Evaluation of Approaches to Summarize Dialogue Transcripts with Focus on User-friendliness

School of Engineering
Zurich University of Applied Sciences

Pascal Aigner, Maurice Gerber and Basil Rohr

Spring term 2021

Abstract This thesis is concerned with the summarization of dialog transcripts. Most research in the field of Natural Language Processing (NLP) focuses on written texts which are structurally different to dialog transcripts. The goal is to evaluate different summarization approaches and make a proposal on which ones prove suitable. The transcripts used for this thesis are an US vice presidential debate transcript and four meetings conducted during the progress of this thesis. The keyword extraction algorithms TF–IDF, RAKE, YAKE! and KeyBERT are investigated as well as the topic segmentation algorithms TextTiling and Textsplit. A new measure is introduced to evaluate the performance of topic splits by an algorithm compared to a reference split. The topic segmentation algorithms show good performance on structured transcripts. However, transcripts with a lack of clear structure, unfinished sentences and use of multiple languages lead to poor results. Several visualization approaches are worked out and an animation proves to be an intuitive way to display topic transitions over time. All discussed methods are implemented in a web application.

Keywords: Transcripts, dialogues, topic segmentation, keyword extraction, text summarization, segmentation performance

Contents

1 Introduction

Skimming a long text or dialog to find a specific piece of information is a laborious task. If it were known beforehand in which section of a text or dialog to search, it would save a lot of time. To take this into account, books and articles often have a table of content or an index. Usually, this does not apply to dialog transcripts. Imagine one was unable to attend a meeting, but a transcript is available. Going through a long transcript utterance by utterance is time-consuming. There might be an agenda, but it still does not indicate at what point in time a particular topic came up. Therefore, there is no efficient way to extract a specific piece of information other than manually skimming the transcript. Challenges such as those described above are investigated in the field of Natural Language Processing (NLP). However, most research about keyword extraction and topic segmentation algorithms focuses on written texts. These are structurally different to dialog transcripts. Former tend to have a more linear composition whereas dialogues can be more chaotic. Even for an individual it can be challenging to follow the topic shifts in brainstorming meetings or debates.

The motivation for this thesis is to propose an approach on how to summarize dialogues effectively with focus on meeting transcripts. This is done by evaluating various existing algorithms and comparing their results in terms of performance and user-friendliness. The structure of this paper reflects the thought process during the investigation. That means, after a method is described, its results are directly discussed. From this examination, new ideas are introduced and determine the further course of the investigation. The main methodological approaches presented are keyword extraction, topic segmentation and visualization. First, the four keyword extraction algorithms TF–IDF [1], RAKE [2], YAKE! [3] and KeyBERT [4] are explained and compared in terms of their properties and usability. All four follow a different approach and some prove to be more suitable for dialog transcripts than others. Second, the performance of the topic segmentation algorithms TextTiling [5] and Textsplit [6] is examined. Former applies a more linear approach in determining subtopics whereas latter leverages word embedding. A new scoring measure is introduced to evaluate the accuracy of the subtopics. Further, multiple visualization approaches are considered to display keywords and assessed in terms of their user-friendliness and comprehensibility. Along with these investigations, a web application is developed which implements the discussed algorithms and visualizations as prototypes.

2 Related Work

This thesis is related to the work done by Davina Golomingi and Luca Rüegger in 2020 [7] and Loran Avci in 2021 [8].

Davina Golomingi and Luca Rüegger enhanced the transcription tool Interscriber. For a more detailed description of what Interscriber is refer to chapter 3. They evaluated multiple use cases for additional features and implemented prototypes in the existing code. Some use cases focused on statistics such as speech distribution per speaker or most used words. Further, they developed a coherent design across all features making them visually appealing and intuitive to use. User tests were performed to obtain feedback and enhance the user experience as well as to make recommendations for upcoming releases. For future work they suggested amongst other points also the implementation of more statistics such as sentiment analysis.

Loran Avci investigated the use of text summarization in dialog transcripts. He divided the utterances in different groups depending on their length. For each group a different keyword extraction approach is applied. The results obtained through this procedure are easy to interpret and have a high information density. In his conclusion and outlook he touched upon the idea of applying topic modeling to receive information about when which topic is discussed in a dialog.

As already mentioned in the introduction, one key aspect of this thesis is to investigate topic segmentation on dialogues and therefore build upon the work by Loran Avci. The thesis by Davina Golomingi and Luca Rüegger serves as motivation for our web application and to get a feeling for user-friendliness.

3 Data

As data for this thesis serve multiple dialog transcripts. These are obtained through Interscriber[1], a platform developed by SpinningBytes AG for automatically converting speech to text, specifically designed for dialogues. An audio file can be uploaded and is then converted to text as well as segmented by speaker as shown in illustration 1. In addition to the audio file, one does also have to provide the language and the number of speakers.

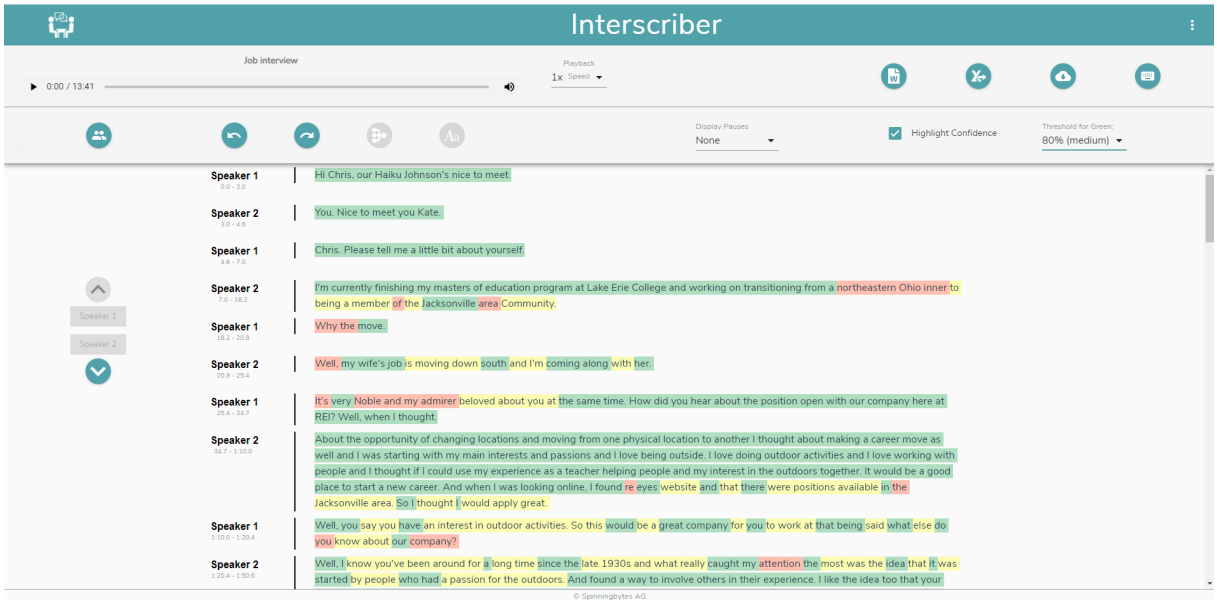


Figure 1: Interscriber interface

Some limitations of the Interscriber are cases where the utterance breaks between the speakers are not identified correctly. For example, the last word said by speaker A is recognized as the first word of speaker B. Names and strong accents can also lead to worse performance of the algorithm. However, even a human sometimes misunderstands an utterance or does not know how to spell a name.

To give indications of the confidence of transcribed words, they can be highlighted. In figure 1 the threshold is set to 80%. This means that all words which are highlighted green have a confidence score of 80% and above. Yellow and red words have lower confidence scores. The interface allows to go through the audio file again and correct wrong transcribed words or utterance breaks. For this thesis, this functionality is not used, meaning the transcripts are not manually corrected. This is a laborious task and to create summaries, minor incorrect transcriptions do not significantly impact later results.

The obtained dialog transcripts can be downloaded as csv files for further processing. The structure of such a csv export is shown in table 1.

[1] https://interscriber.com

Table 1: Transcript csv export structure

Speaker	**Start time**	**End time**	**Duration**	**Utterance**
Speaker 1	00:00	00:03	2.5	Hi Chris, our Haiku Johnson's nice to meet.
Speaker 2	00:03	00:04	1	You. Nice to meet you Kate.
Speaker 1	00:05	00:07	2.4	Chris. Please tell me a little bit about yourself.
...	...	...	...	...

By default, the speakers are consecutively numbered like *Speaker 1*, *Speaker 2* etc. It is also possible to edit these names in the Interscriber interface to match the names of the real speakers.

The transcripts used in this thesis are listed in table 2.

Table 2: Transcripts used in this thesis processed by Interscriber

Transcript	**Speakers**	**Length**	**Source**	**Ref.**
Vice presidential debate	Susan Page, Mike Pence, Kamala Harris	44 min	[9]	A.2.1
Sample job interview	Kate, Chris	13 min	[10]	A.2.2
BA meeting 2021-03-11	Pascal Aigner, Maurice Gerber, Basil Rohr, Mark Cieliebak, Don Tuggener	47 min		A.2.3
BA meeting 2021-03-25	Pascal Aigner, Maurice Gerber, Basil Rohr, Mark Cieliebak	25 min		A.2.4
BA meeting 2021-04-23	Pascal Aigner, Maurice Gerber, Basil Rohr, Mark Cieliebak, Don Tuggener	32 min		A.2.5
BA meeting 2021-05-07	Pascal Aigner, Maurice Gerber, Basil Rohr, Mark Cieliebak	33 min		A.2.6

The first two transcripts are in English and the four BA meetings in German. The vice presidential debate transcript shows a clear structure and well-spoken English which leads to a very accurate transcript. Therefore, it is used as a reference or good example when analyzing the performance of algorithms. To put it in other words, it has a low noise.

The BA meetings are four transcripts of the weekly meetings with our bachelor thesis supervisor Mark Cieliebak and sub supervisor Don Tuggener conducted during the writing of this work. However, they do not have a high degree of structure. It is to note that some transcripts start while the meeting was already in progress, as recording it from the beginning was missed. Additionally, while the language is mostly German, some parts and many words are spoken in English. The effect of this is that the transcription process failed to transcribe words or phrases correctly at some points. This results in the transcripts having a higher noise than the vice presidential debate transcript. The effects this noise has on the performance of algorithms becomes visible in later chapters.

4 Application Dialog Analyzer

For this paper, a web application is built to analyse different approaches in regard of user-friendliness. The application is constructed as an additional prototype tool for the Interscriber but in an independent environment. In this section, a brief overview regarding the Dialog Analyzer[2] is given with the necessary explanation in order to use it. The main page of the web application is depicted in figure 2.

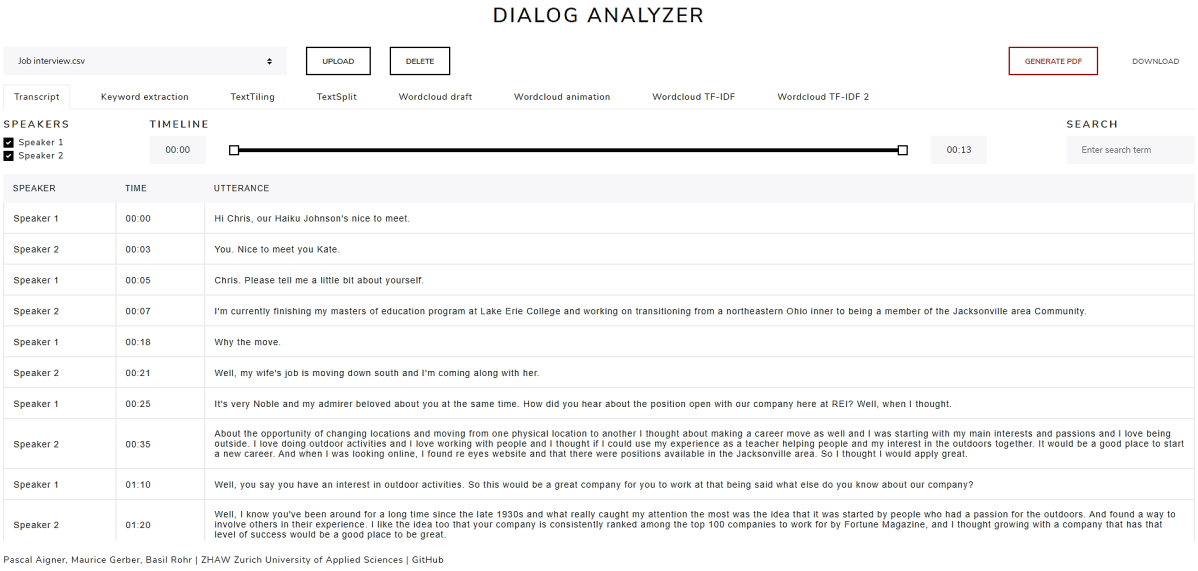


Figure 2: Main page of web application

On the main page in the top left corner there is a drop down menu where the transcript can be selected. At the right side, there is an upload and a delete button where individual files which meet the format requirements described in table 1 can be added or deleted. In the same row on the right-hand side an option is implemented where the current selected and filtered transcript can be converted to a PDF and downloaded.

The transcript tab is regarded as the main page. Besides the main tab there are several additional tabs with specific functions discussed in later chapters. With the web application, those methods can be tested out with a given transcript or a self uploaded one. The transcript tab contains some parameters which can be adjusted. On the left side there is the speaker selection, where speakers can be selected in the transcript below. The timeline limits the transcript at an individual range. The search function on the right highlights words or text passages that have been typed. The table with the transcript directly illustrates the set filters and the highlighted words. The filtered transcript can be converted to a PDF file, with the *Generate PDF* button. The Python version and used libraries can be found in A.1.1. A step by step documentation on how to start the app by oneself can be found in A.1.2.

[2]https://projects.pascalaigner.ch

5 Keyword Extraction

The goal of keyword extraction is to extract the most relevant words from a text in order to derive possible topics. An individual would do this task based on intuition, experience and background knowledge. There are multiple approaches how an algorithm identifies relevant keywords from a given text. In this chapter, four different keyword extraction algorithms are analyzed and their outputs compared based on a BBC article extract [11]. It consists of four sentences, is about the US-Russia presidential meeting in Geneva on 16 June 2021 and displayed below.

The first US-Russia summit of the Biden presidency will take place in Geneva, Switzerland, on 16 June. That comes at the tail end of Biden's already scheduled trip to the United Kingdom for the G7 summit and Brussels for a meeting of Nato leaders, giving the president plenty of time to hear from US allies before sitting down with Putin. White House Press Secretary Jen Psaki, in a statement announcing the meeting, said the summit would cover a full range of pressing issues as the US seeks to restore predictability and stability to its Russian relations. That echoes comments Secretary of State Antony Blinken made during a meeting with his Russian counterpart, Sergei Lavrov, in Iceland last week, as he said Biden's goal was a predictable, stable relationship with Russia.

5.1 TF–IDF

Term Frequency–Inverse Document Frequency or short TF–IDF is an algorithm to evaluate how relevant a word is to a document in a collection of documents. Term frequency (TF) refers to the frequency of a word in one document. Inverse document frequency (IDF) on the other hand, is in how many documents a word occurs. A word that occurs in all or most documents receives a lower score, whereas a word that occurs in one or a few documents receives a higher score. The higher the score, the more relevant a word is to its document. It is to note that a document does not necessarily has to be a whole article, it can also be a paragraph or an utterance. Following from this, multiple high scored words should give an overview of a document's topic.

There are slightly different implementations of TF–IDF. For example, the Python library `sklearn` uses an implementation which calculates the inverse document frequency differently to the formula conceived by Karen Jones in 1972 [1]. Former gives words that occur in all or many documents a higher score compared to latter. Consequentially, stopwords like *the* receive a high score despite not being relevant words. As this outcome it not desired, in this thesis TF–IDF is implemented according to the IDF formula by Karen Jones.

TF–IDF calculates a separate score for the term frequency and the inverse document frequency which are then multiplied together. The TF part calculates the frequency of a word in one document in relation to this document's total word count as shown in eq. 1.

$$tf(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \tag{1}$$

t: word
d: document
$f_{t,d}$: *count of word t in document d*
$\sum_{t' \in d} f_{t',d}$: *sum of all word counts in document d*

The IDF part calculates in how many documents a word occurs as shown in eq. 2.

$$idf(t, D) = \log \frac{N}{| \ d \in D : t \in d \ |} \tag{2}$$

t: word
d: document
N: total number of documents $N = | \ D \ |$
$| \ d \in D : t \in d \ |$: *number of documents d where word t occurs*

Eventually, TF is multiplied with IDF as shown in eq. 3 to obtain the final score. The IDF score of a word is constant across all documetns whereas its TF score varies from document to document.

$$tf\text{-}idf(t, d, D) = tf(t, d) * idf(t, D) \tag{3}$$

The algorithm is explained step by step below.

Step 1 Define the collection of documents.

Remark: What is considered as a document in the context of TF–IDF depends on the use case. In regard of the BBC article extract, a document can be defined as one sentence and there are a total of four sentences, respectively documents. In the context of dialog transcripts, a document can be considered as a single utterance or multiple utterances. The collection of documents would therefore be the whole transcript.

Step 2 Compute the term frequency (TF) as in eq. 1 for every word in a document. Do this for all documents.

Remark: Therefore, every word has a frequency for each document. For example, the word *secretary* might have a frequency of 0.05 for document A and one of 0.03 for document B. If it does not occur in document C, its frequency would be 0 there.

Step 3 Compute the inverse document frequency (IDF) as in eq. 2 for every word.

Remark: Hence, every word has one inverse frequency across all documents. For example, the word *secretary* might occur in two out of three documents and would therefore have an IDF of $log(3/2) \approx 0.4$. A stopword like *the* is most likely to occur in every document and would therefore have an IDF of $log(3/3) = 0$.

Step 4 Compute the TF–IDF score as in eq. 3 for every word in a document. Do this for all documents.

The highest score is achieved if a word occurs only in one of many documents and has a high frequency in this specific document. Such high scored words are most likely to provide relevant information about the topic of the document. As emerged from the explanation above, stopwords are not filtered before applying TF–IDF. Stopwords occur many times in a document and therefore have a high TF score. However, as they are also likely to occur across all documents, their IDF score will be 0, as briefly touched upon in step 3 above. As as result, the multiplication of TF and IDF also yields a score of 0. Consequentially, stopwords will never appear as relevant words with a high score. Removing stopwords would even impact the TF scores of other words which is not desired. As a result, TF–IDF is language independent as there is no need for a stopword list.

The output of TF–IDF can be displayed as a $m \times n$ matrix with the rows m being the words and the columns n being the documents. The BBC article extract is used to generate such an output where each sentence represents a document. This is a brief example and normally one would choose longer texts as a representation for a document. Table 3 depicts the TF–IDF output matrix for six arbitrarily chosen words.

Table 3: TF–IDF scores for six words from the BBC article extract

	Sentence 1	**Sentence 2**	**Sentence 3**	**Sentence 4**
Switzerland	0.081547	0.000000	0.000000	0.000000
of	0.000000	0.000000	0.000000	0.000000
the	0.033845	0.026761	0.023974	0.000000
summit	0.016922	0.006690	0.007991	0.000000
secretary	0.000000	0.000000	0.019254	0.020387
Russian	0.000000	0.000000	0.019254	0.020387

The word *Switzerland* is only mentioned in the first sentence and hence receives a high score there. The stopword *of* receives a score of 0 across all sentences meaning it occurs in all of them. On the other hand, the stopword *the* appears multiple times in three out of four sentences and therefore has rather high scores in those. It even has higher scores than the word *summit* which would be more relevant to derive a possible topic. This is owed to the fact that a document is only represented by one sentence. With longer documents, the stopword *the* will most likely receive a score of 0 across all documents. The scores of *secretary* and *Russian* are equal across all sentences which could indicate co-occurrence.

5.2 RAKE

Rapid Automatic Keyword Extraction or short RAKE is an algorithm developed by Stuart Rose et al. in 2010 [2]. The motivation of the authors was to develop a keyword extractor with a high computing speed and which does not have to be trained on a text corpus. How the algorithm works in detail is explained below with the help of the BBC article extract.

Step 1 Split the text by stopwords and phrase delimiters and remove those.

Remark: The stopword lists are from the Python library `nltk` and the phrase delimiters from the library `string`. Both lists can be found in the appendix A.3.1, A.3.2 and A.3.3. After removing the stopwords and phrase delimiters, the possible keywords are left. In the case of RAKE, keywords can also consist of multiple words. To give an example, the possible keywords of the third sentence in the BBC article extract are shown below.

'white house press secretary jen psaki', 'summit would cover', 'us seeks', 'statement announcing', 'russian relations', 'restore predictability', 'pressing issues', 'full range', 'stability', 'said', 'meeting'

The longer a sequence of words is without containing a stopword or phrase delimiter, the longer the possible keyword is.

Step 2 Calculate the frequency and degree of every word.

Remark: This process is illustrated with the help of the matrix in table 4. It consists of eight words out of all 70 unique non-stopwords from the BBC article extract. The displayed words are part of the two possible keywords *white house press secretary jen psaki* from sentence three and *echoes comments secretary* from sentence four.

Table 4: Word co-occurrence matrix for eight words from the BBC article extract

	comments	**echoes**	**house**	**jen**	**press**	**psaki**	**secretary**	**white**
comments	1	1	0	0	0	0	1	0
echoes	1	1	0	0	0	0	1	0
house	0	0	1	1	1	1	1	1
jen	0	0	1	1	1	1	1	1
press	0	0	1	1	1	1	1	1
psaki	0	0	1	1	1	1	1	1
secretary	1	1	1	1	1	1	2	1
white	0	0	1	1	1	1	1	1

The values on the diagonal are the word frequency $freq(w)$. For example, *secretary* occurs twice in the text and therefore $freq(w) = 2$. The other values are the co-occurrence of word i with another word j within a possible keyword. For example, the word *echoes* appears once together with the words *comments* and *secretary* in the possible keyword *echoes comments secretary*. Therefore, the values in the matrix at the corresponding positions equal 1.

The sum of each column equals the word degree $deg(w)$. This means, the word degree is the sum of the word frequency and the co-occurrences with other words within a possible keyword. In the case of the word *secretary* this results in $deg(w) = 7 * 1 + 2 = 9$.

Step 3 Calculate the final score of all possible keywords.

Remark: Each word degree $deg(w)$ is divided by the corresponding word frequency $freq(w)$. This ratio is considered as score for each word and illustrated in table 5.

Table 5: Word frequency, word degree and their ratio

	comments	**echoes**	**house**	**jen**	**press**	**psaki**	**secretary**	**white**
deg(w)	3	3	6	6	6	6	9	6
freq(w)	1	1	1	1	1	1	2	1
deg(w)/ freq(w)	3	3	6	6	6	6	4.5	6

The score of a possible keyword is determined by summing up the individual word scores within a keyword. For example:
white house press secretary jen psaki: $6 + 6 + 6 + 4.5 + 6 + 6 = 34.5$
echoes comments secretary: $3 + 3 + 4.5 = 10.5$
The higher the score, the more relevant a keyword is considered.

5.3 YAKE!

Yet Another Keyword Extractor short YAKE! is an algorithm developed by Ricardo Campos et al. in 2020 [3]. The YAKE! algorithm does not need to be trained on a text corpus. It does only analyse the inputted text and gives a final score to each keyword. The lower the score, the more relevant the keyword. How the algorithm works is explained below.

Step 1 Pre-process the text.

Remark: The text is split into individual words whenever a delimiter is found. Stopwords are also filtered according to a stopword list provided by the Python library `yake`. It can be found in the appendix A.3.4 and A.3.5.

Step 2 Apply the feature extraction to each word.

Remark: Feature extraction determines the importance of a word, taking into account five different features. Each feature assigns a certain score to a word based on different aspects, which are explained below.

Casing: This feature focuses on capital letters. Words starting with an uppercase letter or acronyms are considered more important than words only in lowercase letters, except at the beginning of a sentence.

$$T_{Case} = \frac{max(TF(U(t)), TF(A(t)))}{ln(TF(t))} \tag{4}$$

TF(U(t)): number of occurrences of the word t starting with an uppercase letter
TF(A(t)): number of times the word t is marked as an acronym
TF(t): frequency of word t

Word position: The position of a word is considered as an indicator of how important it is within a document. If the keyword is located at the beginning, it tends to be more important compared to words in the middle or at the end of a document.

$$T_{Position} = ln(ln(3 + Median(Sen_t))) \tag{5}$$

Sen_t: Set of positions of the sentences where the word t occurs

Word frequency: This feature calculates a score for each word based on its frequency. It is to note that a higher frequency means a higher importance. However, the importance is not proportional to the frequency of a word. For long documents this behavior prevents a bias towards high frequencies.

$$TF_{Norm} = \frac{TF(t)}{MeanTF + 1 * \sigma} \tag{6}$$

TF(t): frequency of word t
MeanTF: mean of all word frequencies
σ: standard deviation of all word frequencies

Word relatedness to context: This feature is based on the assumption that the more different words appear together with a specific word on either side, the less significant this word is. For example, the word *the* is surrounded by all kinds of different words and therefore not significant. A word such as *white* might occur before the word *house* multiple times across a document and might therefore be more important.

$$T_{Rel} = (DL + DR...) * \frac{TF(t)}{MaxTF} \tag{7}$$

DL: dispersion of left side words related to a specific word t
DR: dispersion of right side words related to a specific word t
TF(t): frequency of word t

Term different sentence: This feature identifies how often a word occurs in different sentences. The assumption is made, that a word which appears in many different sentences is considered more important.

$$T_{Sentence} = \frac{SF(t)}{\#Sentence} \tag{8}$$

SF: number of sentence where the word t appears
#Sentence: total number of sentences

Step 3 A total score is calculated from the scores of each feature according to eq. 9.

$$S(t) = \frac{T_{Rel} * T_{Position}}{T_{Case} + \frac{TF_{Norm}}{T_{Rel}} + \frac{T_{Sequence}}{T_{Rel}}} \tag{9}$$

Step 4 N-gram generation and keyword score calculation.

Remark: The *n-gram* is the maximal size of a keyword, but not restricted to this number. For example, *n-gram* = 2, words like *White House* or *United Kingdom* are considered as keywords. However, keywords of length 1 like *congress* can also occur. If *n-gram* = 1, only keywords of length 1 are considered.

The final score for the keywords is calculated according to eq. 10. The numerator multiplies scores of individual words $S(t)$ that are part of the keyword. Whereas the denominator multiplies the keyword frequency $KF(kw)$ by 1 plus the sum of the individual word scores $S(t)$ that are part of the keyword.

$$S(kw) = \frac{\prod_{t \in kw} S(t)}{KF(kw) * \left(1 + \sum_{t \in kw} S(t)\right)} \tag{10}$$

Step 5 Data deduplication and ranking.

Remark: In the last step the algorithm decides if similar possible keywords will be displayed or avoided in the output. The corresponding parameter is θ which is a deduplication threshold and equals a value within the interval $[0, 1]$. If $\theta = 1$, it allows words to be used several times, for example *press secretary* and *comments secretary*. If $\theta = 0$, then duplicated words are avoided.

In table 6 the output of the YAKE! algorithm for our BBC article extract is shown. It consists the two highest ranked keywords for each sentence and their corespondent scores. The input parameters are *n-gram* = 2 and $\theta = 0.1$.

Table 6: Two most important keywords per sentence ranked by score calculated with YAKE!

Sentence	**First keyword and score**	**Second keyword and score**
1	'biden presidency', 0.0138	'us-russia summit', 0.0257
2	'united kingdom', 0.0033	'scheduled trip', 0.0092
3	'jen psaki', 0.0032	'white house', 0.0052
4	'sergei lavrov', 0.0033	'antony', 0.0859

5.4 KeyBERT

The KeyBERT algorithm implemented by Maarten Grootendorst in 2020 [4] leverages the Bidirectional Encoder Representation from Transformers (BERT) developed by Google employees Jacob Devlin et al. in 2018 [12]. While keyword extractors such as RAKE and YAKE! are based on statistical properties of a text, KeyBERT focuses on semantic similarity. Hence, words are represented as vectors in a high-dimensional vector space, referred to as word embedding. Generally, words that are close to each other in this vector space are expected to have a similar meaning. There are multiple approaches to produce such word embeddings. BERT is one of the most recent developments at the time of writing using Long Short–Term Memory and Transformer neural network architectures. The topic of word embedding will reappear in this thesis in section 6.2 where a more in-depth discussion is provided. How the KeyBERT algorithm operates is explained below.

Step 1 Split the text into candidate keywords according to the pre-defined parameter *n-gram range* and remove stopwords.

Remark: The parameter *n-gram range* requires a tuple where the first entry is the minimum length of the keyword and the second entry is its maximum length. To give an example, the keyword candidates of the third sentence in the BBC article extract with *n-gram range* $= (1, 2)$ are shown below.

'announcing', 'announcing meeting', 'cover', 'cover full', 'full', 'full range', 'house', 'house press', 'issues', 'issues us', 'jen', 'jen psaki', 'meeting', 'meeting said', 'predictability', 'predictability stability', 'press', 'press secretary', 'pressing', 'pressing issues', 'psaki', 'psaki statement', 'range', 'range pressing', 'relations', 'restore', 'restore predictability', 'russian', 'russian relations', 'said', 'said summit', 'secretary', 'secretary jen', 'seeks', 'seeks restore', 'stability', 'stability russian', 'statement', 'statement announcing', 'summit', 'summit would', 'us', 'us seeks', 'white', 'white house', 'would', 'would cover'

The list is quite extensive as every non-stopword appears on its own and with its neighbours in a pair. The stopwords are filtered according to the stopword list provided by the Python library `nltk`. It can be found in the appendix A.3.2 and A.3.3.

Step 2 Convert the original text and candidate keywords list to BERT sentence embeddings.

Remark: Sentence embedding follows the same logic as word embedding. Sentences that are closer to each other in the vector space are expected to have a similar meaning. The model used to create the BERT sentence embeddings is `bert-base-nli-mean-tokens` by Nils Reimers and Iryna Gurevych [13].

Step 3 Calculate the cosine similarity between the original text and each candidate keyword. Choose the n candidate keywords with the highest similarity to the original text.

Remark: As stated in step 2, the closer two sentences are in the vector space the more similar is their meaning expected to be. This similarity can mathematically be described by the cosine similarity shown in eq. 11 for two vectors $\vec{A}$ and $\vec{B}$.

$$similarity = cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\left\|\vec{A}\right\| \left\|\vec{B}\right\|} \tag{11}$$

$\vec{A} \cdot \vec{B}$*: dot product of* $\vec{A}$ *and* $\vec{B}$
$\left\|\vec{A}\right\| \left\|\vec{B}\right\|$*: magnitude of* $\vec{A}$ *times magnitude of* $\vec{B}$

Resulting from this, there is a distance between each candidate keyword and the original text. Candidate keywords which are more similar to the original text tend to be a good representation of its topic.

Step 4 (Optional) Apply a diversification procedure parallel to step 3.

Remark: Depending on the parameter *n-gram range* there could be candidate keywords which are similar. For example, with *n-gram range* $= (2, 3)$ there might be a candidate keyword *white house* and another *white house press* which are very similar and could both occur in the selected n keywords. Although these similar keywords might represent the topic well, it may not be desired to have such a high duplication of words. To account for this, a parameter for the diversification in the interval $[0, 1]$ can be set. The higher, the more diverse the keywords are. The algorithm implements this with a threshold of maximum similarity between keywords. If two are too similar, one is excluded. There are a few additional restrictions in the process of diversification which are not further elaborated on.

In table 7 the output of the KeyBERT algorithm for the BBC article extract is shown. The input parameters are *n-gram range* $= (1, 2)$ and *diversity* $= 0.5$. The higher the score, the better a keyword is.

Table 7: Two most important keywords per sentence ranked by score calculated with KeyBERT

Sentence	**First keyword and score**	**Second keyword and score**
1	'geneva switzerland', 0.5406	'june', 0.3994
2	'giving president', 0.4267	'meeting nato', 0.3873
3	'announcing meeting', 0.457	'russian relations', 0.3061
4	'meeting russian', 0.4999	'iceland week', 0.3279

5.5 Discussion

Each keyword extraction algorithm is applied to the vice presidential debate and BA meeting 2021-03-11 transcript. To accomplish this, all utterances of a transcript are concatenated into one text block to which the different algorithms are then applied. A distinction has to be made between TF–IDF and the other three, RAKE, YAKE! and KeyBERT. Former cannot be applied to only one text or document. There needs to be at least two documents such that an IDF score can be calculated. Therefore, TF–IDF is not applied to the vice presidential debate transcript as a whole.

The keywords for the vice presidential debate are shown in table 8 for the RAKE, YAKE! and KeyBERT.

Table 8: Keyword extractor comparison based on vice presidential debate transcript

Algorithm	**Top 3 keywords and their scores**	**Exec. time**
RAKE	('cut taxes roll back regulations unleashed american energy', 38.463), ('existing condition heart disease diabetes breast cancer', 37.667), ('greatest national mobilization since world war two', 31.583)	0.057 s
YAKE!	('joe biden', 0.000), ('trump', 0.002), ('kamala', 0.050)	0.658 s
KeyBERT	('approves vaccine election', 0.408), ('cancer coming love', 0.407), ('president trump doctors', 0.398)	36.919 s
TF–IDF	*NA*	*NA*

What is noticeable at first glance is that the keywords of RAKE are up to eight words long. This is owed to the fact that the length of a possible keyword has a large effect on its score. Therefore, longer possible keywords tend to be the ones with higher scores. For the YAKE! output the parameters $n\text{-}gram = 3$ and $\theta = 0$ are used. Note that $n\text{-}gram = 3$ does not necessarily correspond to keywords of length 3 as seen from the output. All three keywords are names of persons. Since the YAKE! algorithm takes into account the capitalization of words, uppercase words such as names tend to be scored higher than lowercase words. The KeyBERT output is produced with the parameters $n\text{-}gram\ range = (1, 3)$ and $diversity = 0$. Note that all three keywords are of length 3 and do not make much sense from a semantic point of view. Further, the execution time is multiple times higher than for the other two keyword extractors.

The keywords for the BA meeting 2021-03-11 transcript are shown in table 9 for the RAKE, YAKE!, KeyBERT and TF–IDF. To obtain the ones for the TF–IDF, all four available meeting transcripts are used as the collection of documents.

Table 9: Keyword extractor comparison based on the BA meeting 2021-03-11

Algorithm	**Keywords and their scores**	**Exec. time**
RAKE	('hi would like to note post participants name harmonie people hm exakte meeting', 138.857), ('standard o meeting protokoll punkt bitch please access', 50.750), ('schreiben leute meetings amis a meeting minutes setzen einfach', 45.886)	0.037 s
YAKE!	('meeting', 0.002), ('hm', 0.006), ('einfach', 0.015)	0.530 s
KeyBERT	('gegenüber niedergeschriebene eingeschrieben', 0.706), ('filmaufnahme zusammenschneiden dafür', 0.699), ('einmaliges müssen ausgewöhnlich', 0.696)	47.797 s
TF–IDF	('business' , 0.004), ('fassung', 0.002), ('meetings', 0.002)	0.211 s

The same parameter settings are applied as for the vice presidential debate. Again, the keywords of the RAKE have a significant length and seem to be arbitrarily chosen words concatenated together. There is also a mix between German and English. Although one could limit the maximal length of a keyword, this would only filter the highest scored keywords from the output and display overall lower scored keywords, which is not desired. The YAKE! shows one reasonable keyword and has a fast execution time. The KeyBERT shows less useful keywords and has a very high execution time. TF–IDF is limited in the fact that it can only be applied if there is a collection of documents. However, if this is given, the computation time is fast and the keywords are generally well chosen.

What emerges from this section is that it might not be optimal to extract keywords from a transcript as a whole. It could be more useful if the transcript is divided into coherent segments and keyword extraction is applied to each segment. This can result in more specific keywords and give insights in the topical shift of a transcript. Additionally the TF–IDF algorithm can be used in order to extract keywords, if the text is splitted into paragraphs. This subject matter is investigated in the following chapter 6.

The four keyword extraction algorithms presented in this chapter are also implemented in the tab keyword extraction within the web application as illustrated in figure 3.

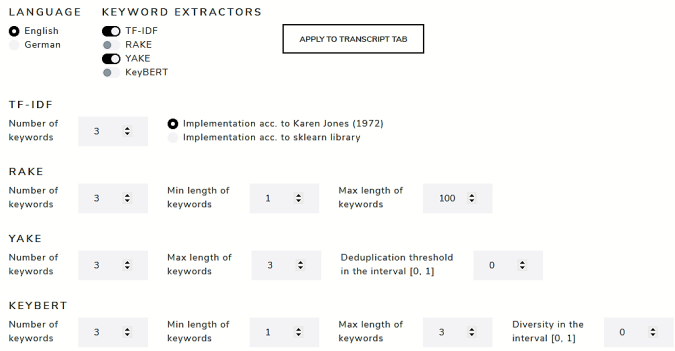


Figure 3: Keyword extraction tab in web application

The toggle button allows to enable and disable an algorithm. The parameters discussed in the explanation of the algorithms can also be altered. When clicking *Apply to transcript table*, the enabled algorithms calculate the keywords per utterance of the currently selected transcript. These are shown in the transcript table in the first tab as depicted in figure 4.

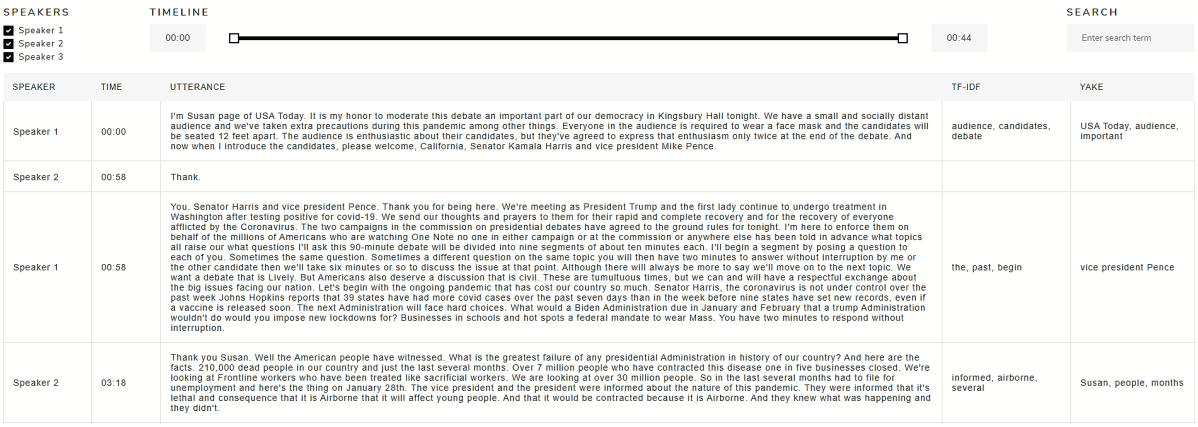


Figure 4: Transcript tab in web application

This feature allows to directly compare the performance of the four keyword extraction algorithms. Note that for TF–IDF the collection of documents are all utterances. This means that there are as many documents as there are utterances in a transcript.

6 Topic Segmentation

The goal of topic segmentation is to split a given text into coherent subtopics. There are several approaches how an algorithm recognizes subtopic shifts. In this chapter, two different topic segmentation algorithms are analyzed based on the dialog transcripts introduced in chapter 3. A new measure is introduced to evaluate the topic splits determined by the algorithms in comparison to reference splits.

6.1 TextTiling

TextTiling is a technique proposed by Marti Hearst in 1997 [5] for dividing texts into multi-paragraph units that represent subtopics. The cues for detecting major subtopic shifts are patterns of lexical co-occurrence. In the context of this thesis, the goal is to apply TextTiling to divide transcripts into multi-utterance units that represent subtopics.

6.1.1 Algorithm

The algorithm is described in detail, along with some remarks to better understand the procedure and why some decisions are made.

Step 1 Concatenate all utterances in the transcript into one continuous text block.

Remark: This keeps the chronological order of all utterances but removes the speaker dimension. In other words, the text block does not contain information on which speaker said which sentence. It is assumed that all speakers impact subtopic shifts equally and therefore all utterances are concatenated without prefiltering.

Step 2 Divide the text block into pseudosentences of a predefined length w.

Remark: The best practise is to use $w = 20$ as stated in [5]. This parameter will be investigated in more detail in section 6.1.2. The reason for breaking up the original sentence structure and forming pseudosentences is the variation in sentence length. Later, the pseudosentences are compared to find lexical co-occurrences. For the best possible results, all sentences must have the same length.

Step 3 Remove the stop words from each pseudosentence.

Remark: Stop words distort the process of finding lexical co-occurrences between pseudosentences. Therefore, they are removed beforehand.

Step 4 Apply the block comparison method with a predefined block size k.

Remark: The best practise is to use $k = 10$ as stated in [5]. This parameter will be investigated in more detail in section 6.1.2. As this is the core step of the algorithm, it is explained in detail with the help of the visualization depicted in figure 5.

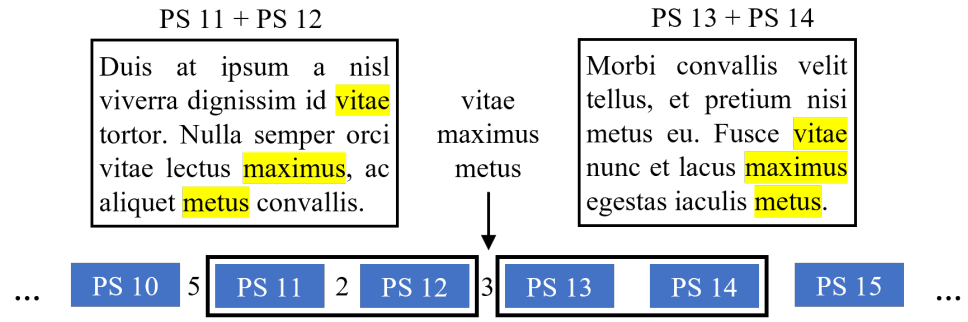


Figure 5: Block comparison method

The blue boxes represent pseudosentences (PS). The numbers between the pseudosentences are gap scores. These are determined using the count of common words in the two adjacent blocks. In this example, $k = 2$ and therefore one block consists of two pseudosentences. These two adjacent blocks move along all pseudosentences and determine the score for each gap.

The actual formula for calculating the gap scores is a normalized inner product as shown in equation 12. The simple approach of just using the count of common words as displayed in figure 5 was just for illustration purposes.

$$score(i) = \frac{\sum_{t=1}^{n} w_{t,b_1} w_{t,b_2}}{\sqrt{\sum_{t=1}^{n} w_{t,b_1}^2 \sum_{t=1}^{n} w_{t,b_2}^2}} \tag{12}$$

The summation index t ranges over all terms that exist across all pseudosentences. The variable $w_{t,b}$ is the frequency of word t in block b. This formula yields a score in the interval $[0, 1]$. A higher score means more lexical similarity between the k pseudosentences to the left and to the right of the gap, and a lower score less. This score can be plotted as done in figure 6.

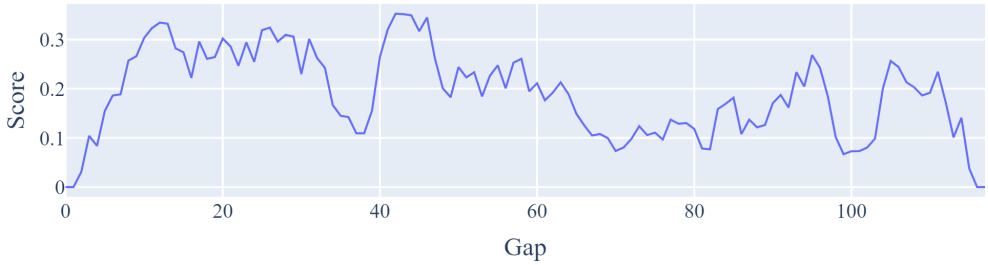


Figure 6: Gap scores of the job interview transcript

The pseudosentence length is set to $w = 20$ and the block size to $k = 10$. Note that for n pseudosentences there are $n - 1$ gaps. At the start of the pseudosentence sequence, the gap score gradually increases and at the end it gradually decreases. This has to do with the block size which is not initiated at $k = 10$ right from the start. The first pseudosentence of the sequence is compared to the second which yields the score for the first gap. The block size at this point is just $k = 1$. Then, the first and second pseudosentences are compared to the third and fourth which yields the score for the second gap. The block size at this point is $k = 2$. This continues until the block size reaches the predefined value of $k = 10$ and can then be regarded as a moving window across the whole pseudosentence sequence as shown in figure 5. At the end of the pseudosentence sequence, the same schema applies, just in reverse order. Therefore, the gap scores at the start and end are smaller as there is not as much vocabulary compared as with the full block size of $k = 10$.

Step 5 Calculate the depth scores and identify the boundaries.

Remark: The depth score is an inversion of the gap score from the previous step. As already discussed, a high gap score corresponds to a high lexical similarity. Contrary, a low score corresponds to a low lexical similarity. Such low scores are valleys in the gap score graph in figure 6. Two major valleys can be seen at around gap 40 and gap 100. This indicates possible subtopic shifts at theses positions. The depth score stands for the depth of a valley in the gap score graph and is calculated as shown in equation 13.

$$depth\ score(i) = (score(l) - score(i)) + (score(r) - score(i)) \tag{13}$$

The difference between the nearest highest peak to the left of the valley ($score(l)$) and the valley itself ($score(i)$) is calculated. Equally, the difference between the nearest highest peak to the right of the valley ($score(r)$) and the valley itself ($score(i)$) is calculated. The sum of these two differences gives the depth score at gap i. Note that the wording nearest highest peak to the left means iteratively going gap by gap to the left of the current gap i. The first time a gap is found, is where the score of the next gap is lower than the score of the current gap i, the iteration stops and the nearest highest peak to the left is found. The same process applies for the nearest highest peak to the right. The depth score is depicted in figure 7.

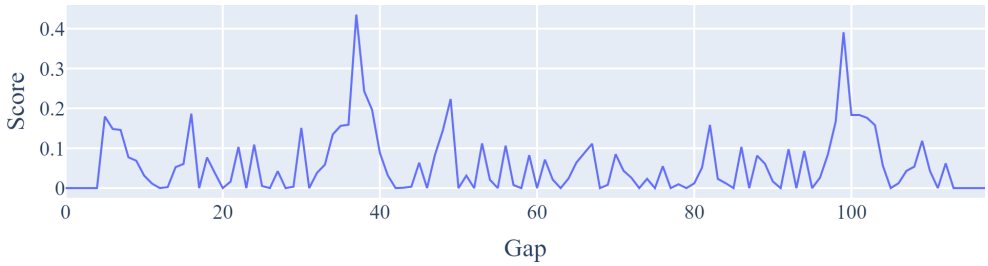


Figure 7: Depth scores of the job interview transcript

The previously seen valleys in the gap score graph correspond to the highest peaks in the depth score graph at about gap 40 and gap 100. For a better understanding of their relationship, figure 8 shows both graphs in the same plot.

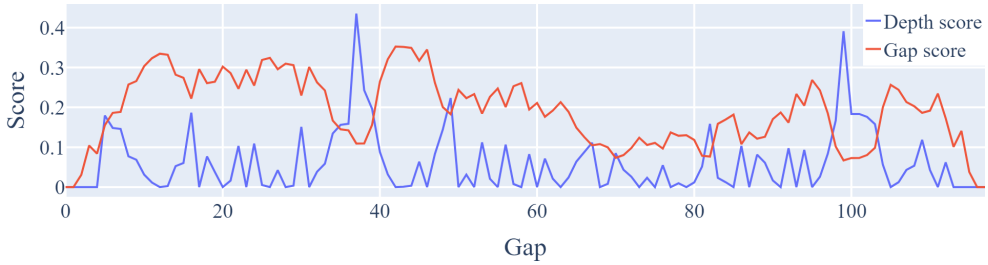


Figure 8: Gap and depth scores of the job interview transcript

Whenever there is a large valley in the gap score graph, this leads to a high peak in the depth score graph. High peaks indicate a subtopic shift in the text and there are multiple approaches how and how many significant peaks are identified. The paper [5] proposes two slightly different cutoff functions to determine which peaks are included. The assumption is that the scores are normally distributed. These are shown in the equation 14 and 15.

$$significant\ peaks > \overline{s} - \sigma \tag{14}$$

$$significant\ peaks > \overline{s} - \sigma/2 \tag{15}$$

The liberal measure (LC) in equation 14 includes all peaks which are greater than the average depth score $\overline{s}$ minus their standard deviation σ. The conservative measure (HC) in equation 15 includes all peaks which are greater than the average depth score $\overline{s}$ minus their standard deviation σ divided by 2. A third approach is to determine a desired number of subtopics n and therefore include the $n-1$ highest peaks. These approaches are analyzed in more detail in section 6.1.2. It is to note that there is a minimum distance which two peaks have to be apart. For example, if one peak at gap 30 is included and another one would be included at gap 45, it is skipped. The minimum distance between two peaks is set to 20 gaps. This means that if for example the six highest peaks are selected, the algorithm might only return four or five if they would be too close together.

The number of peaks returned in any approach mark the boundaries of the text. If for example 5 peaks are returned this translates to 5 boundaries in the text and therefore the text is split up into 6 subtopics.

Step 6 Normalize boundaries.

Remark: The boundaries detected in the previous step are always gaps between pseudosentences. This means, these boundaries do not necessarily translate into a meaningful position in the text. For example, a boundary could be in the middle of an original sentence. Therefore, the boundaries are normalized so that when a boundary is in-between an original sentence, that sentence is added to the subtopic it started in. This results in cleaner boundaries as each subtopics ends with a full original sentence.

In the case of dialog transcripts, the normalization step is narrowed down a bit more. A boundary could not only be in-between an original sentence but also in-between an utterance. This means that a subtopic shift could be detected in the middle of an utterance of a speaker. In this paper, we defined this behavior as not ideal and therefore the boundaries are normalized so that they are not only between sentences but also between utterances. If a boundary is identified in-between an utterance, this utterance is added to the subtopic it started in.

6.1.2 Discussion of Parameters

The parameters pseudosentence length w and block size k as well as three different approaches to determine the significant peaks allow for fine-tuning of the algorithm. To exemplify the effect of the parameters, the vice presidential debate transcript is used as it is sufficiently long to investigate higher parameter values. Also, all plots show the depth score graph. The gap score graph is not investigated. Note that the optimal parameters according to the paper [5] are pseudeosentence length $w = 20$ and block size $k = 10$. However, in the context of dialog transcripts, other parameter settings might prove more optimal.

In a first step, three different values for the parameter pseudosentence length w are set as shown in figure 9. The block size remains at $k = 10$.

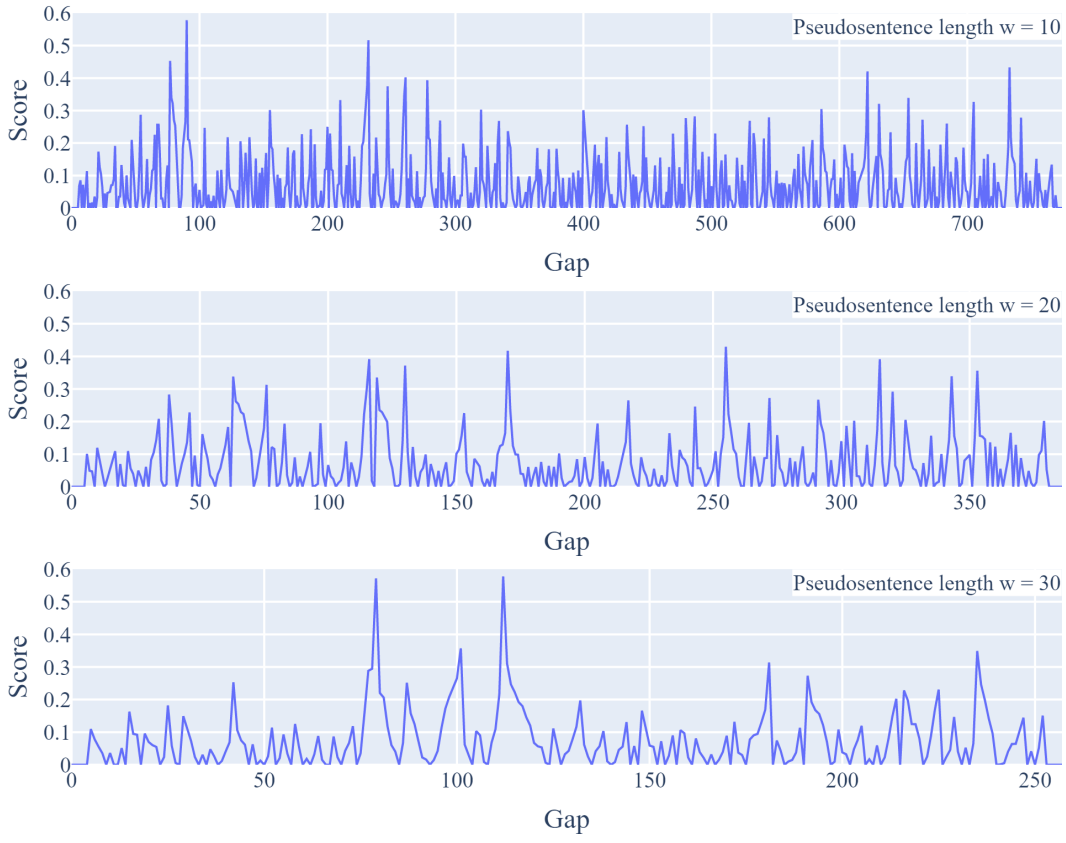


Figure 9: Depth scores of the vice presidential debate transcript with corresponding pseudosentence length

The lower the value of the pseudeosentence length w, the more pseudosentences are formed. Although the values on the x-axis differ, the proportions are still the same. In other words, the first third of all three plots approximately corresponds to the first third of the transcript. The most homogeneous result would be, if the peaks of all three plots would align vertically. However, it is clearly visible that this is not the case. The left peak in $w = 10$ is not visible in $w = 20$ or $w = 30$. The peaks around the first third of the pseudosentence sequence are more or less prominent in all three parameter settings.

In a second step, three different values for the parameter block size k are set as shown in figure 10. The pseudosentence length remains at $w = 20$.

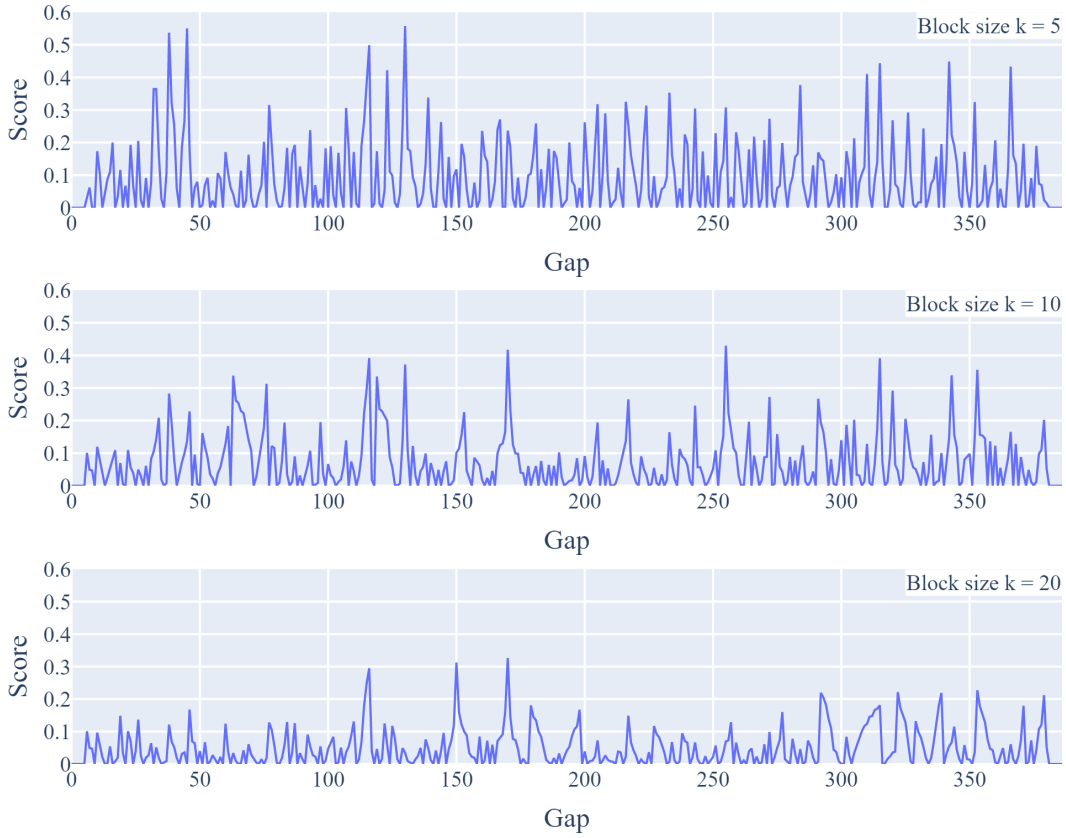


Figure 10: Depth scores of the vice presidential debate transcript with corresponding block size

The higher the block size k, the smaller the overall values of the depth score. Similarly to figure 9, there is also no homogeneous result across all three plots. The peaks around the first third of the pseudosentence sequence are more or less prominent on all three parameter settings.

In a third step one could investigate the effect of different parameter pairs like $w = 10$ and $k = 5$, $w = 30$ and $k = 20$ and so on. The depth score graph would look different for every configuration and just by simply analyzing the graph, one would not find the optimal parameter settings. A more meaningful approach would be to compare the subtopics detected by the algorithm to subtopics marked by an individual. This is further elaborated on in section 6.4.

In a last step, the three approaches to determine the significant peaks are compared in figure 11.

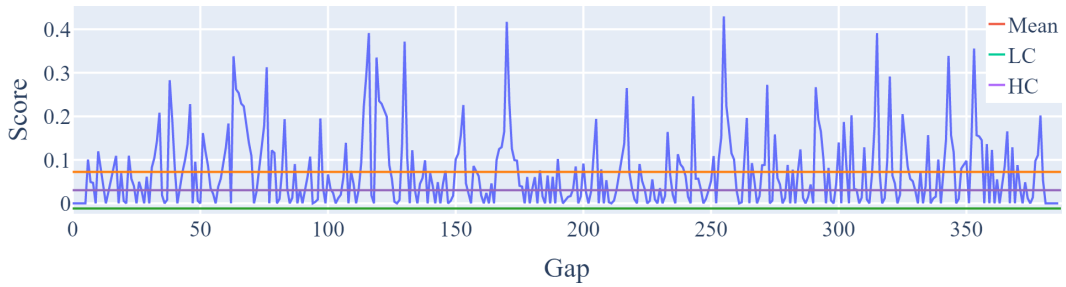


Figure 11: Depth scores of the vice presidential debate transcript with mean, LC and HC

The liberal measure (LC) from equation 14 has a slightly negative threshold and is therefore unfeasible. The conservative measure (HC) from equation 15 has a slightly positive threshold and would therefore include almost all peaks which is also not feasible. This suggests that the depth scores are not normally distributed and these two measures relying on the mean and standard deviation do not perform well. When looking at the distribution of the depth scores in the histogram in figure 12, this assumption confirms.

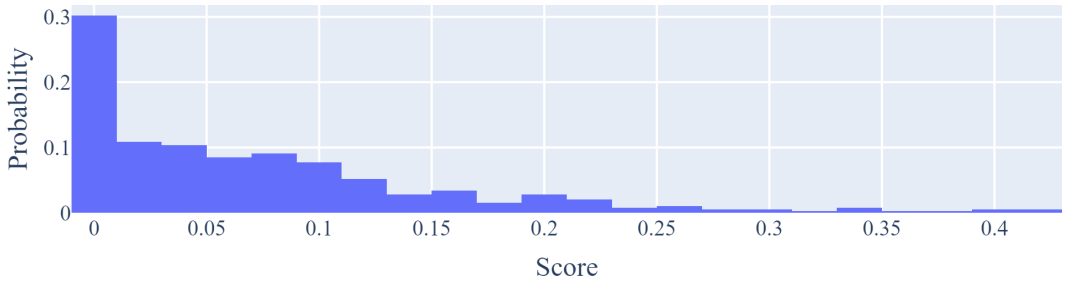


Figure 12: Depth scores histogram of the vice presidential debate transcript

The histogram shows a right skewed distribution of the depth scores. Consequently of the suboptimal performance of the LC and HC, we use the third approach of selecting the n highest peaks in the depth score graph.

6.2 Textsplit

Textsplit is an algorithm developed by Christoph Schock in 2020 [6] which is based on a paper by Alexander Alemi and Paul Ginsparg from 2015 [14]. The algorithm works with vector representations of words and uses similarity measures such as the Euclidean norm to split a text into coherent sections.

The algorithm is described in a way that gives a good intuition on how it works. For a fully mathematical description we refer to the paper [14]. For the following explanation, a sample text of the following five sentences is used.

This is the first sentence of the document. In this text, this is the second phrase. Is this the fourth sentence? The sky is blue and trees are green. Water is blue too but trees can also be red.

Semantically, the above sample text does not have much depth. The intention was to create a sample text consisting of words from two totally different domains. For the first three sentences, the domain is texts and documents and for the last two sentences nature, e.g. sky, trees and water. This example allows to give a good visual understanding of the algorithm in a three dimensional vector space. The most reasonable segmentation of this sample text would be to group the first three sentences in a subtopic and the last two in another.

Step 1 Concatenate all utterances in the transcript into one continuous text block.

Remark: The same reasoning applies as for step 1 in the TextTiling algorithm in section 6.1.1.

Step 2 Split the text block into its sentences.

Step 3 Train a `word2vec` model on a suitable corpus.

Remark: During training, `word2vec` parses a large amount of text referred to as text corpus and evaluates which words frequently co-occur with each other. This leads to a vector representation of each distinct word. The vectors are chosen in a way that a mathematical similarity measure such as the Euclidean distance indicates their semantic similarity. For this paper, the trained English `word2vec` model has 71'291 distinct words and each word is represented as a vector of length 200. This results in a large 71'291 x 200 matrix. The text corpus used to train this model is based on texts from Wikipedia [15]. The trained German `word2vec` model has 64'515 distinct words and also a vector length of 200. The text corpus is also based on texts from Wikipedia [16].

It is impossible to visualize such a high dimension. To give an intuition on how such a multi-dimensional space of words looks, a principal component analysis (PCA) is performed to reduce the 200 dimension to 3. Further, the 71'291 words are reduced to the nouns of the sample text which results in the visualization shown in figure 13.

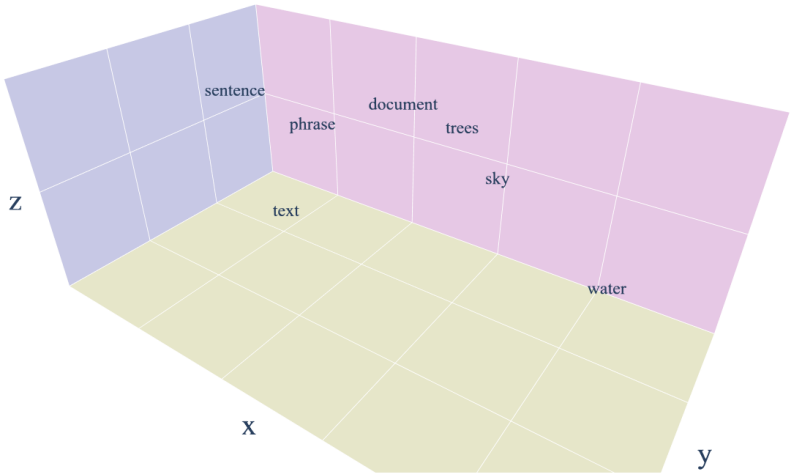


Figure 13: 3-dimensional vector space with unique words

The words *water*, *sky* and *trees* are to the right of the plot and the words *sentence*, *phrase*, *text* and *document* to the left. This gives only a glance at the remarkable performance of `word2vec` models. Recall that the dimension was reduced from 200 to 3 and the 3-dimensional result already provides a solid result.

It is also possible to download a pre-trained model. For example, Google offers a model which was trained on roughly 100 millions words from a Google News dataset [17]. It has a size of 1.5 GB and each vector has a length of 300. From a performance standpoint, working with such an extensive model can be very slow. Therefore, for this paper, the `word2vec` models with a dimension of 71'291 x 200 for English and 64'515 x 200 for German are used.

Step 4 Given all words in the `word2vec` model, count their occurrences in each sentence.

Remark: This results in a count matrix with the dimensions 5 x 71'291, whereas 5 is the number of sentences of the text and 71'291 the total vocabulary. The first sentence in the sample text *This is the first sentence of the document.* contains a total of 8 words and 7 unique words, as *the* occurs twice. Therefore, the first row of this count matrix contains 71'284 zeroes. The words which appear in the sentence, contain their counts in the corresponding column. In this example, this would be six times the value one for the words *this, is, first, sentence, of* and *document* and the value 2 for the word *the*.

Step 5 Multiply the count matrix with the `word2vec` model matrix.

Remark: The dimensions of the count matrix is 5 x 71'291 and the one of the `word2vec` model matrix is 71'291 x 3 when the one with the reduced dimensions is used. The resulting matrix has the dimensions 5 x 3. The word counts of each of the 5 sentences are essentially element-wise multiplied with the corresponding word positions in the `word2vec` vector space and then summed up. The resulting matrix in the case of the sample text is shown in eq. 16.

$$count\ matrix * word2vec\ matrix = \begin{bmatrix} 0.173 & -0.200 & 0.963 \\ 0.129 & -0.364 & 0.727 \\ 0.083 & -0.195 & 0.502 \\ -0.008 & -0.453 & -0.223 \\ 0.039 & -1.408 & -0.027 \end{bmatrix} \tag{16}$$

This can be imagined as every sentence has its own position in the vector space. These positions can again be plotted as shown in figure 14.

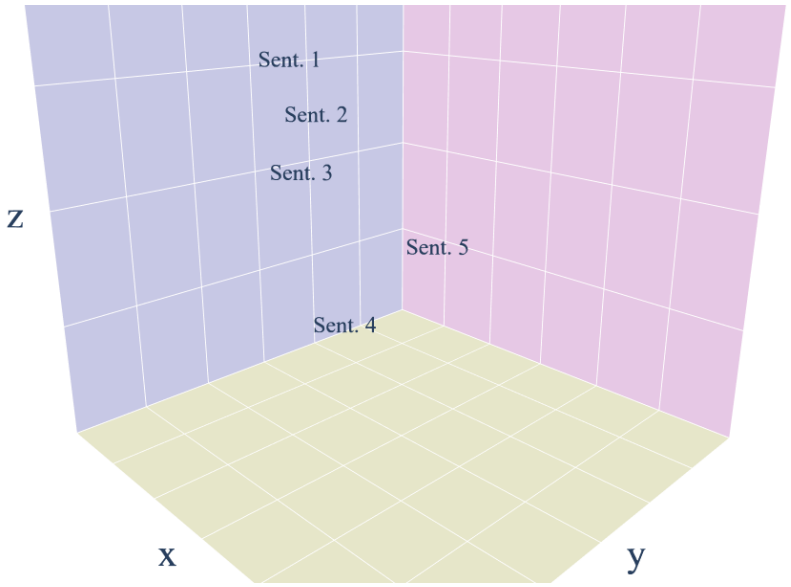


Figure 14: 3-dimensional vector space with sentences

It is clearly visible that sentences 1 to 3 are grouped together and sentences 4 and 5 are somewhat close together. This is again a solid result for a vector space that is reduced from 200 to only 3 dimensions. This also hints the further steps of the algorithm. Graphically, one could already perform a topic segmentation from this visualization. The goal is to this numerically with similarity measures.

Step 6 Add a row of zeroes on top of the resulting matrix from the previous step and form the cumulative sum along all rows.

Remark: This gives the matrix displayed in eq. 17.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0.173 & -0.200 & 0.963 \\ 0.308 & -0.563 & 1.690 \\ 0.386 & -0.758 & 2.192 \\ 0.378 & -1.211 & 1.969 \\ 0.418 & -2.619 & 1.942 \end{bmatrix} \tag{17}$$

This can be visualized in a 3-dimensional vector space again as done in figure 15.

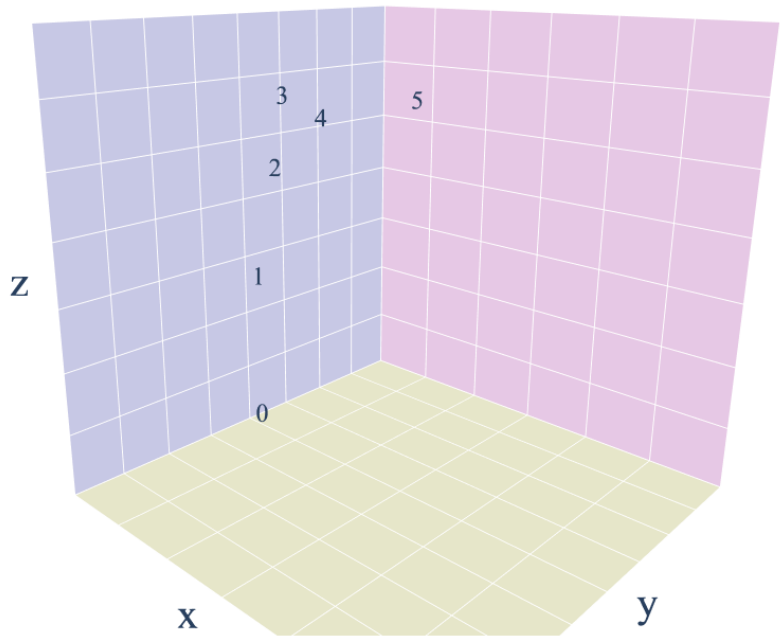


Figure 15: 3-dimensional vector space with cumulated matrix points

The numbers 1 to 5 mark the sentences and 0 refers to the origin. The position of the sentences with their cumulative coordinates show some linearity. Sentences 1 to 3 are more a less on a straight line and 4 and 5 are on a line which is approximately perpendicular to the before-mentioned.

Step 7 Calculate the distance from the origin to the last sentence via every sentence in-between. The first topic split is set after the sentence which achieves the longest distance.

Remark: For a better clarification, the distances are depicted in figure 16

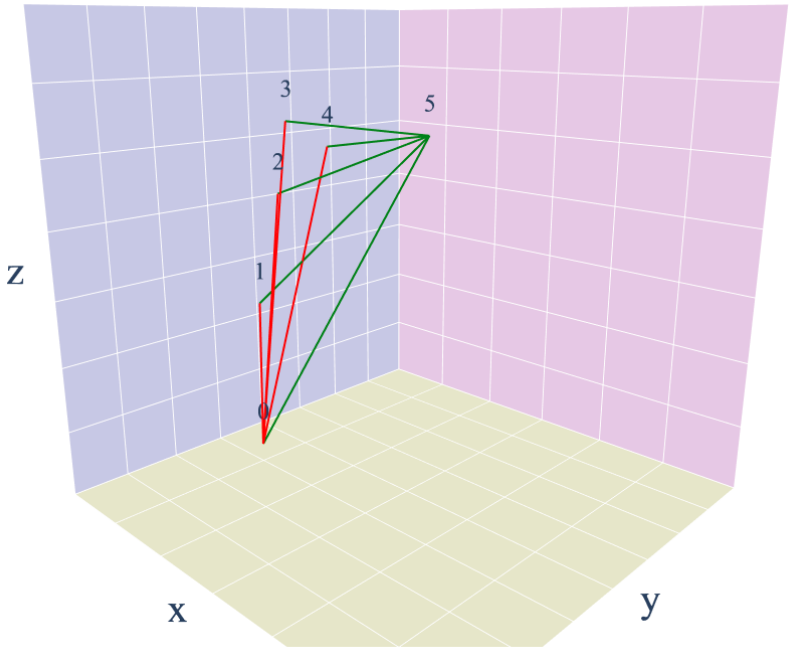


Figure 16: 3-dimensional vector space with cumulated matrix points, iteration 1

The distance from the origin to sentence 5 can be regarded as an amount of coherence for the whole text. The longer this distance is, the more information is captured. The goal when determining the first split is to maximize this distance in choosing the combination of red and green line, which has the longest distance from the origin to sentence 5. In this example, the longest distance results from the origin to sentence 3 plus from sentence 3 to sentence 5. This means that sentence 3 is chosen as the first topic split. This means that sentences 1 to 3 are one segment and sentences 4 and 5 are another segment.

Step 8 Set the split chosen in the previous step as an additional anchor. Now there are two segments. From the origin to the split and from the split to the last sentence. Again, calculate the maximum distance from the origin to the first split via every sentence in-between. Compare this maximum distance to the direct distance from the origin to the first split. Likewise, calculate the maximum distance from the split to the last sentence via every sentence in-between. Compare this maximum distance to the direct distance from the first split to the last sentence. Whichever maximum distance yields the highest gain compared to the direct distance, is taken as the next split.

Remark: To better understand what this means, it is shown in figure 17

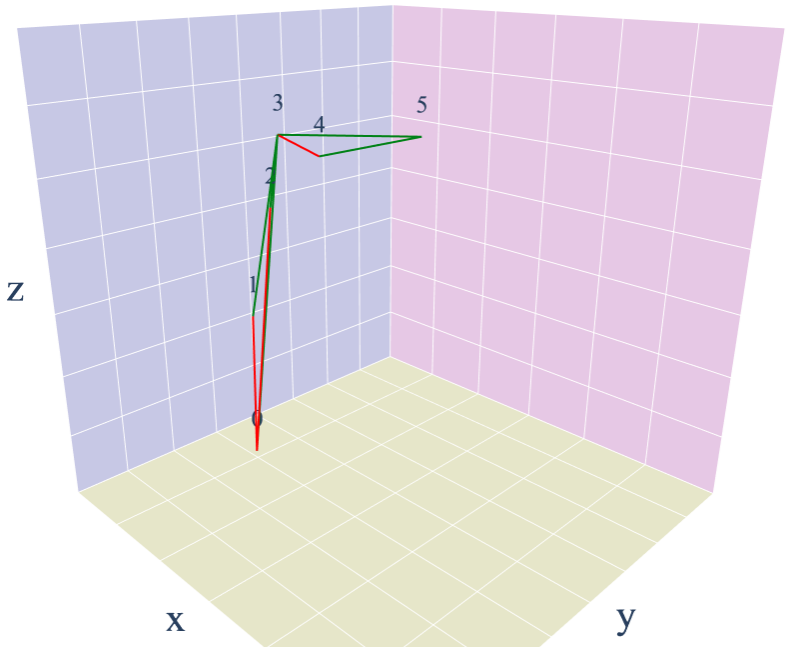


Figure 17: 3-dimensional vector space with cumulated matrix points, iteration 2

As seen, sentence 3 acts as an additional anchor. There is now a segment from the origin to this anchor and from this anchor to sentence 5. Now there are three gains which have to be compared. First, there is the distance from the anchor at sentence 3 via sentence 4 to sentence 5. The gain is how much longer this indirect distance is compared to the direct distance from the anchor to sentence 5. Then, there are two distances from the origin to the anchor via sentences 1 and 2. This is hard to see in the plot, as they are very close together. Again, the gain is how much longer these indirect distances are compared to the direct distance from the origin to the anchor. The plot already hints the result. As the distances via sentences 1 and 2 are almost the same as the direct distance, the gain is very minimal. The gain of the distance via sentence 4 to sentence 5 compared to the direct distance is comparatively large. Therefore, sentence 4 would be selected as the next split.

Step 9 Set the split chosen in the previous step as an additional anchor. Now there are three segments. From the origin to the split whichever comes first. Then, from this first split to the second split. Finally, from the second split to the last sentence. Perform the same procedure of comparing the distance gain of the indirect distances to the direct distances. Continue with this iteration until one of the following criteria meets:
1. The maximum gain is below the predefined threshold.
2. The number of splits reaches the predefined threshold of maximum number of splits.

Remark: This algorithm would theoretically continue until every sentence is its own segment. This is not desired, therefore there are two break conditions which are specified with hyperparameters. Either, the maximum gain of one iteration is below a certain threshold or the predefined maximum number of splits is reached.

Step 10 Normalize the boundaries.

Remark: Contrary to the TextTiling algorithm, the boundaries are already between original sentences. However, as an utterance in the transcript can contain multiple sentences, the boundary may be in-between an utterance. We defined this case as not suitable and therefore the boundaries detected by the textsplit algortihm are normalized so they are between utterances. If a boundary is in-between an utterance, the utterance is added to the subtopic segment where it started in.

6.3 Performance Evaluation

As described in the previous sections, the boundaries returned by both algorithms are normalized to the utterance breaks. This means, splits occur only between two utterances of which each has a timestamp. As a result, each boundary returned by the algorithm matches one of these timestamps. For example, the first split is made after five utterances at timestamp A and the second split is made after 10 utterances at timestamp B. These split timestamps returned by the algorithms are compared to such marked by an individual. A measure is introduced to quantify this difference. Figure 18 provides an illustration of this differences.

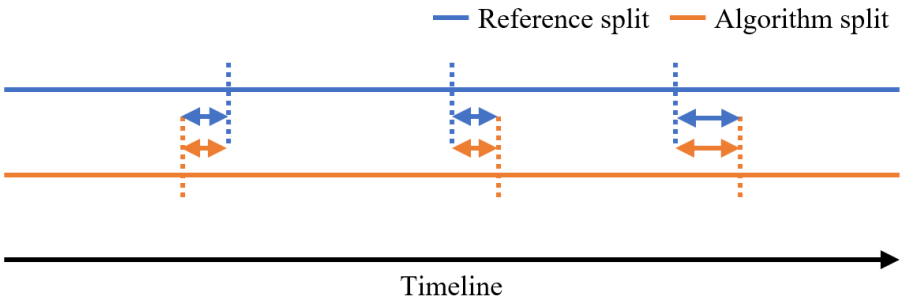


Figure 18: Differences between reference and algorithm splits scenario 1

The dashed blue lines are the reference splits marked by an individual and the dashed orange lines the algorithm splits. The arrows refer to the difference in seconds between the blue and orange splits. A step by step walk-through of the score calculation is shown below.

Step 1 For each blue split, find the closest orange split and calculate the difference in seconds. Accordingly, for each orange split, find the closest blue split and calculate the difference in seconds.

Remark: These differences are already illustrated in 18. It is to note that the number of blue and orange splits does not need to be equal as showcased in figure 19.

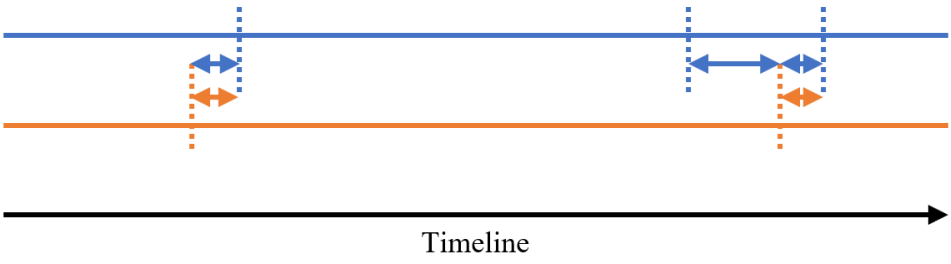


Figure 19: Differences between reference and algorithm splits scenario 2

Step 2 Divide each difference by the length of the transcript in seconds. Then, multiply these scaled values with 100.

Remark: This scales the differences to the interval $[0, 100]$ and therefore does not lead to very small numbers.

Step 3 Divide each scaled difference by $\sqrt{n}$ where n refers to the total number of splits of the corresponding color.

Remark: The reason for this division is explained after the following step. Mathematically, the calculations up to this point can be described as shown in eq. 18.

For each $i \in \{1, \ldots, n\}$ define

$$\begin{aligned} d_{1,i} &= min(\{\mid x_i - y_j \mid * \frac{100}{l * \sqrt{n}} \ : j \in \{1, \ldots, n\}) \\ d_{2,i} &= min(\{\mid y_i - x_j \mid * \frac{100}{l * \sqrt{n}} \ : j \in \{1, \ldots, n\}) \end{aligned} \tag{18}$$

x: list of reference splits in seconds
y: list of algorithm splits in seconds
l: length of the transcript in seconds
n: number of splits of the corresponding list of splits

Step 4 Calculate the median of all differences $d_{1,i}$ and $d_{1,i}$ as shown in eq. 19

$$score = med(d_1, d_2) \tag{19}$$

Remark: To give an example why the division by $\sqrt{n}$ is performed, consider figure 18. Assuming that there is only the left blue and orange split. The median of both of these differences equals the length of the depicted arrow. Now assuming that there are all three splits. The median of all six arrows still equals approximately the length of one of the arrows. Although the second scenario does identify two more splits with a great accuracy, its score is the same as if there is only one split identified. It is defined that the performance of an algorithm should be rated higher if more splits are identified correctly or almost correctly compared to only one split. By dividing through the number of splits, former scenario yields a better score than latter. As dividing by the total number of splits directly might have a too high impact, the square root is introduced to mitigate this effect.

A score of 0 indicates a perfect match of the splits returned by an algorithm and the ones marked by an individual. Contrary, a score of 100 indicates the highest possible difference between both sets of splits.

6.4 Discussion

To evaluate the two discussed algorithms, their results are compared with splits done by an individual. Marking topic shifts is subjective, therefore every individual would define slightly different splits. In this evaluation, splits are defined according to our opinion referred to as reference split and displayed in the following figures with the label *Human*.

Figure 20 shows the vice presidential debate transcript with the splits by an individual and the TextTiling and Textsplit splits. Note that for all transcripts the parameters used for TextTiling are pseudeosentence length $w = 20$ and block size $k = 10$.

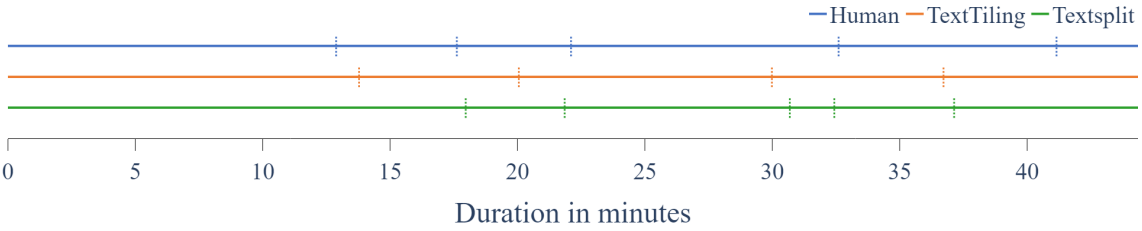


Figure 20: Topic segmentation of the vice presidential debate transcript

It can be seen that an individual made five splits which is also given as input parameter for the number of subtopics to the TextTiling algorithm. However, due to the structure of the algorithm the input parameter and the final number of splits can differ if two splits would be too close together. The Textsplit algorithm on the other hand outputs the same number of subtopics as given as input. In general, the splits do not differ much which can be evaluated more precisely with eq. 18 and 19. The corresponding scores are shown in table 10. The figures 21, 22, 23 and 24 show the splits of the four BA meeting transcripts.

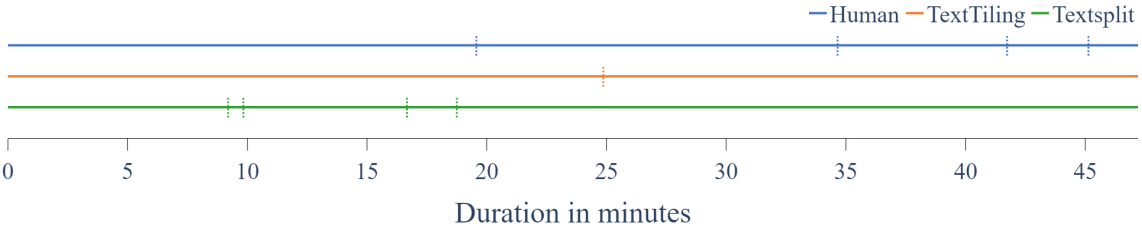


Figure 21: Topic segmentation of the BA meeting 2021-03-11 transcript

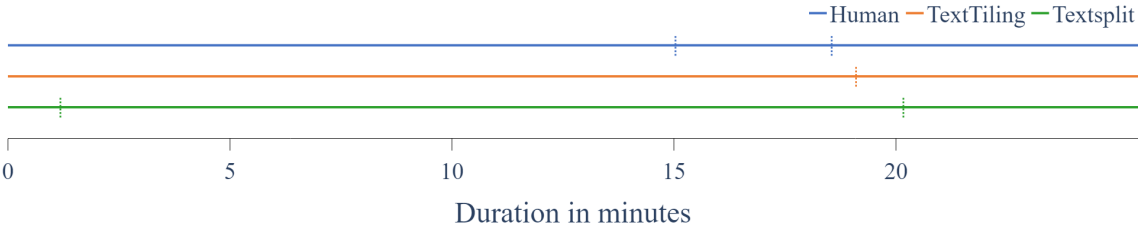


Figure 22: Topic segmentation of the BA meeting 2021-03-25 transcript

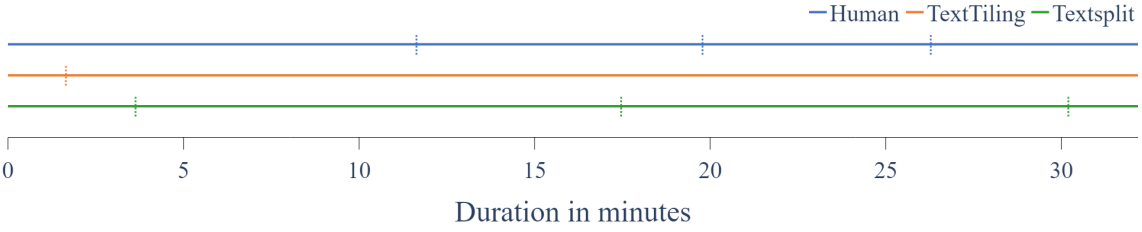


Figure 23: Topic segmentation of the BA meeting 2021-04-23 transcript

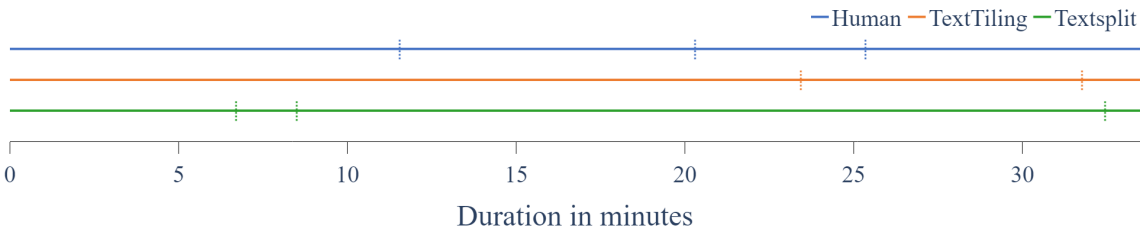


Figure 24: Topic segmentation of the BA meeting 2021-05-07 transcript

It is clearly visible that the amount of topics from the TextTiling algorithm much differ compared to the ones an individual made. As well the area where the split is made is also quite different. This can be explained by the rather poor structure quality of the texts, as mentioned in chapter 3. The corresponding scores are shown in table 10.

Table 10: Different texts compared with their accuracy scores

Text	**Individual–TextTiling Score**	**Individual–Textsplit Score**
Vice presidential debate	2.45	0.35
BA meeting 2021-03-11	11.23	10.65
BA meeting 2021-03-25	2.16	9.38
BA meeting 2021-04-23	31.78	7.03
BA meeting 2021-05-07	5.40	10.30

With the formula explained in chapter 6.3, the topic splits made by an individual and each algorithm can be calculated. It can be seen that the vice presidential debate has a much lower score compared to the four BA meetings. Especially the individual to Textsplit score with 0.35 is close to zero indicating a very good quality of the splits. The scores of the BA meetings vary greatly between 2.16 and 31.78.

Now the keyword extraction methods from chapter 5 can be combined with the topic segmentation algorithms. Table 12 shows the extracted TF–IDF and YAKE! keywords for each of three subtopics determined by the TextTiling algorithm.

Table 11: Topic segmentation and keywords with the TextTiling algorithm

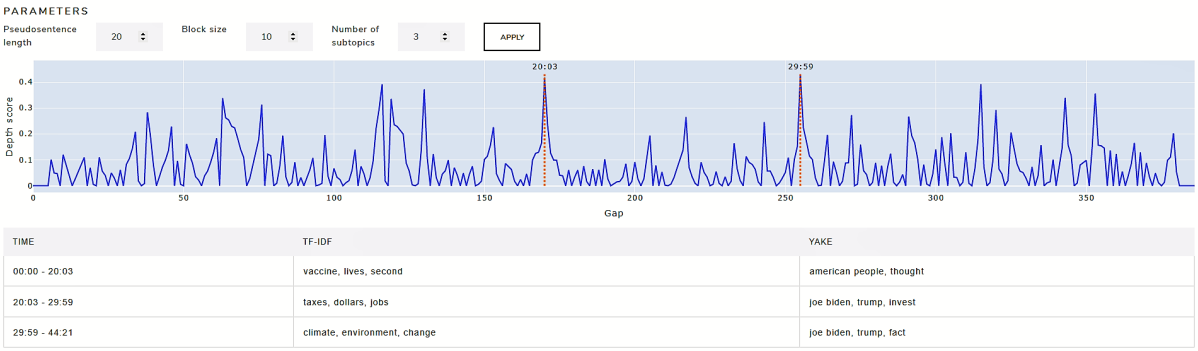
Time stamp for topic split	**TF–IDF keywords**	**YAKE! keywords**
00:00 - 20:03	vaccine, lives, second	american people, thought
20:03 - 29:59	taxes, dollars, jobs	joe biden, trump, invest
29:59 - 44:21	climate, environment, change	joe biden, trump, fact

The keywords obtained by TF-IDF are of high relevance. One sees that the topics in the first 20 minutes revolves around coronavirus and vaccination. The next ten minutes focus on taxes and jobs. The last subtopic is environment and climate change. When reading through the transcript manually, the three subtopic splits chosen by the algorithm are suitable. The keywords obtained by YAKE! focus mostly on names of people which does not provide the same level of relevance as the keywords by TF–IDF. Table 12 shows the same for the Textsplit algorithm.

Table 12: Topic segmentation and keywords with the Textsplit algorithm

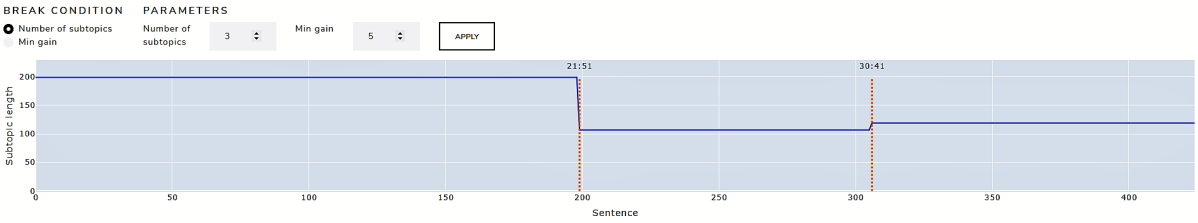
Time stamp for topic split	TF–IDF keywords	YAKE! keywords
00:00 - 21:51	vaccine, reality, lives	american people, thought
21:51 - 30:41	taxes, dollars, tax	joe biden, trump, invest
30:41 - 44:21	climate, environment, change	joe biden, trump, war

The splits determined by the Textsplit algorithm match the ones of TextTiling quite well and the keywords are almost identical. Topic segmentation is also implemented in the web application as shown in figure 25 for TextTiling and figure 26 for Textsplit. One can select a transcript, adjust the parameters and return the topic splits with their corresponding timestamps marked in a plot and table. The keyword extraction settings described in chapter 5 can also be applied.



TIME	TF-IDF	YAKE
00:00 - 20:03	vaccine, lives, second	american people, thought
20:03 - 29:59	taxes, dollars, jobs	joe biden, trump, invest
29:59 - 44:21	climate, environment, change	joe biden, trump, fact

Figure 25: Screenshot of TextTiling output with the marked splits in the web application



TIME	TF-IDF	YAKE
00:00 - 21:51	vaccine, reality, lives	american people, thought
21:51 - 30:41	taxes, dollars, tax	joe biden, trump, invest
30:41 - 44:21	climate, environment, change	joe biden, trump, war

Figure 26: Screenshot of Textsplit output with the marked splits in the web application

7 Visualization

This section discusses different visualization strategies to display keywords. The aim is to create comprehensible and user-friendly graphics.

7.1 Wordcloud

A wordcloud is a collection of words which have a high absolute frequency in a specific text. The higher the frequency, the bigger the word is displayed in the cloud. In this thesis, wordclouds are considered to illustrate the discussed topics in a transcript based on the most used words. Concretely, a transcript is divided into blocks of equal length and for each block a wordcloud is created. These wordclouds should give an idea of how the topic shifts over time in a transcript. In the following chapters different wordcloud approaches are considered to find the most suitable and user-friendly illustration.

7.1.1 Wordclouds with Dedicated Library

The library `wordcloud` is considered to evaluate the feasibility of this approach. In figure 27 the wordcloud from a specific block within the vice presidential debate is presented.

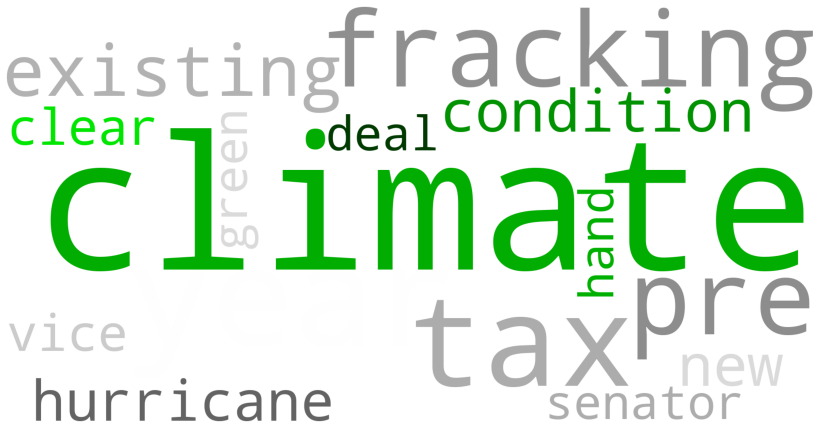


Figure 27: Wordcloud created with library `wordcloud`

Green colored are the words which are new in the current block compared to the previous block. Grey colored are those which already appeared in the previous one. Hence, the topic transitions from block to block become visible.

7.1.2 Wordcloud from Scratch

For a more sophisticated approach, the wordcloud is built from scratch. This allows to tailor it better to the need of being able to navigate through the wordclouds of each text block. The size of the words still depends on their frequency as in figure 27. The position of a word is randomly chosen. Additionally, a parameter is implemented to give the user the possibility to vary the text block length in minutes in which the transcript is subdivided. It is important to note that the transcript is only divided at utterance breaks. To navigate through the different text blocks a slider is implemented. This wordcloud from scratch approach is displayed in figure 28.

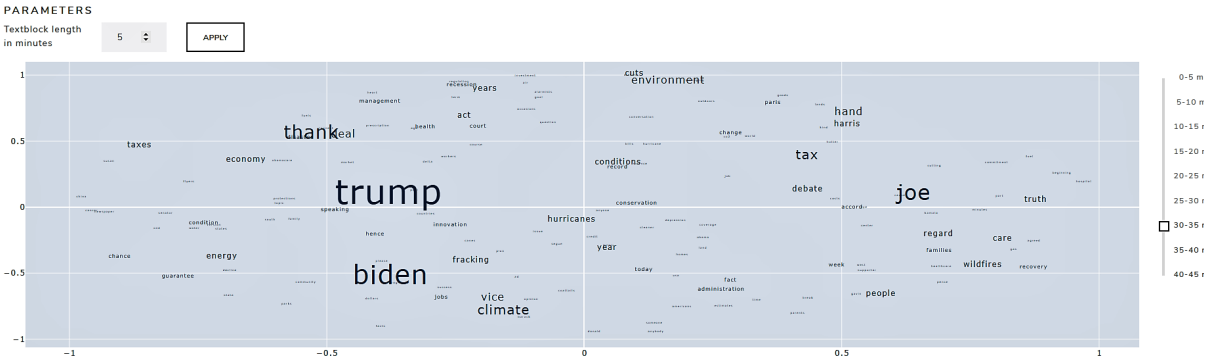


Figure 28: Wordcloud with randomly distributed words and text block slider

7.1.3 Wordcloud Animation

In this section, an animation is considered to automate the text block slider and provide the user a smooth change of images and information. To overcome the chaotic word representation, they are arranged in a tabular form. Also, only the 15 most used words per block are depicted. With a circle, laying in the background of each word, the frequency is displayed. Thus, the smooth transition is ensured. The animation has an automatic slider that cycles through the different blocks and adjusts the circle size accordingly to its frequency. The user is able to pause and change the slider and as before, set the length of the blocks. Through the animation new words of the next block show up at the right side of the wordcloud. Words which already appeared in the previous block among the 15 most mentioned, do not get a new position. This animation is displayed in figure 29.

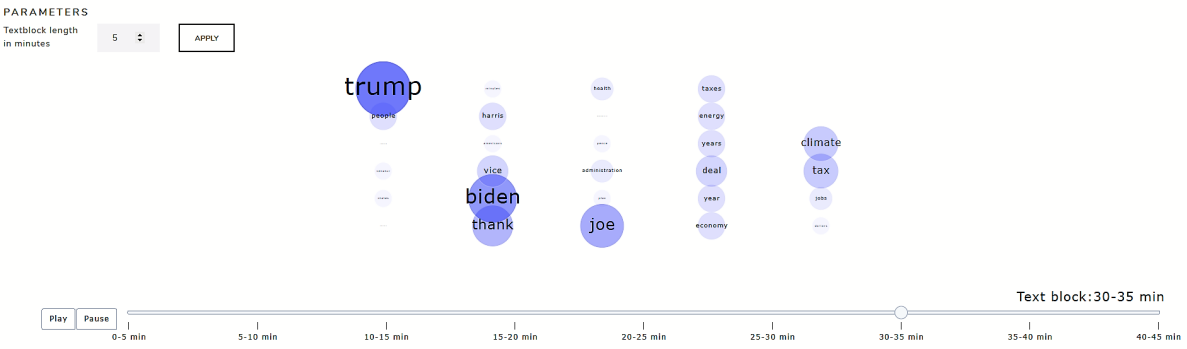


Figure 29: Excerpt of the animation

7.2 Wordcloud with TF–IDF

The TF–IDF algorithm, as described in chapter 5.1 is implemented in the wordcloud animation of chapter 7.1.3 as shown in figure 30. Besides the text block length parameter, the user is also able to adjust the amount of displayed keywords. With the n-highest score setting, one can select up to which score a word is displayed. For example if it is set to 5, all keywords until the fifth highest score are illustrated. The reason why this is not done directly in number of words, is that the keywords can have the same score and therefore be on the same rank. Consequently, there can be more than 5 words in the graph if the previously mentioned value is set.

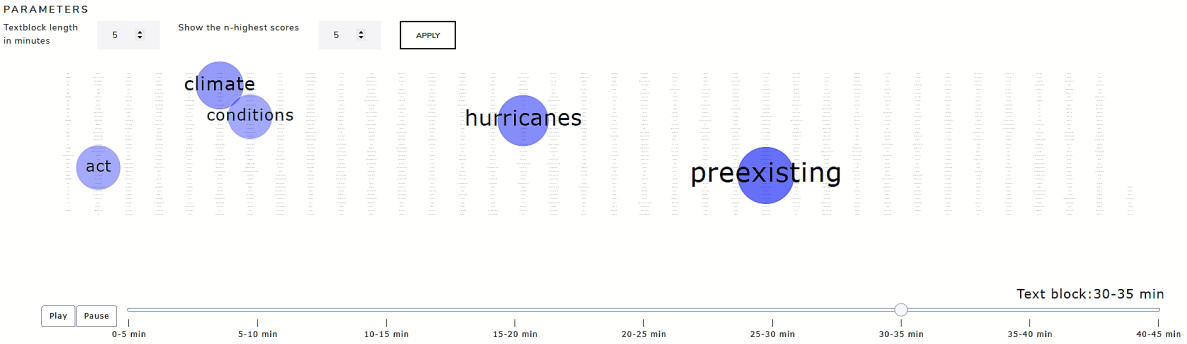


Figure 30: Wordcloud animation based on TF–IDF scores

As the TF–IDF consists of two parts, another approach is to involve the x- and y-axis as shown in figure 31. Thus, the position of the word also has an informative aspect compared to the previous wordcloud graphics.

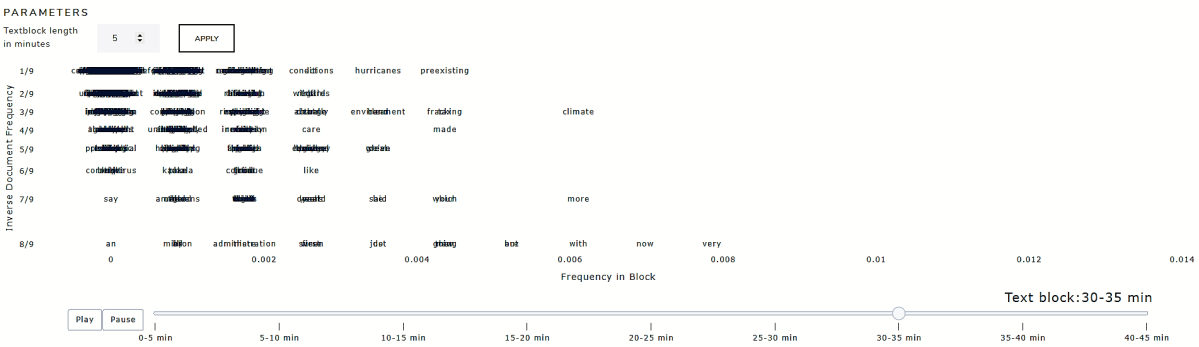


Figure 31: 2 dimensional animation of the TF and IDF scores

The x-axis illustrates the TF score of a word. Meaning, the absolute frequency of a word in the selected block. The y-axis shows the inverse document frequency of the word. It is displayed as a fraction of occurrence in different blocks or paragraphs. At the top of the y-axis, the word occurs in only one block, which gives it the highest IDF score. This means that the term with the highest TF–IDF score is located in the upper right corner.

7.3 Discussion

The library `wordcloud` creates a wordcloud in a minimalist and compact way. However, it has limited space for own implementations. Furthermore, one graphic per text block is created which leads to too many individual images. The wordcloud from scratch stands out because it gives the ability to implement own ideas more specifically. As in the library `wordcloud`, the position of the words is chosen randomly at every execution. This makes it hard to keep track of a single word which occurs in multiple text blocks. In one block it might be positioned in one corner and in another block in a different corner. Therefore, fixed positions in the graph results in a more pleasant overview. The animated approach additionally gives a smooth transition over all different text block wordclouds. One can follow the change of keywords easily. The graphic is eye catching and simple to read. It is very user-friendly and therefore a considerable option to implement. A negative aspect is that it depends on absolute frequency of words. An assumption of the topic can be made, however the most discussed words do not necessarily have to lead to the topic of a text block.

Therefore, the variation with TF–IDF is much more promising in this regard. The parameter n-highest words allows for reducing the number of keywords resulting in a mostly uncluttered overview. The second approach evaluates the use of the x- and y-axis as additional information. The negative aspect is that multiple keywords have equal scores, which leads to overlapping items. This is not a very picturesque way to give information to the user. Therefore, it should be reconsidered, if only the top right corner with the important words is depicted or the pile of words is being divided.

In general, visualizing the keywords over time proves an intuitive way to present an overview of the subtopic shifts in a transcript. It is very simple to extract the important information, which is essential if a quick summary of the transcript is desired. However, the results depend on the keyword extraction method used. Here, TF–IDF proves to be a suitable solution.

8 Conclusion and Outlook

The detailed analysis of the applied algorithms proved to be valuable in interpreting their output. TF–IDF is a simple yet powerful approach to extract keywords. The drawback is that a collection of documents or paragraphs needs to be defined. Therefore, extracting keywords from a whole transcript is not possible. However, if the transcript is split into multiple sections, TF–IDF returns representative keywords. RAKE turned out to be unsuitable because it is not robust to bilingual transcripts. The text is split by stopwords which are defined by a stopword list. If the dialog switches to another language, those stopwords are not recognized. This leads to very long and not meaningful keywords. YAKE! involves multiple features to calculate the score of keywords. Particularly interesting is the involvement of capitalized letters. However, in transcripts where names of speakers are mentioned many times it might return their names as keywords. If one already knows the speakers, these keywords do not represent valuable information. To overcome this problem, the stopword list can be manually extended by the speaker's names. KeyBERT leverages the very powerful BERT model with the drawback of a very slow computing speed. This makes it unsuitable for our web application. A clear benefit of the keywords obtained by KeyBERT could not be identified. The performance of the topic segmentation algorithms TextTiling and Textsplit depends strongly on the structure quality of a transcript. Considering the German BA meeting transcripts, their lack of clear structure, incomplete sentences and plentiful use of English words led to poor results. On the other hand, the clearly structured vice presidential debate led to useful results and a good score in the performance evaluation. The created web application allowed us to get a feeling for potential user needs. The algorithms were implemented in an intuitive way with easy adjustable parameters. The implementation of various visualizations provided a brief overview of the spoken transcript. An animation proves to be a comprehensible solution for a topic transition analysis as far as expressive keywords are chosen by the algorithm.

The findings in this thesis allow to give an outlook on potential future work. In order to achieve better results in the future, attention should be paid to the text quality which has a significant influence on the results. KeyBERT is not considered further in this paper due to its long execution time. However, as it is based on the powerful BERT model it is worth investigating it more deeply. Important to note is that the performance of Texsplit depends on the text corpus it is trained on. In this thesis, smaller and more handy text corpora are used. It is highly likely that when using larger text corpora from Google or Facebook, a better topic segmentation is obtained. One idea could also be to train the model on domain-dependent corpora to address the topic segmentation of specific transcripts such as work meetings. In terms of visualizations, user studies could be conducted to better tailor an animation. Generally, the field of NLP is advanced at a high pace and in the future, better performing algorithms may be developed.

References

[1] K. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, pp. 11–21, 12 1972.

[2] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents," *Text Mining: Applications and Theory*, pp. 1–20, 03 2010.

[3] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, "Yake! keyword extraction from single documents using multiple local features," *Information Sciences*, vol. 509, pp. 257–289, 01 2020.

[4] M. Grootendorst, "Keybert: Minimal keyword extraction with bert," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4461265

[5] M. Hearst, "Texttiling: Segmenting text into multi-paragraph subtopic passages," *Computational Linguistics*, vol. 23, pp. 33–64, 03 1997.

[6] C. Schock, "textsplit," 05 2020. [Online]. Available: https://github.com/chschock/textsplit

[7] D. Golomingi and L. Rüegger, "Analysing transcriptions - an overview over automated transcribed interviews and their statistics," 03 2020.

[8] L. Avci, "Automated text summarization for dialogues with transformer models," 01 2021.

[9] Heads or Tails, "Us election 2020 - presidential debates," *Kaggle*, 2020. [Online]. Available: https://www.kaggle.com/headsortails/us-election-2020-presidential-debates

[10] K. Johnson, "Job interview good example copy," *YouTube*, 11 2016. [Online]. Available: https://www.youtube.com/watch?v=OVAMb6Kui6A

[11] A. Zurcher, "Biden-putin summit: Awkward conversation looms in geneva," *BBC*, 05 2021. [Online]. Available: https://www.bbc.com/news/world-us-canada-57244860

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 10 2018.

[13] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," pp. 3973–3983, 01 2019.

[14] A. Alemi and P. Ginsparg, "Text segmentation based on semantic word embeddings," 03 2015.

[15] M. Mahoney, "text8 corpus." [Online]. Available: http://mattmahoney.net/dc/textdata.html

[16] Wortschatz Universität Leipzig, "Wikipedia 2016 300k corpus." [Online]. Available: https://wortschatz.uni-leipzig.de/en/download/German

[17] Google, "Googlenews-vectors-negative300.bin.gz." [Online]. Available: https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit

List of Figures

List of Tables

A Appendix

A.1 Web application

A.1.1 Python version and libraries

```
Python 3.7.3
```

```
dash==1.19.0
dash-bootstrap-components==0.12.0
dash-core-components==1.15.0
dash-html-components==1.1.2
dash-table==4.11.2
DateTime==4.3
Flask==1.1.2
keybert==0.2.0
nltk==3.5
numpy==1.20.1
pandas==1.2.3
plotly==4.14.3
rake-nltk==1.0.4
reportlab==3.5.67
scikit-learn==0.24.1
textsplit==0.5
word2vec==0.11.1
yake==0.4.8
```

A.1.2 How to start the application

Step 1 Create a virtual environment with the Python version and libraries listed in A.1.1. For guidance check the official Python documentation at `https://docs.python.org/3/tutorial/venv.html`.

Step 2 Download the code at `https://github.com/pascalaigner/summarization.git`.

Step 3 Open the terminal and navigate to the main directory of the repository with `cd .../summarization`.

Step 4 Navigate to the directory of the `app.py` file with `cd Dash`.

Step 5 Start the app with `python app.py`. The terminal output should look like this:

```
Dash is running on http://127.0.0.1:8050

Serving Flask app "app" (lazy loading)
Environment:  production
WARNING: This is a development server.
Do not use it in a production deployment.
Use a production WSGI server instead.
Debug mode:  on
```

Step 6 Open the link `http://127.0.0.1:8050` to access the web application.

A.2 Transcripts

The six transcripts used in this thesis are listed below.

A.2.1 Vice presidential debate

Speaker	Start time	Utterance
Speaker 1	00:00	I'm Susan page of USA Today. It is my honor to moderate this debate an important part of our democracy in Kingsbury Hall tonight. We have a small and socially distant audience and we've taken extra precautions during this pandemic among other things. Everyone in the audience is required to wear a face mask and the candidates will be seated 12 feet apart. The audience is enthusiastic about their candidates, but they've agreed to express that enthusiasm only twice at the end of the debate. And now when I introduce the candidates, please welcome, California, Senator Kamala Harris and vice president Mike Pence.
Speaker 2	00:58	Thank.

Speaker 1	00:58	You. Senator Harris and vice president Pence. Thank you for being here. We're meeting as President Trump and the first lady continue to undergo treatment in Washington after testing positive for covid-19. We send our thoughts and prayers to them for their rapid and complete recovery and for the recovery of everyone afflicted by the Coronavirus. The two campaigns in the commission on presidential debates have agreed to the ground rules for tonight. I'm here to enforce them on behalf of the millions of Americans who are watching One Note no one in either campaign or at the commission or anywhere else has been told in advance what topics all raise our what questions I'll ask this 90-minute debate will be divided into nine segments of about ten minutes each. I'll begin a segment by posing a question to each of you. Sometimes the same question. Sometimes a different question on the same topic you will then have two minutes to answer without interruption by me or the other candidate then we'll take six minutes or so to discuss the issue at that point. Although there will always be more to say we'll move on to the next topic. We want a debate that is Lively. But Americans also deserve a discussion that is civil. These are tumultuous times, but we can and will have a respectful exchange about the big issues facing our nation. Let's begin with the ongoing pandemic that has cost our country so much. Senator Harris, the coronavirus is not under control over the past week Johns Hopkins reports that 39 states have had more covid cases over the past seven days than in the week before nine states have set new records, even if a vaccine is released soon. The next Administration will face hard choices. What would a Biden Administration due in January and February that a trump Administration wouldn't do would you impose new lockdowns for? Businesses in schools and hot spots a federal mandate to wear Mass. You have two minutes to respond without interruption.
Speaker 2	03:18	Thank you Susan. Well the American people have witnessed. What is the greatest failure of any presidential Administration in history of our country? And here are the facts. 210,000 dead people in our country and just the last several months. Over 7 million people who have contracted this disease one in five businesses closed. We're looking at Frontline workers who have been treated like sacrificial workers. We are looking at over 30 million people. So in the last several months had to file for unemployment and here's the thing on January 28th. The vice president and the president were informed about the nature of this pandemic. They were informed that it's lethal and consequence that it is Airborne that it will affect young people. And that it would be contracted because it is Airborne. And they knew what was happening and they didn't.
Speaker 3	04:25	Tell you.

Speaker 2	04:26	Can you imagine if you knew on January 28th as opposed to March 13th, but they knew what you might have done to prepare they knew and they covered it up. The president said it was a hoax. They minimize the seriousness of it. The president said you're on one side of his Ledger if you wear a mask you're on the other side of his Ledger if you don't And in spite of all of that today, they still don't have a plan. They still don't have a plan. Will Joe Biden does and our plan is about what we need to do around a national strategy for contact tracing for testing for administration of the vaccine and making sure that it will be free for all that is the plan that Joe Biden has and that I have knowing that we have to get ahold of what has been going on and we need to save our country and Joe Biden. Is the best leader to do that and frankly this Administration has forfeited. Thank you Ser right to re-election based on this.
Speaker 1	05:25	Thank you, Senator Harris vice president Pence more than 210,000 Americans have died of covid-19 since February the u.s. Death toll as a percentage of our population is higher than that of almost every other wealthy Nation on Earth. For instance. Our death rate is two and a half times that of Canada next door. You had the administration's coronavirus task force. Why is the US death toll as a percentage of our population higher than that of almost every other wealthy country and you have two minutes to respond without interruption.

Speaker 3	0.25	This isn't thank you. And I want to thank the commission and the University of Utah for hosting this event and Senator Harris. It's privileged to be on the stage with you. In our nation has gone through a very challenging time this year. But I want the American people to know that from the very first day President. Donald Trump has put the health of America first. Before there were more than five cases in the United States all people who had returned from China President. Donald Trump did what no other American president had ever done. That was he suspended all travel from China the second largest economy in the world. Now Senator Joe Biden Biden opposed that decision. He said it was xenophobic and hysterical, but I can tell you having led the White House coronavirus Task Force at that decision. Alone by President Trump bought us invaluable time to stand up the greatest National mobilization since World War Two and I believe it's saved hundreds of thousands of American lives because with that time we were able to reinvent testing more than a hundred fifteen million tests have been done to date we were able to see to the delivery of billions of supplies. So our doctors and nurses had the resources support they needed and we began really before the month of February was Art to develop a vaccine and to develop medicines. Ins and Therapeutics have been saving lives all along the way and under President Trump's leadership operation warp-speed We Believe will have literally tens of millions of doses of vaccine before the end of this year. The reality is when you look at the Biden plan, it reads an awful lot like what President Trump and I and our task force have been doing every step of the way and quite frankly when I look at their plan that talks about advancing testing creating new PPE developing a vaccine. It looks a little bit like plagiarism which is something Joe Biden knows a little bit about I think the American people know that this is a president who has put Bank of America first in the American people. I believe with my heart can be proud of the sacrifices. They have made it saved countless American lives.
Speaker 1	08:13	Senator Harris. Would you like to respond?
Speaker 2	08:15	Absolutely, whatever the vice president is claiming. The Administration has done clearly. It hasn't worked. When you're looking at over 210,000 dead bodies in our country American lives that have been lost families that are grieving that loss and you know, the vice president is the head of the task force. And new on January 28th how serious this was and then big thanks to Bob Woodward. We learned that they knew about it. And then when that was exposed the vice president said when I asked well, why didn't y'all tell anybody he said because the president wanted people to remain calm. Well, let's go I have enough but Susan like this is important and I want to add mr. Vice president speaking.
Speaker 3	09:02	I have I'm speaking.

Speaker 1	09:04	Yep, you think more seconds and then we'll give the vice president. So.
Speaker 2	09:07	I want to ask the American people. How come were you when you were panicked about where you're going to get your next roll of toilet paper? How come were you when your kids were sent home from school? And you didn't know when they could go back. How come and do you.
Speaker 1	09:20	Think when your.
Speaker 2	09:20	Children couldn't see your parents? Because you were afraid they could kill them.
Speaker 1	09:24	Miss give vice president Pence a chance to respond vice president. You have one minute.
Speaker 3	09:28	To respond. Oh, there's not a day gone by that. I haven't thought of every American family that has lost a loved one. And I want all of you to know that you'll always be in our hearts and in our prayers. But when you say with the American people have done over these last eight months hasn't worked. That's a great disservice. The sacrifices the American people have made the reality if I'm if I may finish that the reality is, dr. Fauci said everything that he told the president in the Oval Office the president told the American people now President Trump, I will tell you has boundless confidence in the American people and he always spoke with confidence that we'd get through this together when you say it has it worked with dr. Fauci and dr. Burks in our medical experts came to us in the second week of March. They said if the president didn't take the unprecedented step of shutting down roughly half of the American economy that we could lose 2.2 million Americans. That's the reality. Thank you. They also said to us if we did everything right Susan we could still lose more than 200,000 Americans.
Speaker 1	10:32	Vice president One.
Speaker 3	10:33	Life lost is many Susan but the American people I believed it. Of credit for the sacrifices that they have made putting the health of their family and their neighbors first our doctors our nurses our first thank you vice presidents and I'm going to speak up on behalf of what the American people have done.
Speaker 1	10:49	Vice president. Since you were in the front row in a Rose Garden event 11 days ago, what seems to have been a super spreader event for senior Administration and Congressional officials. No social distancing few mask and now a cluster of coronavirus cases among those who were there. How can you expect Americans to follow the administration safety guidelines to protect themselves from covid when you at the White House have not been doing so?

Speaker 3	11:16	Well, the American people have demonstrated over the last 8 months. They've been given the facts they're willing to put the health of their families and their neighbors and people they don't even know first President Trump and I have great confidence in the American people and and their ability to take that information and put it into practice in the height of the epidemic when we were losing a heartbreaking number of 2,500 Americans had a we surged resources to New Jersey and New York and New Orleans and Detroit. We told the American people would To be done in the American people made the sacrifices when the outbreak in the Sunbelt happened this summer again American step forward, but the reality is the work of the president of the United States goes on vacancy on the Supreme Court of the United States has come upon us and the president introduced Judge Amy.
Speaker 1	12:09	Coneybear. Thank you. Thank you very much.
Speaker 3	12:12	If I may say that Rose Garden event, then great deal of speculation about it. My wife Karen and I were there an honor to be there many of the people who were at that events. Susan actually were tested for Coronavirus and it was an outdoor event which all of our scientists regularly and routinely advised the difference here is President Trump and I trust the American people to make choices in the best interest of their health, Joe Biden and Kamala Harris consistently talk about mandates and not not just mandates with the coronavirus but a government takeover of Health Care. Thank you Green New Deal all government control. We're about freedom and Respecting the freedom of the American people.
Speaker 2	12:53	Let's talk about respecting the American people. You respect the American people when you tell them the truth, you respect the American people when you have the courage which we be a leader speaking of those things that you may not want people to hear but they need to hear so they can protect themselves but this Administration stood on information that if you had as a parent if you had as a worker knowing you didn't have enough money saved up and now you're standing in a food line. Because of the ineptitude of a Administration that was unwilling to speak the truth to the American people. So let's talk about caring about the American people the American people have had to sacrifice far too much because of the incompetence of this Administration. It is asking too much of the people that we talked too much of the people look at they would not be equipped with the information. They need to help themselves to protect their parents.
Speaker 1	13:47	And their children Kamala Harris. Senator Harris, I mean, I'm sorry.
Speaker 2	13:51	It's fine. I'm coming. No.

Speaker 1	13:52	No your Senator Harris Tony for life to get back to normal. Dr. Anthony fauci and other experts say that most of the people who can be back stated need to be vaccinated but half of Americans now say they wouldn't take a vaccine if it was released now. If the Trump Administration approves a vaccine before or after the election should Americans take it and would you take.
Speaker 2	14:14	It if the public health professionals if dr. fauci if the doctors tell us that we should take it. I'll be the first in line to take it. Absolutely. But if Donald Trump tells us I should say that we should take it. I'm not taking it.
Speaker 1	14:29	Vice president Pence there been a lot of repercussions from this pandemic in recent days. The president's diagnosis of covid-19 has underscored the importance of Job that you hold and that you were seeking that's our second topic tonight. It's the role of the vice president. One of you will make history on January 20th. We would be the vice president to the oldest president. The United States has ever had Donald Trump will be 74 years old on inauguration day. Joe Biden will be 78 years old that already has raised concerns among some voters concerns that have been sharpened by President Trump's hospitalization in recent days. Vice president Pence have you had a conversation or reached an agreement with President Trump about safeguards or procedures when it comes to the issue of presidential disability? And if not, do you think you should you have two minutes without interruption?

Speaker 3	15:25	Well, Susan thank you. Although I would like to go back to move on but I would like to go back because the reality is that we're going to have a vaccine senator in record time in on her. Heard of time in less than a year. We have five companies in Phase 3 clinical trials and we're right now producing tens of millions of doses. So the fact that you continue to undermine public confidence in a vaccine if the vaccine emerges during the Trump Administration I think is is unconscionable and Senator. I just asked you Stop playing politics with people's lives. The reality is that we will have a vaccine We Believe before the end of this year and it will have the capacity to say countless American lives and your continuous undermining of confidence in a vaccine is just it's just unacceptable and let me also say you know, the reality is when you talk about about failure in this Administration. We actually do know what failure looks like in a pandemic. It was 2009 the swine flu arrived in the United States. Thankfully it was ended up me not being as lethal as the coronavirus. But before the end of the year when Joe Biden was vice president of the United States not seven and a half million people contracted the swine flu 60 million Americans contract with the swine flu. If the swine flu had been as lethal as the coronavirus in 2009 when Joe Biden was vice president, we would have lost two million American lives his own Chief of Staff. Ron. Klain would say last year that it was pure luck that they did quote everything possible wrong and in we learned from that they left the Strategic National stockpile empty. They left an empty and Hollow plan, but we still learn from it and I think American people and I see again can.
Speaker 1	17:30	Be priced present tense. I'm sorry.
Speaker 3	17:32	What we have done and Senator, please. Thank you very much for undermining confidence in a vaccine.
Speaker 1	17:37	Senator Harris. Let me ask you the same question that I asked vice president Pence which is have you had a conversation or reached an agreement with Vice President Biden about safeguards or procedures when it comes to the issue of presidential disability. And if not and if you win the election next month, do you think you should you have two minutes uninterrupted?

Speaker 2	17:58	Let me tell you first of all the day I got the call from from Joe Biden. It was actually Zoom call asking me to serve with him on this ticket was probably one of the most memorable memorable days of my life. I you know, I thought about my mother who came to the United States at the age of 19. I'm gave birth to me at the age of 25 at Kaiser Hospital in Oakland, California. And the thought that I'd be sitting here right now, I know would make her proud and she must be looking down on this, you know, Joe and I were raised in a very similar way. We were raised with values that are about hard work about the value and the Dignity of public service and about the importance of fighting for the Dignity of all people and I think Joe asked me to serve with him because You know, I have a career that included being elected the first woman district attorney of San Francisco where I created models of innovation for law enforcement in terms of Reform of the Criminal Justice System. I was elected the first woman of color and black woman to be elected Attorney General of the State of California where I ran the second largest Department of Justice in the United States second only to the United States Department of Justice and there I took on everything from transnational criminal organizations to The big banks that were taking advantage of homeowners to for-profit colleges that were taking advantage of veterans and then of course now I serve in the United States Senate is only the second black woman ever elected to the United States Senate I serve on the Senate intelligence committee where I've been in regular receipt of classified information about threats to our nation and hotspots around the world. I've traveled the world. I've met with our soldiers in are in war zones and I think Joe has asked me to serve with him because he knows that we share. We share a purpose which is about lifting up the American people and after the four years that we have seen of Donald Trump unifying our country around our common values and principles.
Speaker 1	20:03	Thank you, Senator Harris. You do need that. Neither President Trump nor Vice President. Biden has released a sort of detailed health information that have become the modern Norm until the 2016 election and in recent days President Trump's doctors have given misleading answers or refuse to answer basic questions about his health and my question to each of you. In turn is is this information voters deserve to know vice president Pence. Would you like to go first?

Speaker 3	20:31	Well, I'm Susan. Thank you and let me let me say on behalf of the president and the first lady how move we've all been by the outpouring of prayers and concern and for the president and I do believe it's emblematic of the prayers and the concern that have ushered forth forever. American impacted by the coronavirus but the care of the president received at Walter Reed hospital White House doctors was exceptional and the transparency that they practiced all along the way we'll continue the American people have a right to know about the health and well-being of their president and will continue to do that. But I'm just extremely grateful and was more than more than a little moved by the broad and bipartisan support and sin. I want to thank you and Jill Biden for your Expressions genuine concern and I also want to congratulate you as I did on that phone call on the historic nature of your nomination. I never expected to be on the stage four years ago. So I know the feeling but the reality is we've got an election before the American people in the midst of this challenge a year and the stakes have never been higher. Basically the choice has.
Speaker 1	21:51	Never been as they want to. Senator Harris a chance to respond to the same question I asked which is do voters have a right to know more detailed health information about presidential candidates and especially about presidents especially when they're facing some kind of challenge.
Speaker 2	22:06	Absolutely and that's why Joe Biden has been so incredibly transparent and certainly by contrast. The president has not both in terms of health records, but also, let's look at taxes. We now know because of great investigative journalism that Not paid 750 dollars in taxes when I first heard about it. I literally said you mean 750,000 dollars and it was like no 750 dollars. We now know Donald Trump owes and is in debt for 400 million dollars in just so everyone is clear when we say in debt. It means you owe money to somebody. And it'd be really good to know who the president of the United States the commander-in-chief owes money to because the American people have a right to know what is influencing the president's decisions. And is he making those decisions on the best interest of the American people of you or self-interest? So Susan I'm glad you asked about transparency because it has to be across the board. Joe has been incredibly transparent over many many years. The one thing we all know about. Joe he puts it all out there. He he is honest he is forthright. But Donald Trump on the other hand has been about covering up everything. Thanks.
Speaker 1	23:27	Thank you, Senator Harris. I want to give you a chance to respond vice president.

Speaker 3	23:30	Will look I respect the fact that Joe Biden spent 47 years in public life. I respect your Public Service as well. Thank you. The American people have a president was businessman. It's a job creator. Who's paid tens of millions of dollars in taxes payroll taxes property taxes? He's created tens of thousands of American jobs. The president said those public reports are not accurate and the president's also released literally stacks of financial disclosures. The American people can review just as the law allows. But the distinction here is that Joe Biden 47 years in public service compared to President Donald Trump who brought All of that experience four years ago. Thank you very much. Thank you for trying to this economy around by cutting taxes rolling back regulations. Thank you. Thank you guys for energy fighting for free and fair trade and all of us a nice present. If Joe Biden and Kamala Harris, you.
Speaker 1	24:28:00	Know, that's a good segue into our third.
Speaker 2	24:30:00	Topic which is.
Speaker 1	24:31:00	About the economy. This has been another aspect of life for Americans. It's been so affected by this coronavirus. We have a jobs crisis Brewing on Friday. We learned that the unemployment rate. Rate had declined to seven point nine percent in September, but the job growth has stalled and that was before the latest round of layoffs and furloughs in the airline industry at Disney and elsewhere hundreds of thousands of discouraged workers have stopped looking for work nearly 11 million jobs that existed at the beginning of the year. Haven't been replaced those hardest hit include Latinos blacks and women Senator Harris, the Biden Harris campaign has proposed new programs to boost the economy and you would pay for that new spending by raising four trillion dollars in taxes on Wealthy individuals and corporations some economists warn that could curb entrepreneurial Ventures that fuel growth and create jobs would raising taxes, but the recovery at risk and you have two minutes to answer uninterrupted. Thank you.

Speaker 2	25:35:00	On the issue of the economy. I think there couldn't be a more fundamental difference between Donald Trump and Joe Biden. Joe Biden believes you measure the health and the strength of America's economy based on the health and the strength of the American worker and the American family. On the other hand, you have Donald Trump who measures the strength of the Economy based on how rich people are doing which is why he passed a tax bill benefiting the top 1% and the biggest corporations of America leading to a two trillion dollar deficit that the American people and I have to pay for on day one, Joe Biden will repeal that tax bill. He'll get rid of it. And what he'll do with the money is invested in the American people and through A plan that is about investing in infrastructure something that Donald Trump said he would do I remember hearing about some infrastructure week. I don't think it ever happened. The Joe Biden will do that. He'll invest in infrastructure. It's about upgrading our roads and bridges but also investing in clean energy and renewable energy Joe was going to invest that money in what we need to do around Innovation. There was a time when our country believed in science And invested in research and development so that we were in an innovation leader on the globe Joe Biden will use that money to invest in education. So for example for folks who want to go to a two-year Community College, it will be free. If you come from a family that makes less than a hundred and twenty-five thousand dollars. You'll go to a public university for free and across the board will make sure that if you have student loan debt, it's cut by $10,000. That's how Joe Biden thinks about the Which is it's about investing in the people of our country as opposed to passing a tax bill, which had the benefit of letting American corporations go offshore to do their business.
Speaker 1	27:29:00	Thank you. Dr. Harris vice-president pinch. Your Administration has been predicting a rapid and robust recovery. But the latest economic reports suggest that's not happening should Americans be braced for an economic comeback that is going to take not months but a year or more you have two minutes to answer. Interrupted.

Speaker 3	27:48:00	When President Trump and I took office America gone through the slowest economic recovery since the Great Depression when Joe Biden was vice president, they tried to tax and spend and regulate and bail our way back to a growing economy President Trump cut taxes across the board despite what Senator Harris says the average American family of four had two thousand dollars in Savings in taxes. And with the rise in wages that occurred most predominantly for blue collar. Are hard-working Americans the average household income for a family of four increased by four thousand dollars following President Trump's tax cuts. But America you just heard Senator Harris tell you on day one, Joe Biden's going to raise your taxes. It's really remarkable to things right after a time where we're going through a pandemic that lost twenty two million jobs at the height. We've already added back 11.6 million jobs because we had a president who cut taxes Road. Back regulation Unleashed American Energy fought for free and fair trade and secured four trillion dollars from the Congress of the United States give direct payments to families. Say 50 million jobs through the paycheck Protection Program. We literally have spared no expense to help the American people in the American worker through this Joe Biden and Kamala Harris want to raise taxes. They want to bury our economy under a two trillion dollar Green New Deal, which you were one of the original co-sponsors of in the United States Senate. They want to abolish fossil fuels and ban fracking which would cost hundreds of thousands of American jobs all across the Heartland and Joe Biden wants to go back to the economic surrender to China that when we took office half of our international trade deficit was with China alone. And Joe Biden wants to repeal all of the tariffs that President Trump put into effect of fight for American jobs and American workers. Joe Biden says democracies on the ballot make no mistake about it season. The American economy the American comeback is on the ballot with four more years of growth. Thank you, sir opportunity for more years of prison at Donald Trump 2021 Vegas economic year in the history of this.
Speaker 1	29:59:00	Country. Thank you very much. President Pence Senator Harris.
Speaker 2	30:02:00	Well, I mean, I think that we saw enough of it in last week's debate. But I think this is supposed to be a debate based on fact and truth and the truth and the fact is Jill Biden has been very clear. He will not raise taxes on anybody who makes less than $400,000 a year peel.
Speaker 3	30:15:00	The Trump tax.
Speaker 2	30:16:00	Mr. Vice president speaking. I'm speaking.
Speaker 3	30:20:00	The important as you said the truth. Joe Biden's V twice in the debate last week that he's going to repeal the Trump tax cuts. That was tax cuts that gave the average working family $2,000 in a tax break every single.

Speaker 1	30:34:00	Job Senator.
Speaker 2	30:35:00	That is absolutely not true. That's not only.
Speaker 3	30:38:00	Cutting is the only going to repeal part of the Trump tax cuts.
Speaker 2	30:41:00	If you don't mind letting me finish we can then have a conversation. Okay, please okay. Joe Biden will not raise taxes on anyone who makes less than four hundred thousand dollars a year. He has been very clear about that. Joe. Biden will not end fracking. He has been very clear about that. Joe Biden is the one who during the Great Recession was responsible. For the Recovery Act that brought America back and now the Trump hence Administration wants to take credit. When they ran when they rode the court coattails of Joe Biden success for the economy that they had at the beginning of their term. Of course. Now, the economy is of complete disaster. But Joe Biden on the one hand did that on the other hand you have Donald Trump who has reigned over ad recession that is being compared to the Great Depression on the one hand. You have Joe Biden who was responsible with President Barack Obama with the Affordable Care Act, which brought health Over 20 million Americans and protected people with pre-existing conditions and what it also did is it saved to those families who otherwise were going bankrupt because of hospital bills, they could not afford on the other hand. You have Donald Trump who's in court right now trying to get rid of thank you for trying to get rid of the Affordable Care Act, which means that you will lose protections if you have pre-existing conditions, and I just this is very important Susan. Yes, and it's a we need to give we need to give vice president just like that. He interrupted me. I'd like to just finish please. If you have a pre-existing condition heart disease diabetes breast cancer, they're coming for you if you love someone who has a pre-existing condition. Thank you. Thank you so much for you. If you are under the age of 26 on your parents coverage, they're coming for you Senator.
Speaker 1	32:33:00	Harris. Thank you. You're welcome me give you a chance to respond.

Speaker 3	32:36:00	Well, I hope we have a chance to talk about health care because Obamacare was a disaster the American people remember it well. President Trump and I have a plan to improve Healthcare and for present protect pre-existing conditions for every American would look Senator Harris you're entitled to your own opinion, but you're not entitled to your own facts. You yourself said on multiple occasions when you were running for president that you would ban fracking. Joe Biden looked his supporter in the eye and pointed and said I guarantee I guarantee that we will abolish fossil fuels they have a two trillion dollar version of the green New Deal Susan that your newspaper USA Today said really wasn't that very different from the original Green New Deal more taxes more regulation Banning fracking abolishing fossil-fuel crushing American Energy and economic surrendered China is a a prescription for economic decline President Trump and I will keep America growing the v-shaped recovery that's underway right now. We'll continue with four more years of President Donald Trump.
Speaker 1	33:43:00	Thank you very very much vice president hence. Once again, you've provided the perfect segue to a new topic which is climate change and vice president Pence. I'd like to pose the first question to you this year. We've seen record-setting hurricanes in the South another one hurricane. Delta is now threatening the goal and we have seen Record-setting wildfires in the west. Do you believe as the scientific Community has concluded that man-made climate change has made wildfires bigger hotter and more deadly and have made hurricanes wetter slower and more damaging. You have two minutes uninterrupted.

Speaker 3	34:21:00	Thank you, Susan. First I'm very proud of our record on the environment and on conservation according to all of the best estimates are our air and land are cleaner than any time ever recorded or water is among the cleanest in the world and just a little while ago. The president signed the outdoors actually the largest investment in our public lands and public parks in a hundred years. So President Trump has made a commitment to conservation into the environment now with regard to climate change the climate. Is changing the issue is what's the cause and what do we do about it? President? Trump has made it clear that we're going to continue to listen to the science. Now, Joe Biden and Kamala Harris would put us back in the Paris climate Accord. They impose the green New Deal which would crush American Energy would increase the energy costs of American families in their homes and literally would crush American jobs and President Trump and I believe that the progress that we have made in a cleaner environment has been happening precisely because we have a strong free market economy. You know, what's remarkable is the United States has reduced CO2 more than the countries. They're still in the Paris climate Accord, but we've done it through Innovation and we've done it through natural gas and fracking which Senator the American people can go look at the record. I know Joe Biden says, otherwise now as You do but the both of you repeatedly committed to abolishing fossil fuel and banning fracking and so by creating the kind of American innovation. We're actually steering toward a stronger and better environment with regard to wildfires President Trump and I believe that Forest management has to be front and center and even Governor Gavin Newsom from your state as agreed. We've got to work on Forest management and with regard to hurricanes the National Oceanic Administration tells us that actually Ali is different as they are there are no more hurricanes today than there were a hundred years ago. Thank you, but many of the climate alarmists use very canes and wild Flyers to try and sell the goods of a green New Deal and President Trump and I are going to always put American jobs and American workers first.
Speaker 1	36:43:00	Senator Harris as the vice president mentioned who co-sponsored the green New Deal and Congress, but Vice President Biden said in last week's debate that he does not support the green, New Deal. But if you look at the Biden Harris campaign website, it describes the green new deal as a crucial framework. What exactly would be the stance of a Biden Harris Administration toward the green New Deal. You have two minutes uninterrupted.

Speaker 2	37:08:00	Or so. First of all, I will repeat and the American people now that Joe Biden will not ban fracking that is a fact that is a fact I will repeat that Joe Biden has been very clear that he thinks about growing jobs, which is why he will not increase taxes for anyone who makes less than four hundred thousand dollars a year, Joe Biden's economic plan Moody's which is a reputable Wall Street firm has said will create seven million more jobs than Donald Trump's and part of those jobs that will be created by Joe Biden are going to be about clean energy. And renewable energy because you see Joe understands that the west coast of our country is burning including my home state of California Josie's what is happening on the Gulf States which are being battered by storms Joe has seen and talked with the farmers in Iowa whose entire crops have been destroyed because of floods and so Joe believes again in science. I'll tell you something Susan I served when I first got to the Senate on the committee. That's Responsible for the environment. Do you know this Administration took the word science off the website and then took the phrase climate change off the website this we have seen a pattern with this Administration, which is they don't believe in science and Joe's plan is about saying we're going to deal with it. But we're also going to create jobs Donald Trump when asked about the wildfires in California, and the question was, you know, the science is telling us is you know, what Donald Trump said science doesn't know. So let's talk about who is prepared to lead our country over the course of the next four years on what is an existential threat to us. As human beings Joe is about saying we're going to invest that in renewable energy with going to be about the creation of millions of jobs. We will achieve net zero emissions by 2050 carbon neutral by 2035 Joe has a plan. This has been a lot of talk from the Trump Administration and really it has been to go backward instead of forward. Third we will also re-enter the climate agreement with pride.
Speaker 1	39:22:00	Senator Harris just said that climate change is an existential threat vice president Pence. Do you believe that climate change poses an existential threat?

Speaker 3	39:33:00	I said Susan climate is changing will follow the science. But once again, Senator Harris is denying the fact that they're going to raise taxes on every American Joe Biden said twice in the debate last week and on day one. He was going to repeal the Trump tax cuts. Those tax cuts delivered two thousand dollars in tax relief to the average family of four Across America with regard to Banning for Gagging I just recommend that people look at the record. You yourself said repeatedly that you would ban fracking you were the first Senate co-sponsor of the green New Deal and while Joe Biden denied the green New Deal Susan, thank you for pointing out. The green New Deal is on their campaign website. And as USA Today said, it's essentially the same plan as you co-sponsored with AOC when she submitted an in the Senate and you just heard the senator say that she's going to resubmit America to the Paris climate Accord look, The American people have always cherished our environment will continue to cherish it. We've made great progress reducing CO2 emissions through American innovation and the development of natural gas through fracking. We don't need a massive two trillion dollar Green New Deal that would impose all new mandates on American businesses and American families. Thank you. Joe Biden wants us to retrofit.
Speaker 1	40:56:00	4 million.
Speaker 3	40:58:00	Business buildings. It makes no sense. It will cost. Rob's President Trump. Thank you for America First he's going to put Jobs first and we're going to take care of our environment and follow the science.
Speaker 2	41:09:00	They've been on Senator here. Let's talk about that you the vice president earlier referred to as part of what he thinks is an accomplishment. The the president's trade war with China. You lost that trade War you lost it what ended up happening is because of a so-called trade war with China. America lost 300,000 manufacturing jobs Farmers have experienced bankruptcy because of it. We are in a manufacturing recession because of it and when we look at where this Administration has been there are estimates that by the end of the term of this Administration, they will have lost more jobs than almost any other presidential Administration and the American people know what I'm talking about. You know, I think about 20 year olds. As you know, we have a 20 year-old 20-something Ural who are coming out of high school and college right now and you're wondering is there going to be a job there for me? We're looking at people who are trying to figure out how they're going to pay rent by the end of the month almost half of American renters are worried about whether they're going to be able to pay rent by the end of the month. This is where the economy is in America right now and it is because of the catastrophe and the failure of Leadership of this Administration.

Speaker 1	42:37:00	Thank you, Senator Harris vice president pants. Let me give just 15 seconds to respond because then I want to move on.
Speaker 3	42:42:00	To well, I love to respond. Look lost the trade war with China. Joe Biden never fought it Joe Biden has been a cheerleader for communist China through over the last several decades. And again, Senator Harris, you're entitled to your opinion. You're not entitled to your own facts. We Joe Biden is Vice President. We lost two hundred thousand manufacturing jobs. And President Obama said they were never coming back. He said we needed a magic wand to bring them back in our first three years after we cut taxes roll back regulations Unleashed American Energy this Administration saw 500,000 next jobs created and that's exactly the kind of growth. We're going to continue to see as we bring our nation through. Thank you very much pandemic Green New Deal you guys massive him and ate your Paris climate Accord. It's going to kill jobs this time. Just like it killed jobs.
Speaker 2	43:32:00	When I just need to respond.
Speaker 1	43:35:00	Thank.
Speaker 2	43:35:00	You. Thank you. Joe Biden is responsible for saving America's Auto industry and you voted against it. So let's set the record straight. Thank.
Speaker 1	43:46:00	You. I'd like to talk about China we have as our next topic. We have no more complicated or consequential foreign relationship than the one with China. It is a huge market for American Agricultural Goods. It's a potential partner in dealing with climate change in North Korea and in a video tonight President Trump. And blamed it for the coronavirus saying China will pay vice president Pence. How would you describe our fundamental relationship with China competitors adversaries enemies? You have two minutes.
Speaker 3	44:21:00	Thank you, Susan. Well, let me before I leave that let me let me speak to voting records if I can, you know, everybody knows that NAFTA cost literally thousands of American factories to close. We saw automotive jobs go South of the Border President Trump fought to renegotiate NAFTA and the United States Mexico. Canada agreement is now the law of the land. American people deserve to know Senator Kamala Harris was one of only ten members of the Senate to vote against the USMC a it was a huge win for American Auto Workers. He was a huge win for American farmers, especially dare.

A.2.2 Job interview

Speaker	**Start time**	**Utterance**
Speaker 1	0:00	Hi Chris, our Haiku Johnson's nice to meet.
Speaker 2	0:03	You. Nice to meet you Kate.
Speaker 1	0:05	Chris. Please tell me a little bit about yourself.

Speaker 2	0:07	I'm currently finishing my masters of education program at Lake Erie College and working on transitioning from a north-eastern Ohio inner to being a member of the Jacksonville area Community.
Speaker 1	0:18	Why the move.
Speaker 2	0:21	Well, my wife's job is moving down south and I'm coming along with her.
Speaker 1	0:25	It's very Noble and my admirer beloved about you at the same time. How did you hear about the position open with our company here at REI? Well, when I thought.
Speaker 2	0:35	About the opportunity of changing locations and moving from one physical location to another I thought about making a career move as well and I was starting with my main interests and passions and I love being outside. I love doing outdoor activities and I love working with people and I thought if I could use my experience as a teacher helping people and my interest in the outdoors together. It would be a good place to start a new career. And when I was looking online, I found re eyes website and that there were positions available in the Jacksonville area. So I thought I would apply great.
Speaker 1	1:10	Well, you say you have an interest in outdoor activities. So this would be a great company for you to work at that being said what else do you know about our company?
Speaker 2	1:20	Well, I know you've been around for a long time since the late 1930s and what really caught my attention the most was the idea that it was started by people who had a passion for the outdoors. And found a way to involve others in their experience. I like the idea too that your company is consistently ranked among the top 100 companies to work for by Fortune Magazine, and I thought growing with a company that has that level of success would be a good place to be great.
Speaker 1	1:51	Why why do you want this particular position here with on within our customer service department?
Speaker 2	1:57	Because I want to start in a position that allows me to learn everything about the company from the very basic level of interacting with the customers all the way up throughout the sales process the marketing process and the production of goods and services when I saw the opportunity for customer service, I thought of the line very well with the skill set that I have and also with some skills that I could bring to the company that might be unique.
Speaker 1	2:20	You tell me a little bit more about these skills that you would transfer from your previous line of work to our department here.
Speaker 2	2:26	Sure in education. It's all about making the customer or student feel comfortable and helping them grow. I feel that within a company like REI. It's the same philosophy getting a customer comfortable with the product and helping them grow as they advance in whatever sport they happen to be participating in.

Speaker 1	2:45	Why do you think that our company should hire you specifically amongst our pool of Candidates that are interviewing for this position.
Speaker 2	2:53	Sure, I think one skill set that I have involves my background in education teachers are among the highest stressed Professionals in the workplace and REI offers a number of products that help alleviate stress by getting people outdoors and teachers also have a lot of downtime in the Summers. My thinking is that once I learn the ropes and Advance within the company, I'll be able to market the services and products that are I offers specifically to the teaching. Profession great.
Speaker 1	3:24	I have to ask this question given that you are making a career change here, especially with finishing a master's degree in education. If a position becomes available in your current field down here in the Jacksonville area. How do you approach being offered this job versus being offered a position in your field and I'm asking that in the now and Future.
Speaker 2	3:51	Sure. That's a great question. I'm not pursuing further educational opportunities within that profession. I feel like this customer service opportunity the chance to work in a field that I enjoy as much as teaching is a place that I could grow into long-term. So what I get out of teaching is working with people and helping people become a little bit better than they were when I first met them and that skill translates very well into our apis mission and so I Don't anticipate leaving a successful career for something. I've already done. I feel like this is a natural Confluence of my to interests.
Speaker 1	4:32	Chris. What would you identify as your greatest professional strengths?
Speaker 2	4:40	I think patients is probably my greatest strength that translates across the careers being able to listen to a student or customers difficulty and help them overcome that Quality with patience and compassion is my greatest strength.
Speaker 1	4:54	Would you consider any weaknesses that you have to be detrimental to the job?
Speaker 2	5:00	I think there's a risk when it comes to compassion. I think people can misinterpret compassion as being easy or willing to roll over when it's important to understand that compassion is something that is earned and it's interactive and I feel like if I establish clear boundaries He's with customers clients students. They understand that I am understanding. I am patient. I am kind yet. I'm also going to hold the people I interact with to a high standard.
Speaker 1	5:32	To date. What would you say is your greatest professional achievement?
Speaker 2	5:38	Personally becoming a teacher earning a master's degree. Those have all been wonderful experiences for me. But the greatest achievement would have to be working with young people and helping them get a little bit better at communicating with the world around them.

Speaker 1	5:57	Chris can you tell me about a challenge or conflict that you faced in the workplace and how you would deal with it? Sure.
Speaker 2	6:04	Within the education profession. There are challenges every day. You're dealing with hundreds of different hundreds of different personalities and interests and levels of enthusiasm. So being able to engage students with content and enriching way is part of overcoming a conflict or difficulty. I think building professional relationships is another way to deal with conflicts understanding where people come from Prior to teaching I was involved with a photography company photographing events, like weddings and bar mitzvahs and very very important events in people's lives and a lot of times conflicts that arise. So again, just like in teaching overcoming those conflicts with with patients and listening to the client or customers needs is very important.
Speaker 1	6:47	Now you are interviewing for a position that is not an entry level position and there is the possibility that those in our company that were overlooked for this position may have some animosity regarding the new hire for this position. So how do you approach this potential challenge here at the workplace?
Speaker 2	7:06	The important thing is to listen to the people around me and understand that I am a newcomer and that they know more about the work environment then I will coming in. I think because I have a friendly demeanor and I'm not confrontational when meeting new people that in time will have a chance to earn each other's mutual respect and understand that we both want the same thing the company to grow and us to grow within the company.
Speaker 1	7:32	Where do you see yourself in five years, specifically he within our company.
Speaker 2	7:36	Hopefully I'll be able to develop a market that allows me to interact with people within the teaching profession and and grow our brand using that market as a Target client base. It's great.
Speaker 1	7:51	Are you interviewing currently with any other companies and if so what interests you about those particular companies as well.
Speaker 2	7:58	Since I'm new to the area. I haven't had the opportunity to interview with a wide variety of companies. I have looked at Bass Pro Shops and I'm currently scheduled to interview with them. They have a similar Mission and similar goals to REI. It's just that their product line is more specifically tailored than re is more diverse offerings.
Speaker 1	8:19	We here at REI, you know, it is being in customer service. There are aspects of your job that are going to be a little less active than I could expect you being in a teaching position where I assume you're on your feet more one-on-one face-to-face with individuals. Do you prefer a specific type of work environment? Do you thrive in a variety of work environments? And what does this look like?

Speaker 2	8:46	Well, I thrive in Variety of work environments I wouldn't expect to be sitting in a cubicle interacting with customers unless I'm handling a phone call and that's fine. That's part of the job. But what I would expect is to be actively engaged with the customers who come into the store and helping show them the products that might best suit their needs and introducing them to products that they might not have considered before. So what I've expected to be up moving around and being consistently engaged with a customer or trying to find ways to improve what we offer to our customers.
Speaker 1	9:16	Can you describe a time for me ever that you may have disagreed on a decision that was made for you and staff members at work. And how did you deal with this?
Speaker 2	9:26	Sure within the field of Education. There are new initiatives being rolled out all the time. And sometimes those initiatives get mandated you have to teach this way this many times a week and when something like that happens, I think it's important to go back to what you know about your profession or your skill set and be able to provide research that supports why you do what you do. And I think when faced with a difficulty again, it's important to do what the boss says and it's an important to also advocate for what might be best for the customer or student.
Speaker 1	9:60	All right, we're going to move on to some cognitive behavioral questions. And basically these questions are designed. They're not necessarily about your past your experience. There are sort of out of the box.
Speaker 2	10:12	Questions. And.
Speaker 1	10:13	So just do your best at answering them as her go through this. So if you were an animal, what would you be and why?
Speaker 2	10:23	Tough question, I think. Considering the the traits of my dog if I had to be an animal I would certainly want to be a domesticated dog. They just have that ability to love unconditionally, they're intensely loyal and they just provide so many benefits to people something. I'd like to do in my life as a human to Great.
Speaker 1	10:47	The next question how many tennis balls can you fit into a limousine.
Speaker 2	10:53	Tennis balls in the limousine? I guess to effectively answer that question. I would need some help. I would need to know what type of limousine because if we're talking about a stretch SUV. That's a much larger volume than a smaller limousine. Are we talking to door for door? I also don't know what's in the limousine already other people in there or is it empty once I had those questions answered I'd have to do some calculations about the size of a tennis ball how much area or how much volume it takes up calculate the inside of the limousine and then And make sure the math works out I could get that answer to you. I just can't do it off the top of my head, right?

Speaker 1	11:36	Let me ask this final question are do you have any questions for us?
Speaker 2	11:41	Well, I mentioned a couple times about being interested in helping grow a market for REI. I'm just curious as a new employee or someone starting out on the lower level. How open is upper management hearing the ideas of an employee.
Speaker 1	11:56	We do try to have monthly meetings where we meet with a variety of employees at random. So these aren't things that are necessarily Scheduled regularly, they're very they're very semiannual. They're very informal. We do like to hear our employees opinions about our products, especially and about marketing. One thing that we don't necessarily do is the lower level positions that are dealing maybe with collections or you know, the intake and distribution of customer service calls. We don't necessarily talk to that group of employees. He's very often but I'd be very interested to hear your opinions on that. Should you be hired for the position and maybe ways that we could do better as well by reaching a larger group of our employees to benefit our company.
Speaker 2	12:49	Okay. And then last question for me what from your perspective is the best part about working for REI?
Speaker 1	12:55	I think the work atmosphere with our company is very welcoming and even in my position and somewhat upper management those that I work for that are above me. They're very We're very good to our employees and I think because of the treatment of our employees we retain employment. It's very easy for us to get things done because people want to do their jobs. I have worked myself in positions where I worked underneath people who were more tyrannical in nature and it's very difficult to want to do more and go above and beyond what you need to do in those situations. So I believe that that's the number one benefit of working in our particular company is the atmosphere itself.
Speaker 2	13:36	Great. Thank you so much for your time K. I appreciate much.
Speaker 1	13:38	Take care.

A.2.3 Meeting 2021-03-11

Speaker	Start time	Utterance
Speaker 4	0:02	Und so ein bisschen ist die Frage. Wir sollten jetzt nicht einfach nur Technologie entwickeln sondern ähm immer News Case im Kopf haben und dann etwas bauen was diese newscast tatsächlich löst das ist irgendwie ganz wichtig. Und so. Beim nachdenken ist mir aufgefallen dass es da verschiedene Szenarien gibt oder verschiedene potenzielle User für so eine summary oder was was auch immer wieder jetzt bauen für unseren overview. Dokument. Und zwar. Kann man dich so Fragen wenn so ein Meeting stattfindet einmal im teammeeting oder jetzt hier das Bea Meeting und wir machen eine Aufnahme erstellen ein transkript davon. Wer schaut dann hinter nochmal da rein? Und was füRN von ähm mit was von Bedarf was können Bedürfnis hat die Person die da reinschaut und was will sie da rausholen und das sind ganz verschiedene Sachen je nachdem Wer das ist. Also mal überlegen Wir nehmen das Meeting hier auch im transkribieren das. Würdet ihr hinter in das transkript reinschauen.
Speaker 2	1:27	Also in das erste transkript hab ich schon mal reingeschaut einfach. Um äh mal schauen wie das so abläuft oder wie das so transkribiert wird und die nachfolgenden wahrscheinlich eher weniger jedoch wenn wir sie brauchen um irgendwelche. Algorithmen zu testen oder mal.
Speaker 4	1:52	Hier macht jetzt das Thema äh translation aber nehme andere Thema jetzt über ähm weiß ich nicht ein neues Google Maps bauen oder so also nicht das Thema. Installation und hat euer Meeting mit euern Bea Betreuer würdet ihr hinter nochmal in das transkript reinschauen.
Speaker 3	2:11	Vielleicht. Wenn wenn wenn man über etwas spezielles gesprochen hat dann hat er ein Problem mit ich sag jetzt einfach mal irgend einer mathematischen Formel oder so das erklärt das ist alles klar und dann morgen treffen wir uns wieder denke nach wie war das schon wieder was hat er auch gesagt. Wenn man denn dann irgendetwas hätte wo man wirklich mit einem Ich weiß nicht such Wort oder so nochmal das schnell eingeben könnte wüsste man genau was sie darüber gesagt haben Ich denke soetwas ja vielleicht. Ja noch noch noch hilfreich ja.
Speaker 4	2:45	Okay genau das heißt aber das wäre das Setting irgendwas noch mal kurz nachschauen was besprochen wurde genau ja mhm.

Speaker 1	2:54	Okay ähm ja Ich denke es ist relativ schwierig. Ähm denn so Meeting transkripte sind nicht so lange um den jetzt weiß also das Video wiederschauen würde statt das transkript zu lesen Ich denke die Leute werden wahrscheinlich eher das Video nochmal schauen in 2 Wochen. Schwindigkeits mhm wenn es wirklich etwas ist das kurz wäre wenn es Natürlich 3 Stunden Interview sind dann wahrscheinlich weniger wird wenn er das transkript lesen dort nach Keyboard suchen mhm aber Ich denke für viele Leute ist doch das Video. Die Wiederholung davon ziemlich zentral und wenn das in höherer Geschwindigkeit wieder anschauen kann. Der kriegt das die Leute damit vielleicht sogar effizienter wäre als ich deinen Text transkript aber das kommt wahrscheinlich auch auf den Anwendungsfall um die Länge drauf an.

Speaker 4	3:43	Aber ich glaub also ich glaub ich immer irgendwie schlau sich zu überlegen wir selber. Können wir noch mal reinschauen. Ähm und also Ich bin so ein bisschen zu dem Schluss gekommen Ich würde wahrscheinlich nicht nochmal reinschauen also erstens weil ich nicht die Zeit dafür habe aber auch weil wir beim nächsten Meeting mich nochmal auf die Punkte eingehen wir heute gesprochen haben. Und ihr seid so tief im Thema drin dass sie sowieso wissen was wir besprochen haben und beim nächsten Meeting dann die Sachen wieder anspricht also Ich habe für mich festgestellt ich bräuchte kein transkript davon weil es wird mich glaub ich nicht interessieren. Ja ähm. Den anderen Punkt den du Grad gesagt dass Maurice ich würd nochmal was suchen das ist was ganz anderes also nicht nur jetzt nächste Woche sondern in 4 Wochen wenn es darum geht auch Wir haben doch mal über sein Papa gesprochen. Wie hieß das noch und Stadt jetzt äh das Mail zu schreiben und zu beschreiben a wie hieß dieses aber Wir haben doch mal vor 4 Wochen dann könnten wir nochmal gucken ooh irgendwo in den Vergangenen 5 Stunden müsste doch dieses Paper werden worden sein? Ähm und dann danach suchen also das ist ein ganz anderer Ansatz ähm. Ich versuche ein bisschen dahin zu kommen welches Problem wollen wir eigentlich lösen und wie lösen wir das Problem also das eine ist wirklich such Problem und das hat man häufig man hat über irgendwas gesprochen erinnert sich so grob dran. Oh Mann muss das wieder rausfinden. Da ist ja die Frage wie Suche ich in einem oder auch mehreren transparenten die penzel Jeweils 10 Seiten lang sind. Und also im Video suchen möchte ich garantiert nicht weil da Dreh ich durch weil ich nicht mehr weiß wann das besprochen wurde also würde ich irgendwie anfangen mit 23 stichwörtern ähm und dann gucken wo die stichwörter auftauchen in der Hoffnung dass ich da in der Umgebung irgendwie. Das finde was ich gesucht habe und dann guck ich vielleicht das Video an das ist was anderes um genau zu wissen was besprochen wurde aber ich brauche gute suchen zu. Wenn ich mir jetzt vorstelle die Suchfunktion findet nichts dann könnte so ein topic Verlauf mir helfen dass ich bei so da haben wir über sammelstation mit Word gesprochen haben überdies. Jenes und ich such dann vielleicht so so in dem overview aber eigentlich das wissen brauche ich eine schlaue Suchfunktion die mir irgendwie über mehrere transkripte möglicherweise sagt was du auftauchte mit dem Kontext. Wer was gesagt hat oder so? Es gibt aber eine Anwendung wo ich sehe dass man so ein over you extrem dringend braucht und das ist der Fall dass das Meeting stattgefunden hat und einer von euch nicht da war. Weil der kriegt dann die ganzen Diskussionen und die ganzen Sachen nicht mit. Ähm und entsprechend wäre es für den wahrscheinlich extrem hilfreich hinter die summary anzugucken wie auch immer die aussieht und sagen okay äh was eigentlich gelaufen. Ich glaub das ist so ein bisschen für die Unterscheidung bauen wir etwas für jemand der dabei war der weiß was passiert ist oder bauen wir etwas für jemand der nicht dabei war. Ist ganz relevant. Weil der hat ganz andere Bedürfnisse der wilden oben overview haben was es gelaufen der dabei war will wahrscheinlich eher was nachschauen will was.

Speaker 5	7:21	Finden. Ja sie denkt wenn das. Das kann man ja auch so formulieren zu sagen ja wenn wenn ich jetzt sind wir wissen will was das zehnte im zehnten letzten Meeting war dann ist es ungefäHerr der gleiche die gleiche Situation wenn ich nicht dabei war. Also wenn ich wissen will was wir vor 2 Monaten gesprochen haben war dann hab ich eh keine Ahnung mehr.
Speaker 4	7:53	Quasi oder also der sagen wir mal jemand der nicht weiß was in den Medien besprochen wurde also mhm. Ist das Vergessen hat oder physisch nicht da war oder war Füße Start hat seine Emails gecheckt dass es irgendwie für mich irgendwie das gleiche im Effekt er weiß nicht was wirklich vorkam und ein anderes Bedürfnis. Jemand der dabei war der weiß worum es in dem Meeting ging etwas nachschauen will. so bei ganz verschiedene Anwendungen kann mich erinnern Wir haben letzte Mal so ein bisschen Dr.über gesprochen ich hab so den Editor und dann könnte man so rund rum zeigen wie der Verlauf ist und dann wieder rein gehen wieder editieren und so Die Frage ist ein bisschen. Das überhaupt jemand ähm oder ist das so ein Misch aus verschiedenen Anwendungen nur bald irgendwie technisch möglich wäre uncool aussehen würde aber benutzen wir es keiner weil derjenige der den Überblick kriegen will. Der geht nicht da rein und editiert dann das transkript also das irgendwie so unplausibel ganz konkret ein use Case festlegen und sagen Wir gehen davon aus Person x war dabei oder nicht. Hat folgendes Bedürfnis. Und jetzt lösen wir dieses Bedürfnis für die Person.
Speaker 5	9:12	mhm mhm. Ja genau also ich meine das war so User Stories quasi oder. Oder irgendwie. Genau und Ich denke das ist das ja. Das ist auch etwas das wir am Anfang ganz gesprochen haben dass wir Ich glaube sie macht das was für die Arbeit jetzt von der Einsatz von der Domäne und vom vom muss Kaeser irgendwie fokussieren das denke ich auch das Noch 10. Alles kann so viele verschiedene Richtungen gehen. Ähm. Wenn ich das sinnvoll wenn man glaub ich ist hilfreich wenn wenn man da mal sein Fokus äh zulegen.

Speaker 4	9:50	Probiert hab. Das müssen wir jetzt irgendwie entscheiden ähm was is das Patrick ist es ein Job Interview ist eine faire Meeting is es im teammeeting ist es da ne politisches Interview mit irgendein Politiker ist. Diskussion von äh von Trump und ko und Wer will was daraus Wer will Informationen rausziehen. Und übersetzen das eine was wir vielleicht noch diskutieren sollten das zweite ist. Ähm. Könnt ihr den Begriff ghosting. Jumping out of the Box. Ähm das ist so die Idee dass es ähm so sogenannte disruptive Technologien gibt oder destruktive Erneuerung die quasi mit den Konventionen brechen und was ganz neues entwickeln. Aber das gleiche Problem lösen also typisches Beispiel ist Air BNB die das Problem gelöst haben ich muss irgendwo übernachten und die nicht ein neues Hotel gebaut haben mit etwas größeren Zimmern sondern ein ganz neues Konzept. Wie du irgendwo übernachten kannst oder uber das Problem gelöst haben Wie komme ich von A nach b und mich einfach ein neues Taxi Unternehmen irgendwo aufgebaut haben und dann überlegt haben mit neuen Methoden mit den verfügbaren Mitteln? Wie kann man das eigentlich schlau lösen? Und ich glaub das ist jetzt eine Chance wo man sagen kann ich hab Meeting ich hab da Texte und bis jetzt schreiben Leute Meetings Amis a Meeting minutes setzen einfach hin und schreibe entweder strukturierte warum Truppe etwas gelaufen ist in Meeting. Aber das ist einfach bald bis jetzt nicht automatisch war das jetzt automatisch machen und zum Beispiel Setting haben Ich möchte gerne. Einen nicht Teilnehmer informieren über das Meeting Wie kann ich das eigentlich schlau machen sodass er in minimaler Zeit in maximal über gekriegt. Und. Die Idee die wir ursprünglich mal hatten die ich auch in der letzten Arbeit noch so ein bisschen verfolgt habe war wir erzeugen einen Text wie ein normales Meeting mit eine zusammen Fassung der fließtext hintereinander Weg ist und sagt was passiert ist. Das ist aber nicht irgendwas neues und dann wird bauen das nach der Mensch bis jetzt in den Letzten 100 Jahren gemacht hat. Schlauer wäre vielleicht zu überlegen mit all den Methoden Wir haben was können wir jetzt eigentlich machen um einen schnellen Überblick zu geben über ein Meeting. So dass man seine informationsbedürfnis gestillt hat. Einfach so für wen machen wir das eigentlich das zweite war die Aufforderung mich einfach in den Standard O Meeting Protokoll Punkt Bitch please Access zu denken sondern wirklich zu überlegen. Für den use gehst du gleich aus ixen was könnte man eigentlich cooles machen viell und was könnte man irgendwie neu und revolutionär destruktiv Andenken was ganz anders ist als bisher aber vielleicht viel mehr hilft. Jetzt hab ich irgendwie erschlagen.
Speaker 2	13:40	Zu.

Speaker 1	13:40	Deinem ersten Punkt also Ich denke auch dass der Anwendungsfall für Personen die nicht Wissen um was es ging oder nicht mehr wissen was es ging das sicher der idealste ist. Mhm und denke dass wir sicher auf dem quasi auf dieser Annahme weiterarbeiten werden. Das ist auch spannend für beispielsweise Interviews die publiziert werden weiß jemand was gesprochen wurde aus der Interview und die andere Person ja also denke ich dass der Anwendungsfall dass jemand nicht weiß um was es da ging wahrscheinlich der breiteste ist. Denk ich auch vermietet jemand abwesend war Wie kann in 1015 Sekunden einen Überblick bekommen was gesprochen wurde. Mhm so grob gesagt.
Speaker 5	14:27	Ja. Also vielleicht noch ein anderes Projekt also wenn die um hätte News Geld vielleicht ein bisschen äh herauszukristallisieren per sie denn die Leute überhaupt die Industrie jetzt benutzen. Und was machen die damit und was brauchens für was brauchen Steam. Also und vielleicht gibt es so wie ne ne auch nen Richtung wo wo dann irgendwie praktikabel ist und tatsächlich auch jetzt äh das Produkt an die hilfreich ist. Aber ich kann mir 100 use Case vorstellen sind sie alle Mega cool also die Frage wie entscheidet man das auf welcher Grundlage oder.
Speaker 4	15:08	Und. Also genau die Frage kann ich äh ziemlich beantragen Wir haben noch keine User die wirklich effektiv das nutzen das heißt Wir haben irgendwie ein paar Journalisten diesmal ausprobiert haben die gesagt haben ist cool. Aber auch das noch nicht regelmäßig benutzen weil wir halt vor allem Schweizer Journalisten angesprochen haben die alle schweizerdeutsch brauchen Wir sind noch in so einer warteposition glaub sobald der Schweizer durch haben wird das anders sein. Äh in Deutschland haben wir noch keine Kunden angesprochen die USA die wir hatten bis jetzt ähm sind völlig abstrus äh Wir haben mal einen filmschaffenden gehabt ein Regisseur. Ein Regisseur sorry und die hat ganz viele Filmaufnahmen musste dann ihren Kater sagen welche filmaufnahme zusammenschneiden soll und hat dafür unser Tool benutzt um das zu transkribieren um dann zu gucken wo welche filmaufnahme. Anfängt und endet ähm in den Text um dann zu sagen jetzt musst du mir Sekunde 13 bis 25 zusammenschneiden mit Nach 2 Minuten. Voll die ist einfach sequentiell durchgegangen sagt dem Teil will ich haben den nicht voll die fand das Mega cool aber sie hat gesagt sonst muss sie halt das Video vor und zurückspulen. Wie sie den Kunden genau weiß wo es geschnitten werden soll so konnte sie einfach auf den Knopf Dr.ücken und?
Speaker 5	16:38	Okay sag mir mal das ist so nett Käse.

Speaker 4	16:40	Echt nicht die die die die die einzige wirklich produktiv schon benutzt hat ansonsten aber das Setting ist tatsächlich. Ähm das worauf wir Zielen primär ist ein Journalist ein Interview macht ne Stunde lang und Hintern transkript aber das ist so der braucht das transkript der hat das gemacht das Interview der kennt das in und auswendig der setzt sich eine Stunde später hin und schreibt was. Und will einfach den Text nachlesen können das ist ziemlich klar. Das wo ich gerne hin möchte deswegen fragt sag ich jetzt ein bisschen das wo ich gerne hin möchte ist tatsächlich Business Meetings mhm. Und zwar die Arbeit die eine Sekretärin Sekretär macht im Anschluss ich schreib eine zusammen Fassung und die wird irgendwo auf dem Server abgelegt und niemand liest sie. Ja also ich ich glaub Es gibt 3 Arten von Anwendungen von den Meeting Minuten Zusammenfassung der erste ist ich schreib eine Zusammenfassung der entscheide. Damit hinterher alle zustimmen können dass die Entscheidung waren das ist so ein bisschen abhaken und absichern dass alle irgendwie da zugestimmt haben das zweite ist die schreibt zusammen Fassung die an den Chef oder an nichtteilnehmer geht. Das dritte ist ich schreibe zusammen Fassung die archiviert wird damit man hinterher so ein bisschen äh Finger painting machen kann wenn irgendwas schief gelaufen ist dass man sagen kann okay ich hab das doch damals schon in der Meeting gesagt sie hat nicht auf mich gehört. Weil die ist kein Mensch mehr. Die ganzen zusammenfassen Es gibt ganz viele zusammen Fassung die abgelegt werden die kein Mensch anschaut. Ähm wo ich nicht genau weiß warum man die eigentlich macht wo man jetzt sagen könnte wenn ich sowieso die aufnahmen speichern die transkripte dann brauch ich zusammenfassen gar nicht mehr weil ich kein hinter der nachgucken Wer dann was wo gesagt hat.
Speaker 5	18:45	Mhm also die sind in der Realität in meiner Erfahrung sind Diese 3 Arten von von skripten Signal sind genau das gleiche also Es gibt einfach ein ein transkript oder ein Protokoll ich sag mal Protokoll ist jetzt mal Protokolle Protokolle nicht transkript. Von der Art stichwortartige Zusammenfassung von äh Peter hat das gesagt in stichworten bla bla bla so einfach chronologisch durch Gespräch durch oder das ist ja das was du sagst. Als Sekretärin Sekretärin typischerweise erfasst mhm ähm ja und das ist ja jetzt auch das was wir jetzt so ein bisschen dahin gesteuert habt also mit dem keyword extraction für jeden Kommentar. Und ähm. Genau.
Speaker 4	19:34	Also. Also Es gibt Schon 2 ich finde wenn.

Speaker 5	19:38	Wir also ich ich finde super wenn man zum Beispiel sagt okay wir machen Business Meetings weil dann haben wir schon die Domäne irgendwie dann haben wir nicht meine wissen wo wir arbeiten äh quasi das finde ich gut. Und dann in der nächste schritt für mich wäre dann zu bringen ja eben was ich der Jungs Gäste dabei braucht das und ich glaub. Für mich persönlich ähm wenn ich wenn ich jetzt zum Beispiel nicht dabei war aber auch oder irgendwas anschauen von 14 Wochen wird so wichtig zu wissen was sind die Dinge die besprochen wurden oder falsch entscheidet. Mich betreffen ich will ja eigentlich nur wissen was mich angeht und aus diesem aus diesem Meeting ist also ja und sonst irgendwas diskutiert dass ich mir das alles egal ja. Also. Aber sie ist wahrscheinlich auch. Ziemlich schwierig denke ich dass das irgendwie äh hinzukriegen war klar wenn du wenn ich mit namentlich erwähnt werde okay dann ist klar aber wenn jetzt irgendwie ne gruppenzugehörigkeit ist so ja die wissenschaftlichen Mitarbeiter das wäre dann ich. Da muss man wissen dass ich zu den wissenschaftlichen Mitarbeiter.

Speaker 4	20:47	Genau so Personalie fizierte äh mit den Protokolle sind wahrscheinlich dann nochmal ne Stufe mehr aber ähm ich glaub was man schon unterscheiden muss Es gibt ja verlaufsprotokolle. Ja geht so der hat was gesagt was war die wichtigen Themen und so und Es gibt die ergebnisprotokolle war nur drin steht der dann muss nächste Woche Kaffee mitbringen Punkt aha. Ja und ähm. Mein Ergebnis Protokolle sind was anderes sie sind auch einfach verdient sorry. Die sind auch einfach anzufertigen weil da geht irgendwie nur darum nur die Fotos zu erkennen und die Entscheidung und das irgendwie aufzuschreiben das ist auch typischerweise das Wasser in 5 Minuten hast ja. Ja also dieses verlaufsprotokoll über wurde gesprochen äh ähm ist ein ganz anderes Setting nochmal. Ah das is irgendwie mühsam anzufertigen ähm äh verstanden steht noch Leute Stunden rein. Aber Ich denke so die also vielleicht noch. Ids gibt dann tatsächlich hinter Auch 2 Anwendung von so einem Protokoll das eine ist ich will mich informieren was gelaufen ist so wie der dann sagt Ich war nicht dabei das anderes ich muss nochmal was nachschauen Ich möchte gern den Überblick haben und das irgendwie nochmal was finden. Also. Suchen und und Überblick kriegen Sind 2 verschiedene Sachen und Ich glaube wir können da also beides hat sinnvolle Anwendung viell. Denn gerade diese Szenario ich will noch mal nachschauen was wir damals entschieden haben was wir damals diskutiert haben viell irgendwelche begrifflichkeiten nochmal suchen oder so ist genauso eine sinnvolle Anwendung. Vielleicht noch ganz kurz bevor du meine festlegen wenn so ein Journalist natürlich ein Interview macht dann gehts bei ihm auch darum ähm schnell Sachen wiederzufinden also das ist eher so in der Richtung vom letzten Mal. Ja meine Stunde transkript das sind Ungefähr 10 DIN A4 Seiten Hintereinander Weg 10 1.000 Wörter und jetzt will ich wissen Wann haben wir jetzt eigentlich äh über das Thema Wahlen gesprochen. Oder wahlmanipulation jetzt kann ich natürlich mit Kontrolle 11 das Wort suchen aber wenn ich sozusagen das ganze irgendwie komprimiert darstellen kann und und irgendwie so mit den Top Hits dadurch gehen und in 2 Sekunden sie auch da ging es um Wahlen zu diesem jenes. Ähm ist natürlich hilfreich. Ich glaub trotzdem wir sollten irgendwas nehmen wo wir selber auch ein bisschen Ahnung haben oder eine Vorstellung was wir erwarten würden oder wo wir sagen würden wenn es das gibt dann würde ich das benutzen. Und was mir da einfällt ist sowas Wie ist Pia Meeting ne werdet nicht immer Alle 3 teilnehmen würde es da etwas geben wo er sagte Wir nehmen das auf und hinter guck ich eben rein und was würde ich dann daran erwarten. Bisschen wieder dann sagt wo geht um meinen Teil wie finde ich das raus. Ähm die Alternative ist sowas wie teammeetings dass wir jede Woche haben dann nehmen wir 10 Leute teilen meistens sind nicht alle da und hinterher muss ich nochmal Mail Schreiben an alle wo ich sage was wir besprochen haben. Also sorry Ärger ich mich dann immer wieder oder ich mach es nicht und dann kriegen wir heute nicht mit dass sie jetzt irgendwie die Buchung ihrer Stunden anders machen müssen weil. Es gibt kein Protokoll. Also das ist irgendwie so ein Setting jemand war nicht dabei ist Mega spannend finde ich. Für Business Meetings. Jetzt nehme ich mal Bea Meetings auch als Business.

Speaker 1	24:52	Meetings. Ja also ein paar gute Punkte denk ich aber. Ich glaube du hast ja schon mal eine Arbeit im Bereich gemacht wo es darum ging tust zu extrahieren mhm genau und Ich denke. Eine Zusammenfassung eines Business Meetings muss wahrscheinlich eine ziemlich hohe Genauigkeit haben. Also es ist ziemlich relevant dass das was da rauskommt wirklich auch so war quasi dass die Leute einen Mehrwert haben nicht plötzlich verwirrt sind von einem transkript oder eine zusammen Fassung mhm und auch quasi noch viel dann in welchem Teil ginge. Ähm Ich glaube der Name erwähnt wurde dann ist das relativ simpel aber wie dann quasi ein gegenbeispiel gebraucht hat wenn nur eine Gruppe von Personen erwähnt wird also Ich denke quasi das Business Meeting Setting. Das lässt nicht so viel Spielraum quasi für Fehler. Außer es ist mir relativ wichtig dass ich ein Meeting transkript habe das genau ist das ich weiß worum ging es Wer für wen wurde was quasi entschieden Wer ist für was jetzt ja einfach so quasi das ziemlich exakt sein muss dass ich wirklich. Davon überzeugt bin einen Mehrwert habe. Und Ich denke dass das vermutlich eine relativ große Challenge auch ist. Ausgeht darüber bis quasi ein Protokoll ist für Tattoos oder so Action Items oder das wirklich gerade ein transkript ist Wer hat was gesagt. Ich denke das muss man auch noch quasi. Durch denken was hier praktikable wäre und was die fehlertoleranz ist der Leute die das zu sehen bekommen ja. Dass sie zu meine Gedanken gerade.
Speaker 2	26:57	Ja also für mich wären eigentlich schon Business Meetings am interessantesten also Ich denke nicht dass ich. Äh eben äh Interviews auswerten oder vielleicht in einem Interview noch mal nachschauen wo hat er was gesagt das wäre etwas für Journalisten aber bei Business Meetings Kind könnte ich jetzt auch sagen dass. ÄHWäre eigentlich ein äh sehr sinnvoll für mich oder. Wo ich mich jetzt am Essen sehen kann dass ich das wieder äh anwende oder benutze aber ja wie Pascal schon gesagt hat dieses sehr schwierig ein so ein bisschen? Mit den aufzusetzen oder ein transkript oder eine zusammen Fassung die sehr genau ist und Ich denke da ist es schon sehr schwierig denk ich. Weil es ist schon wichtig dass ich aus dem Meeting eben diese wichtig also diese Information herausnehmen was was zu tun ist.
Speaker 5	28:06	Ja also nur weil ich nicht mehr Weg lösen könnten sie ist kein Grund es nicht zu tun das ist der Bachelor Arbeit und Forschung also ich meine äh Hallo erwartet hier niemand irgendwie Google Resultate oder wie weiß ich was es geht immer darum zu schauen. Also du schon.
Speaker 2	28:28	Aber also sonst wäre sicher denk ich am interessantesten oder für mich jetzt persönlich an Business Meeting. Ah.

Speaker 4	28:36	Okay. Also nehmen wir doch mal also ich find Business Meetings immer noch ein sehr weiter Begriff also der Griff der wen ähm ich meine was ich mir gut vorstellen kann es jetzt im teammeeting also von mir aus das Bea Meeting hier oder das Team Meetings. In der heute Andi diskutiert und ein paar Themen Chef sagt irgendwie paar Sachen die neu sind in der Administration und das irgendwie niemand neu eingestellt wird und der Donner hat gekündigt wird und aha sorry habe viel wichtiger. Nachmittag genau und jetzt war jemand nicht dabei. Und will wissen was gelaufen ist. Ich glaub nicht dass man dann sagen kann das und das sind die Entscheidung gewesen und das ist abschließend ja. Aber Ich glaube man kann. Das Meeting darstellen also darstellen im Sinne von dem zeigen wo vermutlich spannende Punkte für ihn waren. Ach für den nichtteilnehmer also stell mir vor ich Krieg danach ich hab noch ein Master Studenten mit dem ich das heute morgen auch schon gesprochen deswegen hab ich so ne ganz grobe Idee ich Krieg danach irgendwie so eine Art PDF. Also entweder Peter statisch oder interaktiv Tool Interactive natürlich viel cooler weil der Stadt spät und da steht drauf in dem Meeting waren folgende Personen ähm das Meeting ist so verlaufen dass am Anfang Person x. 30 Minuten Monolog geführt hat und da ging es um Folgende 3 Themen und danach haben dann die 2 Teilnehmer B und C noch ne heiße Diskussion Geführt 5 Minuten mit dem Schlagabtausch zu Folgenden 5 stichpunkte. Ah da weiß ich schon ungefäHerr was gelaufen ist und dann kann man irgendwie sagen okay die. Action Items ****** die wir erkannt haben automatisch sind Folgende 5 statements. Das sind alles die Wörter wo das Wort to do machen oder Action alten vorkommt. Und Es gibt irgendwie so ein zeitlichen Verlauf folgende Themen kam vor und wurden dann und dann besprochen oder so weil das so in Weiß ich nicht in verschiedenen Kästchen auf dem Blatt Papier und mal gucken. Auf es ist vielleicht schön visualisiert das sieht irgendwie so eine Word Cloud und so wenn es statisch macht in PDF guck mal das an und dann muss man halt irgendwie da reingehen und dann suchen wo das ist. Wenn das jetzt interaktiv macht dann sieht man halt irgendwie die Ersten 30 Minuten hat Person x dominiert mit folgenden Themen dann klickt man drauf und dann sieht man irgendwo mit highlighting in den transkript wo die Themen sind. Jetzt ist ein bisschen das jetzt so mein Bild was ich vor Augen hab jetzt könnt ihr daran Ei aber ich will doch auf jeden Fall weiß ich nicht ähm. Was auch immer sehen eine dreidimensionale Volker mit den Wörtern gemischt auf die Personen oder was auch immer ähm wie Stelle ich jetzt da was gelaufen ist? Und ein paar Sachen davon sind ganz einfach weil das ist reines Wörter zählen Statistik ein paar Sachen davon sind Rocket Science und muss man irgendwie bauen. Und da muss man vielleicht auch sagen das gibts noch nicht also von den von dem DIN A4 steht was ich mir vorstelle gibt es halt so ein böckli eine volltext Zusammenfassung von dem Meeting. Armes können nicht bauen ist zu aufwendig Na to do irgendwann. Aber so ein bisschen was will ich eigentlich sehen.

Speaker 3	32:32	Also Ich denke der Punkte mit dem interaktiven das ist sicher der größte Vorteil gegenüber einem niedergeschriebene eingeschrieben zusammenfassen wenn das Beispiel anschaut eben die Assistentin oder Assistenz. Der dann daneben sitzen einfach alles zusammen fast ich mein das ist dann an PDF ein Blatt Papier ähm darum dass mit dem interaktiven denk ich sicher der größte Vorteil oder etwas auch anderes vielleicht revolutionär im Vergleich zu einer normalen zusammenfassen was. Was genau dort hinein kommt oder kommen sollte finde ich jetzt so auf Anhieb noch ein bisschen eine schwierige Frage also Ich denke da müssen wir uns sicher vielleicht nochmal zu dritt zusammen setzen uns ausführlich überlegen eben wenn man weiß? Ist ein Business Meeting Es geht um zusammen Fassung jemand der nicht dabei war möchte mehr Informationen darüber das ganze sollte interactiv sein am dann wirklich nochmal vielleicht zusammen sitzen und auch uns überlegen eben was würden wir da raus wollen oder was wäre für uns der. Große Vorteil wo sind wir den Atlantik gegenüber einer normalen zusammen Fassung ähm ja.
Speaker 4	33:50	Es ist. Ich denk das ist so mein teammeeting oder jetzt das Bea Meeting ist das gleiche man auch hier treffen irgendwelche Entscheidungen auch hier gibt es irgendwie eine heiße Diskussion verschiedene Sachen werden am Anfang über. Bird und hier PDF und solche Sachen gesprochen jetzt geht es eher darum Was ist die Domäne was die Anwendung Was ist das Auto kam äh nachher sagen vielleicht noch auf den Termin verschieben nächste Woche oder nicht. Das sind ganz verschiedene Sachen und sich zu überlegen Was ist es was ich eigentlich hinterher also jetzt der Basel nicht dabei gewesen wäre wenn ich den Anruf und sage he Wer hat heute das Meeting. Das ist passiert also was würdet ihr dann sagen. Und das muss ja irgendwie lang rüberkommen mehr oder.

Speaker 5	34:39	Weniger. Ich frag mich noch Ops n Unterschied macht und man sagt das ist quasi ein wiederkehren des Meeting ist also wenn sie gern das Business Meeting ist. Oder ob es quasi so ein einmaliges oder müssen wir ausgewöhnlich Meeting ist. Weil Anja weil bei dem da. Hab ich wahrscheinlich anders Interesse dran. Und zusammen Fassung oder informationsextraktion aus von einem. Einmaligen Business treffen oder so also bei einem Weekly Meeting weiß ich aber da war ich weiß wahrscheinlich. Ja das ist die Bär zum Thema ganz ganz viel ich weiß Wer dabei ist normalerweise und und und ich weiß wahrscheinlich auch noch grob ungefäHerr so die Ziele usw sind. Da würde ich also es ist in einem solchen Setting bis dann wirklich mehr für mich. Yeah. Den entscheide gehen und um vielleicht auch um die Dinge die mich betreffen jetzt da will ich eigentlich kein so erzähl Protokoll Ich will nicht alles nachlesen was alles besprochen wurde der der halbstündige monologe am Anfang interessiert mich eigentlich nicht. Sondern in diesem Kontext werden mich äh im Nachhinein wahrscheinlich in die entscheidende Dinge die mich betreffen und. Und das ist aber ein anderes Setting also wenn ich irgendwie. Äh keine Ahnung wenn wir jetzt mal zum Beispiel mit einem wirtschaftspartner an sein erstes Mal treffen und seine projektidee ich mir ausarbeiten wollen sagen die wollen irgendwie. Keine Ahnung sentimentanalyse zu buchrezession machen oder sowas dann weiß dann weiß ich nachher eigentlich nicht mehr da war ich weiß auch nicht was besprochen wurde was der Idee war zum Beispiel was die wollen. Und so also ist so wie ein an der hab ich ein breites informationsbedürfnis oder beim Informationen und vieles gar nicht wissen und bei so einem anderen beim Mann seitdem ich eigentlich viel mehr. Mhm vielleicht echt auch explodieren können also ich sag mal so gefühlsmäßig bei diesem reduzierten Meetings da da will ich mich auch nicht lang drum kümmern das will ich in 2 Minuten wissen was jetzt da Was ist jetzt los. Muss ich wissen bei den anderen Meetings vielleicht ein bisschen mehr explorieren und länger mich damit beschäftigen vielleicht.
Speaker 4	37:22	Kommt ja.

Speaker 3	37:23	Also Ich denke sehr sicher auch hilfreiche das sinnvoll wenn wir möglichst viele use cases abdecken könnten also Ich denke das würde das ganze dann auch erfolgreicher machen dass man dann gerade wenn es interaktiv ist dass man zum Beispiel wie du gesagt hast. Man möchte gar nicht wissen Wer dabei war das da einfach ein Kästchen macht man anklicken kann äHI would like to Note Post participants Name Harmonie people ähm exakte Meeting dass man da einfach. Ja ein paar Einstellungen vornehmen kann wo man genau sagen kann auch ich muss gar nicht wissen Wer dabei war ich muss die Namen nicht wissen gib mir einfach das und das. Und dann doch noch die Möglichkeit aber halt wenn es jemand ist ja genau möchte wissen möchte Wer dabei war gib mir die Namen dass man das dann auswählen kann dass wir das auch ausgeben also dass man vielleicht auch. Mehrere unterschiedliche use cases damit abdecken kann.

Speaker 4	38:23	Also ich find den Punkt wirklich cool mit dem einmalig oder mehrfach weil du hast ein ganz anderes Setting das stimmt und du hast auch ganz andere Erwartungen und auch anderes vorwissen. Also eben wenn ich schon an 10 Meetings war dann weiß ich ungefäHerr wie es läuft und so es geht für mich nur noch darum ebenso in 2 Minuten den Überblick zu kriegen was jetzt das Update gelesen und das andere ist irgendwie ich Krieg so Trump Interview. Und will wissen wo drum ging's eigentlich oder ich Krieg irgendwie so ein Job Interview vorgelegt von irgendeiner Person wo ich keine Ahnung hab und soll dann das nächste Interview führen und will mich eben von mir was wollte eigentlich schon alles besprochen. Vielleicht muss man dann. Auch pro Domäne also so ein bisschen wie du das gesagt hast Movies ko mene sagen Es gibt jetzt verschiedene informationsbedürfnisse nicht nur pro User sondern auch pro Domäne beim jobinterview interessiert mich ganz ganz andere Dinge. Die ich extrahieren muss als beim Weekly Meeting. Das hat man vielleicht von der App template macht und sagt beim jobinterview versuchen wir raus zu extrahieren ähm das Gehalt was gesprochen wurde der einstellungstermin. Äh aufenthaltsbewilligung was auch immer das sind so mehr oder weniger die Sachen die mich wirklich interessieren die ich wissen sollte dann strukturierten daten und dann gibt es irgendwie so vorwissen und was auch immer und. Ich muss ein bisschen sehen worüber haben sie schon gesprochen also haben wir über übergreifen gesprochen oder haben sie eher über tennisspiel gesprochen äh ähm das sind die Themen die besprochen wurden im Verlauf der Themen vielleicht. Aber es ist anders wenn ich nwt Meeting hab das stimmt. Also ich hab verschiedene informationsbedürfnis uns nochmal völlig anders wenn ich äh Sales Meeting hab am Anfang wo ich gar nicht weiß wo es darum geht wo brainstorming gemacht wird viell dass man vielleicht dann sagt he also nicht mehr auf Business. Von der Sache ist das Weekly Meeting in folgenden Setting und dafür wollen wir das machen ich ich find jetzt so ja oder es ist halt ein Job Interview und dafür hat man ganz anderes Bedürfnis. Ich muss bei den Job Interviews nicht nochmal reingucken ich mach sowieso Notizen aber in der tollen das zweite Interview führt könnte man mal überlegen ob es Sinn macht dass das aufgenommen ist du guckst eben rein und weiß das schon besprochen wurde. Hey das letzte Mal schon über über das Thema Preis und lass gesprochen wollte ich nochmal drauf zurückkommen oder so. Aber ich kann die Notizen geben erstens weil sie nicht verständlich sind zweitens steht da irgendwie so was drin wie Ich heiße Kandidaten für den absoluten Idioten oder was auch immer ähm was auch immer da reinschreiben oder steht. Wie äh was drin was der dann gar nicht wissen sollte weil ich über unsere Gruppe geredet also was auch immer ja also so ein bisschen ähm hat verschiedene Anwendungen und verschiedene? Sag mal äh die extrahiert je nachdem was es für ne Art von Meeting ist.

Speaker 5	41:44	Mhm also Ich bin jetzt sorry dass ich darauf anspringe aber Ich bin d News jetzt mit Interview äh natürlich Mega spannend also weil also wir jetzt auch viel damit beschäftigt sind aber. Wenn wir jetzt ja genau gleich also ich schreib irgendwelche Notizen oder und. Äh schick sie dann dir oder sonst irgendwem zum so so ja das war so mein Eindruck und so aber dann eben das umsetzt auf so eine Entscheidung zu treffen wir bestellen jetzt überhaupt an und dann irgendwie so. Die Notizen noch zu den zu den Kandidaten oder die transkripte Protokolle von Interviews zu haben das werde ich auch sein also eben wie du sagst glaub ich echt anders ist als ebenso an. Also völlig anders ist das Meeting.
Speaker 4	42:35	Mhm aha da ist es einfach schwierig an beispieldaten zu kommen. Ja weil.
Speaker 5	42:42	Also allgemein bei den Business Meetings ist wahrscheinlich schwieriger ist dann auch Teil an der Domäne würd ich denk ich. Oder wir treffen uns jetzt einfach jeden Tag Schmerzen eine Stunde.
Speaker 4	42:60	Ja. Ich mein selbst das teammeeting von uns äHHätt ich jetzt schon bedenken das einfach als Aufnahme euch zu geben und ihr könnt dann da irgendwie rein hören und Analysen machen. Und Job Interviews erst recht.
Speaker 5	43:18	Also. Du hast du hast doch dieses eine dieses. Das im Moment is Krieger Demo zeigt wurden.
Speaker 4	43:30	Die doch die davon ja das ist 2 Minuten welche Schauspieler diesen Job Interview spielen das ist nicht. Nein also meine realistisch könnte man die Bea Meetings nehmen als als Beispiel so ein Mädchen das ist wirklich Meeting Setting jemand ist nicht dabei und will hinter wissen Was ist. Weil die können wir jedes Mal aufnehmen da haben wir auch Schon 2 3 aufnahmen. Das ist so jetzt Business Case mäßig nicht das spannend aber ich glaub da kann man vieles dran dann aufbauen was auch für normales teammeeting gilt man auch hier werde ich irgendwann mal sagen Hey wir müssen irgendwann eure BA abgeben Es gibt ein Termin dafür. Irgendwo euch bei der Language Beratung melden oder was auch immer. Also. oder guck mal ob irgendwo Beispiele von jobinterviews findet vielleicht gibts da recordings.
Speaker 2	44:31	Ich wollte Grad sagen Ich denke Es gibt sicher viele jobinterview äh. Beispieltexte oder ja von Schauspielern gespielt das gibt es sicher viel. Denk ich jetzt mal ich hab noch nicht nachgeschaut aber Ich nehme an Es gibt sicher. Ein paar ein 2 Interviews. Zudem die wir gebrauchen können.
Speaker 5	45:08	Ich muss jetzt dann los äh ähm. Ja Ich glaube Wir haben viel und breit diskutiert jetzt also Wir haben jetzt auch sowas nochmal aufgewacht so wie viele sind im Fokus und Ich glaube wie ähm Maurice gesagt sicher. Wahrscheinlich hilfreich wenn ihr nochmal überlegt was das ist auch so interessiert und das auch gerne machen würdet weil er mich dann auch wirklich machen und das muss ja dann logischerweise auch irgendwie passen.

Speaker 4	45:41	Ähm. Ja das is irgendwie eher sekundär also Ich glaube wenn wir was cooles machen also mhm sagen wenn was cooles bauen können für Business Meetings oder ein cooles Konzept entwickeln wohinter Leute irgendwie drauf gucken dann wow. Wenn ich das haben kann ja ob das jetzt im Winter Schreiber ist oder nicht dann eher sekundär im Moment Domäne mal verstehen und uns überlegen Wie kann das eigentlich aussehen wie also darauf fokussiert hier eigentlich. Das macht schon noch Sinn.
Speaker 5	46:12	Mhm. Ja Ich glaube es ist ich glaub die Macht wirklich Sinn wenn wir probieren jetzt möglichst Frauüh auf um zu fixen weil sonst eben. Verlag in Richtung entwickelt hast und dann Einsatz wieder die für euch.
Speaker 4	46:26	Dann. Ja ja. Macht ihr euch Gedanken bis nächste Woche.
Speaker 1	46:34	Ja genau und dann werden wir das nochmal besprechen und ähm eingrenzen genau.
Speaker 4	46:40	Mein parallel ich mein das was sie sicherlich weitertreiben könntest du das technische mhm äh rauchen in verschiedenen Richtungen ähm auf das hinter das Tool der Wahl. Ein anderes aber so wie die ganzen Methoden mal kennst du nicht was das bei euch irgendwie spannend ist dass wir das alles mal kennenlernen. Gut. Alles klar. Woche wieder.
Speaker 2	47:12	Jo ja dann ja gut tschiss.

A.2.4 Meeting 2021-03-25

Speaker	**Start time**	**Utterance**
Speaker 3	0:02	Müsst ihr nicht irgendwas hypothetisches machen sondern er sagt okay Wir haben die Aufnahme einer war nicht dabei was möchte dir sehen was nützt ihm und was nicht. Die kannst durchspielen also ihr könnt wir können das nächste BA Meeting einfach rauskicken und am Ende muss er sich das angucken und überlegen ob das was geholfen hätte. Also dann ist jetzt zum Beispiel nicht da. Das wäre jetzt der perfekte Käse eigentlich.
Speaker 4	0:32	Mhm das stimmt.
Speaker 3	0:38	Ich denke das kann man ich mein das wär jetzt gut wenn du da sagt okay Wir bauen das jetzt mal mit dem Fokus BA Meeting das ist so Setting was häufig vorkommt der fehlen ab und zu Leute wie wie machen wir das. Was braucht es dafür was würden wir brauchen? Also ihr 3 was würdet ihr eigentlich haben wollen. Braucht ihr eine such Funktion.

Speaker 2	1:11	Also Wir haben uns das schon ein paar Mal überlegt was was was denn für uns interessant wäre ja sein ein bisschen das wichtigste was wir gedacht haben ist. Man sieht das Meeting kann schnell mal ne Stunde gehen oder vielleicht in anderen beispielen geht es noch länger Sagt 2 Stunden und anstatt anstelle von 4 Leute sind Dann 10 Personen anwesend und dass wir uns überlegt haben sodass grundsätzlich wichtigste. Ist mal wenn man etwas sucht das man sicher schnellstmöglich dorthin kommt also Wir haben uns überlegt eben zum Beispiel mit dem helf jetzt wenn wir über über Gesundheit gesprochen hätten innerhalb von 2 Stunden. Was ist die die beste Möglichkeit so schnellstmöglich an den auch zu kommen den man eigentlich sucht?
Speaker 3	2:03	Ja ja.
Speaker 2	2:05	Dass man dort mit möglichst viel Voreinstellungen oder Parameter die dann der suchen der auswählen kann eigentlich die lösung des möglich eingrenzen. Kann das schnellstmöglich an die Informationen kommt die man eigentlich möchte.
Speaker 3	2:20	Da gehst du davon aus dass du schon weißt was vorgekommen ist und nach etwas konkretem suchst zum Beispiel.
Speaker 2	2:28	Ja oder dass man weiß oder man weil man weiß was man möchte sozusagen also irgendwo muss es um etwas spezifisches gegangen sein das Suche ich was war dort die Kern aus. Sage Was ist genau gegangen aber man hat schon einige suchwörter oder Ideen um das eingrenzen zu können.
Speaker 3	2:47	Ja. Reicht es da wenn man einen Suchbegriff hat.
Speaker 2	2:52	Also Wir haben uns auch überlegt das eben das ist jetzt die einfache Funktion aber Wir haben uns auch überlegt dass man mehrere Sachen eingeben kann und dann sollte es ungefäHerr ähm auf der Time Line zeigen Wo ist es um das gegangen also wir sagt Abgabe Termin. Den Bachelor Arbeit und Ich weiß nicht Referenzen zum Beispiel dass man mehrere Wörter eingeben könnte das dann anzeigt wo in etwa sich das befunden hat das.
Speaker 1	3:25	Ja ja. Für das braucht es dann das äh Text teilen oder das topic segmentation dass wir das dann implementiert haben.
Speaker 3	3:34	Aber es ist nicht wirklich oder dafür braucht es nicht. Oder wenn du sagst ich hab irgendwie die 3 Begriffe Abgabe Termin und Referenzen und Bachelor Arbeit dann.
Speaker 1	3:46	Dann kann man.
Speaker 3	3:47	Auch mit kleinen er gerade nicht. glaub das topping tailing ist für die Leute die ungefäHerr verstehen wollen was passiert ist mhm die Suche ist ne ganz andere User Story.
Speaker 4	4:03	Genauso wie 2 Seiten die Suche braucht man wenn man weiß worum es etwas gegangen ist mhm oder auch schon dort und es wieder vergessen hat genau und das textteile wäre für jemanden der gar nichts weiß der 15 Sekunden kurz die besten. Themen Keyboard haben möchte mhm genau das wäre die 2 Käses quasi.

Speaker 3	4:25	Ja das das muss man sich überlegen was einem da sag mal wichtig ist beziehungsweise ähm ja auch worauf wir fokussieren wollt mhm also. Und nicht alles machen glaub ich. Ihr könnt doch sagen gemacht nur eine von den beiden Richtungen quasi eine von den beiden blicks Seiten. Mhm mhm mhm. Ja Suche is jetzt is zwar net also ich finds auch cool mit den mit den strichen drunter das ist sicher hilfreich. Ähm das ist immer die Frage ob wir Quasi 2 Oder 3 Themen parallel verbrannt schreibt noch weiter oder ob wir ähm euch dann auf eine Richtung bisschen mehr fokussiert und das dann müssen wir in der Tiefe anschaut und dir überlegen. Mhm sind die die Ergebnisse sind spannend.
Speaker 4	5:15	Any way ja ja denke technisch interessanter ist die Richtung der Text eiling mhm quasi also der Käse für Personen die quasi nicht wissen warum es ging und einen Überblick über das gesamte haben möchten. Mhm genau für jemand der etwas sucht. Das ist relativ einfach in der Implementierung. Aus dem Grund ist die Suchfunktion noch erweitern zum Beispiel mit Trecker expressions dass man jetzt Noch 5 Werte gleichzeitig suchen könnte oder lass doch etwas intelligenter filtern. Genau aber Ich denke das ist technisch nicht ganz so anspruchsvoll und der andere use Case brauchen wir Zeit.
Speaker 3	5:54	Und wissen also ich glaub auch die Suche also das ist ja ein ganzes forschungs Gebot Information triebel und. Da gibts halt ganz ganz ganz viele Sachen also Synonym Suche und ähnliche Wörter und ähnliche Themen und und so ähm. ich glaub schlussendlich das was ich mir da vorstelle ist dass ich da so ein term eingebe und dann bisschen spielen kann mit wie genau such ich den eigentlich also es ist eine exakte Suche oder ist es das Thema was ich Suche. Aber ich glaub visuell ist da nicht so viel was man da tut also ja weiß nicht so viel Neues da gibts halt schon ganz viel. Also. Finde jetzt auch den anderen Fall Ich möchte Überblick kriegen ähm natürlich. Auch innovativer also mhm halt irgendwie noch nicht so abgegrast das Feld oder zumindest kenne ich noch nicht genau. ja. Wenn es so NBA Meeting gibt es angenommen jetzt jemand von euch wäre diese Woche nicht dabei ganz spezifische Themen die euch immer interessieren würden.
Speaker 2	7:14	Vielleicht so die Tattoos was es auf nächstes mal zu tun gibt oder wenn wenn Aufgaben Task verteilt werden wenn wenn du jetzt sagen würdest Pascal mach das was sie macht das Maurice macht das auf nächstes mal und Ich war nicht dabei. Dass ich dann wissen wissen würde was was ich mein Task Was ist meine Aufgabe.

Speaker 3	7:34	Nächste Woche genau was dich sicher interessieren würde ist wenn du erwähnt wirst also egal mit was aber. Ähm das ist sicherlich etwas was jeden interessiert von der persönlich erwähnt wird. To do sind sicherlich spannend ja oder Entscheidungen. Also sowas wie wir entscheiden Wir schreiben die Arbeit auf griechisch. Ja sicherlich wichtig mhm mhm also dazu gab es ja die Arbeit von dem archie und das ist technisch eine echte Herausforderung Entscheidungen zu finden. Ähm. denk. Das könnt ihr jetzt nicht lösen aber ihr könnt Andenken wie man sie einbauen würde wenn man es könnte. Also. Das sind ja 2 verschiedene Schritte so. Ich will wissen welche Entscheidungen wurden getroffen. Und wie zeig ich dir denn an. Selbst wenn ich noch nicht weiß mit welchem Verfahren ich genau finden kann da gibt es ein paar Ansätze aber die funktionieren alle nicht so gut wie wir gerne möchten. Mhm mhm mhm. Kann man in dem jetzt nur Sachen gelb markiert in dem Preview. Ähm wo ich ja auch. Glaube was was helfen kann ist wenn man Sachen größer und kleiner darstellt. Wenn man das da in dem Tool kann also einfach sagt ahh ja.
Speaker 4	9:12	Ja das ist im Prinzip einfach eine CSS Formatierung und da könnt ich auch größer kleiner kursiv Fett oder so einstellen.
Speaker 3	9:20	Also man kann auch mal Dr.über nachdenken wenn ich den Überblick kriegen will mhm. Ob es noch andere Möglichkeiten gibt langen Text quasi zu komprimieren oder. Übersichtlicher darzustellen. also hatte die keyword extraction und wenn ich die richtig in Erinnerung hab dann schreib das einfach seitlich die Keyboards raus oder.
Speaker 4	9:45	Genau das haben wir damals implementiert aber das ist auch nicht also da muss man durch scrollen und. Nicht die schnellste Möglichkeit Ich denke lieber selbst ja über eine Time Line unsere intelligenter.

Speaker 3	9:59	Ja aber wenn wir jetzt bei der Suche bleibt mhm. Also ist das ja auch noch nicht so richtig übersichtlich wenn ich da einfach so ein bisschen Geld markiert hab wenn ich jedes Mal hingucken muss und lesen muss worum es eigentlich geht mhm ja. Was was ich jetzt im Kopf hab ist dass man irgendwie so wie so ne Welle um das das Stichwort macht und die Wörter immer größer werden lässt rechts und kleiner werden lässt rechts und links? Ah ja er macht das eigentliche Wort ganz Fett und. Also vielleicht ist das schlau oder vielleicht nicht oder nicht irgendwie so 5 Wörter rund rum rechts und links und macht sie alle gleich Fett. Vielleicht ist das gar nicht so schlau wenn du das eigentliche Wort am nächsten machst im größten macht weil das heißt ja gesucht das kennst du sowieso naja interessiert wahrscheinlich der Kontext also. Vielleicht der Satz in dem es gesagt wurde oder eben plus Minus 5 Wörter oder so. Dann siehst du auch relativ schnell was es is aber du hast natürlich recht ich muss trotzdem komplett durchscrollen das heißt auch bei der Suche. Könntest du sagen ich extra hier halt die ganzen Sätze wo es drin vorkommt und zeigt die irgendwie an. Mhm ich allen ihre irgendwie das Wort untereinander und zeigt dann jeweils die die statements davor und dahinter an wo das Wort drin vorkommt und von wem es war und weiß ich nicht.
Speaker 4	11:27	Also so ein bisschen. Mhm ja ich weiß was du meinst ja ja. Das war quasi nur die Sätze rausnimmt und die ganze Aussage.
Speaker 3	11:39	Ja weil. Sie hat jetzt die ganzen Aussagen. Und die anderen weggelassen zwischendurch oder.
Speaker 4	11:46	Genau sind jetzt nur diese Dieter auch. Gekennzeichnet sind.
Speaker 3	11:51	Also Ich denke da muss man dann schon so ein bisschen gucken dass man nicht auf auf aktionslevel bleibt sondern eher. Wieviel Kontext brauch ich davor und danach also ich hab jetzt kein Beispiel wo jetzt nur das Wort mit 2 Wörtern davon danach vorkommt wo die Adresse sehr kurz sind aber das das könnte auch passieren.
Speaker 1	12:08	Mhm.
Speaker 3	12:11	Genau denk Mamas denk. Na jetzt. Ja ja. Und jetzt ist so ein bisschen die Frage was nützt es mir dass der Tank you steht wenn ich nicht weiß worum es geht also ich brauch mehr Kontext. Aber ich brauche jetzt nicht ne halbe Seite Kontext mhm. Also da jetzt zum Beispiel.
Speaker 4	12:38	Conny Bär.
Speaker 3	12:44	Okay hilft mir das jetzt zu verstehen worum es ging natürlich.
Speaker 4	12:48	Nein. Aber Ich werde auch nicht nach Tränke suchen den Krieg.
Speaker 3	12:52	Ja so ja Nein vielleicht stimmt das ja dann ist das eine blöde blödes Beispiel.
Speaker 4	12:60	Mhm mhm. Aber ja das ist eine gute Idee finde ich dass man vielleicht etwas sucht. Das doch intelligent darstellen kann und dass man gerade Eine 6 zahlen Aussage einblendet oder dort einmal vorkommt mhm genau. Ja viel hat noch mit so intuitivität zu tun weil seine Oberfläche.

Speaker 3	13:20	Ja. Genau das ist nicht nur Technologie ist auch. habt ihr mal geguckt ob man diese Timeline kippen kann und vertikal darstellen kann.
Speaker 4	13:33	Äh Nein noch nicht okay aber ich kann mir vorstellen dass das das ist immer etwas schwierig aber möglicherweise geht das Recht einfach oder es geht gar nicht dass du diese Möglichkeit.
Speaker 3	13:45	Genau entweder. Irgendwie dass die Option oder.
Speaker 4	13:49	Mhm Oh das ja genau. ja. Mhm mhm ja aber nächste Woche wollen wir sicher Taste ausbauen. Quasi auch wirklich den Text dazu anzeigen und überprüfen ob es für uns als Menschen auch Sinn macht. Wie man das segmentieren würde ja dann auch noch äh das andere Paper aus 2019 genauer anschauen mit der Implementierung ja und dort mal ein paar einfache Tests machen um den Vergleich zum Text eilig zu haben? Mhm mhm mhm. Pfaffinger sagt mir dann dass ich das gar nicht mehr Lohn weil es viel schlechter ist als das der einfach ein paar Aussagen dazu machen können mhm genau. Und dann vielleicht auch noch die Suchfunktion etwas intelligenter gestaltet und mit den Vorschlägen die du jetzt gebraucht hast. Einfach mal anschauen wie wäre es wenn es jetzt so umgesetzt wäre das nur ein paar Wörter rundherum angezeigt werden werden. Mhm genau.

Speaker 3	15:02	Jetzt habt ihr. Viele Möglichkeiten so für einzelne Tools ähm. Ich denk irgendwann so in in 34 Wochen müsste dann vielleicht noch so ein Schritt zurück gehen und sagen wie zeige ich das jetzt eigentlich alles auf einmal an mhm. Also. Ich öffne jetzt das transkript oder diesen Dialog Analyser für das Meeting von heute. Was seh ich denn eigentlich? Weil ich ja so 5 verschiedene Tabs oben und ich muss mir dann so aussuchen was ich angucken will ist wahrscheinlich nicht der richtige Weg mhm ja. Und ich glaub da muss man so ein bisschen dann wirklich Stories definieren was braucht der User in dem Moment wo er zum ersten Mal von so einem Meeting das öffnet. Sag mal so ganz lapidare Sachen wie ich will wissen Wer teilgenommen hat. Ich will wissen wie lang das Meeting also passt so ganz rudimentäre Statistiken ähm. Gehören sicherlich dazu weil wenn ich das nicht sehe und erst irgendwie so durchscrollen muss bis zum Ende und dann sich und in den letzten Time stamp okay aber das ist trivial einfach so ein bisschen was würdet ihr eigentlich brauchen. ja. Gibt es oben eine Selektion wo ich sage mich interessieren die to dos am meisten oder ich hab das vorkonfiguriert und das ist mein Dashboard und ich kann sozusagen das selber anordnen was ich jemals sehen will. Ja ist für mich die äh Time Line mit mit den segmenten von dem Text teilen am wichtigsten oder ist es für oder kommt das erst später wenn ich ein bisschen genauer angucken will. Brauch ich erstmal so eine Word Cloud die das gesamte Meeting darstellt. Mhm mhm mhm. Das äh. Ja ja. Ist dann sozusagen wie reduziere ich jetzt all die Möglichkeiten die Ich habe. Auf das was der User wirklich sehen sollte. Also will muss kann oder darf. Mhm. Ja ja. Macht ihr schon so Notizen von dem was ihr euch überlegt und rausgefunden habt welche paypal gelesen habt und so.
Speaker 4	17:36	Ja ja ja ja bei uns vor zu unsere Quellen aus quasi. Ich hab schon ein Lotte Credo comment mit der Gruppen äh Verlauf der Arbeit.
Speaker 3	17:46	Okay gut genau. Ja also ihr müsst ja nicht ein verlaufsprotokoll schreiben also muss jetzt nicht hinterher sagen dass er in der ersten Woche das und der oja ja genau aber Ich denke auch so Überlegungen ähm die wir jetzt machen was es jetzt eine sinnvolle Darstellung dass man halt. Vielleicht so als Notiz hat übrigens Alternative Darstellung wäre noch gewesen alles mit highlighting mit gelb aber Wir haben uns dagegen entschieden aus folgenden Gründen. Ähm das hinter nicht so so völlig. Also nicht hinterher sagen okay das jetzt die Darstellung Punkt sondern muss man meistens ein bißchen diskutieren warum jetzt gerade die. Und das wäre die Alternative gewesen ne. Gut super. Noch was von eurer Seite.
Speaker 4	18:33	Äh von meiner Seite nicht.
Speaker 1	18:35	Einfach noch äh wegen den Termin in. Im Monat April also jetzt jetzt geht es nicht um die Arbeit Ich habe einfach weil äh Pascal und.

Speaker 4	18:46	Rio genau ja.
Speaker 1	18:48	Noch immer am Donnerstag jetzt eine segelkurs. Und da findet genau von 11 bis 5 statt den ganzen April. Äh und dann müssen wir den Termin vom Meeting verschieben wenn das geht.
Speaker 3	19:06	Ja ja das geht. Ähm. mhm mhm. Wann könnte denn hier?
Speaker 4	19:19	Relativ flexibel glaub ich.
Speaker 1	19:22	Außer Donnerstag ja mhm oder Donnerstag Frauüher.
Speaker 3	19:27	Würde ein Moment also ich könnte Donnerstag dann nicht mehr eigentlich oder.
Speaker 4	19:32	Ähm. Höchstens Frauüher morgen.
Speaker 1	19:37	Ja das.
Speaker 4	19:38	Geht bei mir nicht. Stimmt.
Speaker 3	19:42	Kann nicht sein Frauüh am Morgen ist sehr relativ und irgendwann wurde es gehen muss am ja also das heißt Donnerstag gar nicht eigentlich. Okay gut äh. An. So Mittwoch.
Speaker 1	20:10	Vormittag ähm Mittwoch geht für mich den ganzen Tag nicht Abend.
Speaker 2	20:17	Vielleicht Mittwoch aber ich weiß.
Speaker 1	20:20	Aber dann muss es Frauühestens um 6 sein.
Speaker 3	20:24	Ja ungern also okay gehen wir durch dann wäre noch der Freitag.
Speaker 4	20:32	Mhm Nachmittag. Das würde gehen.
Speaker 2	20:37	Ja nach nach 10 Uhr ja.
Speaker 3	20:41	Einfach ausschlafen morgen.
Speaker 1	20:43	Ich seh schon ja ja Nein Donnerstag 10 Würde vielleicht auch gehen aber dann hätten wir wie eine an den Termin nur bis 11.
Speaker 2	20:55	Immer so Meetings bis 11 Uhr eigentlich Freitag Donnerstag und Mittwoch daher erst ab 10 ab 1011 Uhr würde für mich.
Speaker 3	21:04	Gehen okay dann können wir Freitag um 10 machen.
Speaker 4	21:09	Ja das passt.
Speaker 1	21:11	Ja. Also ich meine my könnten wir wieder Donnerstag machen ist einfach ja der der ganze April jeden Donnerstag.
Speaker 3	21:26	Freitag ab neunten.
Speaker 4	21:29	Ja weil mhm genau. Oh
Speaker 2	21:38	10 Uhr.
Speaker 3	21:41	Ja genau.
Speaker 4	21:43	Also ab 2. April lästern.
Speaker 2	21:45	Leider Kaffee.
Speaker 3	21:46	Das ist.
Speaker 4	21:47	Karfreitag okay ja.
Speaker 3	21:52	Genau ich hab ab neunten genau wie sieht es nächste Woche aus da habt ihr schon euren Kurs oder.
Speaker 1	21:58	Ja erst ab 1.

Speaker 3	21:60	April. nächste Woche. Seh ich nicht dass ich am anderen Termin kann weil da is dann Karfreitag mhm am Mittwoch geht bei euch nicht mein Dienstag is komplett schon zu dass äh. Ja dann hat das ist doch super dann haben wir das Setting wir machen das Meeting Ohne 2 Teilnehmer und ihr müsst euch hinter informieren aus dem Recording können überlegen ob es ausfallen lassen. Also mhm und ihr könnt einfach kurz ne Mail schreiben wenn irgendwas is also wir können dann immer noch machen wenn jetzt irgendwie was dringendes besprochen werden muss aber wir können uns auch Mails mit Screenshots schicken oder so und wir müssen nicht jede Woche. Meeting machen also ist das nicht ja is auch wenn ihr wisst was ihr tut dann macht er einfach. Gut dann lass uns nächste Woche ausfallen. Gut und danach dann ab freitags. Ähm schreib dir noch den Don nachher kurz. respektive macht ihr wieder ne Einladung fürs Teams oder hatte ich die.
Speaker 4	23:19	Gemacht die können wir uns machen.
Speaker 3	23:22	Aha Wir haben wir eigentlich auch noch wir können gleich ja wenn das gleiche Teams nehmen ja. Gut ja.
Speaker 4	23:34	Ähm wann ist die genaue Abgabe der bachelorarbeit.
Speaker 3	23:39	Weißt du was 14. Juni aber Ich würde der spüren ich hab mir mal eingetragen. Also Ich glaube dann wäre es wahrscheinlich eher der 11. weil typischerweise ist freitags aber das geht hier in den BA Terminen. Ihr habt ihr auch so ein Formular was ausfüllen musst und. Kurz gucken.
Speaker 2	24:11	Ob ich das auch verstehen.
Speaker 3	24:13	Hab. Nee also ich meine es ist der zehnte oder 11. Juni mhm. Und danach ist dann so ein Oder 2 Wochen später dann typischerweise die Präsentation. Das heißt da ist dann irgendwie eine Stunde da könnt ihr glaub ich noch angeben wann ihr das gerne hättet. Ähm da kommt ein externer Experte der hört sich die Präsentation an liest eure Arbeit und mit dem zusammen mach ich dann die Note. Und theoretisch weiß der sonst nichts von eurer Arbeit das heißt der liest nur euren Text. Und das heißt das was ihr wirklich wichtig findet sollte in diesen Text dann auch auftauchen. Der guckt auch nicht auf den sourcecode. Ja ja. ja.
Speaker 2	25:18	Is.
Speaker 1	25:19	Gut ja das passt.
Speaker 3	25:21	Ja gut dann sehen wir uns in 2 Wochen wieder.
Speaker 4	25:24	Genau genau perfekt alles klar.
Speaker 1	25:27	Tschüß.

A.2.5 Meeting 2021-04-23

Speaker	**Start time**	**Utterance**
Speaker 4	0:11	Okay. Gut okay also hat den PDF Export gemacht und was noch.

Speaker 5	0:20	Äh wir henno dass äh Word Cloud das haben aber noch nich jeder apine. Das ist einfach mal so Test wies gebe die Word Cloud mit äh äh farblichen ab stimmig. Das eigentlich. Alli grüne such lädt das die neue Wörter sind wo im. Äh nach Sticks Block 1 der Suche mit also mich an die Stelle wieviel Minuten sind Xbox erziehen okay hab ich jetzt einfach mal 5 Minuten Knopf und der Kali ähm. Die 5 Minuten du da eigentlich zämme. Als Text. Äh ne und dann wird die Welt Analyse unter Teck Spruch verglichen mit dem nächste textblock und so haben wir eigentlich den ähm. Äh also durch die farbliche Abstimmung sehr gewähltes Wort oder welche Wörter sind der wo jetzt viel gebucht werden Willi Wörter sind jetzt User kite. Und. Ja jetzt hab ich einmal eigentlich da ähm. Für jede textblock ein einziges Einzel spielt.
Speaker 4	1:39	Will was heißt Frau.
Speaker 5	1:41	Grau ich eigentlich dass es ähm im letzten Text Spruch schon Fach euch und jetzt eigentlich bin ich im neue a ihr sind hier wo neue zu hoch sind. Aber rot wär eigentlich die wo sich keit sind aber die Zeit natürlich nicht da will ich dann nur der Output vom neue textblock da müssen wir den Wien oder an der Text kriegen wir auch noch Ich bin damit. Und ja also wie gesagt einfach dauert das äh zum Beispiel people es geht eigentlich nur noch relativ Süßigkeiten. Äh wie gesagt einfach das people Minuten fahre ich aber dem textblock hä ähm viel ausgesprochen worden ist. Und das zieht sich eigentlich äh du dort du die nächste und da das d ich jetzt ähm. Äh eigentlich eine wo sich keit wird aber sicherlich komisch dass der jetzt im neu Auer zeigt wird das ist noch das ist aber der einzige glaub. Und dann haben wir eigentlich für jede Text Spruch oder für Jede 5 Minuten vom Gespräche besonders eine Word Cloud mit äh verbliebene Wörter und der neue. Und Wir sind uns noch nicht ganz sicher Übersetzer tätig bin und wie das muss jetzt nur wieder verfolge. Y haupten Mehrwert hat somit das. Zum jetzt äh überall dir Word Cloud H oder die Word Platz Reinartz Lüge. Das ist sicher leid bis wo is mein ist überlege aber sei zum Teil wirklich also. No relativ aufschlussreiche outputs. Aber das Maxi das äh kleine zum Beispiel da Fracking.
Speaker 4	3:38	Auf den davor gehen würdest dann wäre der Fracking. In in grün oder grau richtig. Mhm wenn wir.
Speaker 5	3:47	Aber es wenn es war nur einmal brauch ich vielleicht gerne diese Word claudine vorher willst natürlich auch sind jetzt glaube ich Also 15 Wörter limitiert das sind eigentlich die 15 Mai Spruch der Wörter. Wo vorher mit und Tracking schon mal benutzt wurde im vorherige Text aber sie ja genau es wird nicht anzeigt wenn es halt nicht so häufig gebucht werden darum sind sie das dass ich es gar nicht Rede? Vorherige Word Cloud.

Speaker 1	4:20	Genau. Ich finde äh irgendwie in das is cool äh ähm das Problem Ich bin auch Darstellung ist noch nicht optimal der äh. So und bin ich nicht gut verdaubar ähm aber äh Ich denke dass äh normale Liste oder der Ball oder so jetzt neue übersichtlich. Ähm äh und mein Problem mit dem das Wort scho angezeigt wird er. Dann verfolgt Logik von dem ganz also wenn du halt nur 15 aber fragt er irgendwie. Vorab bevor man das jemand bist du Grad äh ich muss Hochdeutsch sprechen weil das ja transferieren.
Speaker 4	5:23	Ne der schneller voran.
Speaker 1	5:29	Äh.
Speaker 4	5:31	Äh.
Speaker 1	5:32	Äh was ich sagen wollte ist dann vielleicht kann man ja schon vor im also ein einmal durch den Text durchgehen das ist doppelt schmeißt ja wahrscheinlich raus. Und dann äh schauen was sind überhaupt insgesamt die wichtigen Wörter auch immer und dann kann man ja wenn das erste Mal Fracking vorkommt dann weiß man schon das ist ein Wort das wichtig ist füRFüRFür den. Schau mal so für den Text. Bekommen sie dort schon äh quasi äh in die Liste nehmen weil man weiß das ist ein wichtiges Wort insgesamt wird es kommt dann später nochmal also so vokabel Larry äh Sound macht über den ganzen Text bevor man. Dann diese segmentweise Auswahl der Keywords macht.
Speaker 4	6:20	Zum Beispiel ich finde auch das ist das hat extremes Potential Word Clouds Hermanns ähm. Wenn man irgendwie so macht dass es hilfreich ist für den User und da gibt es 2 Aspekte das eine ist die Auswahl also welches sind die wichtigen Wörter. Ähm jetzt glaub ich einfach ein wordcount gemacht oder also ich glaub du wirst rausgeschmissen wordcount auf den segmenten.
Speaker 5	6:47	Nehm ich an es ist eine Word Cloud Funktion und ja die zählt einfach die die Wörter. Ja genau so seh kreiert eigentlich schon das ganze Bild Ich habe einfach noch ein paar Parameter also Wir haben einfach paar Parameter eingestellt welche Farbe diese Wörter haben.

Speaker 4	7:04	Sind ja jetzt seine Workout sind ja 2 verschiedene Dinge. Das eine ist du fehlst aus deinem Text aus welche Wörter erscheinen sollen. Ähm gibt's quasi für jedes Wort irgendein Score an. Den kann man selber berechnen um jetzt hier einfach ne fertige Funktion genommen die hat einfach gezählt wie häufig kommt das Wort vor und das zweite ist wie wird visualisiert also das ist ja getrennt voneinander. Das einzige was man hier irgendwie jetzt höRFür die Berechnung des Bildes braucht ist die ins Korb pro Wort und dann wird es mir relativ gerechnet und dann entsprechende Größe bestimmt. Was man jetzt machen könnte wäre in 2 Richtungen weitergehen das eine ist? Die Auswahl welches sind die hilfreichen wichtigen Wörter kann jetzt das machen wir dann gesagt hat man irgendwie über den gesamten Text die die wichtigsten oder man macht irgendwie so tief IDF Kurs oder macht irgendwie pro pro Segment irgendwelche schlauen Sachen. Kann auch topic modeling machen was auch immer also kann da ganz verschiedene Dinge tun. Das zweite ist wie visualisiert man das. Und. So ein bisschen die Frage Wie kann man eigentlich dem User zeigen wie der Verlauf ist jetzt könnte man ich finde deinen Vorschlag nicht so optimal dort einfach ne Tabelle untereinander es ist äh. Vielleicht funktional aber nicht sexy ähm wenn ich jetzt etwas wo man. Quasi visuell sich durch navigieren kann. Viel spannender. Ähm. Sag mal dass das einfach was man sich vorstellen könnte wäre wenn jetzt all diese Bilder quasi auf der zeitleiste legt und ich hab so eins leider und kann dann von links nach rechts gehen. D dann jeweils das Bild. Ich kann irgendwie ablaufen lassen seh ich so die Bilder nacheinander und dann Krieg ich gerade mit ähm also kann das einfachste ich Pack alle Bilder untereinander und kann sie mir angucken. Er er. E wir dann die nächste Version wäre ich neben. Ein Sohn View und ich hab hier unten drunter unter China hab ich so ein rechts links Button und kann mich durch die Bilder durch klicken und sie dann jeweils was sich geändert hat also müssten überlegen was hier anzeigt. Bin nicht sicher ob man die Wörter sehen muss die verschwinden sondern die Wörter die neu dazukommen finde ich halt viel spannender dann weiß ich halt worum es neu gegangen ist. Das andere ist ein bisschen redundante Information. Und. Das dritte was ich persönlich jetzt eigentlich Mega cool fände wäre hier berechnet diese Word Clouds für jedes Segment das gibt uns euch hinterher quasi das Vokabular was vorkommt in der Word Cloud. Und dann. Macht ihr ein. Ein Bild was animiert ist. Und man kann quasi dann sehen wie die Bedeutung von einem Wort zu und wieder abnimmt aufgrund von der Größe von dem Wort. Ähm. jetzt jetzt muss man dann immer also ich stell mir wirklich hart Film vor der automatisch generiert wird wo das Wort Job was hier oben ist da steht immer am gleichen Ort. Und abhängig davon wie hoch der Score ist jetzt größer oder kleiner repräsentiert. Das heißt wenn ich diesen Film ablaufen lasse ist der ersten großer schwarzer Fleck und dann kommt das Wort irgendwann mal kurz vor wieder nicht dann wird es irgendwann kommt raus und geht wieder Weg wenn es dann häufig ist dann kommt halt raus und bleibt auch gleich groß und verschwindet. Ja ja. Nicht so auf den Play Button Dr.ücke dann läuft irgendwie in in 30 Sekunden dieses Video ab und ich seh welche Themen angesprochen wurden. Dadurch das vorstellen.

Speaker 3	11:21	Wäre denn das sind in Abhängigkeit zur Zeit also dass man dann immer noch irgendwie sein Time Bauer hätte wo man den Ablauf des transkripts seht und dann wenn man jetzt zum Beispiel diesen kurzen Film stoppen würde dass man dann auch genau sehen würde Wo ist man überhaupt.
Speaker 4	11:38	Genau okay. Also wie wir das dann genau ist ob man das wirklich als Word Cloud darstellt oder ob man sowie der dann sagt das in der Tabelle macht und dann werden die Wörter halt nach und nach größer und kleiner viel das mächtig. Weiß nicht ob das funktioniert weil man braucht eine relativ große Cloud. Ähm wo man genug Platz hat all diese Wörter anzuzeigen. Ähm man könnte sich auch vorstellen dass man das in der ersten Version nicht als so eine schöne angeordnete Cloud macht sondern macht irgendwie so ein. Viel eicht Versuch das mal eben profitieren darf ich den Bildschirm einmal teilen. Ja ja. Am. Whiteboard.
Speaker 1	12:36	Also ich glaub ich Ich bin Ich.
Speaker 4	12:38	Bin es cool.
Speaker 3	12:39	Wie du sagst sexy.
Speaker 1	12:41	Aber vielleicht.
Speaker 3	12:42	So als kunstprojekt.
Speaker 4	12:44	Gerlicher also überhaupt nicht so schwierig ja ich Ich.
Speaker 1	12:50	Glaube also ja okay also ich glaub ich ähm. Die Frage ist halt einfach was ich dann der use Case irgendwie oder.
Speaker 4	12:56	Also use Case.
Speaker 1	12:58	Aber schaut sich.
Speaker 4	12:58	Das in 30 Sekunden Überblick welche Themen besprochen worden. Und wann.
Speaker 1	13:05	Wir hier geht es um die zeitliche Komponente genau.
Speaker 4	13:08	Ja genau. Das hab ich bis jetzt noch nie gehabt irgendwo also bis jetzt musst du irgendwie so äh die Keywords die ausgeben lassen dann von oben nach unten runterscrollen oder so und das ist einfach nicht nicht intuitiv. Also ich sag mal was man was man am einfachsten machen könnte jetzt. Sag mal den technischen Mitteln die ihr habt ihr seid ja keine kunststudenten da gebe ich den Dorn recht ist macht so eine riesige Tabelle. Und jedes Wort. Kriegt also jedes Wort von denen die hier als wichtig erachtet ne jetzt Irgendwie 16 Äh es könnte noch nie 50 sein und jedes Wort kriegt so ein Bubble hier ist halt der Bubble für das Wort. Jop. Und dann hab ich hier den Times leider.
Speaker 5	14:02	Kannst du noch ein bisschen runter scrollen. Bisher nur die.
Speaker 2	14:07	Aber ich glaub das Rauschen.
Speaker 4	14:09	Over glaub ich ah okay das.
Speaker 5	14:14	Sein okay Nein das hab ich gesehen.

Speaker 4	14:17	Okay. Und jetzt hab ich diesen Times leider und wenn ich hier stehe ähm. Dann ist das Wort Job irgendwie ganz klein. Alles unwichtig wahr. Und wenn ich dann hier hin gehe da ist das Wort irgendwie in dem Segment extrem häufig dann wird das halt irgendwie so riesig angezeigt. Das ist sozusagen der Word Cloud einfach mal in einfachen Raster und ihr müsst nur hab ich von Zeit segmenten die Größe von der Wort Darstellung ändern. Hab ich das vorstellen. Wenn ich hier von ple Button hab noch. Da drauf klicken und dann geht das immer so in 30 Sekunden dadurch dann seh ich was wann besprochen wurde.
Speaker 3	15:12	Ähm vielleicht aus als aus deiner Erfahrung was wäre einfacher zu implementieren effektiv ein ein Video oder interactiv wo man dann zum Beispiel. Mit Klick einfach den den Time bar anklicken kann und dann hat man so ein Schieber könnte man durchfahren und die Wörter verändern sich dann interaktiv oder ist es doch einfacher das als als als Video. Anzuzeigen.
Speaker 4	15:40	Also Ich glaube dass man da jetzt nicht ein Video draus machen würde sondern dass man das eher mit CSS macht. Ähm dass man einfach also sag mal dass wir jetzt die die einfachste Variante die ich mir vorstellen kann die interaktiv ist ähm und für jeden Time stamp den du hier hast viell berechnest du quasi vor welches Kurs haben die entsprechenden Wörter. Also an dem Score Leben hat jetzt das Wort Job. Äh bei dem ersten hat irgendwie ins Tor 3 hier hat den Chor 17 hinterlegst einfach wie groß ein Wort dargestellt wird abhängig von dem Score dass du hast da unten irgendwie weiß ich nicht. Sehen Oder 30 Sekunden Segmente und dann weißt du für jedes Wort den Chor und fiel dann also ich sag mal die die unschöne Variante ist springt halt einfach von der Größe her. Die schöne Variante ist du machst mit CSS Animation die von Größe 3 auf Größe 17 vergrößert das kann man relativ einfach. Soso Animation von Text größer kleiner machen das machst du quasi unabhängig für jedes Wort in dem in dem Raster und siehst dann quasi immer was was hoch und runter kommen dann kannst du sogar noch die Farbe ändern. Das heißt wenn der Sport Job von kleiner groß wird. Am. So Ach so. Ähm machst irgendwie grün dann sieht man dass es Dazu kommt und wenn es viel abnehmen ist der machst irgendwie in grau oder so dann weiß man das geht irgendwie Weg dann hat man so ein bisschen da kann man ein bisschen mitspielen.
Speaker 3	17:28	Die Wörter wären aber schon vordefiniert also man würde zuerst durch den gesamten Text gehen und dann je nachdem wie lange der Text Sich 10 20 30 Wörter raussuchen und dann werden es immer diese diese gleichen Wörter über. Die ganze Timeline eigentlich.

Speaker 4	17:47	Ja das kann man sich natürlich überlegen ob man das schlauer macht man wenn Wörter nur am Anfang vorkommen hinter nicht mehr dann hast du quasi also jetzt sind wir im Detail wie macht man das aber. Jetzt Sport Job nur in in dem Bereich hier vorkommt wirklich danach nie mehr und dann gibt es ein anderes Wort was du hier hinten vorkommt wenn es nicht klar warum man eigentlich die entsprechenden Bubble freihalten sollte. Also dann könnte man auch sagen macht jetzt irgendwie diesen diesen Bubble hier für den mittleren Bereich für Job und dann wenn da hinten irgendwie Karies oder so. Dann nimmt man halt in die gleichen tabelleneintrag die gleiche Zelle für das Wort Car das ist ja nicht schlimm wenn es nicht erlaubt.
Speaker 1	18:39	Aber dann siehst du ja nicht mehr welche Wörter sie überhaupt gibt. Also wenn vielleicht willst du ja dann auch auf das Wort klicken können um 10 in welchen segmenten unten in der teimlein dass es vorkommt.
Speaker 4	18:52	Das stimmt.
Speaker 1	18:53	Ja also vielleicht hören wir uns einfach so machen dass die die Wörter wirklich extrem klein sind die nicht vorkommen oder sowas. Wahrscheinlich gegen Verein tun.
Speaker 4	19:06	Das ist jetzt irgendwie so mit der Tabelle ist das einfachste du könntest auch Weg von Word Cloud machen die sich immer dynamisch dann verschiebt oder so ähm. Ja man könnte auch sagen du zeigst immer alle Wörter an die werden halt größer und kleiner je nachdem. Die Relevanz gesehen du hast ja auch den ganzen Bildschirm zur Verfügung also. Sag mal wenn du jetzt wirklich das Full Screen machst. Kriegst du auch Locker 100 Wörter eben also du hast eine Stunde Audio tendenziell haben wir jetzt schade wenn. Wenn du dich da irgendwie auf 20 Wörter oder so beschränken müsstest.
Speaker 1	19:47	Ja ja das kann ja ein Parameter sein wieviel wird er oder sowas bestimmen oder keine Ahnung dann der.

Speaker 4	19:58	Ja das hat so verschiedene Aspekte. Das eine ist die berechnet man eigentlich diese Keywords also einfach wie welches sind die wichtigen Wörter in den Text. Da habt ihr jetzt irgendwie einfach Wörter gezählt könnte das Rad nehmen wir könntet ä FIDF nehmen das kann man einfach im backend machen. Dann kriegt jedes Wort für jedes Segment was ihr habt uns Chor. Dann ist die Frage wie visualisiert man das jetzt also eine Möglichkeit viel. Ich glaube Ich würde es nicht als Video machen weil bei Video kannst ja nichts dynamisch machen da kannst nicht reinklicken und so würd glaub ich eher dann in der webapplikation mit mit CSS teilen oder so. Also viel mehr Möglichkeiten da die interaktiv was zu machen der dann sagt du kannst reinklicken und siehst da die statements dazu meine feels Mega cool ich stehe angekommen zum eintragen für das Job ich kann da drauf klicken. Ne dann hier rechts irgendwie oder unten drunter gratis statements. Was is? Mit Highlight in resse. Und wie gesagt wenn er baut jetzt nur Prototypen also prototypisch könnte auch sagen das ganze ist jetzt nicht Gruß mit vergrößern verkleinern sondern viel ihr macht da unten wie 20 solche. Nächsten Zeit Schritte und dann rechnet das Ding hat jedes Mal aus ich meine ist auch eine Frage der Performance und so
Speaker 2	21:36	Ja Ich denke das ist machbar.
Speaker 4	21:38	Ja genau das ist es hilfreich.
Speaker 2	21:43	Also. Das wahrscheinlich auch denk ich.
Speaker 4	21:47	Was für euch wenn wir jetzt wieder sagen Wir haben die Aufnahme und lassen das abspielen? Wär das hilfreich. Würdet ihr wenn ihr einer Meeting nicht teilgenommen habt da drauf klicken um zu sehen worum gings eigentlich.
Speaker 2	22:13	Ja gut den im schlimmsten Fall hätte man nur 30 Sekunden verloren wenn es jetzt einen nicht weiterbringen würde denke ich ja aber im besten Fall sieht man etwa den zeitverlauf und weiß wo man etwas genauer nachschlagen könnte wenn man noch etwas sucht.
Speaker 3	22:28	Ja denke gerade im Vergleich zu einem ähm zu einer such Funktion wenn man jetzt einfach Job eingibt sieht man Wo ist es vorgekommen mehr nicht wenn man aber so etwas wie hier hätte. Ja würde man gerade auch noch die Wichtigkeit oder die Häufigkeit sehen also wenn der Nebenjob viel größer wird weiß man es ist wichtiger oder häufiger gebraucht. Ähm das hilft einem sich auch wenn man nach etwas speziellem sucht vielleicht zum einen und zum anderen wenn man keine Ahnung hätte sieht man hier Am Anfang war Job sehr groß in der Mitte war vaccination sehr groß am Ende kleine change dann weiß man schon okay. Es ist vielleicht zuerst um Arbeitsplätze gegangen danach um vielleicht Cockpit Impfungen am Schluss um um Klimawandel einfach Ich denke für einen groben Überblick aber auch wenn man etwas suchen würde während Diese 30 40 Sekunden sicherlich sehr hilfreich.

Speaker 4	23:30	Ja wenn ich sehr schlecht Begriffe habe ich speziell interessieren. Könnte man das ja auch noch zu lassen also das was ihr als Suchfunktion jetzt schon gehabt. Hier einfach sagen kann irgendwie. Meine Begriffe. Eine Therme soll Ich kann nicht so gut auf den Bildschirm schreiben Ich hoffe ihr könnt trotzdem raten was ich meine ähm und kann hier von Hand ein paar Griffe eingeben die dann auch in der Tabelle. Auftauchen. Wo ich dann einfach vorgeben kann ich will weiß ich Nicht 5 Oder 10 Wörter von Hand angeben mein Namen Ego Shooting und so ähm und vielleicht dass mich das gute interessiert und das Produkt Inter Schreiber? Ähm egal obs jetzt wichtiger Begriff ist oder nicht will einfach sehen wenn das Ding überhaupt vorkommen. Es ist so ein bisschen. Rede und dann weil das kann ich ja auch wenn ich irgendwie eingebe sehe ich dann quasi in der Suchfunktion in dem leider den gehabt unten drunter schauen wo das vorkommt. Aber. Man kann ja auch das irgendwie so ein bisschen kombinieren wenn das was ihr jetzt habt viel wär ja sowas gibt hier ein das Wort ähm. Aus oder so. Und im Moment zeigt er dann auch in den Slider in den in den zeitverlauf mit den strichen an wo das vorkommt. Also ich glaub da kann man dann relativ viel spielen wenn man mal so die grundfunktionalität hat. Ja. Und ich glaub ein dritter Schritt is dann wenn das mal funktional geht und wir sehen dass es nutzen hat das dann schön zu designen also das ist das was der Bahn vorhin meinte ihr seid keine Design Studenten. Also wenns mal funktioniert dann kann man sehen ob es überhaupt was nützt.
Speaker 2	25:35	Ja. Ja Ich denke wir können ja sicher bis nächste Woche einen einfachen Prototyp in diese Richtung aufbauen und dann eine erste Beurteilung machen. Was man noch verbessern könnte wie es bis jetzt aussieht genau?
Speaker 4	26:17	Was mir noch eingefallen ist es kommt ja immer wieder also in den Diskussionen immer wieder ganz viele Idee der wird ja nicht alles implementieren können gab es? Gut wenn ihr zumindest dann dokumentiert sagt Alternative den wäre noch dies und dies und dies viell vielleicht einfach so was gibts mäßig ähm. Wenn du noch da okay also Skizzen in die Arbeit rein nehmen und sagen okay Wir haben uns mal auch überlegt man könnte irgendwie weiß ich nicht Word Klaus anzeigen jetzt nicht. Diese Tabelle sondern echte workouts und dann halt von Screenshots von eurer Word Cloud und wie Mans machen könnte also Aldi den irgendwo noch so in so einer ideensammlung haben damit sie nicht verloren gehen ich denk. Da muss man dann hinterher sowieso auswählen was man dann echt macht aber das äh ist sicherlich spannend mhm. Jetzt in eurer Applikation. Ähm. Da habt ihr ja Sachen auch schon wieder rausgenommen das Öl. Richtig gesehen hab also Funktionalität hier Frauüher mal hattet. Die glaub ich wieder ausgeblendet oder.
Speaker 2	27:37	Ja beispielsweise die Keywords pro aus Sorge.
Speaker 4	27:41	Ja genau.

Speaker 2	27:42	Das hatten wir mal drin aber es hat nicht wirklich es war nicht intuitiv und teilweise auch gar keinen Sinn gemacht da haben sie rausgenommen.
Speaker 4	27:51	Ich glaub gut ich finds gut wenn ihr eure Applikation so Tabs hättet auch mit der alten Funktionalität erstmal die alle mal ausprobieren kann und dann kann man halt wirklich sagen okay das bringt nichts aber wenn jetzt die nächste Arbeit kommt. Wer ist cool wenn die Funktionalität noch da wäre.
Speaker 2	28:08	Ja das können wir machen das ist kein Problem ja? Also eine Sammlung von Tabs.
Speaker 4	28:15	Ja genau alles.
Speaker 2	28:16	Was man da genau mhm. Ja dann denke ich haben wir das meiste gesagt und wir würden bis nächste Woche diesen Vorschlag hier ähm. Aus Prototyp implementieren.
Speaker 4	28:44	Wie gespannt geht?
Speaker 2	28:49	Das gibt sicher Lösungen.
Speaker 4	28:52	Ja bist du nicht in den Ferien.
Speaker 1	28:54	Nächste Woche.
Speaker 4	28:54	Mag Ich bin in den Ferien ja. Also ich äh hab die ganze Woche frei stimmt weil du bist da du bist dann danach die Woche in den Ferien. Genau. Wechseln uns ab ja. Also ihr könnt ja mal. Was was man tatsächlich auch überlegen kann ist so im Back End wie berechnet man eigentlich die wichtigen Wörter? Starte jetzt verschiedenste Algorithmen angeschaut ja für die Keyboard Experten schon gemacht und so viel. Da könnt ihr euch auch super aufteilen dass einer mal entscheidet ich nehm verschiedene Algorithmen ähm und guck mal was gut oder schlecht funktioniert in der Praxis. Genau. Und ja. Sonst noch was.
Speaker 1	30:04	Wann wird es nächste Woche aber Ich glaube es macht wenig Sinn jetzt anschauen weil du nicht dabei bist Mark oder?
Speaker 4	30:12	Also könnte schon Ich weiß nicht wie weit in einer Woche kommt also das das könnt ihr entscheiden klappt ihr seid noch nicht so weit denkt das ist jetzt eher größere Block.
Speaker 1	30:24	Das genau. Also genau lass mich lass mich doch wissen also Ich würde sagen wir lassen uns grundsätzlich lassen wir es ausfallen also Ich denke es macht Sinn den Markt weißt aber wenn ihr das besprechen wollte mit mir dann dann könnt ihr können wir gerne machen. Mhm Ich glaube nicht Ich weiß nicht sinnvoll ist dass sie mir ist das Resultat dann schon präsentiert mit der Bahn ist Mark. Oder ja ich kann dann gebe ich einfach meinen Senf dazu und dann gehts in meine Richtung muss ich nicht.
Speaker 4	30:55	Können.
Speaker 1	30:55	Wir die Idee gehabt oder. Gib mir das und.
Speaker 4	30:59	Und dann. Ja aber also das ist ja nicht schlimm ich find wenn wenn ihr dann findet äh die defa blöd wir machen jetzt was ganz anderes oder sieht nicht aus oder es funktioniert nicht also entscheidet doch ob ihr was zu zeigen habt oder nicht.
Speaker 1	31:14	Mhm okay ja ja also Ich bin hier genau kennt mich.
Speaker 4	31:18	Ja sehr gut. Mhm gut Laptop Top alles klar ja dann sehen wir uns in 2 Wochen wieder ja genau das machen wir.

Speaker 1	31:33	Und du weißt ja nicht was nächste Woche. In 2 Wochen dann sehen wirst.
Speaker 4	31:38	Oder ja ja ich seh das schon du. Ja es ist schwierig sind ja 3 Leute sonst hätte ich gesagt wir teilen sie uns irgendwie wieder ne okay. Gut.
Speaker 1	31:59	Alles klar machen wir wollten uns noch sprechen nachher oder.
Speaker 4	32:03	Ja Minuten ich muss noch kurz rein gucken.
Speaker 1	32:09	Wenn Ich würde sagen wir machen um 11. Geht es.
Speaker 4	32:11	Für dich ja ich Ruf dich gleich an so kurz vor 11 okay gut bis dann tschüß.

A.2.6 Meeting 2021-05-07

Speaker	**Start time**	**Utterance**
Speaker 3	0:03	Also kann ich loslegen.
Speaker 2	0:07	Perfekt. Ähm Wir haben uns mit der der Word Cloud befasst. Also die Idee dass man quasi für gewisse zeitabschnitte im transkript eine Word Cloud darstellt. Und die haben wir mal 3 unterschiedlich aufwendige implementationen gemacht. Um auch ein bisschen ein Gefühl dafür zu bekommen das technisch umsetzen würde in diesem desch Framework. Und äh die erste Variante war so etwas. Das sind jetzt einfach die Wörter beziehungsweise die die nomen. In diesem transkript und Je größer sie geschrieben sind umso häufiger werden sie genannt. Mhm und das für unterschiedliche zeitintervalle. Diagramme dann einfach so durch klicken und sie dann die entsprechende Word Cloud für den jeweiligen Abschnitt. Okay das ist noch nicht super intuitiv aber es war mal unsere erste Ansatz um das zu implementieren dass man quasi durch so eine Friedens durchklicken kann. Und dann sieht welche Wörter wurden in diesem Abschnitt am häufigsten genannt. Ja Was.
Speaker 3	1:23	Ist die Skalierung da hat das Grundlage für das ist einfach nur.
Speaker 2	1:28	Diese würde man idealerweise raus blenden das haben wir in unserer letzten Version auch gemacht das sind im Prinzip nur die Koordinaten der Wörter. Okay also wo ein Wort ist halt keine Bedeutung ja okay im Moment ist das auch noch nicht so intelligent implementiert also die Koordinaten der Behörde sind zufallszahlen. In einem nächsten Schritt würde man alle Wörter mehr so. Nebeneinander platzieren mhm aber weil umso provisorisch sich anzuschauen wie dat aussieht er relativ simpel zu implementieren und die textbook.
Speaker 1	2:03	Lang das noch einstellen kann.
Speaker 2	2:06	Genau hier könnte man jetzt zum Beispiel sorgen. 10 Minuten. Ja dann würde dies Karl hier neu berechnet und dann würde man ein 10 Minuten Fragen sehen genauso wie jetzt.
Speaker 3	2:21	Okay.

Speaker 2	2:22	Das war doch etwas Flexibilität hat je nach unterschiedliche dialoglänge ja vielleicht etwas anders sind wir genau. Dann unser nächster Versuch war eine Animation zu machen. Mhm mhm mhm. Genau dass man quasi eine Playback hat welche dann die Friedens automatisch abspielt. Und hier beispielsweise das Wort people ist immer am gleichen Ort aber ihr größeres geschrieben ist umso häufiger wird es in diesem Abschnitt genannt beispielsweise hier ist es ziemlich groß dann wird es wieder kleiner. Wieder ein bisschen größer kleiner. So war sie das war mal eine Implementierung das ist in dieser Variante jetzt noch nicht so elegant weil es ziemlich viele Wörter sind die auch praktisch kaum genannt wurden das könnte man noch etwas eingrenzen und dann auch die Positionierung noch etwas aufhübschen. Mhm und was uns auch nicht so gefallen hat ist das. Die Übergänge nicht fließend sind also die textgröße ändert sich abrupt von Beispielsweise 24 auf 12 mhm und Dadurch kann man nicht so gut die erhalten genau. Aber sonst von der Performance her funktioniert das besser also ziemlich schnell und effizient zwischen den Frames switchen was hier noch etwas mehr Zeit benötigt. Und dann unsere letzte Implementierung war. Ähm wenn man quasi noch so Kreise. Mit den Wörtern ähm verbindet und jene die kaum genannt werden transparent macht und so hat man dann eine etwas einfließende Animation. So sieht man relativ schnell welche Wörter jetzt relevant sind. In einem gewissen textabschnitt. Genau beispielsweise hier China oder Jobs. Dann hier Joe beiden da könnte man diese Wörter wird man natürlich herausfiltern weil das die Namen der Teilnehmer sind. Aber Alles in allem gibt allein besseren Überblick als jetzt noch diese Version. Ja okay diese Kreise haben auch die also Möglichkeit dass sie sich quasi vergrößern und verkleinern und dann das sieht was man bei den Text nicht. Machen kann.
Speaker 3	4:55	Warum kann ich mit dem Text nicht?
Speaker 2	5:00	Ähm das ist einfach keine einstellungsmöglichkeit die Kreise sind quasi Daten also ihre Kreise deine Größe und die verändert sich quasi von Größe. Von Durchmesser 20 auf Durchmesser 24 und das kann quasi. Animiert werden durch diesen Plot aber schriftgrößen können nicht animiert werden einfach schrittweise verändert.
Speaker 4	5:25	Der Text ist eigentlich nur eine Beschriftung.
Speaker 2	5:28	Des Daten okay genau während die Kreise quasi wirklich dort sind. Also diese Animationen ist auch nicht dafür gedacht zum beispielsweise irgendwelche Graphen zu animieren mhm und Wir haben das jetzt ein bisschen umfunktionierten quasi eine Word Cloud nachzubauen. Und einigermaßen flüssige Übergänge zu machen.
Speaker 3	5:50	Jetzt geht nochmal auf das erste die erste Visualisierung also ganz kurz war das alles was wir zeigen wolltet oder habt ihr noch.

Speaker 2	5:59	Mehr bevor ich jetzt anfange Nein das war der Schwerpunkt quasi.
Speaker 3	6:05	Aha spannend also ich finde erstmals funktioniert schon so ein bisschen also kann mir vorstellen dass das was wird auch wenn es noch nicht so weit is glaub ich das ist noch nicht so dass wir jetzt so sieht aber was läuft da eigentlich den.
Speaker 2	6:19	Ganzen. Genau es ist noch in der anfangsphase.
Speaker 3	6:23	Ja. Bei der ersten Darstellung Gedacht wenn man wirklich so eine zweidimensionale Darstellung macht mhm dann könnte man dem ja auch eine Bedeutung geben.
Speaker 2	6:41	Kraus eine räumliche Bedeutung.
Speaker 3	6:42	Der Bedeutung und zwar viel was mich glaub ich interessieren würde ist so ein bisschen die Bedeutung der Wörter für das gesamte Interview. Im Vergleich zu der einzelnen Passagen hm. Sie hat jetzt zufällig platziert glaube vielleicht ist es auch übersichtlicher wenn man jedem Wort quasi seinen eigenen quadranten gibt ja so welche Tabellen mäßig jedes Wort. Eine eigene Platzierungen hat doch gar nicht das Problem mit den überlappungen und so mhm. Und wenn man dann quasi die beiden Richtungen. Interpretierbar machen würde mhm. Mhm mhm mhm. Da muss sich gut überlegen kompliziert dass wir Top läuft das verstehen aber angenommen man nimmt so die x Achse von links nach rechts. Und ordnet die Wörter die. Häufig in vielen. Paragraphen vorkommen also über das gesamte Interview vorkommen. Hier links an und die Wörter die eher spezifisch für einzelne Paragraphen sind rechts. Ja dann wird man wenn man das dann ablaufen lässt. Basi links wenig Bewegung sehen weil die halt Ring sind sozusagen die tauchen häufiger auf die gehen so ein bisschen hoch und runter in der Bedeutung und rechts sieht man dann quasi immer so die die. Mehr oder weniger spezifisch Verein Paragraphen sind mhm das sie jetzt zufällig zu verteilen wenn ich sie zufällig verteilen muss ich immer überall guck.
Speaker 2	8:29	Genau das stimmt das stimmt.
Speaker 3	8:30	Könnte so einige Millionen sein. Die zweite Dimension könnte irgendwie so die zeitachse sein weil es irgendwie so intuitiv wäre dass man die Wörter die eher am Anfang in der Paragraphen vorkommen weiter oben und eher die. Zum Ende hin vorkommen weiter unten platziert oder so mhm. Also dann wenn man das so hinkriegt wie auch immer man das rechnet dann hat man sozusagen auf der rechten Seite müsste so der der Bubble von oben nach unten so sich bewegen. Oder weniger. Rechts oben die Wörter die in den Paragraphen spezifisch sind und am Anfang eher vorkommen und dann bewegt sich so dieses ganze des plätschert so ein bisschen runter auf der rechten Seite und links die sind mal hoch. Darunter aber Wir sind eigentlich immer da.
Speaker 2	9:23	Mhm genau.

Speaker 3	9:28	Dann. Wer das vielleicht übersichtlicher weil so hab ich das Gefühl dass entweder man reduziert die Anzahl Wörter massiv mhm also so muss man halt an 20 stellen gleichzeitig.
Speaker 2	9:41	Sozusagen genau so kann man sich nicht wirklich fokussieren oder keine Bedeutung von der Aktion genau das.
Speaker 3	9:47	Stimmt? Jetzt interessiert wann wurde irgendwie über das Werksteam gesprochen dann kann man sich ja dadurch sollen das vor aber. Ich hab das Gefühl so mit der Idee das ganze als Word Cloud zu machen jetzt eher unübersichtlich mhm während der Mans auch als Tabelle macht sozusagen und dedizierte Position hat die. Schön aufgeräumt sind dann sieht man viel schneller was Groß und Klein ist und so dann macht man sich das Leben vielleicht einfacher mhm das mit den bubbles finde ich gut weil es so. Schön visuell.
Speaker 2	10:25	Ist ja. Genau. Ja das wäre auch unser nächster Schritt gewesen noch zu überlegen wie man die Positionierung intelligenter machen könnte aber dein Vorschlag ist finde ich eine gute Idee dass man so bei dir auch so eine Bedeutung gibt. Nicht weiß diese Wörter haben diese Bedeutung jener hier oben haben eine andere das ist sicherlich eine gute Idee finde.
Speaker 3	10:51	Ich glaub man kann damit so farbkodierung und Position und Größe. Genau die Informationen rüberbringen man muss aufpassen ähm Es gibt so Untersuchung dass die meisten Leute diese komplexen codierung die wunderschön ganz viele Dimensionen darstellen. Gar nicht verstehen sie die meisten User wenn du den so eine heatmap zeigt dann haben sie keine Ahnung was das heißen soll. Am geht mir oft auch so da kommt irgendjemand zeigt mir Tool und dann siehst du so Chart der irgendwie bei dimensional und hat noch verschiedene farben und die Dinger verschiedene Größen ob S da irgendwie verstanden.
Speaker 2	11:30	Also. Ja ja das.
Speaker 3	11:33	Nicht zu viel reininterpretiert aber macht aber. ja.
Speaker 1	11:45	Und von der Anzahl Platz also sollten wir das eher aufteilen oder alles am besten in ein Plot hineinpacken oder wie siehst du das von der. Von der Benutzerfreundlichkeit.
Speaker 3	11:59	Ja. Was meinst du mit aufteilen.
Speaker 1	12:02	Also neben ohman jetzt zum Beispiel den so lassen könnte und das mit der Position bei der Word Cloud implementieren oder sollte es am Schluss ein Platz sein welche alle Funktionen? Beinhaltet oder.

Speaker 3	12:19	Oh das ist jetzt. Ähm. Also Ich denke die Position der Wörter jetzt nicht zufällig zu machen sondern quasi in so ein Raster legen ist ja relativ einfach ja sich dann zu überlegen wie man das Raster auswählt ist wahrscheinlich schon eher schwieriger. Dass man also du hast jetzt da weiß ich Nicht 100 Wörter oder so das wahrscheinlich ähm ich glaub fürs ausprobieren ähm fände ich es gut wenn man die Anzahl Begriffe einstellen kann. Und den den Times die segmentierung also wie lange sind ein Segment. Das das hat er jetzt quasi schon infiziert ähm dann kann man da so ein bisschen mit ausprobieren wie viele kann man eigentlich auch gleichzeitig wahrnehmen viell. Die. Darstellung den bubbles finde ich gut. Aber das ist so ein Plot Framework oder das könnte wahrscheinlich nicht stark modifizieren.
Speaker 2	13:22	Es gibt schon Möglichkeiten beispielsweise die Positionierung das sind im Prinzip einfach Koordinaten also man könnte und so eine Tabelle darstellen indem man einfach systematische Koordinaten nimmt und diese den Wert zuweist einigermaßen. Agent.
Speaker 3	13:41	Ja also wie gesagt ich Will 30 Wörter haben dann kannst ja genau aufteilen. Machste schön glatt draußen.
Speaker 2	13:49	Und genau das das wäre kein Problem.
Speaker 3	13:52	Wahrscheinlich kein Problem. Also ich glaub dadurch jetzt übersichtlicher.
Speaker 2	13:56	Mhm das stimmt dann. Und vielleicht noch allgemein du hast letztes Mal gesagt dass wir Frauüh heurige Funktionen auch wieder rein nehmen sollen. Ja das ist der Tipp topp von unserer Seite Die Frage ist noch wenn wir so eine praktisch unterschiedlich gute oder unterschiedliche ausführliche Versionen von einer Funktion haben sollten wir.

Speaker 3	14:26	Das. Ausprobieren also ja spricht überhaupt nichts dagegen dass ja nur so ein Haus wo wir ich wird wirklich noch so eine Zeile Texte der 2 Zeilen Text darüber schreiben. Ähm vielleicht sogar eine Referenz später eure bachelorarbeit und schreibt die überlauffunktion ist da unter geschrieben und die Methode Code heißt so und so oder was auch immer so dass man die Sachen wieder findet aber. Wenn man wirklich Sachen ausprobieren will dann sind die ganzen Tabs super wirklich erweise werden dann viele musste hinter so ein bisschen anordnen das sind die guten und das sind die Prototypen die oder vielleicht kannst du ein paar trifft man ein paar Feinde oder so also so dass man sie erkennen. Mhm aber finde ich cool. Diktiert zum Ausprobieren ich glaub was man super machen könnte wenn du nochmal auf das Word Cloud gehst dass man auch da wie bei den anderen oben Noch 2 3 weitere Parameter hat nämlich die Anzahl Wörter. Und vielleicht sogar die tabellengröße. Also wenn du sagst ich hab 15 Wörter dann kannst du irgendwie eine Tabelle 3 mal 5 machen Oder 5 mal 3 oder Einmal 15 Ähm dass man einfach sagt man bietet dann verschiedene Tabellen. Aufteilung an da kann man wirklich damit spielen dann kann man gesagt okay ich Will 30 Wörter aber ich will sie mehr oder weniger untereinander haben. Oder ich will die 30 Wörter auf fl 7 mal 4 und äh. Na dann 5 was auch immer und dann bleiben Palais oder was auch immer also da so ein bisschen mitspielen kann ähm ob man eher breit oder eher hoch. Was ein spannendes will man die Wörter da drin anordnet Nächsten 2 Dimensionen das ist sicherlich ein bisschen? Er muss mal Dr.über nachdenken dann Schlaf hatte auch was die 2 Versionen sind das ist so. Die änderte jetzt die Schrift Größe. Und von dem Text.
Speaker 2	16:44	Genau die Schriftgröße ändert sich ähm bei allen Versionen und haben zusätzlich noch die bubbles die sich ein bisschen dynamischer vergrößern verkleinern.
Speaker 3	16:53	Ich nehme an die die Farben auch anpassen.
Speaker 2	16:55	Oder ja ja das ist kein Problem. Ja ja. Das ist auch so die die voll fahre.
Speaker 3	17:01	Ja genau. Und einfach so die die semantik von es ist zum Beispiel ein Board was in der Bedeutung zugenommen hat im Vergleich zum vorhergehenden wird man halt irgendwie farblich kodieren. Sagst wird grün wenn es zunehmend ist es wird rot wenn es abnehmend ist oder so der grau wenn es abnimmt also irgendwie so.
Speaker 2	17:27	Was ist der farbverläufen das war richtig cool.
Speaker 3	17:31	Ja? Ja gibt es hier. Einfach verschiedene blaue Töne die sind jetzt abhängig davon.
Speaker 2	17:40	Das ist eine ein Transparenz. Also was sie Je größer umso weniger transparente kleine umso transparenter.
Speaker 3	17:49	Ja das kann man separat Steuern.
Speaker 2	17:51	Sozusagen ja genau genau.

Speaker 3	17:54	Genau das heißt man könnte jetzt irgendwie äh quasi aufsteigende Begriffe die mehr. Jetzt neu sind. Ähm größer die anderen kleiner Tasten irgendwas mal ausprobieren ob das funktioniert oder ob das einfach dann fahren wir war wird. Könnt ihr nur Kreise machen oder könnt ihr auch rechtecke machen.
Speaker 2	18:16	Es geht auch andere Formen richtig Krieg aber Ich werde das das Wort etwas besser einräumen.
Speaker 3	18:24	Also ich wär glaub ich positive vor allem wenn man so Tabellen mäßig.
Speaker 2	18:28	Macht genau dann so richtig.
Speaker 3	18:31	So groß machen das ist halt das Wort in der entsprechenden Schriftgröße einrahmt mhm genau Ja aber cool. Habt ihr das Gefühl dass könnte funktionieren.
Speaker 2	18:48	Ja.
Speaker 3	18:52	Jetzt hast du gerade gesagt man kann das eigentlich nicht. Also angenommen ihr würdet da draus mhm. 10.000 Fans machen. Mhm mhm mhm. Also nicht nur 8 Sondern 10 1.000 dann könnte man da mehr oder weniger wie so n Film.
Speaker 2	19:13	Durchgehen oder. Ja also quasi den äh das transkript in 10.000 Einheiten splitten.
Speaker 3	19:23	Ne oder die Einheiten immer noch lassen oder quasi so ne ähm. Mit der Installation machen also quasi von einem Bild zum nächsten ähm immer wieder zwischenschritte berechnen sodass das Mut animiert ist kurz da. Im Moment springt es ja.
Speaker 2	19:44	Ah ja also ja ich weiß was du meinst hab ich auch schon überlegt. Quasi die Schritte interpolieren mhm um dann mehr also mit flüssig darzustellen. Ich denke es wäre grundsätzlich auch möglich im Hintergrund von diesem Plot ist im Prinzip nur ein ein Data Frame und in diesem Daten könnte man beispielsweise Zwischen 2 Zeilen eine interpolierte Zeile einfügen. Ja und dann würde das spät wie sie aussieht.
Speaker 3	20:16	Mit zwischen beiden messwerten und.
Speaker 2	20:18	Ja ja ja das mhm. Aber Ich denke das lohnt sich auch von unserer Seite weiter zu verfolgen da kann man auch ziemlich viel rausholen genau.
Speaker 3	20:35	Jetzt nochmal so. Kannst du relativ schnell einstellen das mal weniger Begriffe ein anzeigt oder ist das schwierig zu ändern im Moment geguckt.
Speaker 2	20:49	Ähm. Revenge schwierig ist OK ich Verlass auf das nächste Mal noch etwas intuitiver laut geht es schnell.
Speaker 3	21:08	Er ist ja bei.
Speaker 4	21:10	Den Animationen noch.
Speaker 3	21:11	Nicht ähm. Ist nochmal so Schritt zurück. Löst das das Problem was eigentlich los war. Also Krieg ich so einen schnellen Überblick über die Diskussion.

Speaker 2	21:37	Ich denke in einer aufbereitete Version wahrscheinlich schon beispielsweise nur wenn ich hier im dritten Frame das Wort vexin groß sehe dann weiß ich dass ich um Impfungen geht. Oder das Wort Text hier oder Dollars. Ja genau da wieder Text und Texas. Und wenn wir natürlich noch die stopwords entsprechendem besser filtern würde beispielsweise eben die die Namen der Sprecher die natürlich häufig genannt werden rausnimmt. Mhm ja könnte man auch so eine texteingabe machen wir hier. Wo der User dann eingeben kann okay diese Personen werden Alle 3 Sätze genannt dann nehm ich sie raus dann würd ich wieder mehr Platz haben für aufschlußreiche Begriffe wie beispielsweise kleine Tier mhm oder wild fires beispielsweise in weiland? Hurricane. Also Ich denke das ist schon zum einen Überblick bekommen während hier noch so um Impfungen geht geht es hier mehr um die Umwelt und das Klima und irgendwo dazwischen noch um die die Steuern und Infrastruktur.
Speaker 3	22:44	Genau die richtige Auswahl der Begriff ist da noch was. Ähm was man sich hier verfeinern muss mhm. Ja Ich werde jetzt auch. Aha Wir nehmen jetzt die Aufnahme von unserem Gespräch vom letzten Mal als eigentlich könnten wir das ja mal machen oder muss ja nicht funktionieren. Mhm und jemand ist nicht dabei gewesen und sie dann diese Word Cloud. Ja weil das schon fertig vom letzten Mal.
Speaker 2	23:26	Äh Nein das ist noch nicht aus transkript vorhanden okay oder beispielsweise das jetzige könnten wir quasi für Duden. So war dann würde dann mal so eine Animation sehen von unserem jetzigen Dialog. Oh warte mal sagen ob es für ihn schlüssig ist die Begriffe dir hier groß sieht.
Speaker 3	23:50	Eben hab ich eins.
Speaker 1	23:53	Ohh. Sollte alles auch auf Deutsch gehen oder.
Speaker 2	23:59	Was ja ja genau das ist sprachunabhängig diese Implementierung also nicht ganz beispielsweise Stop words müssen mit einer. Ja werde mit einer Liste abgeglichen.
Speaker 3	24:10	Also ihr könnt ihr das CSV von intersky integrieren.
Speaker 2	24:13	Oder yap ow wir also die Integration Interview Seite.
Speaker 3	24:23	Genau. Guck mal eben ob ich das transkribieren kann. lange es dauert? Wir sind leider noch nicht so dass wir viel parallel machen sprechen. Dauert das bei einer Stunde Audio natürlich auch nicht mhm. Okay das ist schon sehr lang. Okay also das könnte mal ausprobieren so ein echtes Meeting der.
Speaker 2	25:21	Hochladen er könnte das auf nächste Woche vorbereiten mit dem transkript mhm und dann könnten wir uns das anschauen können auch mehrere nehmen dieses Meeting. Von letzter Woche zum Beispiel hat mir Grad ein paar paar quasi.

Speaker 3	25:35	Sauber dann sieht man auch so ein bisschen ist in der Wirklichkeit ist sich mein für so ein eine Diskussion ist gar nicht so spannend Also vielleicht schon aber ist ja nicht unser use gestern. Aha. Fahre so ein bisschen. Ihr habt jetzt ein fertiges Friede genommen für diese Animation für die Visualisierung Wenn man das jetzt nach und nach ausbauen will. Könnte sich wahrscheinlich lohnen wenn man selber baut oder wenn man irgendwas nimmt wo man ganz ganz viel selber anpassen kann wenn das nächste was sicherlich kommt ist dass man irgendwie drauf klicken will um dann erstelle in dem Interview zu springen. Glaub ich.
Speaker 2	26:21	Ja das diskutiert. Mhm also Ich denke Es gibt hier diese hovereffekt da müsste man theoretisch auch einen Klick Effekt vielleicht imitieren können der dann eine. Ein Fenster auslöst wenn man jetzt sieht wo jetzt beispielsweise das Jo verwendet wird in diesem Abschnitt. Mhm genau genau. Aber sonst denke ich das jetzt von unserer Seite ist glaub ich alles gesagt und gezeigt mhm und auf nächste Woche würden wir mal die tabellarische Sicht implementieren mhm und auch etwas an der schriftlichen Teil der bachelorarbeit. Vorfahren.
Speaker 3	27:14	Genau. Kauf das müsste er mal anfangen Wenn man also ich finde die Sprachsteuerung mit den 2 Dimensionen finde ich spannend wenn ich zum Beispiel vorstellig mach nur. 2 Spalten ist irgendwie so das einfachste richten vorstellen kann. Ähm die eine Spalte sind Wörter die eher überall im transkript auftauchen die andere Spalte sind Wörter die nur spezifischen einzelnen segmenten sind wie auch immer man das genau berechnet. Soll ich dann überlegen. Am und die andere Dimension ist sozusagen wann tauchen die Wörter auf damit trifft das eigentlich nur die linke die rechte Spalte. Wo Wörter spezifisch für einzelne Segmente sind? Hallo. Gibt man schon mal versuchen.
Speaker 2	28:10	Also ja auf jeden Fall das werden wir uns anschauen.
Speaker 3	28:13	Ne. Ja was hat er mit dem Don noch besprochen letzte Woche.
Speaker 2	28:30	Wir haben kein Meeting durchgeführt.
Speaker 3	28:33	Okay also dann.
Speaker 2	28:34	Wartest genau genau das hat sich nicht gelohnt zu diesem Zeitpunkt schon so zu zeigen.

Speaker 3	28:40	Mhm mhm mhm. Ja ja. Also was wir schon auch mal überlegen könnte wäre so ein bisschen mit dem Framework was ihr jetzt habt ob das zukunftsträchtig ist oder nicht also ob das genügend funktionalitäTMächtigkeit bietet oder ob es einfacher ist irgendwie so ein. Alles Fresh yt Framework sozusagen zu nehmen und da einfach die Begriffe reinzuschreiben zu verändern also eine Tabelle mhm effektiver also gibt es wahrscheinlich auch Sachen. Ich geh noch nicht den Wahnsinn nutzen davon dass in Word Cloud ist. Mhm was hat sie visuell cool aus aber von informationsgehalt verbietet sie ja nicht mehr. Als äh einfach die Begriffe anzeigen. Ein bisschen muss ich das überlegen ob das was bringt oder nicht. Genau. Ja gut. Den Bericht hatte angefangen zu schreiben oder.
Speaker 2	29:50	Mhm nicht so. Nein Wir haben mal eine Vorlage parat und eine grobe Strukturierung. Aber ist doch keinen wirklich konkreten Inhalt darüber mir gedacht eben diese Woche fahren wir definitiv damit an damit wir am Schluss nicht irgendwie in Stress geraten genau.
Speaker 3	30:13	Ja das wär glaub ich gut wenn du mal so ein bisschen über eure Strukturierung reden könnten wenn sie so durchdacht habt nächstes mal spätestens overlap schon speziell weil wir nicht so viele Experimente gemacht habt ihr halt. Sachen in kommentiert und ausprobiert das quasi keinen wissenschaftlichen Nachweis dass irgendwas funktioniert oder nicht was völlig okay ist aber das macht dann auch bei der Arbeit ja ganz verschiedene Richtungen explodiert viell. Schwierig ich denk was. Es angucken müssen ist das so ein bisschen auch tiefe in der Arbeit noch an irgendeiner Stelle habt also mhm ah jetzt hat er ganz viele Sachen angeschaut das ist okay aber jetzt vielleicht mit der Word Cloud dass da auch noch wirklich so ein bisschen. Ähm überlegt dass mit dem Sohn sein wie macht man das wirklich gut jetzt nicht nur implementierungs mäßig sondern auch Konzept mäßig nochmal vielleicht Irgendwie 3 4 Beispiele ausprobieren mit verschiedenen Settings für. Extrahiert man echt die Wörter wichtig wie macht man die segmentierung viell Was ist die voll Setting abhängig von der Text länger also dass man so ein bisschen sagen kann das ist jetzt das würde euch drauf fokussiert habt also Kann 1 von den Themen sein. Aber einfach das sieht auch so in der Tiefe Zahl darstellen könnt Wie funktioniert wie macht man jetzt also. Was sind gute und schlechte praktisch sozusagen? Ich hab jetzt Noch 4 Wochen da könnte relativ viel noch jetzt machen um ein tieferes Verständnis.
Speaker 2	31:46	Zu erlangen also genau.
Speaker 3	31:48	Ja Nächste Woche ist kein Meeting glaub ich oder. Ist doch. Alter an der Donnerstag ist Feiertag. Wie ist das am Freitag seid ihr da wollten wir was machen oder mit dir sowieso frei und brückentag?
Speaker 2	32:17	Echt glaube wären hier oder ja.
Speaker 4	32:20	Sicher auch.
Speaker 2	32:21	Ja Ich denke sehr gut wenn er Mieten hätten.

Speaker 3	32:24	Wir genau muss nachgucken also.
Speaker 4	32:27	Wenn es für dich passt.
Speaker 3	32:28	Ist dann wahrscheinlich im Auto. Tessin verschaut am 2. Juni an aber zur Not machen wir dann im Zug oder im Auto nicht vormittags Auto. Oder wir machen es am Mittwoch.
Speaker 2	32:47	Mittwoch.
Speaker 4	32:48	Wäre es für mich schwierig aber.
Speaker 3	32:51	Ja.
Speaker 4	32:52	Am Donnerstag wäre eigentlich jetzt auch wieder eine Option.
Speaker 3	32:55	Over gefeiert hab. Ja jetzt. Ähm. Ah Nein Quatsch sorry Nein das ist gar nicht ich äh. Ist eine Woche später im sag ich am Freitag die Uni anschauen genauso Quatsch Nein nächste Woche Freitag Freitag ist gar kein Problem der brückentag ja ja.
Speaker 2	33:22	Ja. Okay. Gut Tip Top.
Speaker 3	33:29	Sehen wir uns nächste Woche ja ja. Bis dann ciao bis wochenend.

A.3 Punctuation and stopwords

The punctuation and stopwords in English and German are listed below.

A.3.1 Python library `string` punctuation list

! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [\ ] ^ _ ‘ { | } ~

A.3.2 Python library `nltk` English stopword list

i, me, my, myself, we, our, ours, ourselves, you, you're, you've, you'll, you'd, your, yours, yourself, yourselves, he, him, his, himself, she, she's, her, hers, herself, it, it's, its, itself, they, them, their, theirs, themselves, what, which, who, whom, this, that, that'll, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just, don, don't, should, should've, now, d, ll, m, o, re, ve, y, ain, aren, aren't, couldn, couldn't, didn, didn't, doesn, doesn't, hadn, hadn't, hasn, hasn't, haven, haven't, isn, isn't, ma, mightn, mightn't, mustn, mustn't, needn, needn't, shan, shan't, shouldn, shouldn't, wasn, wasn't, weren, weren't, won, won't, wouldn, wouldn't

A.3.3 Python library `nltk` German stopword list

aber, alle, allem, allen, aller, alles, als, also, am, an, ander, andere, anderem, anderen, anderer, anderes, anderm, andern, anderr, anders, auch, auf, aus, bei, bin, bis, bist, da, damit, dann, der, den, des, dem, die, das, dass, daß, derselbe, derselben, denselben, desselben, demselben, dieselbe, dieselben, dasselbe, dazu, dein, deine, deinem, deinen, deiner, deines, denn, derer, dessen, dich, dir, du, dies, diese, diesem, diesen, dieser, dieses, doch, dort, durch, ein, eine, einem, einen, einer, eines, einig, einige, einigem, einigen, einiger, einiges, einmal, er, ihn, ihm, es, etwas, euer, eure, eurem, euren, eurer, eures, für, gegen, gewesen, hab, habe, haben, hat, hatte, hatten, hier, hin, hinter, ich, mich, mir, ihr, ihre, ihrem, ihren, ihrer, ihres, euch, im, in, indem, ins, ist, jede, jedem, jeden, jeder, jedes, jene, jenem, jenen, jener, jenes, jetzt, kann, kein, keine, keinem,

keinen, keiner, keines, können, könnte, machen, man, manche, manchem, manchen, mancher, manches, mein, meine, meinem, meinen, meiner, meines, mit, muss, musste, nach, nicht, nichts, noch, nun, nur, ob, oder, ohne, sehr, sein, seine, seinem, seinen, seiner, seines, selbst, sich, sie, ihnen, sind, so, solche, solchem, solchen, solcher, solches, soll, sollte, sondern, sonst, über, um, und, uns, unsere, unserem, unseren, unser, unseres, unter, viel, vom, von, vor, während, war, waren, warst, was, weg, weil, weiter, welche, welchem, welchen, welcher, welches, wenn, werde, werden, wie, wieder, will, wir, wird, wirst, wo, wollen, wollte, würde, würden, zu, zum, zur, zwar, zwischen

A.3.4 Python library `yake` English stopword list

dr, dra, mr, ms, a, a's, able, about, above, according, accordingly, across, actually, after, afterwards, again, against, ain't, all, allow, allows, almost, alone, along, already, also, although, always, am, among, amongst, an, and, another, any, anybody, anyhow, anyone, anything, anyway, anyways, anywhere, apart, appear, appreciate, appropriate, are, aren't, around, as, aside, ask, asking, associated, at, available, away, awfully, b, be, became, because, become, becomes, becoming, been, before, beforehand, behind, being, believe, below, beside, besides, best, better, between, beyond, both, brief, but, by, c, c'mon, c's, came, can, can't, cannot, cant, cause, causes, certain, certainly, changes, clearly, co, com, come, comes, concerning, consequently, consider, considering, contain, containing, contains, corresponding, could, couldn't, course, currently, d, definitely, described, despite, did, didn't, different, do, does, doesn't, doing, don't, done, down, downwards, during, e, each, edu, eg, eight, either, else, elsewhere, enough, entirely, especially, et, etc, even, ever, every, everybody, everyone, everything, everywhere, ex, exactly, example, except, f, far, few, fifth, first, five, followed, following, follows, for, former, formerly, forth, four, from, further, furthermore, g, get, gets, getting, given, gives, go, goes, going, gone, got, gotten, greetings, h, had, hadn't, happens, hardly, has, hasn't, have, haven't, having, he, he's, hello, help, hence, her, here, here's, hereafter, hereby, herein, hereupon, hers, herself, hi, him, himself, his, hither, hopefully, how, howbeit, however, i, i'd, i'll, i'm, i've, ie, if, ignored, immediate, in, inasmuch, inc, indeed, indicate, indicated, indicates, inner, insofar, instead, into, inward, is, isn't, it, it'd, it'll, it's, its, itself, j, just, k, keep, keeps, kept, know, knows, known, l, last, lately, later, latter, latterly, least, less, lest, let, let's, like, liked, likely, little, look, looking, looks, ltd, m, mainly, many, may, maybe, me, mean, meanwhile, merely, might, more, moreover, most, mostly, much, must, my, myself, n, name, namely, nd, near, nearly, necessary, need, needs, neither, never, nevertheless, new, next, nine, no, nobody, non, none, noone, nor, normally, not, nothing, novel, now, nowhere, o, obviously, of, off, often, oh, ok, okay, old, on, once, one, ones, only, onto, or, other, others, otherwise, ought, our, ours, ourselves, out, outside, over, overall, own, p, particular, particularly, per, perhaps, placed, please, plus, possible, presumably, probably, provides, q, que, quite, qv, r, rather, rd, re, really, reasonably, regarding, regardless, regards, relatively, respectively, right, s, said, same, saw, say, saying, says, second, secondly, see, seeing, seem, seemed, seeming, seems, seen, self, selves, sensible, sent, serious, seriously, seven, several, shall, she, should, shouldn't, since, six, so, some, somebody, somehow, someone, something, sometime, sometimes, somewhat, somewhere, soon, sorry, specified, specify, specifying, still, sub, such, sup, sure, t, t's, take, taken, tell, tends, th, than, thank, thanks, thanx, that, that's, thats, the, their, theirs, them, themselves, then, thence, there, there's, thereafter, thereby, therefore, therein, theres, thereupon, these, they, they'd, they'll, they're, they've, think, third, this, thorough, thoroughly, those, though, three, through, throughout, thru, thus, to, together, too, took, toward, towards, tried, tries, truly, try, trying, twice, two, u, un, under, unfortunately, unless, unlikely, until, unto, up, upon, us, use, used, useful, uses, using, usually, uucp, v, value, various, very, via, viz, vs, w, want, wants, was, wasn't, way, we, we'd, we'll, we're, we've, welcome, well, went, were, weren't, what, what's, whatever, when, whence, whenever, where, where's, whereafter, whereas, whereby, wherein, whereupon, wherever, whether, which, while, whither, who, who's, whoever, whole, whom, whose, why, will, willing, wish, with, within, without, won't,

wonder, would, would, wouldn't, x, y, yes, yet, you, you'd, you'll, you're, you've, your, yours, yourself, yourselves, z, zero

A.3.5 Python library `yake` German stopword list

a, ab, aber, aber, ach, acht, achte, achten, achter, achtes, ag, alle, allein, allem, allen, aller, allerdings, alles, allgemeinen, als, als, also, am, an, andere, anderen, andern, anders, au, auch, auch, auf, aus, ausser, außer, ausserdem, außerdem, b, bald, bei, beide, beiden, beim, beispiel, bekannt, bereits, besonders, besser, besten, bin, bis, bisher, bist, c, d, da, dabei, dadurch, dafür, dagegen, daher, dahin, dahinter, damals, damit, danach, daneben, dank, dann, daran, darauf, daraus, darf, darfst, darin, darüber, darum, darunter, das, das, dasein, daselbst, dass, daß, dasselbe, davon, davor, dazu, dazwischen, dein, deine, deinem, deiner, dem, dementsprechend, demgegenüber, demgemäss, demgemäß, demselben, demzufolge, den, denen, denn, denn, denselben, der, deren, derjenige, derjenigen, dermassen, dermaßen, derselbe, derselben, des, deshalb, desselben, dessen, deswegen, d.h, dich, die, diejenige, diejenigen, dies, diese, dieselbe, dieselben, diesem, diesen, dieser, dieses, dir, doch, dort, drei, drin, dritte, dritten, dritter, drittes, du, durch, durchaus, dürfen, dürft, durfte, durften, e, eben, ebenso, ehrlich, ei, ei,, ei,, eigen, eigene, eigenen, eigener, eigenes, ein, einander, eine, einem, einen, einer, eines, einige, einigen, einiger, einiges, einmal, einmal, eins, elf, en, ende, endlich, entweder, entweder, er, Ernst, erst, erste, ersten, erster, erstes, es, etwa, etwas, euch, f, früher, fünf, fünfte, fünften, fünfter, fünftes, für, g, gab, ganz, ganze, ganzen, ganzer, ganzes, gar, gedurft, gegen, gegenüber, gehabt, gehen, geht, gekannt, gekonnt, gemacht, gemocht, gemusst, genug, gerade, gern, gesagt, gesagt, geschweige, gewesen, gewollt, geworden, gibt, ging, gleich, gott, gross, groß, grosse, große, grossen, großen, grosser, großer, grosses, großes, gut, gute, guter, gutes, h, habe, haben, habt, hast, hat, hatte, hätte, hatten, hätten, heisst, her, heute, hier, hin, hinter, hoch, i, ich, ihm, ihn, ihnen, ihr, ihre, ihrem, ihren, ihrer, ihres, im, im, immer, in, in, indem, infolgedessen, ins, irgend, ist, j, ja, ja, jahr, jahre, jahren, je, jede, jedem, jeden, jeder, jedermann, jedermanns, jedoch, jemand, jemandem, jemanden, jene, jenem, jenen, jener, jenes, jetzt, k, kam, kann, kannst, kaum, kein, keine, keinem, keinen, keiner, kleine, kleinen, kleiner, kleines, kommen, kommt, können, könnt, konnte, könnte, konnten, kurz, l, lang, lange, lange, leicht, leide, lieber, los, m, machen, macht, machte, mag, magst, mahn, man, manche, manchem, manchen, mancher, manches, mann, mehr, mein, meine, meinem, meinen, meiner, meines, mensch, menschen, mich, mir, mit, mittel, mochte, möchte, mochten, mögen, möglich, mögt, morgen, muss, muß, müssen, musst, müsst, musste, mussten, n, na, nach, nachdem, nahm, natürlich, neben, nein, neue, neuen, neun, neunte, neunten, neunter, neuntes, nicht, nicht, nichts, nie, niemand, niemandem, niemanden, noch, nun, nun, nur, o, ob, ob, oben, oder, oder, offen, oft, oft, ohne, Ordnung, p, q, r, recht, rechte, rechten, rechter, rechtes, richtig, rund, s, sa, sache, sagt, sagte, sah, satt, schlecht, Schluss, schon, sechs, sechste, sechsten, sechster, sechstes, sehr, sei, sei, seid, seien, sein, seine, seinem, seinen, seiner, seines, seit, seitdem, selbst, selbst, sich, sie, sieben, siebente, siebenten, siebenter, siebentes, sind, so, solang, solche, solchem, solchen, solcher, solches, soll, sollen, sollte, sollten, sondern, sonst, sowie, später, statt, t, tag, tage, tagen, tat, teil, tel, tritt, trotzdem, tun, u, über, überhaupt, übrigens, uhr, um, und, und?, uns, unser, unsere, unserer, unter, v, vergangenen, viel, viele, vielem, vielen, vielleicht, vier, vierte, vierten, vierter, viertes, vom, von, vor, w, wahr?, während, währenddem, währenddessen, wann, war, wäre, waren, wart, warum, was, wegen, weil, weit, weiter, weitere, weiteren, weiteres, welche, welchem, welchen, welcher, welches, wem, wen, wenig, wenig, wenige, weniger, weniges, wenigstens, wenn, wenn, wer, werde, werden, werdet, wessen, wie, wie, wieder, will, willst, wir, wird, wirklich, wirst, wo, wohl, wollen, wollt, wollte, wollten, worden, wurde, würde, wurden, würden, x, y, z, z.b, zehn, zehnte, zehnten, zehnter, zehntes, zeit, zu, zuerst, zugleich, zum, zum, zunächst, zur, zurück, zusammen, zwanzig, zwar, zwar, zwei, zweite, zweiten, zweiter, zweites, zwischen, zwölf