Automatic Text Summarization

TECHNISCHE UNIVERSITÄT DARMSTADT

Group 5

Final Results

Content

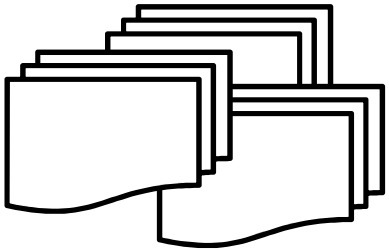
TECHNISCHE UNIVERSITÄT DARMSTADT

- Task
- Implementation
  - Step 1 | Content Selection
  - Step 2 | Creating Hierarchies
  - Step 3 | Creating Summaries from Hierarchies
- Evaluation

Hierarchical Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

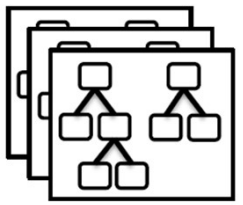
Task | Implementation | Evaluation



documents

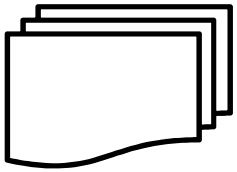
content selection

hierarchical ordering



creating summaries from hierarchies

summaries



Step 1 | Content Selection
Data Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Nugget ratios

~ 14% nuggets per topic

Nugget structure

~ 82% are complete sentences

No words, that occur primarily in nuggets or non nuggets

Step 1 | Content Selection
General Idea

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Approach: Learn a classifier to recognize nuggets

Use binary labeled training data from all 10 given topics

Labeled Data

…	
Diet is often part of the problem.	1
Get up early in the morning.	0
They now have 300 Associate Offices in 25 countries.	0
ATTEND takes about 30-60 days to become effective.	1
…	

Step 1 | Content Selection
Approaches

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

- FastText Classifier
- Naive Bayes
- Fully Connected Neural Network
- Convolutional Neural Network

Step 1 | Content Selection
Approaches

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

- FastText Classifier - Bad
- Naive Bayes - Bad
- Fully Connected Neural Network - Bad
- Convolutional Neural Network - Okay

Step 1 | Content Selection
Convolutional Neural Network – Preprocessing

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

gensim simple preprocess

only sentences between 5 and 50 words, pad them to 50 words

Negative Example	I don't know.
	U.S. History II
	Attendance and tardies!
	All-School Assemblies

Step 1 | Content Selection
Convolutional Neural Network – Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Input words as 300 dimensional embeddings (GloVe) MH1

Random Search to find good Parameters with early stopping

Criteria: MH2

- Recall > 0.05
- Precision as high as possible

Implementation with Keras

Computation on Lichtenberg Cluster

Folie 9

MH1 early stopping to prevent overfitting, but also for efficiency
Markus Hoehn; 12.07.2018

MH2 Begründen:
Antiproportional, Precision ist uns wichtiger als Recall (weil SUmmary am ENde gut sein muss), aber recall muss mindestwert haben
Markus Hoehn; 12.07.2018

Step 1 | Content Selection
Convolutional Neural Network – Architecture

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Parametername	Possible Range	Final Params
Batch size	[120, 180]	150
Number of convolutional layers	[1, 2]	2
Filter sizes per layer	[4, 7]	6, 6
Number of filters per layer	[30, 60]	47, 42
Optimizer	[adam, sgd, adagrad]	Adagrad

Step 1 | Content Selection
Convolutional Neural Network – Results

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Recall = 5.9% MH3

Precision = 52%

Use 30 sentences with highest nugget probability score

- Only most confident sentences if more predicted
- Guarantees enough material if too few predicted nuggets

Folie 11

MH3 Zusammenhang: Da wir ehh immer 30 nehmen ist uns recall nicht so wichtig wie precision

Wir haben entschieden die Nuggets pro topic auf 30 zu setzen, da dies für unser Ziel genug Material ist

Außerdem: Wenn wir mehr predicten würden, dann wählen wir trotzdem die 30 sichersten und können basierend darauf gute, lesbare summaries erstellen:
Falls wir weniger predicten würden, hätten wir nicht genug material und müssen somit mehr predicten
Wir entschieden uns, dass 30 Nuggets ein guter Wert ist
Markus Hoehn; 12.07.2018

Step 2 | Creating Hierachies
Generate Hierachies from Nuggets

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Goal: insert Nuggets into tree

- Insert one Nuggets after another
- compare Nuggets (similarity =/= general/specific)
- find right Bubbles for Nuggets (similar topics)

Prelimitaries:

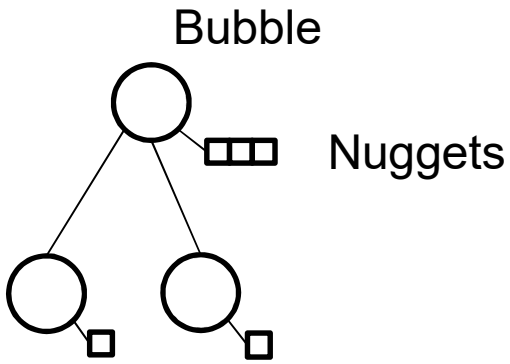
- removed Stopwords, removed words less then 2 characters, Stemming

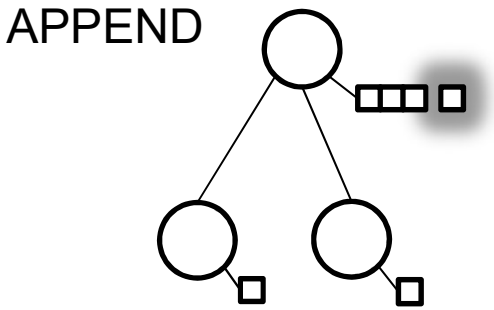
Step 2 | Creating Hierachies
`insert` Function

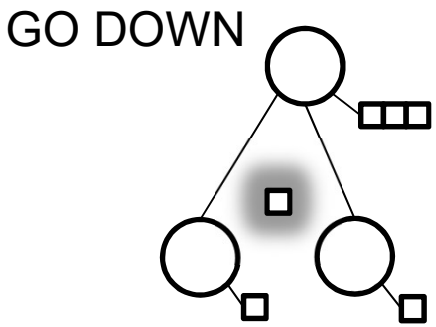
TECHNISCHE UNIVERSITÄT DARMSTADT

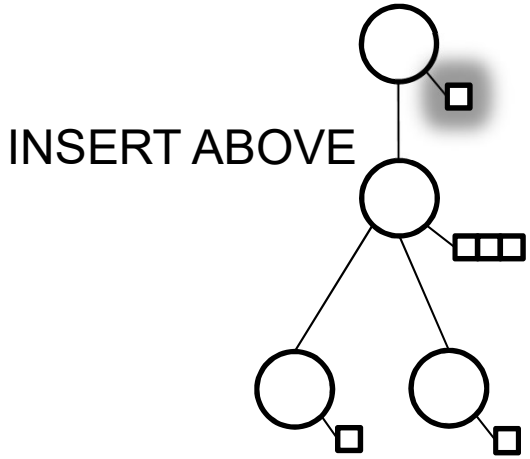
Task | Implementation | Evaluation

Insert()
- recursive function
- inserts new nuggets









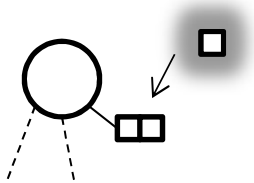
Step 2 | Creating Hierachies
`compare` and `which` Function

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

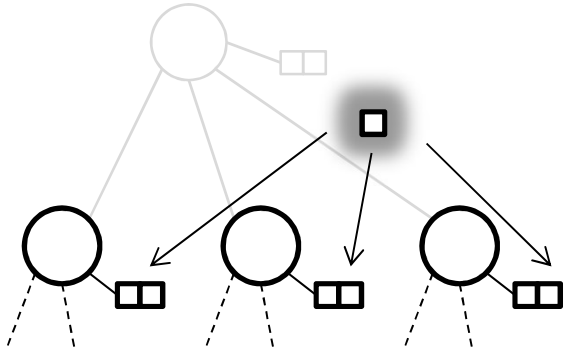
compare()

- find right position for Nugget
- compare TF-IDF scores



which()

- find right Bubble for Nugget
- uses NLTK path_similarity()



Step 2 | Creating Hierachies
Evaluation of Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

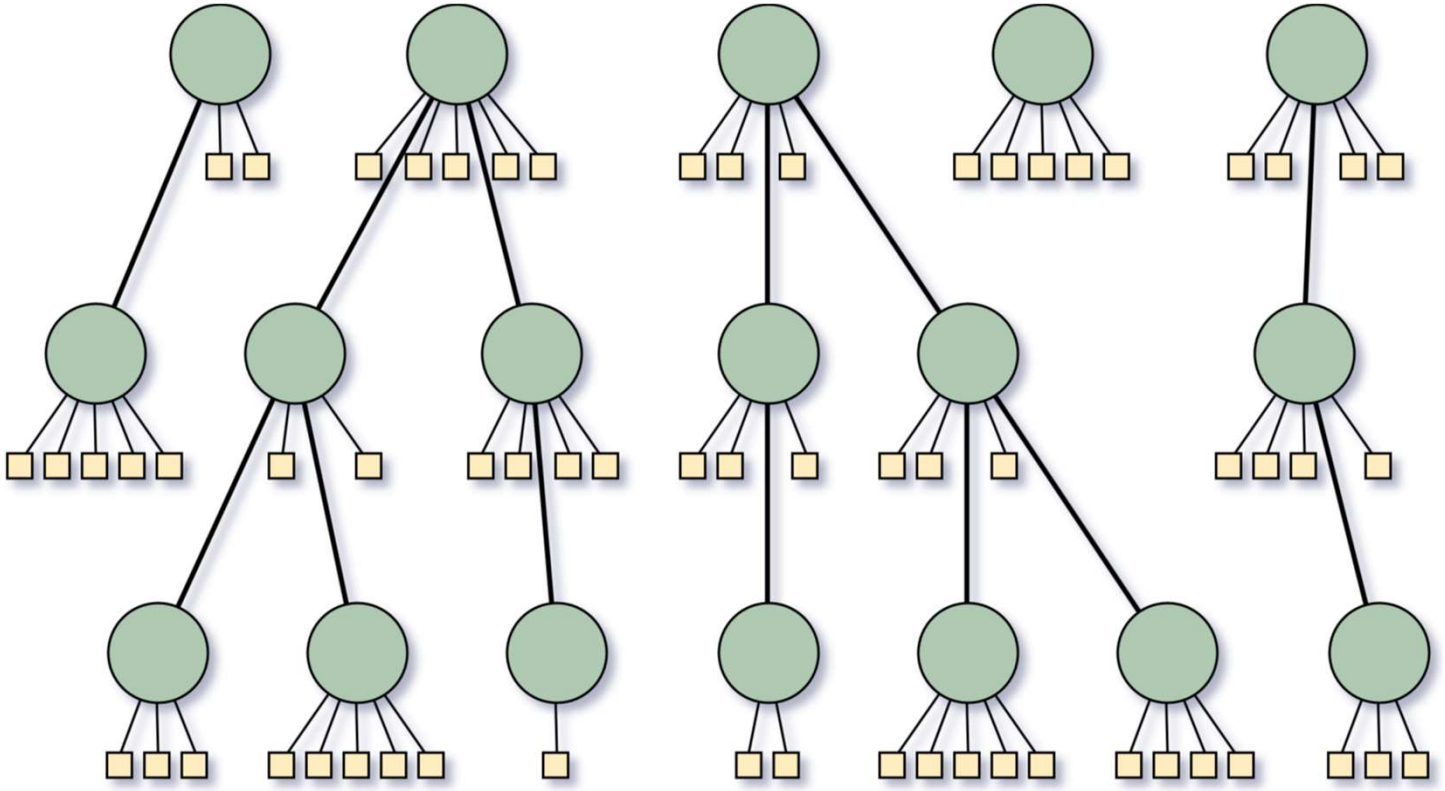
- We tried Annotation Tool from AIPHES
  - 11% similarity average against Gold standard (same as Random Trees)
  - but
    - 1300 Nuggets in gold standard versus 30 Nuggets we used
    - less Nuggets showed more similarity
    - our algorithm is slow (>30 min) with 300+ Nuggets

- Find „right“ balance instead
  - 1-5 Nuggets in each Bubble
  - 5+ Bubbles in root node

Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

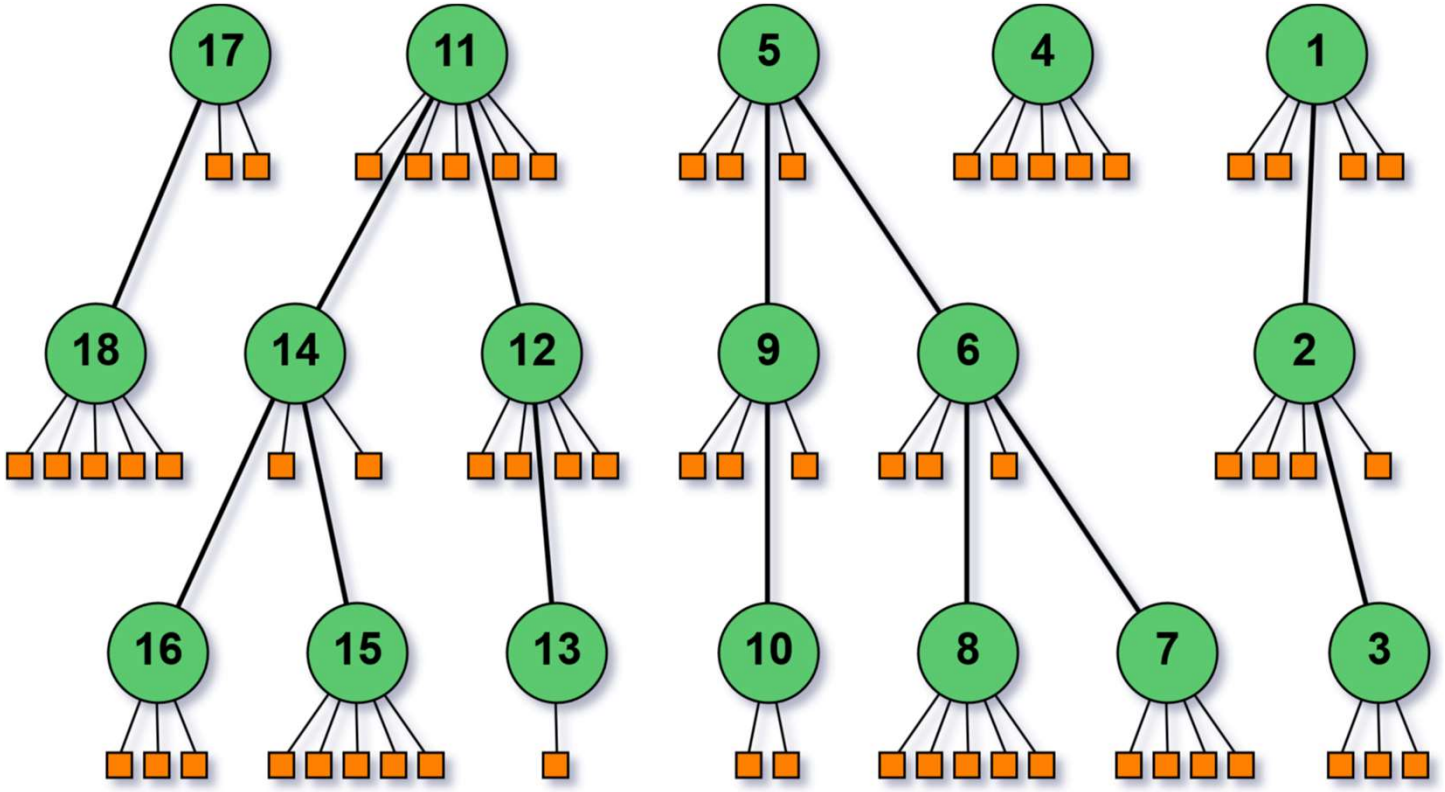


Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

- first approach: Iterate in depth-first search order over the nuggets

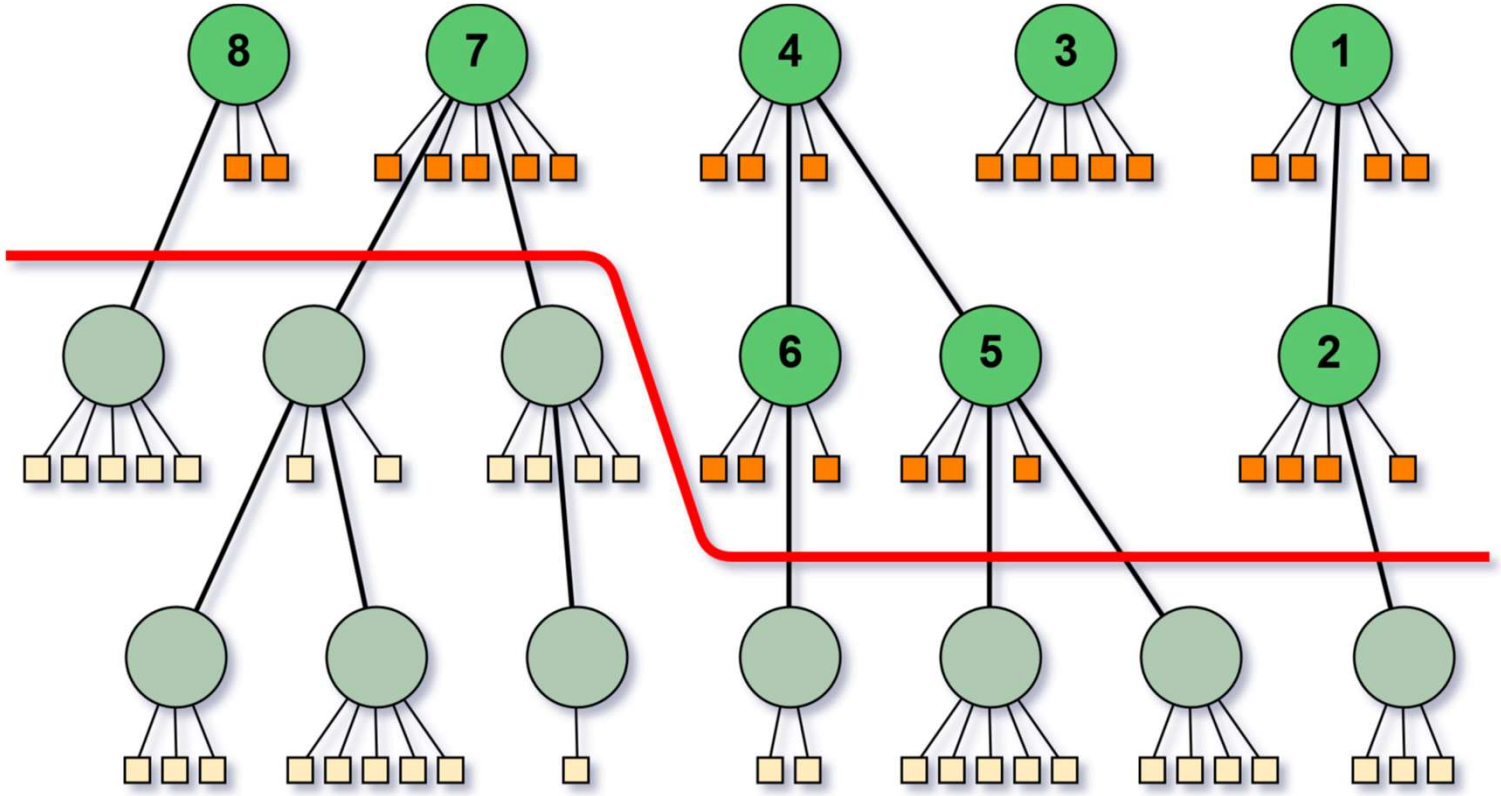


Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

- cut layers off to get shorter summaries

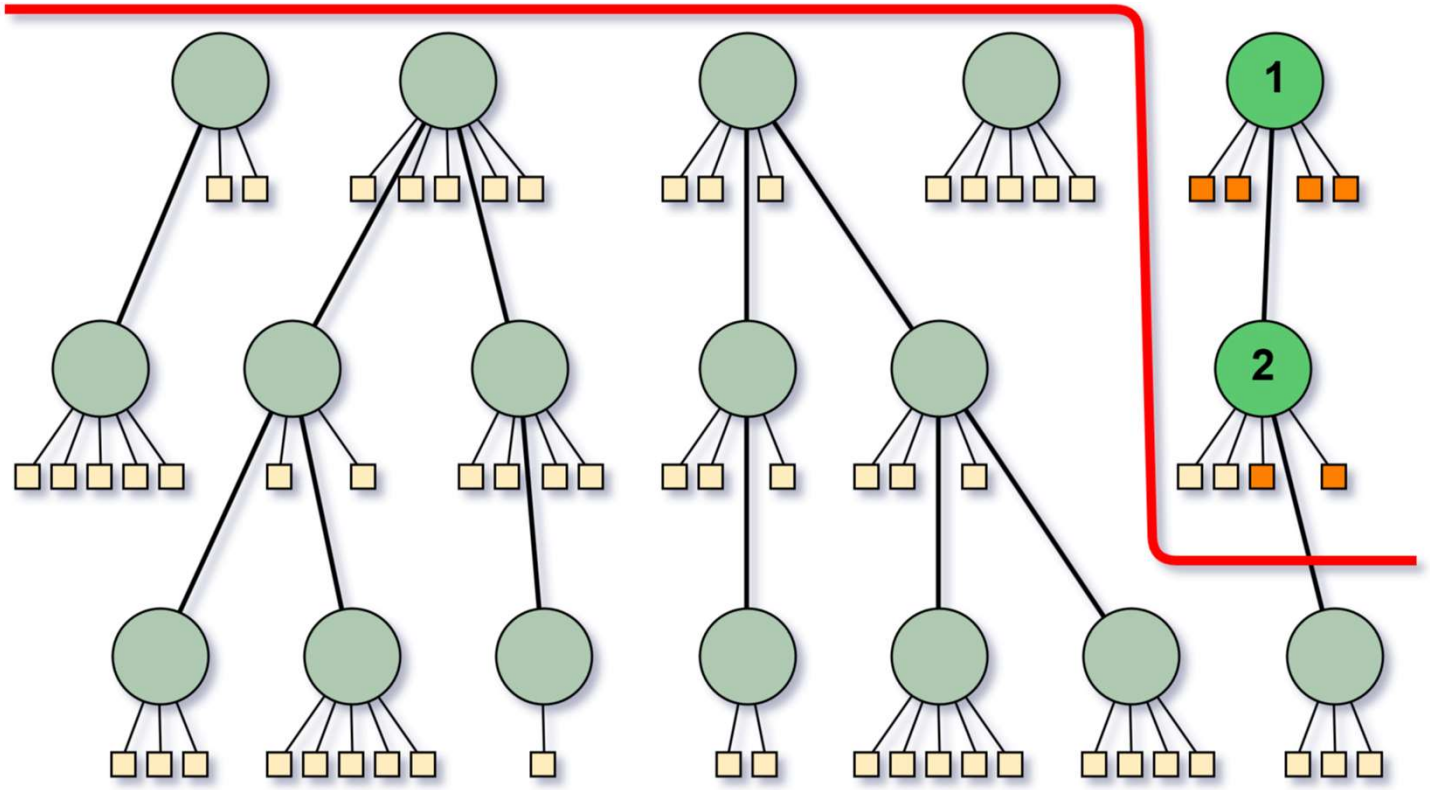


Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

- problem when summaries are very short

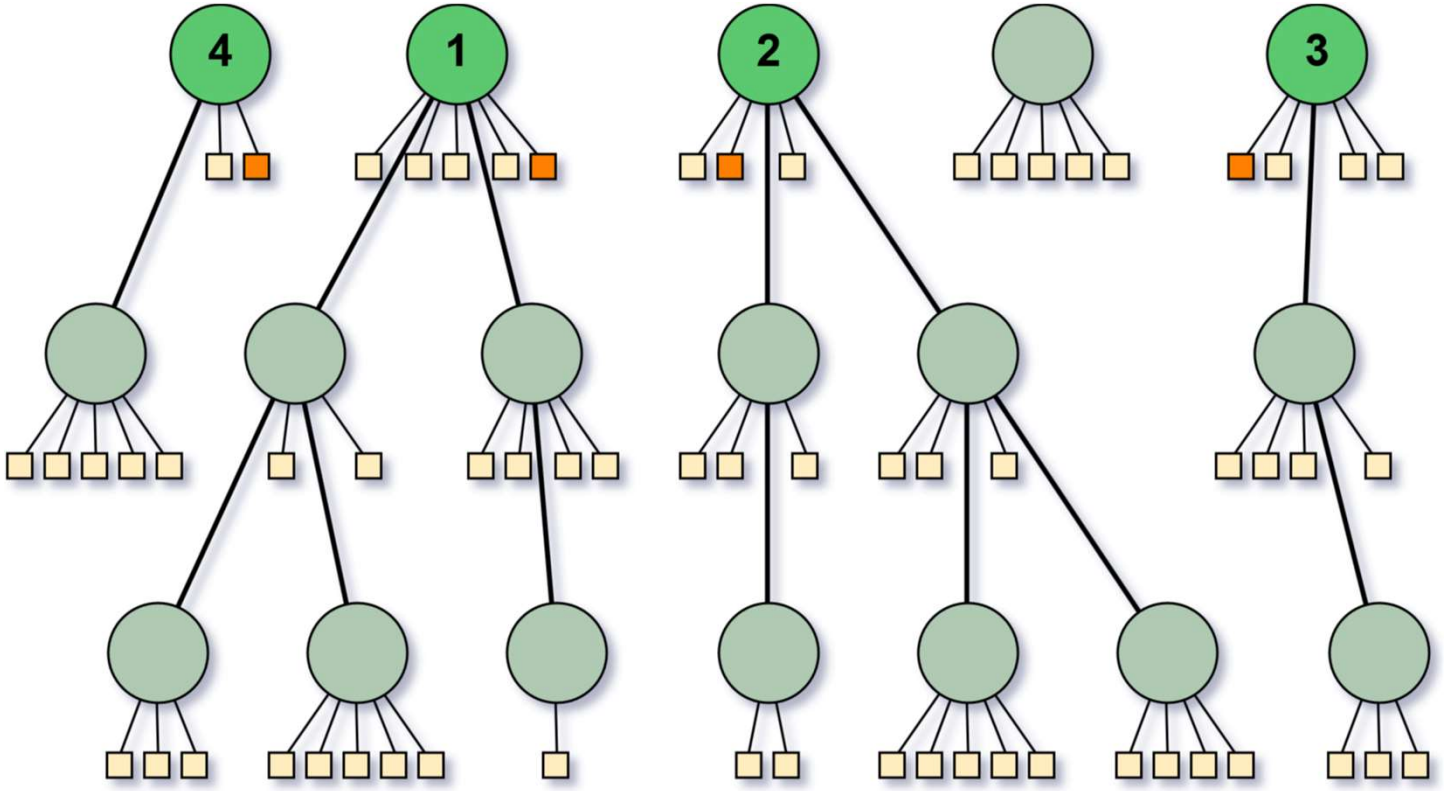


Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

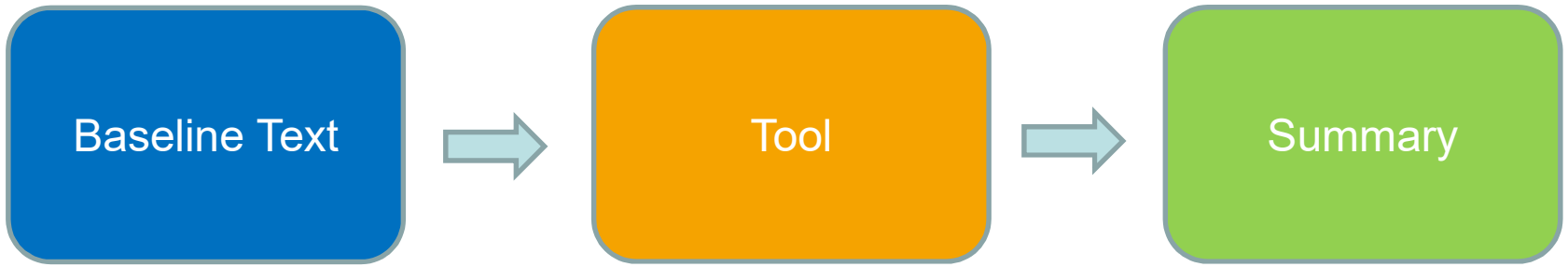
- select shortest sentences from root bubbles of the largest trees



Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

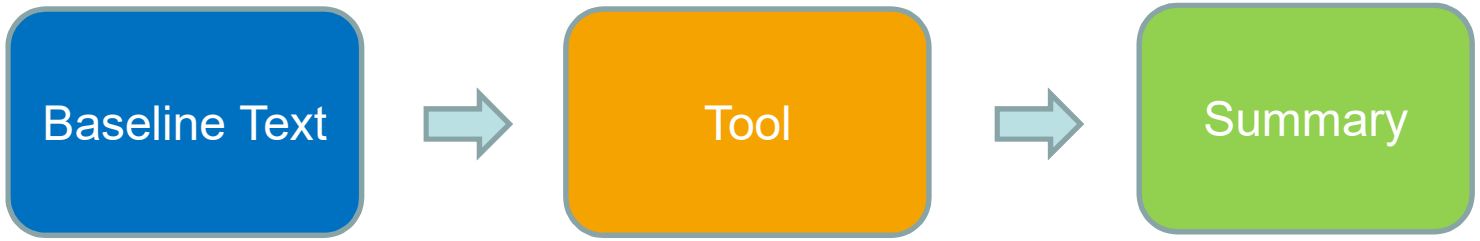
Task | Implementation | Evaluation



Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation



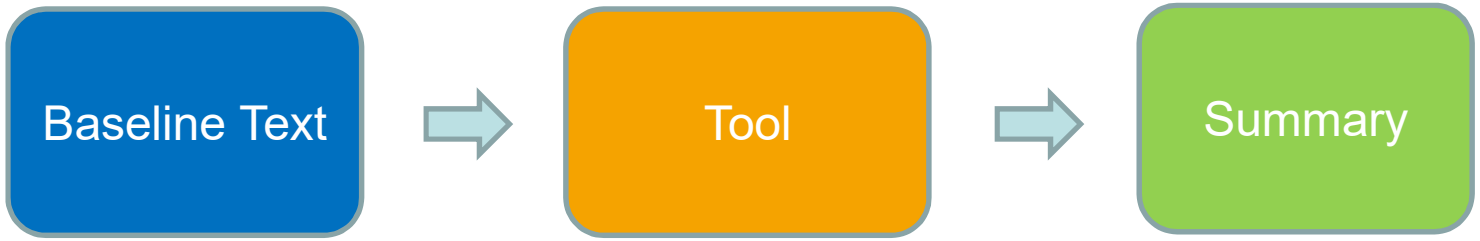
Gensim Summarizer

- Implements variation of TextRank algorithm
- Did not work well at all!

Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation



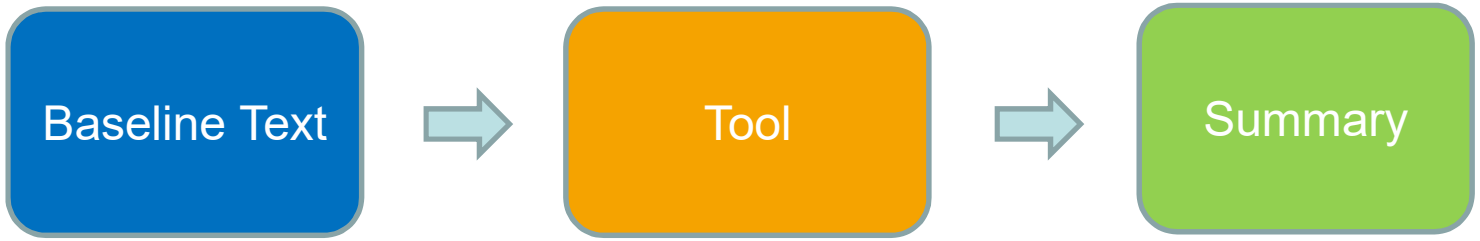
Sumy summariezers

- Implementations of KL divergence, LexRank, LSA, Luhn algorithm, TextRank, …
- Some produced barely acceptable output

Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation



Summa summarizer

- Also implementation of TextRank…
- But with optimized similarity funktion
- Worked really well for many topics

Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Problem

- Implementations mostly based on word count, sentence count or ratios
- Never on character count

Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Solution

Set word count so that approximately 600 characters are reached

Cut those sentences leading to more than 600

Adding sentences from hirarchy starting from the top level bubbles

make sure there are no duplicates

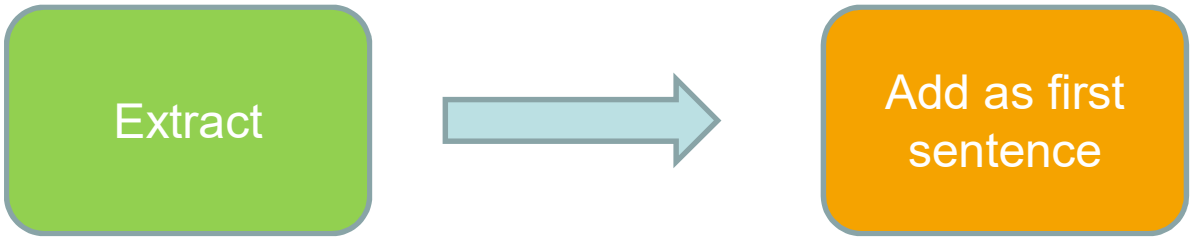
Step 3 | Creating Summaries from Hierarchies

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Observation that there are many definition sentences!

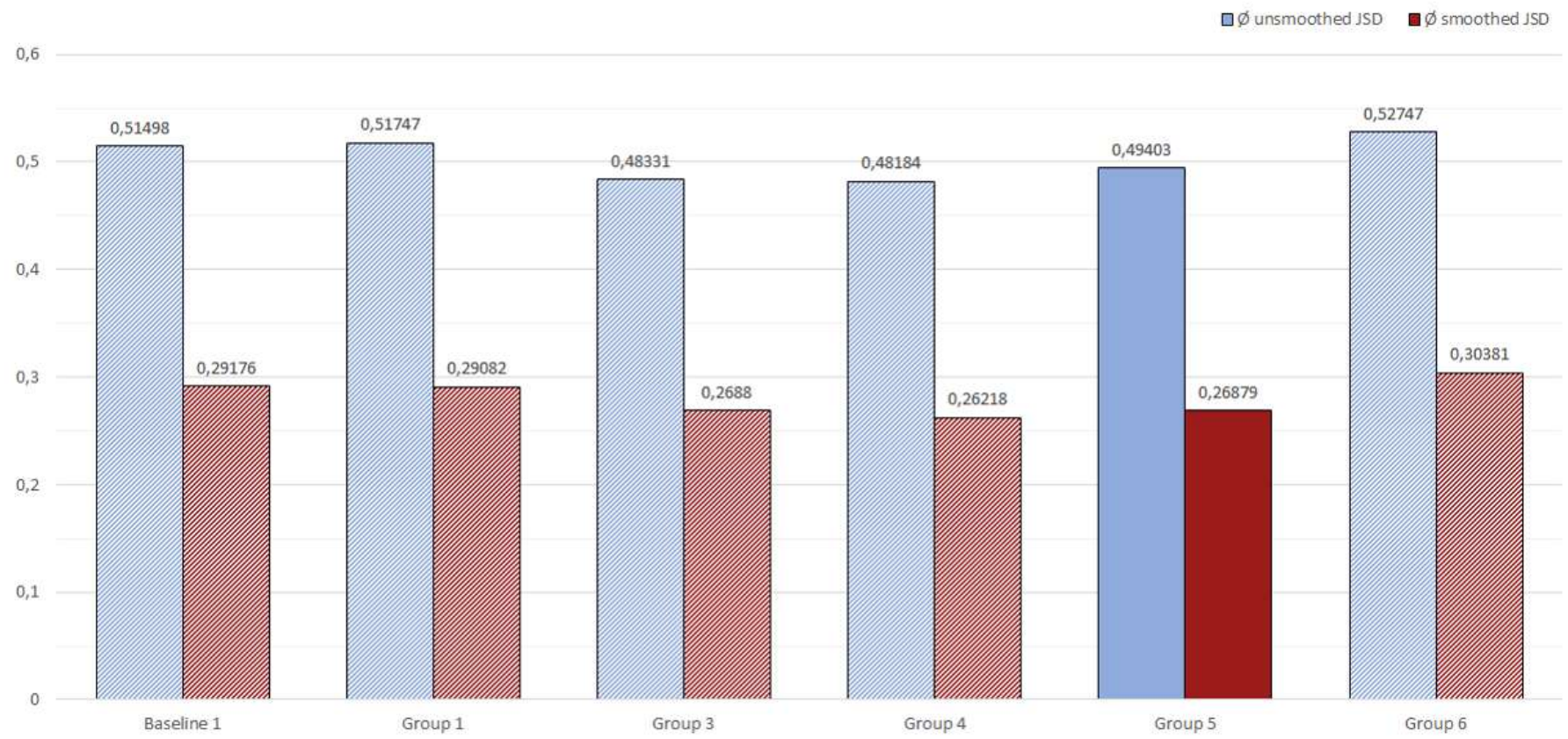
„ADHD is a brain-based disorder where the chemistry of the brain (neurotransmitters) is not functioning as it should.„



Evaluation
Comparison with JSD

TECHNISCHE UNIVERSITÄT DARMSTADT

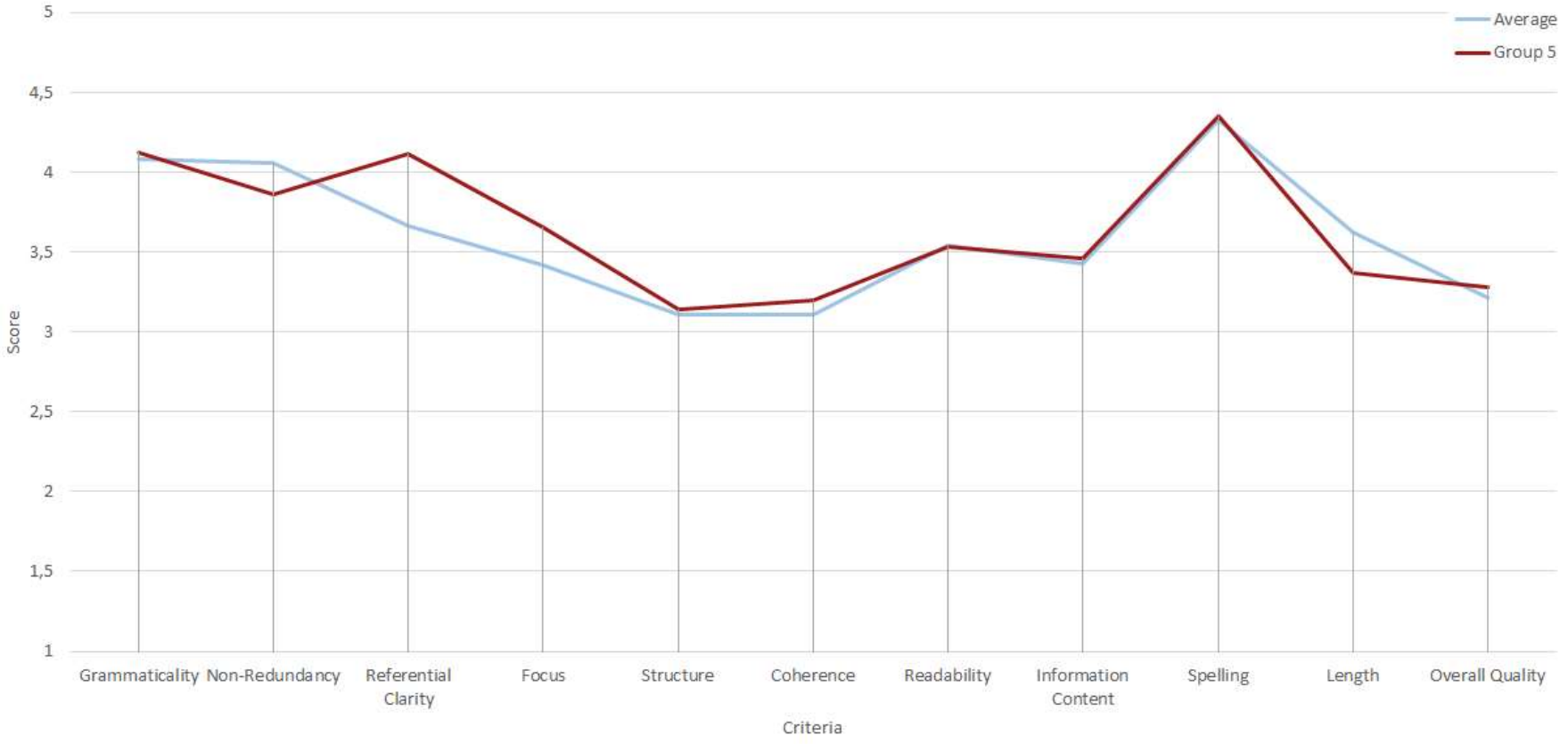
Task | Implementation | Evaluation



Evaluation
Scores per Criteria

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation



Evaluation
Score Calculation

TECHNISCHE UNIVERSITÄT DARMSTADT

Formula for calculating score for each summary evaluation:

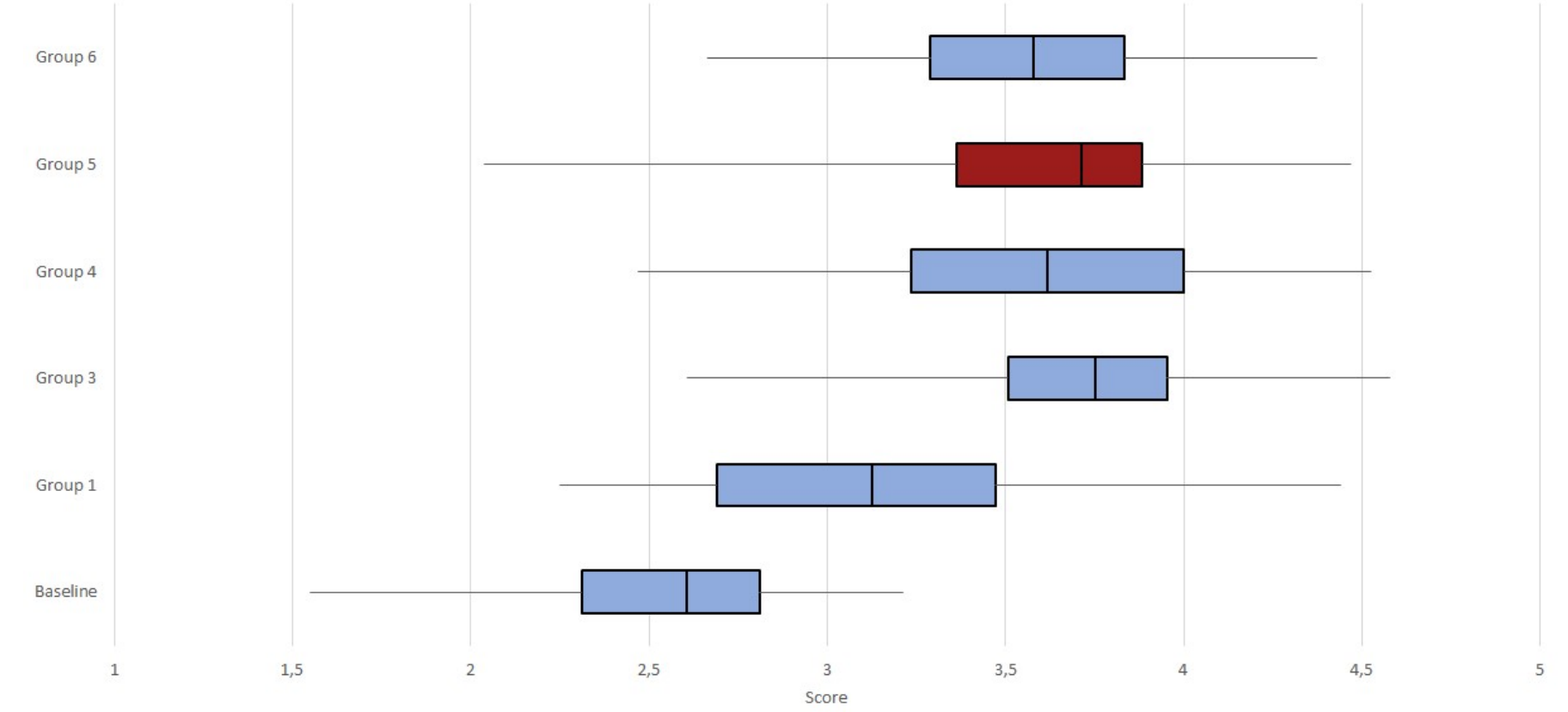
$$score_{eval} = \frac{\sum_i score_i \times weight_i \times confidence_i}{\sum_i weight_i \times confidence_i}$$

Calculate average score for each summary

Evaluation
Statistical Score Values over all Topics per Group

TECHNISCHE UNIVERSITÄT DARMSTADT

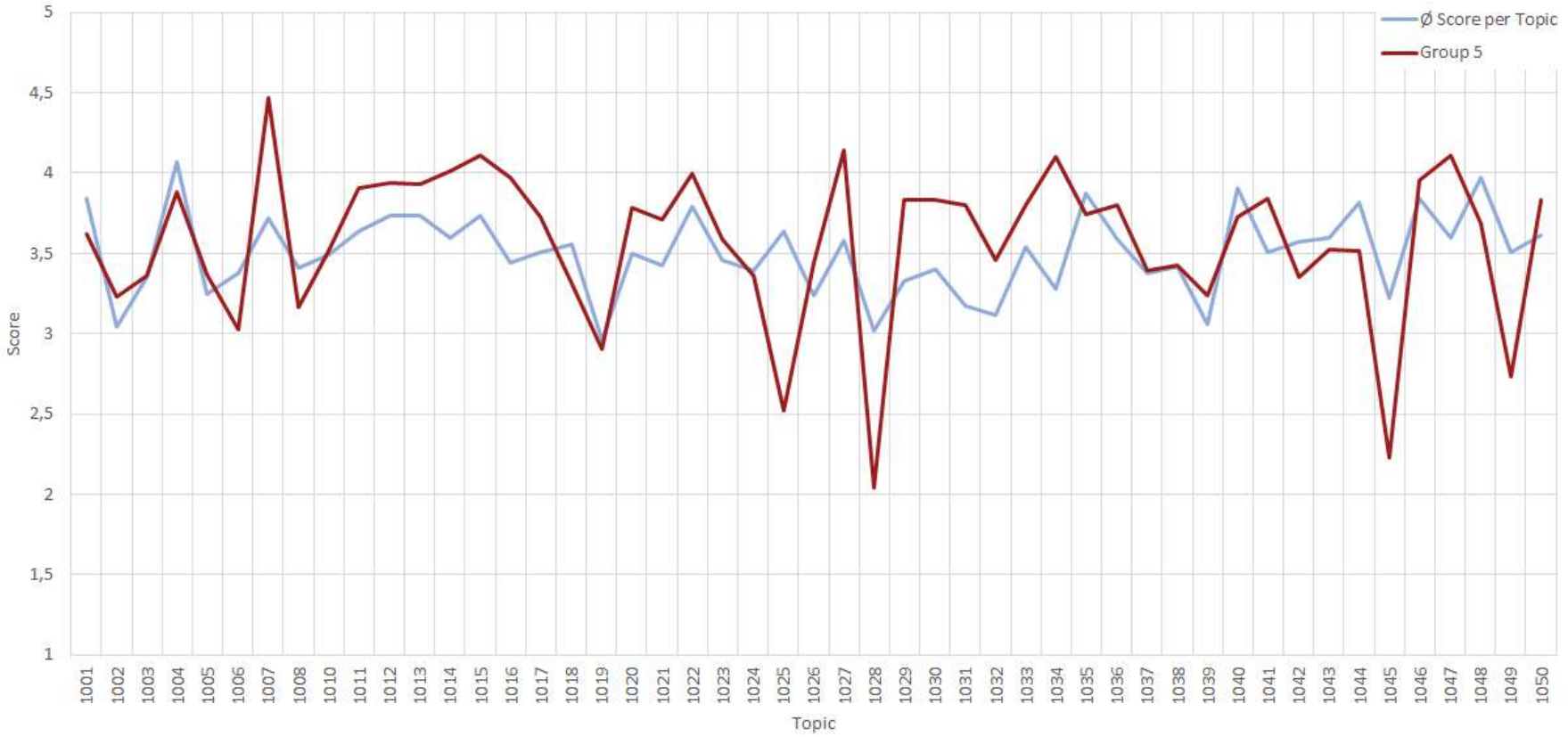
Task | Implementation | Evaluation



Evaluation
Calculated Score per Topic

TECHNISCHE UNIVERSITÄT DARMSTADT

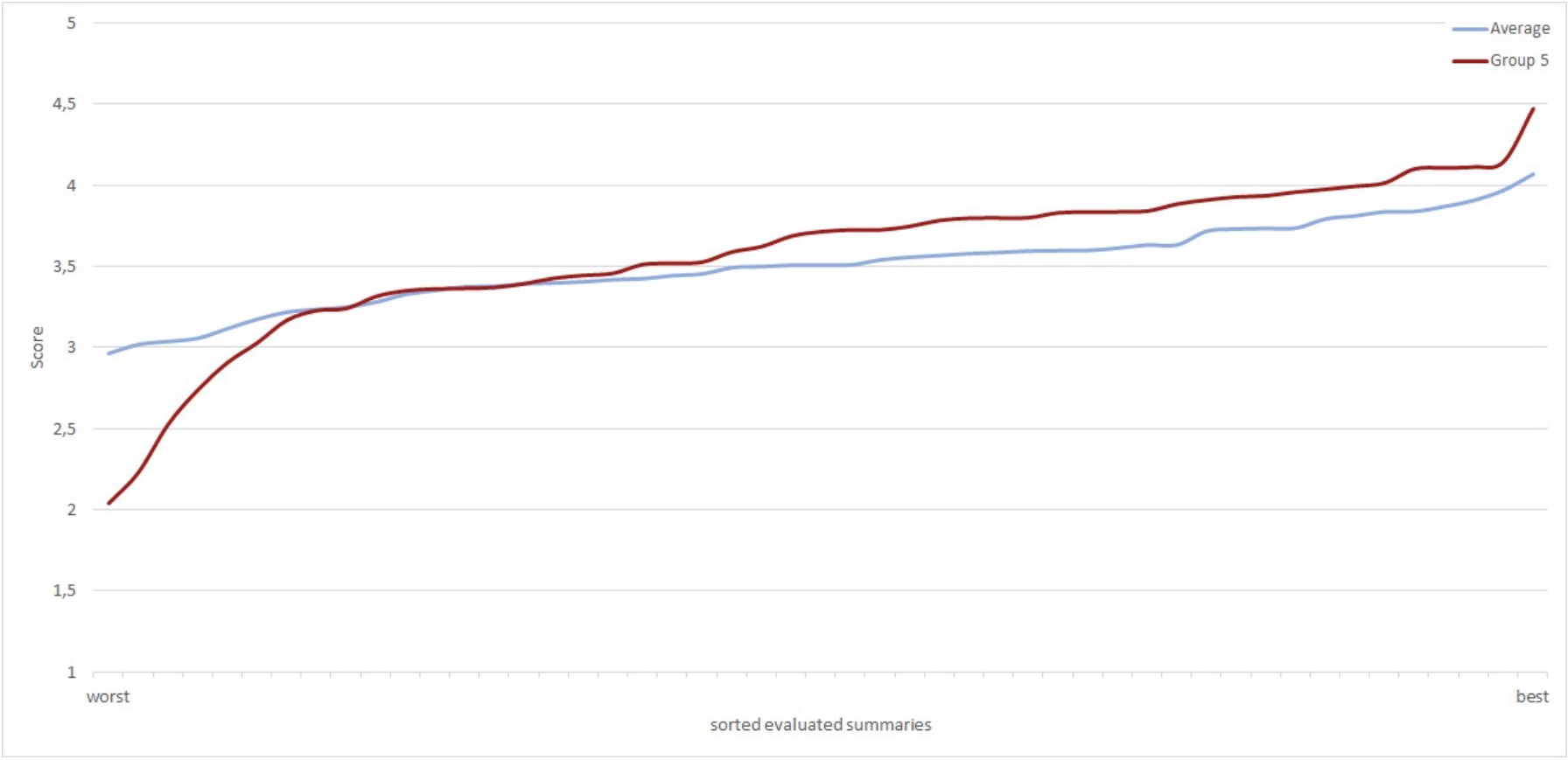
Task | Implementation | Evaluation



Evaluation
Calculated Score per Topic – Sorted

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation



Evaluation

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Our worst summary (Topic 1028):

The parents are exhibiting some serious problems as well, such as drug abuse, alcohol abuse, criminal involvement, and domestic violence.

teen discipline, teen boot camp, alcohol abuse, Binge Drinking, Substance Abuse, Addiction, bad behavior, boot camp, children respect, parenting tips, aggressive behavior, James Lehman, Total Transformation Individuals suffering from mental health disorders may use alcohol and illicit drugs to decrease or mitigate their psychological distress 16 .

Evaluation

TECHNISCHE UNIVERSITÄT DARMSTADT

Task | Implementation | Evaluation

Our best summary (Topic 1007):

Fears and Phobias can be resolved with hypnosis and hypnotherapy Hypnotherapy is an ideal option because it is safe, effective, and non-invasive.

If a parent is afraid of spiders then a child can learn that fear and it could develop into a phobia.

Hypnotherapy is effective at helping you to overcome your fear by treating the anxiety caused by the trigger, and by re-training the mind to remember the original trigger in a way that does not create anxiety.

Medication can be very effective in treating phobias, especially social phobia and agoraphobia.

TECHNISCHE UNIVERSITÄT DARMSTADT

QUESTIONS ?