin, ISI)
<http://www.isi.edu/~cyl/SEE/>
SEE 2.0 is used by DUC 2001 to support its human evaluation protocol