

Summarization Evaluation

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

Overview

- Approaches to Summarization Evaluation
- KTH News Corpus
 - Task Based Evaluation
- KTH eXtract Corpus
 - Sentence Recall
- The Road Ahead

Martin Hassel

2

Approaches to Summarization Evaluation

Challenges in summarization evaluation

- What is a good summary?
- There may be several good summaries for one specific text
- Do different users need different summaries?
- At what compression rate?
- How do we make evaluation repeatable?

Martin Hassel

3

Evaluation schemes

- Manual evaluation
 - Human judges
- Semi-automatic evaluation
 - Task based evaluation
 - Keyword association
- Automatic evaluation
 - Comparison to Gold Standard summaries
 - Sentence Recall
 - Vocabulary Test

Martin Hassel

4

Manual evaluation

- Human judges
 - + Semantically based assessment
 - Subjective
 - Time consuming
 - Expensive

Martin Hassel

5

Semi-automatic evaluation

- Task based evaluation
 - + Measures the summary's utility
 - Subjective interpretation of questions and answers
- Keyword association
 - + No annotation required
 - Shallow, allows for "good guesses"

Martin Hassel

6

Automatic evaluation

Comparison to Gold Standard:

- Sentence Recall
 - + Cheap and repeatable
 - Does not distinguish between different summaries
- Vocabulary Test
 - + Useful for key phrase summaries
 - Sensitive to word order differences and negation

Martin Hassel

7

Corpora types

- Text only
- Text + Abstract
- Text + Extract
- Text => Text + Extract (manually)
- Text + Abstract => Text + Extract (automatic)

Martin Hassel

8

KTH News Corpus

- Directed Web Spider (newsAgent)
- News articles, telegrams and press releases
- 13 news sources
 - E.g. major newspapers, Amnesty, Riksdagen ...
- Collected during one year
- More than 100.000 texts (~200Mb)

Martin Hassel

9

Manually tagged sub corpus

- 100 randomly chosen texts
- Topic central question and answer
- Three persons (33, 33 & 34 Q&A each)

Q: Who is CEO at Telenor?

A: Tormod Hermansen

Martin Hassel

10

Task Based Evaluation

- 10 Human Language Technology students
- 10 texts (+ questions) each
- Summarize the text in compression steps
 - 20% - 30% - 40% - 20% (using keywords)
- Try to answer the question at each step

Martin Hassel

11

Task Based Evaluation, results

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

- SweSum failed on 3 texts, hence 97 texts
- Yes/No questions would have allowed for automatic assessment of answers, but would also have opened up for guessing

Martin Hassel

12

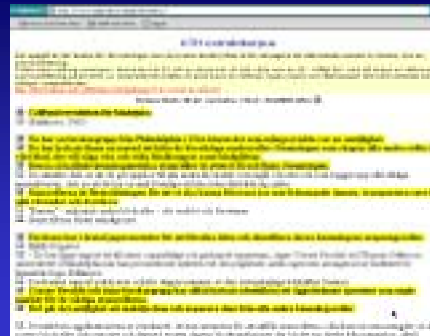
KTH eXtract Corpus

- 10 texts randomly chosen from SvD
- 45 persons submitted a total of 213 extracts
- Mean length of shortest extracts 20%
- Mean length of longest extracts 54%
- Mean length of all extracts 34%

Martin Hassel

13

KTH Extract tool



Martin Hassel

14



Martin Hassel

15

KTHXC Pros & Cons

Pros:

- Reusable
- Time saving
- Extendable
- Multilingual

Cons:

- Supports only one ideal extract based summary (based on majority vote)

Martin Hassel

16

Sentence Recall Evaluation

- 10 texts from KTHXC
- 10 persons submitted a total of 74 extracts
- Precision of votes:
 - 40% at 100% recall
 - 69% at mean length of extracts for each text
- The ten texts were summarized with and without Named Entity Tagging

Martin Hassel

17

Overlap submitted extracts

Ideal (Gold Standard) extracts of mean length compared to SweSum generated extracts:

- Without NE
 - 57% of the generated sentences were found in control extracts of equal length
- With NE
 - 34% of the generated sentences were found in control extracts of equal length

Martin Hassel

18

Summarizing without using NE

- Clusters of interrelated sentences tend to get high ranks
- High frequency of shared words
 - Conserves redundancy
- High cohesion throughout the summary

Martin Hassel

19

Summarizing using NE

- Tends to prioritize elaborative sentences
 - Elaborative sentences seem to contain more NE per sentence than introductory sentences
- Lessens the fluency of the summary
- List of facts (repetitive use of names)

Martin Hassel

20

ORIGINALTEXT

6 veckors baby svårt misshandlad

Pappan misstänkt för misshandeln

En sex veckor gammal bebis kom sent i lördagskväll svårt misshandlad in på akuten i Sundsvall.

Flickan har mycket svåra skall- och lungskador.

- Hennes tillstånd är livshotande, säger jourhavande åklagare Åke Hansson.

Barnets pappa har anhöllits som misstänkt för misshandeln på den sex veckor gamla flickan.

På onsdagsmorgonen avgör kammaråklagare Lisa Eriksson i Sundsvall om han ska begäras häktad.

Lisa Eriksson var knapphändig i sina uppgifter på tisdagen.

- Utredningen är ännu i ett känsligt läge och förhör pågår. Av utredningstekniska skäl kan jag inte berätta mer, säger hon.

Enligt polisens uppgifter var båda föräldrarna med när barnen kom med ambulans till Sundsvalls sjukhus.

-Dagen efter greps pappan av polisen och anhölls senare på sannolika skal misstänkt för grov misshandel, förklarar Börje Ohman, chef för länskriminalen i Sundsvall.

Sex veckor gammal

Flickan - som enligt uppgift till Aftonbladet är sex veckor gammal - kom in till akuten Sundsvalls sjukhus vid 22-tiden i lördags kväll.

Martin Hassel

21

Hennes skador var livshotande. Hon hade bland annat svåra skallskador och svåra skador på lungorna. Sjukvården i Sundsvall räckte inte till för att rädda flickan. Hon fördes därför med ambulanshelikopter till Akademiska sjukhuset i Uppsala där hon under gårdagen fick specialistvård.

Enligt läkare i Uppsala var hennes tillstånd i går fortfarande livshotande.

Flera fall av våld

Den lilla flickans skador har enligt läkarna orsakats av flera fall av våld i hemmet.

Akademiska sjukhuset ville i natt inte lämna någon ytterligare information om hennes tillstånd.

2001 anmäldes nära 7 000 fall av barmisshandeln i Sverige. På Astrid Lindgrens barnsjukhus i Solna upptäckts i dag

ungefär ett spådbarn i månaden som är offer för den form av barmisshandeln som kallas Shaken baby-syndrome.

Skadan uppkommer när barnet utsätts för kraftiga skakningar av en vuxen person - oftast en förälder. De barn som överlever riskerar att bli vårdpaket med grava handikapp.

Petter Ovander

Summary using Named Entities

- - Hennes tillstånd är livshotande, säger jourhavande åklagare **Åke Hansson**. Lisa **Eriksson** var knapphändig i sina uppgifter på tisdagen. Sjukvården i **Sundsvall** räckte inte till för att rädda flickan. Enligt läkare i **Uppsala** var hennes tillstånd i går fortfarande livshotande. 2001 anmälades nära 7 000 fall av barmisshandeln i **Sverige**. På **Astrid Lindgrens barnsjukhus** i **Solna** upptäckts i dag ungefär ett spådbarn i månaden som är offer för den form av barmisshandeln som kallas Shaken baby-syndrome. **Petter Ovander**
- **Lexikon:** Svenska
Ord före 407
Ord efter 121
Sammanfattningsgrad: 29%
Typ av text: tidningstext
Nyckelord: vecka skada uppgift sjukhus pappa flicka sundsvall akademisk tillstånd läkare
Hittade **namn:** 1 Åke Hansson; 1 Astrid; 2 Eriksson; 1 Börje Ohmans; 1 Petter Ovander;
Hittade **platser:** 4 Sundsvall; 2 Uppsala; 1 Solna; 1 Sverige;

Martin Hassel

22

Summary not using Named Entities

- 6 veckors baby svårt misshandlad
- Pappan misstänkt för misshandeln
- En sex veckor gammal bebis kom sent i lördagskväll svårt misshandlad in på akuten i Sundsvall. Flickan har mycket svåra skall- och lungskador.
- Hennes tillstånd är livshotande, säger jourhavande åklagare Åke Hansson.
- Barnets pappa har anhöllits som misstänkt för misshandeln på den sex veckor gamla flickan.
- Sex veckor gammal
- Flickan - som enligt uppgift till Aftonbladet är sex veckor gammal - kom in till akuten Sundsvalls sjukhus vid 22-tiden i lördagskväll. Hennes skador var livshotande.
- Petter Ovander
- **Lexikon:** Svenska
Ord före 336
Ord efter 92
Sammanfattningsgrad: 27%
Typ av text: tidningstext
Nyckelord: vecka skada uppgift pappa flicka sjukhus tillstånd misshandlad misshandeln läkare

Martin Hassel

23

The Road Ahead

- Re-implement SweSum in Java
- Better modularity
- Connect SweSum² to Granska & GTA
 - True lemmatisation
 - Better sentence boundary detection
 - Phrase annotation
- Improved weighting schemes
- Light semantics by context (LSA/RI)

Martin Hassel

24

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
- KTH eXtract Corpus - A small corpus of manually made extracts (*in Swedish*)
 - <http://www.nada.kth.se/iplab/hlt/kthxc/showsumstats.php>
- SweNam
 - <http://www.nada.kth.se/~xmartin/swene/index-eng.html>

Martin Hassel

25

Schematic view of SweSum

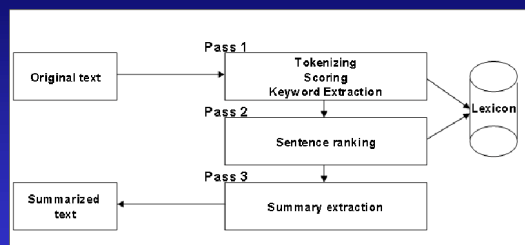


Figure by Nima Mazdak

Martin Hassel

26