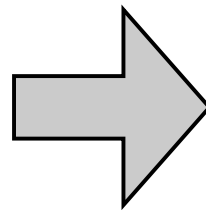


Kvalitet af genereret tekst

Tekst-til-tekst generering

søgning,
spørgsmålsbesvarelse,
simplificering,
opsummering

Sommer var det, midt paa Dagen, i et Hjørne af Hegnet. Lige for stod der et gammelt Egetræ, om hvis Stamme man gjerne kunde sige, at den vred sig i Fortvivlelse over den Mangel paa Harmoni, der var mellem dens ganske ny, gulladne Løv og dens sorte og tykke, krogede Grene, som meest af Alting lignede grovt fortegnede gammelgothiske Arabesker.



Det var sommer. Det var midt på dagen. Det var i et hjørne af hegnet. Overfor stod et gammelt egetræ, hvis stamme nærmest så ud som om den vred sig i fortvivlelse over manglen på harmoni mellem dens gule løv og dens sorte grene.

To slags tekstkvalitet

A. **Kvalitetstab** som følge af maskinelle klippe-klistre-manøvrer



B. **Kvalitetsforskelle** i oprindelige kilder



Om Slægtens to første Baroner er her kun at meddele, at Helmuth Hieronimus i en øjeblikkelig Forlegenhed lod omhugge og bortsælge den pragtfulde, vel en Fjerdingvej lange Lindeallé, ... Og at Hannibal Hieronimus i en mere permanent Trang for Kontanter lod i fem paa hinanden følgende Aar alle sine Udhuse blotte for det alentykke Egetømmer...

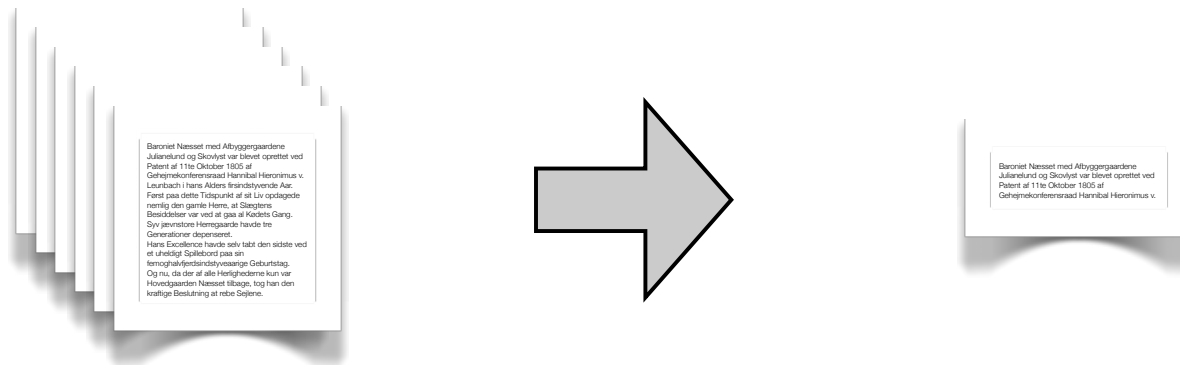


Den var en gang, dengang.
Dreng som fik en.
Drage som fløj
af helved. til ud
i luften.

Mål: forbedre **A** ved hjælp af **B**

Opsummering (Berg-Kirkpatrick et al., 2011)

Hvordan fungerer automatiseret opsummering?



- Et referat har en fast størrelse (typisk 100-200 ord)
- Sammensæt et referat, der indeholder så mange centrale passager fra de oprindelige dokumenter som muligt

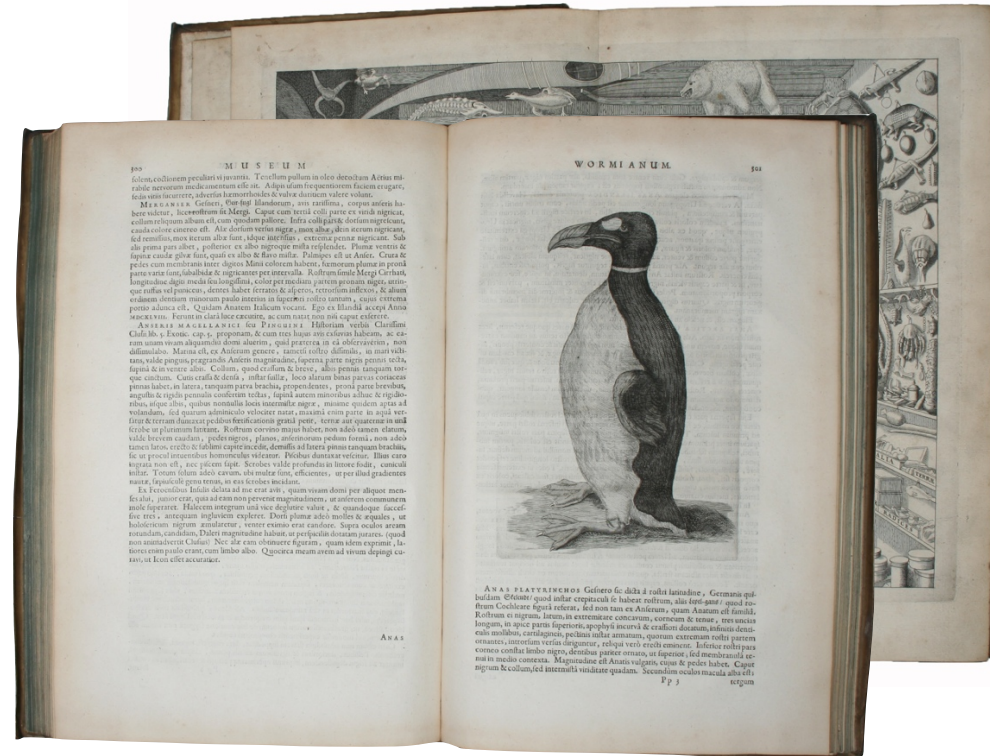
En relateret vanskelighed: at flytte sammen

- Indholdet af to lejligheder skal pakkes sammen til én
- Hver ting har en nytteværdi og et volumen
- Det gælder om at få mest nytte ud af den begrænsede plads



kan løses som et **optimeringsproblem**

Første komplikation: dubletter



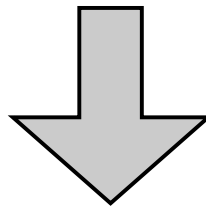
Anden komplikation: overlappende egenskaber



Tilbage til opsummering

- En sætning realiserer nogle informationsenheder (egenskaber)

X var glad for den gamle bog,
som Y smed ud



`glad_for(X,bog); gammel(bog); smed_ud(Y,bog)`

Maskimér værdien af informationsenheder, mens referatets længde holdes under en given størrelse

Når tekstkvalitet overses

The screenshot shows a web browser window with the address bar displaying `newsblaster.cs.columbia.edu/summaries/2012-10-29-05-11-45-068.html`. The browser's bookmark bar includes links like 'Doc', 'Ref', 'Handoff!', 'Cite', 'KB', 'Offline', 'pin', 'read later', 'my pinboard', 'insta', 'kindle', 'astext', and 'Other Bookmarks'. The website header features the 'Columbia Newsblaster' logo with the tagline 'Summarizing all the news on the Web'. On the right, it shows the date 'Monday, October 29, 2012' and the article range 'Articles from 10/26/2012 to 10/29/2012', with a 'Last update: 5:11 AM EST'.

On the left side, there is a search bar labeled 'Search for:' with a 'Go' button and a dropdown menu for 'Offline summarization'. Below this is a vertical menu with categories: 'U.S.', 'World', 'Sci/Tech', 'Entertainment', and 'Sports'. Further down are links for 'View Today's Images', 'View Archive', 'About Newsblaster', 'About today's run', and 'Newsblaster in Press'.

The main content area displays the article title 'Old dogs and new bricks: a night for grown-up Lego fans' in a large, bold font. Below the title, it says 'Summary from United States, from articles in English' and '[UPDATED] (see summary with new information since yesterday)'. The article text begins with 'If receiving a new Lego set is always a good feeling, then consider getting your kids and yourself a unique Lego creation but this time... for charity. (article 1) Nannan Zhang and his team at Creations for Charity have set up a very unique program aiming at raising money to buy Lego sets for underprivileged children. (article 1) To fuel their creativity and make the Lego fans contribute, the movie Directors came up with a little contest spanning over the next 4 months. (article 7) The rules are very simple, design any type of transportation vehicle out of re-purposed Lego pieces, as long as these pieces do not originate from the same type of vehicle as the final creation. (article 7) The Lego addiction certainly did the rest, as Dennis spent about 6 months building this one of a kind Lego Hot Rod, with full remote control and power support. (article 5) The attention to detail in this sculpture gives a fairly impressive result, especially given the reduced amount of bricks - or plates - used to build this space station clock. (article 9) The artist, has put a lot of efforts and attention to detail in this creation, which is part of a bigger collection. (article 8)'

En model for tekstkvalitet

- Anaforbestemmelse er ikke nok
- Vi skal lære at imitere sproglige mønstre i god tekst
- I mit projekt anvender jeg data fra brugerdrevne Q&A-sider, hvor svarenes kvalitet bliver vurderet af de besøgende



Eksempel: brug af pronomener i Q&A

