Text Simplification

Tom Wysocki
Peter Pribil
Margot Guettier
Morgan Ruiz-Huidobro

January 14th 2022

Advancement

- Improved syntactic simplification
- Progression in lexical simplification

Syntactic Simplification

PP removal improvements and challenges

-Sentences with copulas corrected (I was on time. ---> I was on time.)

-Verb type important - transitivity / intransitivity: impossible task

-Limitation on PP number: what to keep?

-Weighting PP based on importance: how to decide?

Complex sentence splitting issues

-Original idea with comp/xcomp not useful: ([*[The doctor said that], [I should cry]])
-Modified original idea too complex: ([The doctor said:] ["you should cry"])

Lexical Simplification

WordNet and Zipf frequency

- Usage of zipf frequency (from a SUBTLEX data of 51 million of words)

- Consideration of multi word expressions

- Threshold frequency fixed to 3.00

- Context consideration

WordNet and Zipf frequency

Original sentence	Simplified sentence
I have a **corgi** and a **bike**.	I have a **dog** and a **bike**.
You visits a beautiful **castle**.	You visits a beautiful **castle**.
We see a **big** and **tall** man. He is wearing a **titfer**.	We see a **big** and **tall** man. He is wearing a **hat**.
This is a real **golden retriever** !	This is a real **dog** !
He did an **abnegation** of what I said.	He did an **act** of what I said.

Issues with context

"This is a real **golden retriever** !" → "This is a real **dog** !"
Golden retriever emphasized by "real".

"He did an **abnegation** of what I said." → "He did an **act** of what I said."
Too much generalization: serious loss of meaning.

Readability Metrics

- Flesch reading ease
- Coleman–Liau Index
  } number syllables/characters per word
- Dale–Chall
  - 3000 most common words
  - length of sentences

Future work :

- Prepare the report
- Improving existing syntactic simplification process
    - implement "and" complex splitting where "x and y" =! S1 U S2
    - limit PP threshold
- Potentially improving existing lexical simplification process
- Improving the software interface
- Custom evaluation method

Thank you

Questions?