UTRECHT UNIVERSITY

Department of Information and Computing Science

Human-Computer Interaction - Master Thesis

An Overseeable Overview:

An Exploratory Study of Automated Timeline Generation within the Open Government Act (Woo)

Daily supervisor:
Max Muller, MSc

First examiner:
Dr. ir. Eelco Herder

Second examiner:
Dr. Christof van Nimwegen

Candidate:
Roy Schoonwater
(6505589)

In cooperation with:
Dutch National Archive

July 1, 2024

Abstract

Gaining oversight from a large dataset is a challenging task. Especially within the Woo process where documents have to be selected without looking into them. This thesis investigates the use of timeline visualisation software within the Dutch Woo request process, aiming to increase its effectiveness by organising the data and providing context for documents. The study involved first the mapping of the distinct parts of the Woo process and its stakeholders. Using this process map, two methods are established that could enhance the process of request specification and document selection. The process specification process focuses on finding out what information the requester is looking for through the use of topic extraction. Document selection focuses on selecting documents by placing them in context within timelines. Semi-structured interviews were conducted with 10 participants. The participants were presented with the proposed methods and insights were gathered through the use of a qualitative analysis. The analysis showed positive results on how the Woo process could benefit from the implementation of the proposed methods. Participants stated that the solutions provide an overview and structure, and allow a conversation topic by providing context. However, they emphasised that these solutions should only be used as tools and not as leading. Therefore, the structured methods could be implemented in software to effectively, efficiently and transparently improve the request specification and facilitate the process of selecting relevant documents.

Contents

1. Introduction

Building trust is often a challenging task. Over the past three years, trust in the Dutch government has seriously decreased, highlighting the complexity of this pursuit. [1]. One of the ways to increase trust is by being transparent and therefore, the government has accepted the Dutch Open Government Act (Wet Open Overheid - Woo) [2]. This law mandates the government to publish information when requested. The purpose of this law is to be transparent as a government to its citizens. The law has had its challenges over the years, and the biggest challenge is time. In 2022, the average processing time for requests was 162 days, significantly exceeding the legal maximum of 28 days, with an allowable extension of up to 14 days [3]. This is caused by problems in information management, including difficulty in establishing a common understanding of what information the Woo requester wants, and what the government has. The number of documents can range from a hundred to thousands of documents. It is therefore necessary to filter out non-essential information.

The need for information management has progressively increased within organisations [4], [5]. With current technologies, it has become easier to grasp vast amounts of data. This is necessary to stay up-to-date with increasing technological norms [6]. Although these technologies have considerably helped companies with their daily tasks, there is still a challenge that companies face: big data [7]. A variety of solutions to handle big data have been introduced to the market such as Amazon Web Services, Hadoop or MapReduce [8], [9]. These solutions implement efficient algorithms to sort vast amounts of data for a complete and understandable visualisation. In the field of data visualisation, much research has been done on what visualisation works best in certain situations [10], [11]. One of the ways to visualise large sets of timestamped data is through timelines.

Representing event-related data on a timeline is an effective method when visualising a story [5]. It provides a chronological display of specific events as they occurred, offering insight into the sequence of events. This can be applied in a variety of scenarios. For instance, during elections to see what possible factors caused an increase or decrease in the polls [12]. Similarly, within an organisation, it is important to determine when things took a turn and who is responsible [13], [14]. Furthermore, timelines can vary a lot on what is shown. Not only events but also multiple parallel timelines, interchanged messages, and relationships can be shown [15], [16]. Therefore, timelines can provide an overview of large datasets in a comprehensible way by visualising data with context.

Therefore, timelines might be useful in certain government processes, like the Woo, as they help information management [17]. One specific example where timelines were required is the Dutch Childcare Benefits Scandal Crisis [18]. The scandal implicated around 36.000 parents who were mistakenly suspected of fraud, as a result of biases in the Dutch information systems related to benefits provision [19]. Consequently, these parents incurred substantial debt and faced denials of opportunities such as employment and housing due to their unjust conviction for fraud. Therefore, the Dutch Parliamentary Inquiry Commission requested a complete timeline of events to see where things went wrong and who was responsible for these mistakes. The timeline was manually drawn up. It was used to identify the ones responsible, and to see what processes within the organisation should be improved.

Creating a timeline manually is quite time-consuming and thorough analysis is required [15], [20]–[22]. However, timelines can be incredibly helpful in certain situations. Within Woo, timelines can give the requester a perspective on available information and provide extra context to the documents. Furthermore, it allows the Woo requester to pinpoint the required information. Hence, timelines might be beneficial to solve parts of the problems in information management. However, to reduce the time and effort it takes to draw up timelines, automated timeline software could offer the solution.

This research will investigate that solution and therefore, the following research question has been drawn up:

> ***RQ:*** *How can the implementation of timeline generation software increase the understanding of document requests and improve the effectiveness of the Woo request process?*

In the implementation, the primary focus will be on the methodology and GUI, not on the algorithms of the automatic timeline generator. To answer this research question, the whole process first needs to be mapped out. This provides a clear understanding of the process itself and what stakeholders are involved. Therefore, the following subquestion is drawn up:

> ***S1:*** *What stakeholders of the Woo request process can benefit from a timeline visualising software to facilitate the speed and clarity of information management?*

By investigating the first subquestion within a preliminary study, the stakeholders and parts of the Woo process are found. One of the parts is request specification, which requires a structured way of finding topics within a dataset. This gives an overview of available information and therefore what can be requested. This leads to the following subquestion:

> ***S2:*** *To what extent can topics be identified to help define the required information and perspective(s) within the Woo-request process?*

This subquestion will be answered with the help of interviews to elicit if and how Woo can benefit from a structural and automated way of finding the different topics in government files. Another part of the process that is identified is the selection of relevant documents. Therefore, a timeline visualisation method will be used. To investigate the requirements of this software, general design principles found in the literature research will be adjusted to the timeline visualisation software and tested. Therefore, the following sub-question is drawn up:

> ***S3:*** *To what extent do general design features of software apply to a timeline generation software that can be used within the Woo-request process?*

This research contributes to both societal and scientific domains. Firstly, the research helps the information management of the Woo process by proposing innovative methods that enhance mutual understanding and aid in understanding how a timeline visualisation software can aid the Woo process. Secondly, the Woo process will be fully mapped. This process can be used for other research or used by the government to remove any inefficiencies within the Woo process. Third, this study contributes to the academic literature by testing the Straussian Grounded Theory as a methodological framework for conceptualizing large datasets during timeline creation. Lastly, it will benefit the research field by mapping what general design principles are beneficial in the case of creating timelines. As a result, a better understanding of the application of general design principles is obtained. This understanding can be used to test other applications of such software not only within the government, but also in other sectors.

Chapter 2, will look at current strategies applied to create a comprehensive overview with a timeline. Therefore, current procedures that are applied, different perspectives within timelines, multiple fields of application, and multiple aspects of timeline visualisation software will be discussed. Chapter 3 will begin by mapping the Woo process to gain an understanding of its components and involved stakeholders. Then participants, materials and the procedure will be discussed in the main research. The main research concerns executing interviews to gather views on topic extraction and requirements for timeline visualisation software. Chapter 4 will discuss the results of these experiments. Chapter 5 will discuss these findings as well as provide limitations, future work and a conclusion. Following that, Chapter 6 will provide a recommendation using the findings. That chapter will give an initial illustration of possible useful software. Ultimately, Chapter 7 will present an epilogue that reflects on the process.

2. Related Work

Now that the potential benefits of creating a timeline have been established, a further in-depth look needs to be taken at the current way of creating timelines, what kind of timeline algorithms exist, and how these timelines should be customisable for the user. Therefore, the initial focus is on examining the current method of creating timelines together with different perspectives, fields of applications, and challenges & limitations. Subsequently, different algorithms, features, customisable aspects, quality assessments and instances involved in automatically generating timelines will be looked into. This literature overview is presented to facilitate a full understanding of timelines from different perspectives.

2.1 Timeline Creation

When retrieving information about a series of interrelated events, it is of utmost importance to get a clear, but more importantly, full overview of the situation. There are multiple ways of gaining the overview but one of the most efficient ways is to set up a timeline [23]. Within the government, timelines are already used for parliamentary inquiries and occasionally in the Woo [15], [24]. To get this comprehensive overview, complete information on the situation, and therefore, large datasets are required. Because of the need for the full extent of the information, the process of creating a timeline is an elaborate and time-intensive task [15], [20], [22]. This section will begin by outlining the process of creating a timeline during parliamentary inquiries. Subsequently, it will discuss various perspectives on timelines and explore different fields of application. Finally, it will address the challenges and limitations currently encountered in the timeline creation process.

2.1.1 Procedure

Making a timeline involves a sequence of steps. People from different sectors might do it a bit differently but there will be an overlap in creating the timeline. Accordingly, the Process Delivery Diagram, created by Muller et al., shows such a process of creating a timeline [15]. This diagram discusses multiple steps that are taken when making a timeline. These steps have been elicited by interviews with governmental employees who make informative timelines on requests.

Step 1 - Data Collection
The process of data collection is divided into two steps:

1. Defining Research Question
2. Retrieving Documents

In the process of creating a timeline, one should first *define a research question*. This research question helps to define a perspective of the timeline. Doing this before further steps gives the timeline a scope and direction, which is crucial for getting an objective overview of the situation [25]–[27]. Generating a timeline with a holistic perspective proves to be a challenging task. Maintaining objectivity is difficult, as events may be described differently by distinct individuals. Furthermore, objectivity is not always required as one might want to inspect a sequence of events from a certain perspective to gain a better understanding. Creating a scope beforehand might help to reduce a potential bias.

Subsequently, *documents need to be retrieved* that fall within the chosen direction and scope. The goal of this step is to gather a considerable amount of documents for a complete overview of the situation. After this set of documents is gathered, the person continues to the second step: document analysis.

Step 2 - Document Analysis
Document analysis reviews the documents and determines if and how they fit within the timeline. Some documents might be more detailed or important than others. This step consists of three substeps:

1. Review Documents
2. Construct Chronology
3. Annotate Chronology

For the first step, *documents need to be reviewed* to decide what concepts, actors and events each document contains. Dividing the documents into those three categories will give a better overview of the documents, and the timeline itself.

After reviewing the documents, they need to be placed in the right order. Therefore a *chronology needs to be constructed*. During this process, the initial research question needs to be taken into account to see whether all the information within the scope is gathered. When the chronology shows information gaps, extra documents can be added.

After constructing a chronology of the document set, the *chronology needs to be annotated* [28], [29]. This is the substep where the documents are explained through events and a good summary is given of each important concept such as events, and actors. Finally, the last step needs to be executed: the visualisation.

Step 3 - Visualisation

The visualisation of the timeline consists of two substeps:

1. Selecting the Topic
2. Situationalising TimeFlow

Selecting the topic is about selecting a perspective that the timeline shows. This selection is guided by the objectives of the timeline such as gaining a deeper understanding of a subject or informing someone in particular [15], [30]. The perspective chosen for the timeline dictates which events are included, shaping its structure and narrative. [31]. This step is essential for tailoring the timeline to specific needs or perspectives.

In the following step *'Situationalise Time Flow'* each event is added based on the topics and perspective that is determined. Therefore, every event needs to be checked whether it fits topics and perspectives and how it fits in [32]. After that step is completed, the timeline is finalised.

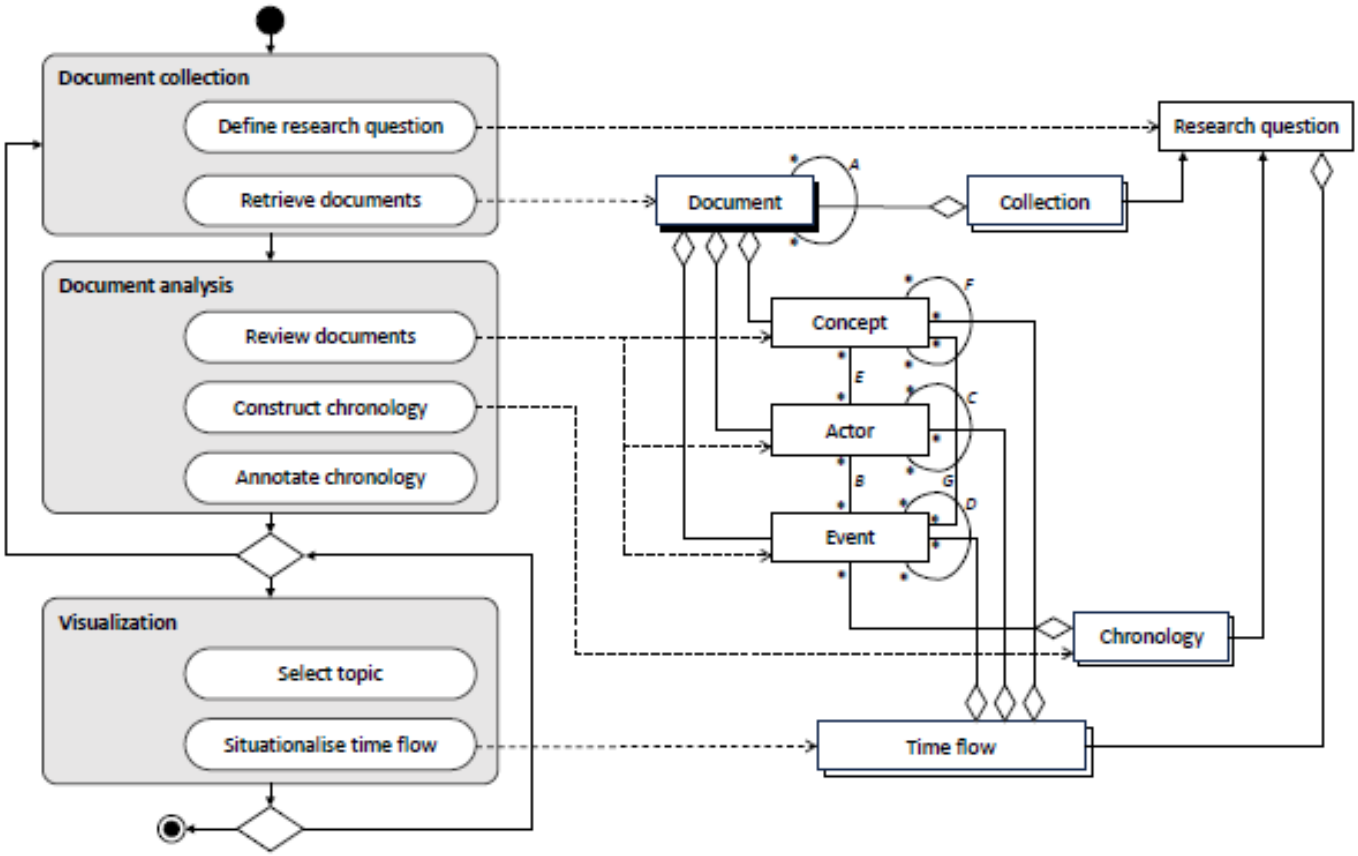


Figure 2.1: Process of creating a timeline [15]

2.1.2 Perspectives

Creating timelines involves determining the important information for a comprehensive overview and deciding how to visualise that information effectively. One should not only worry about the completeness of the timelines, but also about the ability to quickly gain an understanding of the story from a certain perspective. Therefore it is important to understand the purpose behind its creation. This gives an insight into what the visualisation should look like. An important factor that influences different aspects of the perspective is the narrative [33]. Narratives are explained differently per person. These differences are influenced by a variety of factors as will be discussed [34]. These factors and how they can change the perspective of a timeline, need to be discussed. Furthermore, these perspectives can provide insight into how timelines can differ. This research will later discuss whether comparing these perspectives is desirable.

Historical Context

One of the factors that influence a narrative is the period in which the story is told [35], [36]. Stories from different periods often focus on different aspects of the story. This stems from the fact that people from different eras handled things differently and thought in a different manner compared to people from the present. Certain events might stay unclear to people until they understand other aspects of history [37], [38]. Therefore these narratives are often in need of extra background information to understand how things evolved in different periods.

Cultural Perspective

People from various cultures may have different interpretations of the same events, leading to varying timelines [39], [40]. For instance, Western and Eastern narratives of World War II may highlight different events. Cultural viewpoints not only lead to different storytelling styles, but the culture in which a story is narrated also influences how it is told. [41]. When people see differences within a culture, they try to adjust the stories that are being told in order to make them more appropriate or more suited to that culture. People also change stories to ensure that people from different cultures understand each other better [42]. This is also linked to the fact that cultural differences in thinking or doing affect the way that events happen.

Purpose

Another factor that influences a narrative is the purpose of conveying it [43]–[45]. The purpose influences the content and focus of a story or timeline. The impact of an informative or persuasive narrative depends on both the quantity and the specific nature of the information it contains. An example of this would be when someone is showing a timeline within a courtroom and tries to persuade the jury, it will show a different timeline, with different events or descriptions of events than the opposite party [46]. If the purpose of one party would be to inform the jury, a more nuanced and probably complex overview would be given. Another purpose might be to provide a general overview of the situation, resulting in a shorter narrative with less detailed information, as it aims to convey a broad perspective.

The general purpose might change the perspective of a timeline, but you can also look at more specific purposes like blame attribution [47]. One of these examples that has been mentioned earlier is the Dutch Childcare Benefits Scandal Crisis. During this affair, biases within the benefits system resulted in mistakenly suspecting people of fraud. [48]. This resulted in a sequence of negative consequences for the citizens who were the victims of the scandal. Since this was such a complicated matter, the government asked for a timeline to put things more into perspective and to see what information various organisations knew at distinct points in time [49].

Audience

Not only does the purpose influence the method through which a timeline is made but also the audience [50], [51]. The audience can affect the complexity of a narrative. Returning the the example of the courtroom, narratives will probably be based on the things that specifically affect the jury [46]. Certain specific events that occurred might be brought up within a timeline. These events might not have a big influence on the process itself but might be important and relate directly the the audience. Furthermore, conveying a narrative to an academic audience might influence the formality of the language and complexity of information [52].

2.1.3 Fields of application

Having determined the varying perspectives in timeline narratives, it is now necessary to identify the various fields where timeline creation is or can be applied. That shows the importance and adoption of timelines in current society. Additionally, these fields can provide insight into potential applications of automated timeline generation both within and beyond governmental contexts. Given the extensive range of application fields, this section will organise them into different categories.

Academic and Educational

The concept of timelines has been used for about 250 years in a wide variety of applications [53]. One of the primary applications of timelines is in academia and more specifically in the field of history. Learning about history is learning a series of events where the importance lies in knowing why events happened and what consequences they have had [54]. Therefore, timelines are essential when learning the sequence of events and its reasoning. This appears to be not only important for history but also for science [55]. Trying to learn complex subjects without having a solid foundation to build upon would amount to reinventing the wheel.

Business

Within the business world, timelines are used for a variety of reasons. They do not only serve as tools for planning but also for analysis and execution [56]. Therefore they create a sense of trust within employees, as they have a plan that they can follow. This results in a more robust execution of tasks as employees do not have to focus on what to do but rather how to do it [57]. Furthermore, timelines can give a broad overview of why and where a certain turning point is. Given a timeline of a company's activities with a graph of its sales, can stimulate finding patterns in what activities resulted in revenue [13], [58]. It can help project teams, for instance, to pinpoint the specific actions or decisions that led to unfavourable outcomes.

Medical history

The medical field also makes daily use of timelines. It is important to look at events concerning a patient to see past surgeries, diseases or other conditions [59], [60]. This helps physicians make better decisions about future steps. It also increases efficiency, reduces errors and improves decision-making [61]–[64]. This is done through the creation of visually appealing and well-structured timelines which helps the cognitive processes of physicians to act accurately. Additionally, a piece of software named *TimeLine*, which utilises deep learning models on patient timelines, aids in making accurate patient diagnoses [65]. Therefore, doctors, nurses and other medical personnel use timelines regularly as a visualisation of patient history.

Event Planning and Management

In the field of event planning and management, timelines are also widely used. When planning a large event, multiple parties are involved in the process. Ensuring that all partners are aware of the event sequence and how tasks are interconnected will help guarantee successful task completion, prepare for any unexpected events and reschedule when necessary [66]–[68]. For instance, Gantt Charts are used for this purpose [69]. Therefore, large organisations like wedding planners, festival creators or construction sites heavily depend on timelines. Timelines can also be used to discuss certain events retrospectively, to address communication gaps and to avoid future mistakes [70]. Not only do timelines help the management, but also the visitors of events [71]. A program called *TIME* gathers social media data to create timelines to help visitors plan their trips to events.

Personal Geneology

Personal or professionally created genealogies are another sector in which timelines are used. The Central Bureau of Geneologie in the Netherlands creates genealogical timelines. Moreover, individuals can also create them on their own. These genealogical timelines allow people to better trace their family history and possibly understand the reasons behind the decisions made by past family members [72]. There is existing software that can automatically generate a genealogy along with additional information using digital documents. Software like *ChronoTape* and *GeneaQuilts* have been developed, which automatically generate a geological timeline [73], [74]. Therefore, people who are interested in their genealogy can use the timelines to gain insight into their family history.

Entertainment

When creating a story for a movie, book or comic, it is important to create an overview of events that will happen [75]. The webcomic XKCD, created by Randall Munroe, has manually created timelines of large movies to demonstrate the complexity of the movies in an aesthetically pleasing and compact way. Other tools like *CARDINAL* [76] can help create timelines for movie scripts. It is important to create timelines for comics and movies since it puts things into perspective. You can make sure that everything falls into

place and there are no mistakes within the timeline. Specific examples of timelines used for entertainment are live TV shows, cinema movies or large book series.

Now that the fields of application have been discussed, a clearer view has been gained of the usage of timelines. However, manually creating timelines presents its own set of challenges and limitations.

2.1.4 Challenges and Limitations

While establishing timelines, various challenges arise in the process. These challenges must be taken into account when designing software to automatically generate timelines. Therefore Table 2.1 shows a list of different challenges that are faced when creating a timeline. These challenges are taken into account during this research to see how automated timelines can reduce or resolve these challenges.

Table 2.1: List of challenges faced when making a timeline manually.

CHALLENGES AND LIMITATIONS	SOURCE(S)
Reliability of TimeLine	[15], [20]
Relevant Event Extraction	[21]
Complex Visualisation	[15], [43]
Biased Documents	[15]
Subjectivity of Researchers	[15], [40], [77]
Understanding Relations between Elements	[15], [21]
Thorough Analysis Required	[15], [20], [21]
Time-Consuming	[15], [20], [22]

As can be seen in Table 2.1, there are multiple challenges that one might face when creating a timeline.

Reliability of TimeLine

The reliability of the timeline concerns whether the timeline shows the complete overview of what it is meant to show. There are a lot of factors that influence the completeness of the timeline, which, in turn, challenges the reliability.

Relevant Event Extraction

The second challenge is related to the reliability of the timeline. Selecting relevant events for a timeline is a challenging task as it requires determining which events are important and within scope. Therefore careful consideration is required during this process.

Complex Visualisation

Creating a clear visualisation is a big challenge when making a timeline. Often there are a lot of documents, events, actors, and other variables that influence the size and complexity of the timeline. Therefore, connections between these variables need to be carefully visualised and taken into account.

Biased Documents and Subjectivity Researchers

These challenges concern the possibilities of creating a biased timeline. There is a chance that a bias is already within the documents used. Despite the expectation that news organisations uphold impartiality in reporting events, an inherent bias is often present in the articles. Moreover, the personal viewpoints of the researchers could introduce a bias into the timeline. When creating a timeline, people might choose to take a neutral stance, depending on the required perspective. Therefore the chances of the timeline having subjective attributes are almost inevitable.

Understanding Relations between Elements

Relations between events or actors can be fairly complex. A specific type of relationship can be quite similar to each other for instance a correlation and causation or an analogical or comparative relationship.

Thorough Analysis and Time-Consuming

These challenges concern, the difficulty, extensiveness, and time-consuming nature of creating a timeline. Some timelines, like the Dutch Childcare Benefits Scandal Crisis, can become quite extensive. Because of its complexity, it is also quite time-consuming. These are both vital reasons for automating the process of creating a timeline.

2.2 Timeline Visualisation Software

Timeline visualisation software has already been researched for decades [78]–[80]. Considering the possibilities that the latest technologies have created, the quality and speed of creating timelines with software, have become increasingly higher. This section will cover multiple aspects of timeline visualisation software. The first part will focus on algorithms for automatically generating timelines. Thereafter general design features for timeline creation software will be discussed followed by the customisable aspects of such software. Then different ways to assess the quality of timelines will be discussed and this section ends with some examples of timelines that are generated and will be introduced.

2.2.1 Algorithms

The key part of automatically generating timelines is to get the most fitting algorithm. As discussed in section 2.1, many different factors need to be taken into account when creating a timeline. One algorithm can work in a certain situation but might be less effective in another. Therefore, it is important to consider the various types of algorithms that can be utilised.

These algorithms can be divided into six categories of algorithms that are defined by the Multi-Document Summarisation (MDS) [81]–[83]. The six categories of algorithms are:

1. TimeLine Summarisation (TLS)
2. Multiple TimeLine Summarisation (MTLS)
3. Archive TimeLine Summarisation (ATLS)
4. Graph-based Timeline Summarisation (GTLS)
5. Query TimeLine Summarisation (QTLS)
6. Comparative TimeLine Summarisation (CTLS)

TimeLine Summarisation (TLS)

TimeLine Summarisation (TLS) is a relatively new but evolving field in natural language processing, focusing on compressing information over time into coherent and informative overviews. [79], [84]. It therefore focuses on extracting the most important information out of a large dataset. After extracting this information, events are created and subsequently put into chronological order. This chronological overview provides a complete outline of the events or topics as they have evolved [81], [82], [85].

TLS can be used within different sectors. These sectors can be the ones described in Section 2.1.3 or other ones like the news industry. By creating timelines, based on the headlines of news articles, people find the timelines more relevant and understandable [86]. It not only can make it more understandable and relevant for the public, but also for law enforcement; as it can aid in disaster surveillance [87]. By using data from social media, such as *X*, real-time generated TLS that tracks high-impact events, can classify, incrementally cluster, post-process and subevent summarisation. This offers a more efficient, effective, and possibly less biased way of creating timelines.

Multiple TimeLine Summarisation (MTLS)

The Multiple TimeLine Summarisation is based on TLS but differs from it by generating multiple parallel timelines that cover different events and activities [88]. This is based on getting a complete picture of a situation by separating different stories into independent timelines. This method helps to get a deeper and more comprehensive understanding of complex topics that can be interpreted in multiple ways. Instead of looking at specific events, it pays more attention to the time period they happened in [86]. Therefore, these parallel timelines can be differences within the perspectives as described in Section 2.1.2. This way, the generated timeline is more nuanced, and it avoids focusing too much on a user's input that might have some bias.

Furthermore, it might be the case that certain sequences have evolved differently over time. There might be a link between different events but it would be subjective to say that there is a direct effect. Therefore one can create multiple objective timelines and leave the linking up for interpretation. However, this can also be a disadvantage as the user has to switch between timelines to find the links themselves. MTLS can then provide a complete timeline that shows how these have evolved over different parallel tracks.

Archive TimeLine Summarisation (ATLS)

ATLS is a conceptual method based on TLS that deals with creating timelines from archived or historical data [17]. What separates ATLS from the previously discussed, algorithms is the fact that this method involves a vast amount of information and data, and also looks at other historical data. This conceptual method for summarisation is especially useful for understanding the development of topics over long periods and can involve analysing and summarising content from historical documents, archives, or past web events [17]. Therefore this type of algorithm distinguishes itself from MTLS and TLS by including more diverse data over a longer period that does not cover one chain of events. The advantage is that it gives more context on the specific events in relation to the past, although the overwhelming context can cause confusion.

Graph-based TimeLine Summarisation (GTLS)

Another type of timeline summarisation is Graph-based Timeline Summarisation (GTLS), this is a new way of timeline summarisation that creates a timeline for particularly large collections of news articles [81], [89], [90]. This differs from other timeline summarisations by using a specific algorithm for extracting key events that are presented over time. These key events are then presented as event graphs. By representing these key events within one graph, overseeable sub-graphs are created to gather a full overview of a series of news articles [90].

This process ensures the inclusion of critical events as well as maintaining accuracy and coherence. It is a good solution for overseeing the extensive amount of news data, providing a concise yet comprehensive view of events as they unfold over time. This timeline summarisation technique is especially valuable for researchers and analysts who need to quickly understand the main points of large news datasets [90], [91]. However, the level of detail is fairly limited as it focuses on key events.

Query- and Comparative TLS

In addition to the aforementioned algorithms, there are two more algorithms. These are the Query-based TimeLine Summarisation (QTLS) [17], [92]–[94] and Comparative TimeLine Summarisation (CTLS) [89]. QTLS uses queries to create timelines by, for instance using documents that are returned by search engines. QTLS consists of two important steps:

1. Event Detection
2. Event Ranking

Event detection focuses on finding documents with specific terms based on the query [92]. Furthermore, these terms are then clustered by date [93] and then the occurrences are measured to be used in the following step [94]. To rank the items, multiple methods can be used. Learning-to-rank models like Ge et al. [95] can rank events based on importance. Furthermore, one can score events on relevancy and saliency [93], on interest and burstiness [92] or weighting terms in descriptions and using the Levenshtein similarity measure to remove any duplicate events. Furthermore, the CTLS algorithm collects multiple events in a dataset and creates multiple timelines to gather a perspective of one timeline compared to another [89].

Although MTLS and CTLS are quite similar main difference between MTLS and CTLS is that MTLS can create multiple parallel timelines that can cover different or complementary stories, whereas CTLS creates multiple timelines for comparison. For instance, a local timeline can be compared to global events that have taken place. A disadvantage to QTLS and CTLS is the fact they might contain a bias. If certain differences between the timelines are unknowingly larger than they are, certain differences might be noted earlier which causes a bias. Furthermore, queries can also contain biases resulting in a more biased timeline.

To gain a full overview of all the TLS algorithms discussed here, an overview table can be seen in Figure 2.2 where all types of algorithms are discussed with descriptions and formal definitions [17].

	TIMELINE SUMMARISATION	MULTIPLE TIMELINE SUMMARISATION	ARCHIVE TIMELINE SUMMARISATION	GRAPH BASED TIMELINE SUMMARISATION	QUERY-BASED TIMELINE SUMMARISATION	COMPARATIVE TIMELINE SUMMARISATION
DESCRIPTION	TimeLine Summarization (TLS) performs Date Selection and Date Summarisation. They select key dates from documents and summarize events of those dates using techniques like frequent date mentions or graph-ranking models. TLS is mainly extractive but can also be abstractive.	Multiple TimeLine Summarization (MTLS) generates separate timelines for each story in a dataset, detecting events and linking them based on saliency and consistency.	Archival TimeLine Summarization (ATLS) is a framework consisting of two key steps: Timeline Generation and Timeline Presentation. It focuses on processing longitudinal datasets of timestamped documents, often covering more extended periods than typical TLS datasets.	Graph Based TimeLine Summarisation (GTLS) is the process of extracting the key events of news articles to show within a graph. This ensures a more coherent and accurate visualisation of the timeline.	Query-based TimeLine Summarization (QTLS) applies TLS on documents related to a user query. It involves Event Detection and Ranking, using methods like clustering by dates, peak detection in date occurrences, or classifiers for event importance.	Comparative TimeLine Summarization (CTLS) creates comparative timelines highlighting contrasts between two timestamped documents, calculating both local and global event importance.
FORMAL DEFINITION	**TLS:** takes as input a standalone homogeneous dataset of timestamped documents $\mathcal{D} = \{d_1, d_2, ..., d_{|D|}\}$ and generates a timeline $T = \{p_1, p_2, ..., p_{|T|}\}$ of time-summary pairs $p_i = (t_i, s_i)$, where s_i summarizes important events happening at time t_i;	**MTLS:** takes as input a dataset of timestamped documents $\mathcal{D} = \{d_1, d_2, ..., d_{|D|}\}$ that can be standalone or returned using a query $\mathcal{Q} = \{w_1, w_2, ..., w_k\}$, and outputs a set of timelines $\mathcal{T} = \{T_1, T_2, ..., T_m\}$ for each story or topic detected in $\mathcal{D}$, where each timeline T_i is a sequence of time-summary pairs $p_i = (t_i, s_i)$;	**ATLS:** takes as input a longitudinal dataset of timestamped documents $D = \{d_1, d_2, ..., d_{|D|}\}$ taken from an archival collection, either standalone or returned by a query $Q = \{w_1, w_2, ..., W_k\}$. The period of time covered by D is usually much longer than the one typically used in TLS and outputs a timeline T generated from D as a sequence of time-summary pairs $p_i = (t_i, s_i)$, where s_i summarises important events happening at time t_i	**GTLS:** Takes as input news story documents $D = \{d_1, d_2, ..., d_{|D|}\}$. It then outputs a set of key events $E = \{e_1, e_2, ..., e_{|E|}\}$ that are used to create a comprehensive graph G	**QTLS:** outputs a timeline $T = \{p_1, p_2, ..., p_{|T|}\}$ as a sequence of time-summary pairs $p_i = (t_i, s_i)$ from a set of timestamped documents $\{d_1, d_2, ..., d_{|D|}\}$ based on a query $\mathcal{Q} = \{w_1, w_2, ..., w_k\}$ where w_i denotes a word belonging to the query;	**CTLS:** takes as input two datasets of timestamped documents $\mathcal{D}_A = \{d_1, d_2, ..., d_{|D_A|}\}$ and $\mathcal{D}_B = \{d_1, d_2, ..., d_{|D_B|}\}$ and outputs two timelines T_A and T_B made of contrasting events detected in $\mathcal{D}_A$ and $\mathcal{D}_B$, each as a sequence of time-summary pairs $p_i = (t_i, s_i)$;

Figure 2.2: Table with different algorithms explained with definition and formal definition.

2.2.2 Features

There are multiple features of timelines that can change the meaning behind them. The features that will be discussed in this section concern the most important aspects of a storyline. Features primarily concern the timeline's key aspects. In contrast, customisable aspects involve additional customisation options as will be discussed in Section 2.2.3. Features of the timeline should be carefully considered. This section will discuss different features relevant to the timeline creation process.

Topic

First of all, the exact topic needs to be discussed. As mentioned in Section 2.1.1, a research question needs to be defined to clearly state the direction and scope of the research [15]. Not only is this helpful when manually generating timelines, but machines also need detailed input to create a preferred outcome [96].

Timespan

The second feature that is relevant for creating a timeline is the timespan. The timespan concerns the beginning and end date chosen for the timeline [15], [97]. The timeline can be created over days, years, or decades depending on the scope of the subject matter.

Actor, Place or Event TimeLine

Another feature of a timeline is to choose a perspective in which the timeline is shown. Considering the objective of the timeline, it might be beneficial to focus on different concepts such as actors or locations within the timeline rather than sequences of events [15], [16]. These different types of timeline objects can present different pieces of information about the timeline. These objects can either be viewed individually or as a whole.

Orientation

The orientation of a timeline can either be horizontal or vertical. Horizontal timelines are read left to right, commonly used for longer timespans, while vertical timelines present a top-down perspective, which can be useful for shorter timespans or more detailed views [98].

2.2.3 Customisable Aspects

Various customisable aspects can be implemented within an automatic timeline generator. These customisable aspects will be discussed within this subsection. These customisable aspects have been gathered from other visualisation tools and frameworks [99], [100].

Level of detail

The first aspect is the level of detail. This aspect refers to how much information is included in the timeline. It can range from a minimalistic overview with just key events or actors to a detailed report including minor events or less prominent actors [99]. There might be a scale on which the user can select the desired level of detail.

Filtering

This feature allows users to include or exclude, for instance, specific types of events or actors. Filtering can be based on various criteria such as date ranges, event types, or the significance of actors, enabling a more specific view of the timeline [99], [100]. Within creating a timeline, it might be good to add specific keywords to include or exclude within the timeline.

Time Units

The granularity of time units on a timeline should be customisable. The generator might choose a specific time unit to visualise the timeline, but maybe another time unit might be preferred by the user. The types of time units might include years, months, days, or even hours and minutes, depending on the precision required for the timeline's purpose [99].

Interactivity

This aspect determines how users can interact with the timeline. Interactive timelines allow users to click or hover on actors or events for more details. They can possibly also zoom in or out for different levels of granularity, or even rearrange elements to explore different perspectives [100].

Integration of Multimedia

Within some contexts, the timeline might be more comprehensible with the help of integrating photos, videos or audio clips. This multimedia integration can enhance the storytelling aspect of the timeline, making it more engaging and informative [100].

Labels and Markers

Labels and markers can provide extra context and clarity. Annotating within the timeline was also mentioned to be done within manual timeline creation (Section 2.1.1). Labels can include dates, descriptions, or names, helping to identify significant points or periods on the timeline [99], [100]. Markers, such as lines or tooltips, provide extra visualisation or information, aiding in comprehending the sequence and timing of events.

Design Elements

This encompasses visual aspects such as colour schemes, fonts, and graphical styles. These are some basic features to export a timeline with a certain view depending on making it more engaging or formal [99], [100]. Furthermore, it can also be used to convey information, like using specific colours to represent different types of events or actors.

Manual Editing

Lastly, one of the most important aspects is to be able to manually change specific things within the timeline. The generator is intended to function as a supplementary aid to the user, rather than being the predominant element of preference in all instances.

These features and customisable aspects are general principles that have been identified in the research of Price and Siang [99], [100]. These principles are used in different kinds of software. One software that creates semi-automated timelines is the software from Monday blog which helps with marketing planning in group projects [101]. As can be seen in Figure 2.3, different principles as discussed above are highlighted.

The *light green* rectangles show different features as discussed in Section 2.2.2 and *red* rectangles show different customisable aspects as discussed in this section. This software has not been thoroughly tested. However, it may offer additional functionalities, as other settings can be adjusted. These settings are highlighted in red in the top right corner.

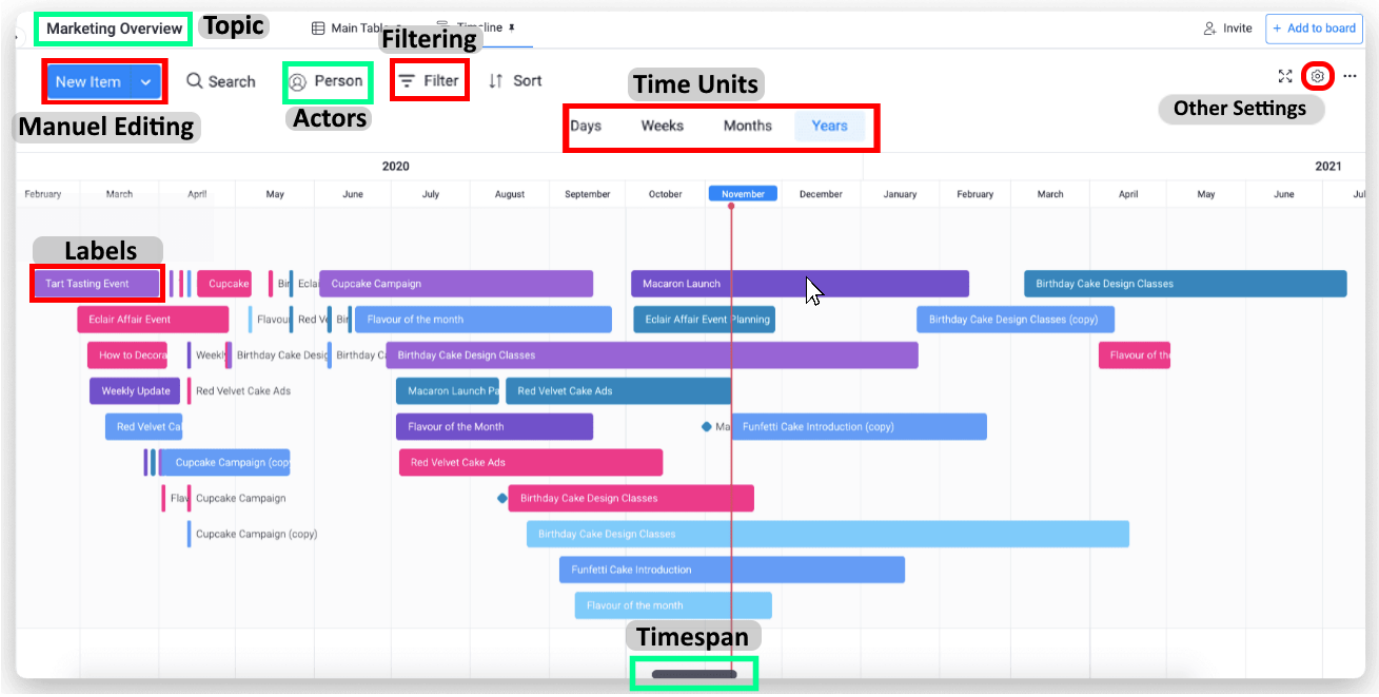


Figure 2.3: Timeline software by Monday Blog containing highlighted principles [101]

2.2.4 Quality Assessment

Now that a good overview has been given of the different types of guidelines, there needs to be looked at how the quality of automatically generated timelines or the software around it, can be assured. There are multiple ways to assess the quality since the overall quality depends on a variety of aspects. Furthermore, the quality of these timeline visualization software can be assessed through both *Algorithmic Evaluation metrics* and *Human-Centric Evaluation metrics*.

Algorithmic Evaluation

To assess the content quality and structure of a specific timeline through algorithmic metrics, five specific aspects can be taken into account. These aspects are relevance, coherence, coverage, connectivity/diversity and utility [102], [103].

- **Relevance** covers the amount of value that the generated events have based on the query given by the user. Therefore it looks at the specific words within the user-given query and the articles [81].
- **Coherence** of articles cover the similarity between documents. Therefore document topics are extracted and compared. This can be measured by analysing word patterns through, for instance, the TF-IDF approach [81], [102], [104].
- **Coverage** focuses on encompassing essential story elements while encouraging the variation of elements. This can be measured by calculating how much a document covers a certain feature. [81], [102]
- **Connectivity/Diversity** covers the amount of overlaps of different lines within articles, compared to the total amount of articles[81], [102].
- **Utility** covers the overall usefulness of the timeline to the user, while looking at the accessibility, the ability to meet user needs and the overall informativeness. This is often measured by taking all other measurements into account to create an overall utility score [81].

A good balance needs to be found within these five measurements. Therefore, trade-offs need to be made. It might be the case that a high relevance score increases the chances of including noise. This association can be attributed to the fact that relevance is directly correlated with the query previously submitted. Therefore the query might be too broad and include some noise within the measurement. Also within multiple measurements, trade-offs need to be made. For instance, increasing the diversity might lower the relevance since a diverse set might not always be relevant to the cause.

Human-Centric Evaluation

Outside of these algorithmic evaluation metrics, there are also human-centric evaluation metrics for checking the quality of a timeline visualisation software. These metrics can not be automatically used to improve timelines but can demonstrate failures in the algorithm, visualisation of the timeline, or software. Human-centric metrics cover scores given by users about the timeline. These metrics can be assessed by ranking them on a scale. Users can be confronted with a survey afterwards when a timeline has been generated. This survey can then be used as input to improve the algorithm that generates timelines or metrics that can improve them. Other human-centric metrics that can be used to determine performance and user experience are interactivity, consistency, perceived ease of use, and many others [105].

These two types of metrics can be utilised separately or integrated into a hybrid approach. A hybrid solution potentially offers an ideal balance between efficiency and effectiveness. It allows for automatic enhancements to the timeline, while the human-centred method provides a comprehensive view of potential improvements. This dual approach leverages the strengths of both automation and human insight for optimal timeline refinement.

2.2.5 Instances

There are already various types of timeline-generating software on the market. Every software has its advantages and disadvantages that need to be taken into account when considering what type to use for visualisation. This section will cover different types of timelines that can be used for visualisation.

Metro Maps. Metro maps, as can be seen in figure 2.4, are compact structured sets of documents that maximise the coverage of prominent pieces of information [102]. These metro maps consist of a set of distinct lines that follow a separate path but can overlap. The shape of this timeline is the same as the ones that can be found in a metro station to show the different metro lines that one can take. Every stop in these metro lines represents a specific event on a metro map. Each line, which normally represents a specific metro, defines an aspect of the story. These aspects could be strikes, specific groups, or financial situations. These maps can serve as an effective tool to help a user cope with an overload of information. Metro Maps are constructed using the GTLS algorithm. An example of a Metro Map can be found in Figure 2.4.

Figure 2.4: Image of Metro Map of the Greek Financial Crisis - [102]

Story Trees. Story trees are a different kind of timeline. They mostly have one primary story with separate branches that go in different directions but can indicate a causal explanation of certain events [106]. Furthermore, each story tree consists of a story, different events, new events and merged events. StoryFlows are developed in adherence to the principles of the MTLS and GTLS algorithms. An example of a Story Tree can be seen in Figure 2.5.

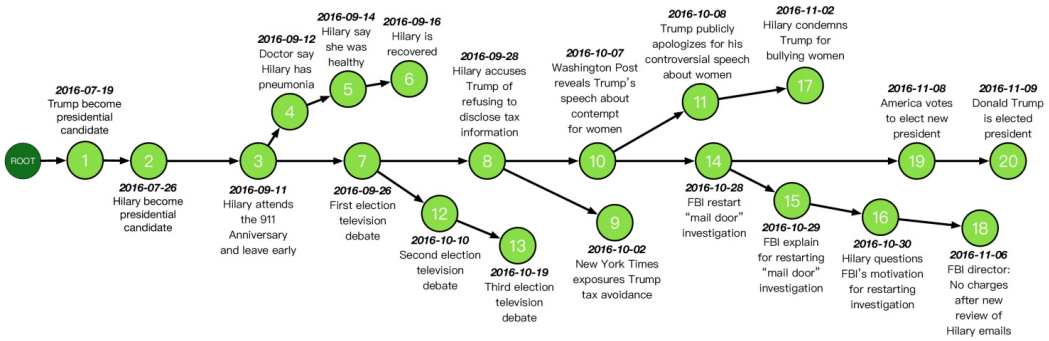


Figure 2.5: Image of StoryForest of "2016 U.S. presidential election." - [106]

Story Graph. A story graph is a type of graph that shows one linear storyline but every event in that storyline can have a further explanation of its own [107]. StoryGraphs are developed in adherence to the principles of the GTLS algorithm. An example of a StoryGraph can be seen in figure 2.6.

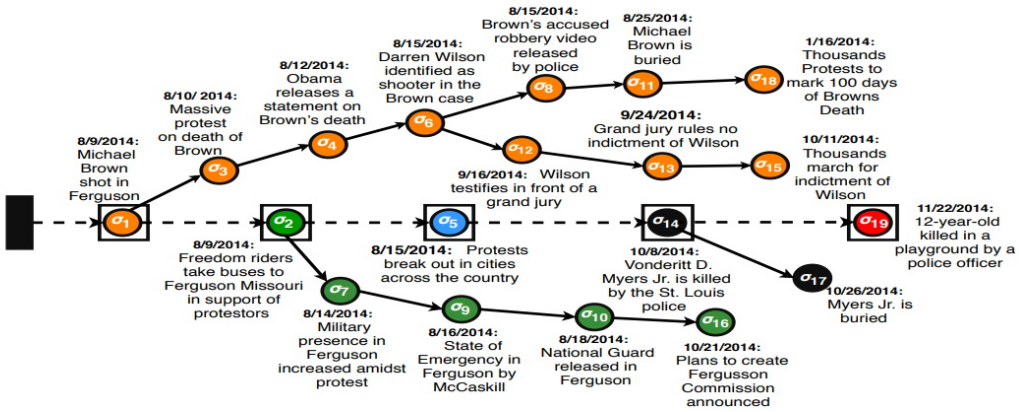


Figure 2.6: Image of StoryGraph of the Black Lives Matter uprise - [107]

Other than activity-based timelines, another kind of timeline can be drawn up. This type does not specifically look at the stories, which each line repre-

sents, but rather at different actors within a storyline. It could be that every actor in a story has his journey with overlapping events.

StoryFlow. One of these storylines is the StoryFlow [16]. This type of storyline generator introduces multiple actors to serve as the different lines to see what actor did what. This type of storyline would be very useful for blame attribution as you can easily see what actor had what knowledge and participated in what events. StoryFlows are created in adherence to the principles of the MTLS algorithm. An example of the StoryFlow is shown in Figure 2.7.

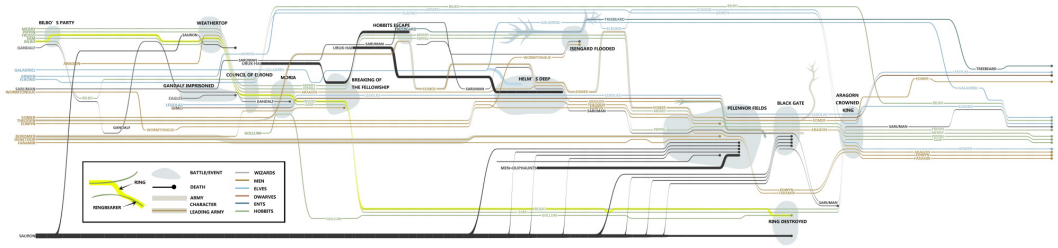


Figure 2.7: Image of StoryFlow of the Fictitious Movie Trilogy The Hobbit - [16]

StoryCake. Another way of showing a storyline in a more extraordinary way is by using the StoryCake [108]. StoryCake visualises the timeline in a 'piece of a cake'-shaped way where the timeline follows the circumference of the cake. StoryCakes are developed in adherence to the principles of the MTLS algorithm. An example of the StoryCake is shown in Figure 2.8.

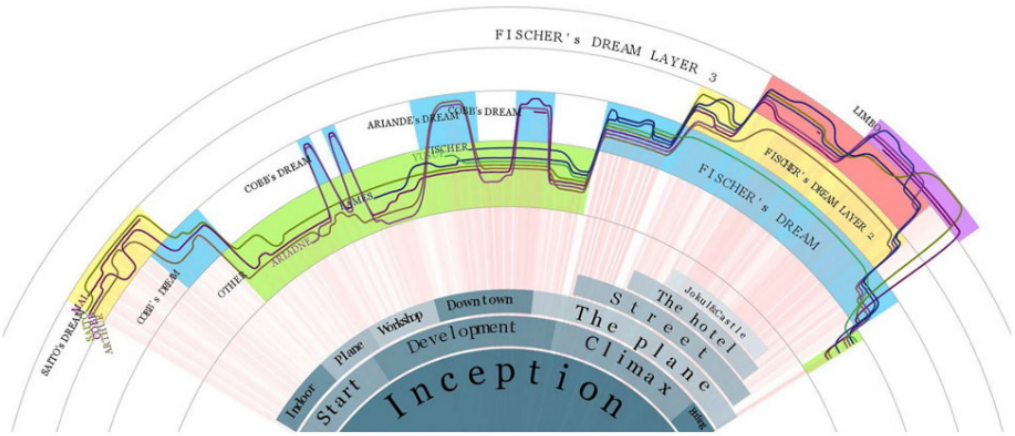


Figure 2.8: Image of StoryCake of the Fictitious Movie Inception - [108]

Following these static timelines, interactive timelines have also been proven

effective. These timelines allow users to open up different paths or gain more information through the use of hyperlinks.

Interactive Timelines. The following timeline does not have a specific name but shows an interesting view of creating interactive timelines. This timeline mainly uses images to show the sequence of events. Once clicked on an image, further explanation will be given regarding it. An image of this timeline can be seen in Figure 2.9.

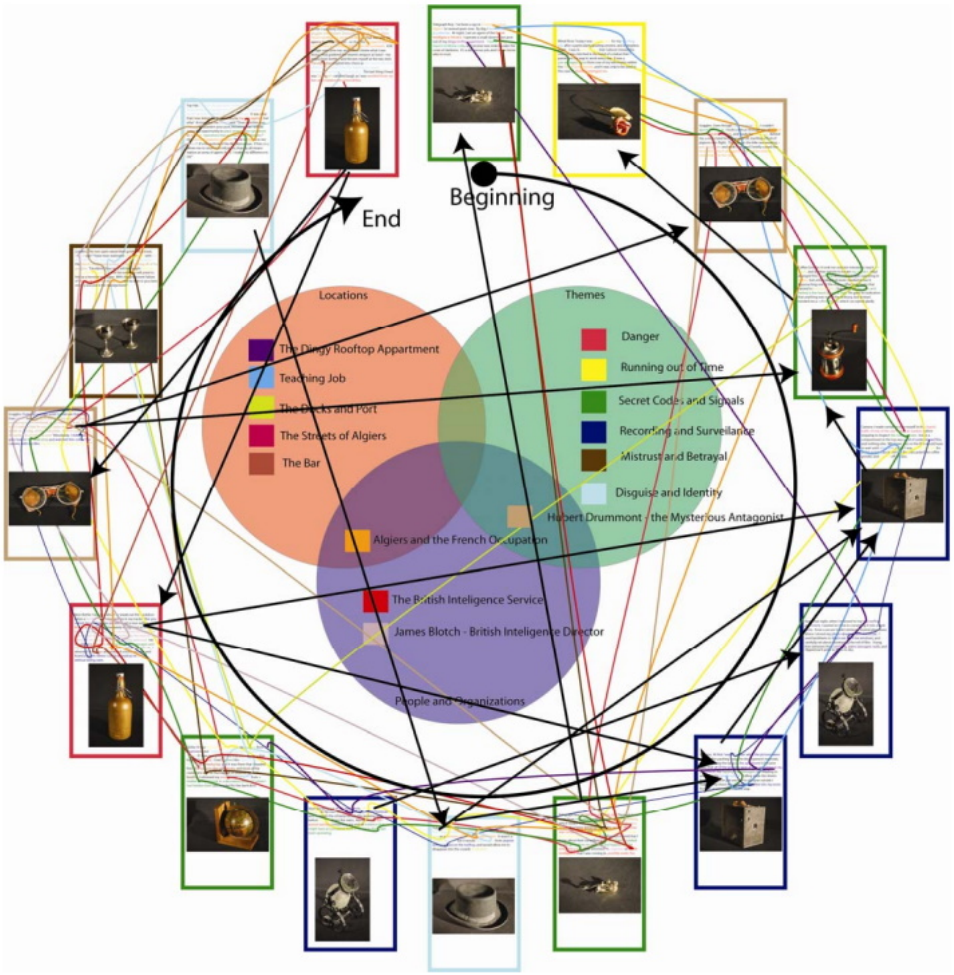


Figure 2.9: Image of an Interactive Timeline Regarding a Fictitious Espionage Story- [109]

3. Method

Having established a foundation on different aspects of timelines, this study will delve deeper into one specific field of application where automatic timeline generation would be beneficial: the Woo-request process. The Woo-request process concerns requesting governmental information, in agreement with the Open Government Act [110]. Information can be retrieved from all governmental instances like ministries, municipalities or provincial councils. In 2022 and 2023, two studies concluded that the process was extremely slow and cost a lot of resources. [3], [111]. On average, ministries take 162 days to respond to a Woo request, despite the legal maximum being 42 days. The three biggest reasons for this delay are the management of information, staff capacity, and cultural change. Furthermore, the scope of the Woo is extremely large since all government files can be requested.

Within the management of information, there is often a misconception about what information is required and therefore more information is given than necessary. This excessive information results in two problems: spending more time preparing and redacting documents, and requiring requesters to spend more time sifting through the documents to find the information they need. Timelines, as seen in the literature review, can provide a solution for this problem by providing clear overviews of large datasets.

Timelines are already manually drawn up within the government for certain situations. This is time-consuming and therefore only used in complicated cases. Therefore automated timeline generation could help to increase efficiency. This method will first discuss the research questions drawn up for this research. Then a preliminary study will be discussed that will help to get an overview of the current Woo process and help identify the stakeholders. After that, the participants who will partake in the main research, the materials, and the procedure of the main research will be discussed. Finally, the data analysis of the study will be discussed.

3.1 Research Questions

Having established the research focus, the following research question has been drawn up:

> ***RQ:*** *How can the implementation of timeline generation software increase the understanding of document requests and improve the effectiveness of the Woo request process?*

In the implementation, the primary focus will be on the methodology and GUI of the process. Therefore the focus will not be on the algorithms of the automatic timeline generator. To answer this research question, the following subquestions have been drawn up:

> ***S1:*** *What stakeholders of the Woo request process can benefit from a timeline visualising software to facilitate the speed and clarity of information management?*

This subquestion will be answered during the preliminary study of mapping the Woo-request process together with its stakeholders.

After finishing the preliminary study, it was found that an important part of the process is the request specification. Therefore a structured method will be discussed that can retrieve a list of concepts or topics from a gathered subset of documents. To test this method, the following sub-question is drawn up:

> ***S2****: To what extent can topics be identified to help define the required information and perspective(s) within the Woo-request process?*

This subquestion will be answered with the help of interviews to elicit if and how the process can benefit from the structural and automated topic extraction in government files. The last subquestion will investigate another part of the process: the document selection part. This part will examine whether the timeline software is beneficial and what design principles, as found in the literature study, can be applied to aid the process.

> ***S3:*** *To what extent do general design principles of software apply to a timeline generation software that can be used within the Woo-request process?*

This subquestion will be answered by proposing adjusted general design principles to stakeholders of the Woo process through the use of interviews.

3.2 Preliminary Study

3.2.1 Process Mapping

Now that the research questions have been established, it is necessary to determine what stakeholders are involved within the Woo and which parts of the Woo process will be further investigated in the main focus of this research.

Process Diagram

In the reports *'Ondraaglijk traag'* and *'Matglas'*, the general process of the Woo has been mentioned [3], [111]. These reports are used to create a foundation for understanding what the process involves. Therefore, the process as described in the reports is mapped to use as a concept. That concept is then discussed, through online meetings, with multiple stakeholders of the process including, journalists, Woo request handlers, government lawyers and other employees of Dutch ministries. Through these conversations, adjustments are made to the first concept to construct a general process overview of the Woo within ministries. The results of this preliminary study are implemented in a diagram that can be seen in Figure 3.1. It should be noted that this is a generalised process flow, and there are variations between different ministries as well as departments within the ministries.

The process flow demonstrates different parts of the process through the use of different colours:

- **Blue** demonstrates the initial phase of submitting a Woo request and specifying the needs of a requester.
- **Green** indicates the part of the process where a request might be rejected for various reasons.
- **Red** demonstrates the process of finding documents and selecting those required to fulfil the Woo request.
- **Purple** is the process of redacting the documents and getting a legal clearance to publish the documents.
- **Turquoise** indicates the process of releasing and publishing the documents.

During the process mapping, the stakeholders already indicated multiple problems. These problems that were mentioned will be discussed here.

Problems within Woo Process

Unclarity. The first problem that is mentioned is the unclarity of some requests. When requesting information, the requesters are not aware of what files are available. Therefore estimates are given on what files they think the government has. People who have already submitted Woo requests, like journalists, already know how specific governmental organisations save their files and what types of documents contain specific information. This facilitates the process.

Signature Line. Another problem is the signature line otherwise known as "parafeerlijn". This is the process of getting clearance from the management to continue to the next phase of the process. Since the management does not know everything about the request itself, it is very time-consuming to wait before they have the time to read through the files and give the clearance.

Inventory List. Once clearance is given to continue to the document selection phase, the requester can often only choose from an inventory list of document titles. These titles often do not reveal the contents of a file, leading to the request of more files than necessary 'just to be sure'.

Document Redaction. This problem leads to another issue, as the overload of documents must be redacted for sensitive information. Even though there is software that can help perform this task, it is still susceptible to errors and there is always a lawyer that has to go through the files.

These problems have been taken into account for the next steps of this research. The main focus of this research will be on the later stages of the blue part, request specification, and the red part, document selection.

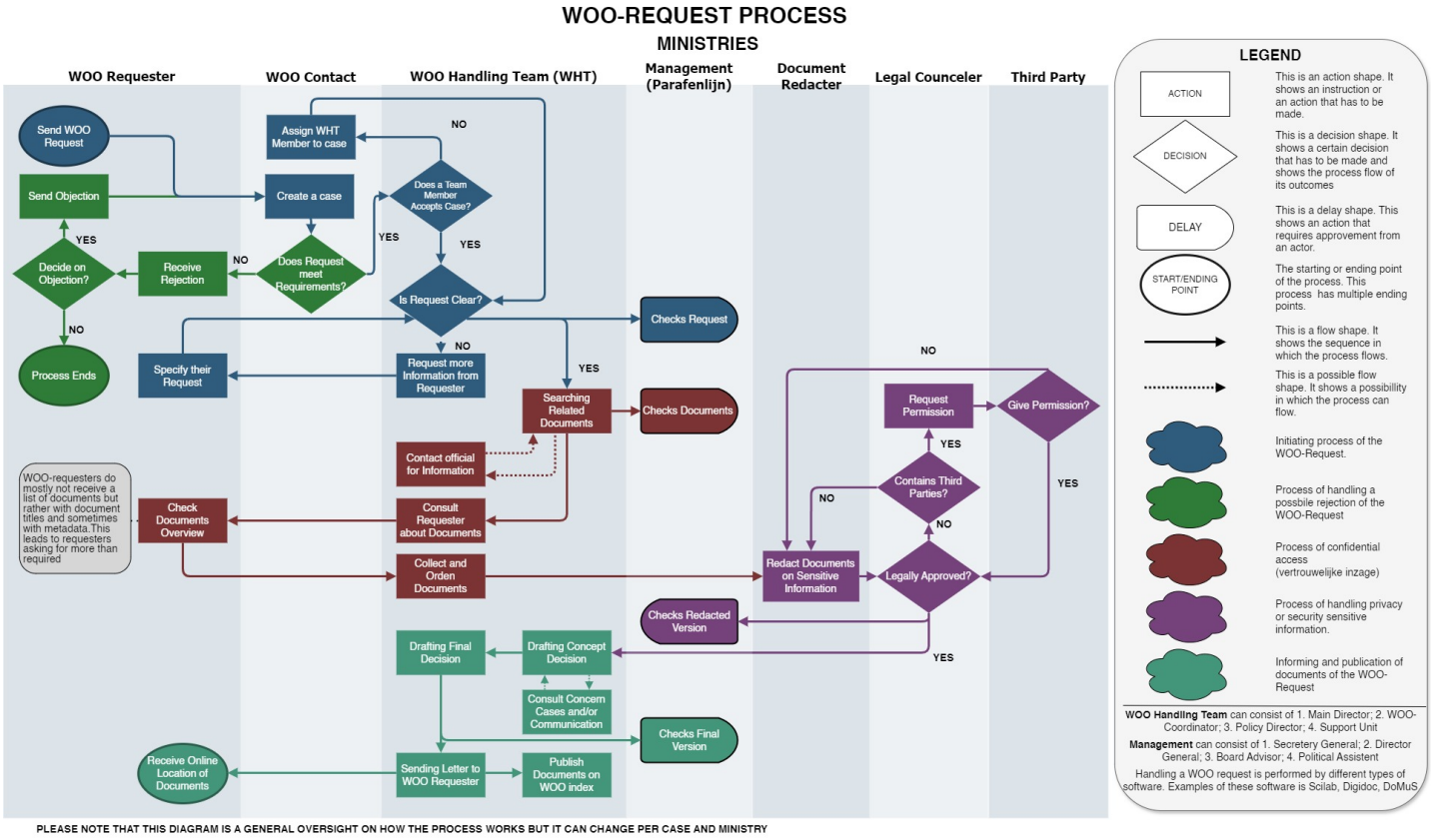


Figure 3.1: The Woo process as mapped during the preliminary study

For the request specification part, shown in the blue part of Figure 3.1, extra research needs to be performed. This research aims to determine whether finding the common ground between the available information and the required information is beneficial. Therefore a structured and automated way of finding topics within a dataset should be applied.

Topic Extraction Methods

Different methods and algorithms for extracting topics already exist like LDA, Word Embeddings and NMF. One method retrieved from a theory that uses these algorithms to extract topics is The Grounded Theory [112]. This theory not only describes the whole process of retrieving information but also describes the process of structuring these topics and creating a hierarchy. Therefore this theory offers the ability to not only extract the topics but also structure them for overview.

3.2.2 The Grounded Theory

Over the years, multiple types of Grounded Theories have been proposed. The biggest differences between the diverse types of Grounded theories are the origins and methodological nuances. The two most used types of Grounded Theories are the Glaserian and Straussian Grounded Theories named after the creators of the Grounded Theory Barney Glaser and Anselm Strauss [113], [114]. These creators had different views on the Grounded Theory and therefore established two different approaches to Grounded Theory. The Glaserian Grounded Theory is closer to the original theory and focuses only on the inductive side of the theory whereas the Straussian Grounded Theory focuses on induction, deduction, and verification and performs it through a systematic execution. Because of this structured approach and deductive and verifying steps, the Straussian theory seems to be a better fit to find out what concepts are covered and what are relevant.

Straussian Method The structured approach of the Straussian Grounded Theory consists of three steps: *Open Coding*, *Axial Coding* and *Selective Coding*. Open Coding is the process of finding the relevant categories within each document and overlapping categories. Axial Coding is the step of finding relationships within different categories to create a hierarchy. Selective Coding is the last part that tries to find the core categories to create a theory based on documents. Since this research aims to find out whether using the Straussian Grounded Theory would be a valid method before creating a timeline, some additional steps need to be added.

Additional Steps The two steps that this research introduces are: *Relevance Coding* and *Hierarchical Coding*. The whole proposed framework for the Grounded Theory can be seen in Figure 3.2. *Relevance Coding* will be added between *Open Coding* and *Axial Coding* and focuses on retrieving the relevant concepts based on the information request that has been submitted. This filtering step ensures that only relevant concepts are retrieved.

Hierarchical Coding is the step that is added between *Axial Coding* and *Selective Coding* to ensure that all concepts are already put into hierarchical order. This step can help the follow-up process of creating a timeline by already creating a hierarchical concept map but furthermore puts things more into perspective.

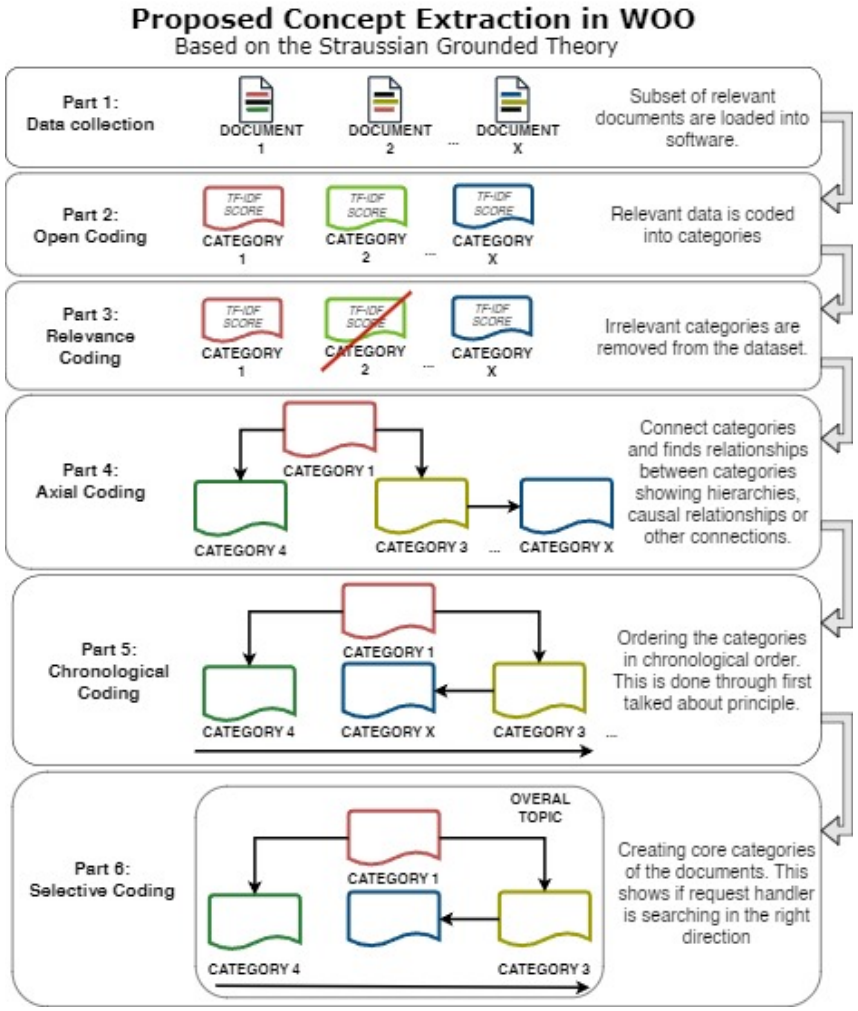


Figure 3.2: Figure of proposed Grounded Theory with added additional steps

One of the aims of this research is therefore to determine whether the Straussian Grounded Theory can aid the topic selection process before the document selection by identifying key concepts within documents. These concepts can then be highlighted to information seekers, potentially reducing information gaps.

3.3 Participants

The participants that are gathered for this research are gathered using a mix of the non-probability sampling methods of purposive sampling combined with snowball sampling. People are contacted from different organisations with an initial set of questions to see whether they met the need to participate in the experiment. Furthermore, they are asked whether they knew other qualified people to participate in the experiment. Although this snowball sampling might influence the generalisability of the results, only recommended people from other organisations are contacted and therefore no direct colleagues. Given that the process mapping identified a broad range of participants, and it was chosen to focus on a specific part of the process, a set of participant requirements was established to define research boundaries. Therefore, the participants need to fall within one of the three categories as described below.

1. **Requester** People that have submitted Woo requests for either work or out of interest.
2. **Handler** People that have handled a Woo request for or in the name of a national governmental organisation.
3. **Advisory Organisation** People who work at organisations that are in contact with both requesters and handlers to advise on improving the Woo process.

Furthermore, all participants had to be 18 years or older to participate. Since the participants in this research need to meet a relatively specific set of requirements, there is a limited sample of people who can participate.

The method of recruiting participants resulted in 10 people participating (n = 10) from different organisations. The participants are divided into three different groups. These groups are *Woo Requesters*, *Woo Handlers* and *Advisory Organisations*. The *Advisory Organisation* group consists of participants who work for external organisations that talk to both requesters and handlers to consult the government on mistakes and suggest improvements to the process. An overview of these participants can be seen in Table 3.1.

Table 3.1: Participants of this study

Participant ID	Participant Group	Job Title
1.	Woo Handler	External Woo Handler
2.	Woo Handler	External Woo Handler
3.	Advisory Organisation	Employee at Open State Foundation
4.	Advisory Organisation	Consultant at ACOI
5.	Woo Requester	Independent Journalist
6.	Woo Requester	Journalist at National Paper
7.	Woo Handler	Woo Lawyer at National Archives
8.	Woo Requester	Journalist at Regional Paper
9.	Woo Handler	Woo Handler at Ministry
10.	Woo Requester	University Professor

As can be seen in Table 3.1, from the 10 participants, 4 are categorised as Woo Handler, 4 as Woo Requester and 2 as Advisory Organisation. Furthermore, within these groups, different types of stakeholders are asked to contribute. Therefore journalists at large and small organisations, an independent journalist and a professor are interviewed for the requester group. For the handler group, there are external Woo handlers, a Woo handler from a smaller national government organisation and a Woo handler from a ministry. Lastly, the two people from an advisory organisation are either from the ACOI, which is an advisory organisation set up by the government or the Open State Foundation a non-governmental not-for-profit organisation. Furthermore, the Open State Foundation has also published the reports *'Ondraaglijk traag'* and *'Matglas'* as mentioned earlier.

3.4 Materials

To conduct this research, multiple resources are used as can be seen in Table 3.3. The table shows all the different types of files or items that are used to conduct the interviews. Since all participants are Dutch, the files are all written in Dutch. The documents used can also be found in the Appendix section 6. The materials that are discussed here will be separated into *Administrative Materials* and *Content Materials*. *Administrative Materials* will contain the Information Sheet and Informed Consent. *Content Materials* will delve deeper into the Information Artefacts containing the illustrated methods as used in the research and the Information Structure, which was used as the outline of the research.

3.4.1 Administrative Materials

Information Sheet First of all, the participants are presented with an information sheet. This sheet states the purpose of the research, the participant ID and further details of the researcher. After the research, the participants are given the information sheet containing the information. This allows the participant to come back to their decision to participate, by having the contact details and their participation ID. The Information Sheet as presented to the participants can be seen in Appendix B.

Informed Consent Secondly, the participants are presented with an informed consent sheet. This sheet contains several requests for the participant to be 18 years or older, for their data to be used, analysed and processed and the permission to record the data for further analysis. The participant have the opportunity to photograph the informed consent when preferred before it is stored safely by the researcher. The Informed Consent as given to the participants can be seen in Appendix A.

3.4.2 Content Materials

Interview Artefacts Then a document is used that contains the interview artefacts. This file contains multiple texts and images that are shown to the participants during the interviews for analysis. The text could give a clear instruction on what that part of the interview is about and present a scenario to the participant. The images within the Interview Artefacts illustrate three conceptual propositions: the process of extracting topics from a dataset for request specification, an image of a possible timeline generated automatically by software, and 12 design principles that could be applied to the automatic timeline generation software. The Interview Artefacts Sheet be seen in Appendix C. These artefacts are used to gain the views of the participants on the proposed methods of topic extraction and timeline visualisation software.

Concept Extraction Processes. The process of extracting topics shows three different possibilities of a 4 step process. The process was based on the adjusted Grounded Theory as can be seen in Figure 3.2. Those are the steps of the method as executed as a whole. This includes the steps that the computer takes as well as the users. Therefore the processes as presented to the participants were reduced to the steps that the handler has to take:

Step 1: Collect documents.
Step 2: Load documents in the program.
Step 3: Remove important concepts.
Step 4: Let the requester select concepts.

Steps 1 and 2 are relatively simple and not relevant to this research, it has been chosen to keep these the same for the three process propositions. Steps 3 and 4 are different from each other.

Step 3 differs from each other as one is in the form of a word web. The second one is a simple list where you have to select the relevant concepts. The last process proposition shows a concept list with metrics that show the relevance of the concepts in relation to the Woo request and the documents themselves.

Step 4 shows the concepts in a chronological hierarchy. These three differ since one allows the user to see a document index related to specific concepts. The second allows the requester to digitally select the concepts. The third has additional possibilities on the second process proposition by allowing the user to collapse concepts within the hierarchy and see extra relationships between the concepts.

Furthermore, the Interview Artefacts sheet presents a possible visualisation of the timeline and 12 design principles that might be beneficial to the automated timeline generator. Figure 3.3 shows one of the three process proposals as shown to the participants. The other proposals can be found in Appendix C. These proposals are used to gain insights into whether the topic extraction is useful within the Woo and how this method should be implemented.

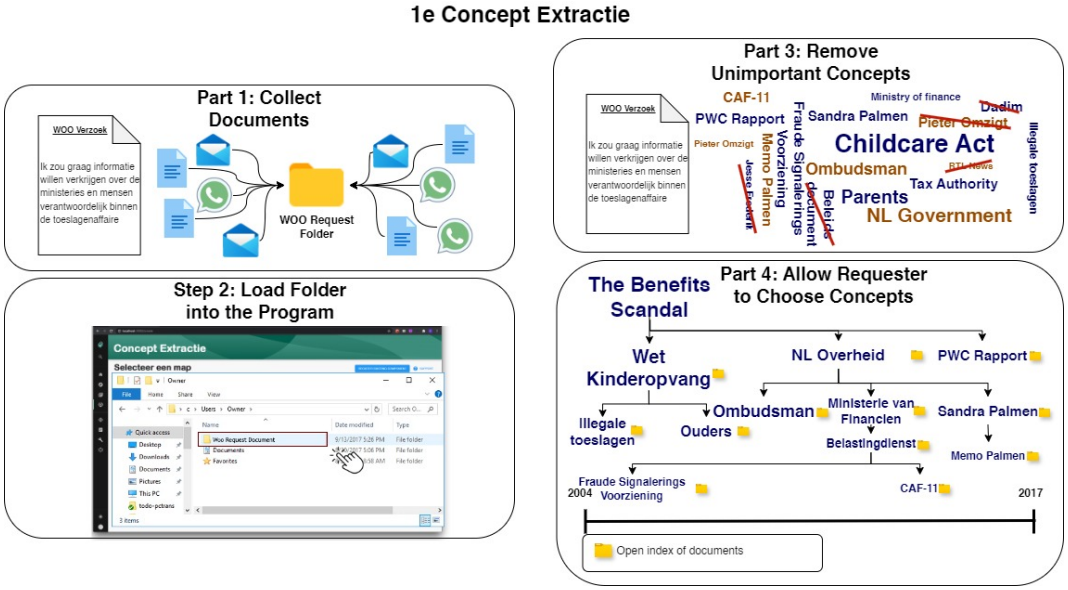


Figure 3.3: First Proposed Concept Extraction Process

Timeline Visualisation and Principles. The visualisation of the timeline shows events with related documents. Furthermore, the actors are highlighted in red and the events are separated into larger groups. Repetitive documents, such as informative letters where the only difference is the salutations, are combined into a folder. This visualisation is based on the requirements mentioned in the TimeFlow document by Muller et al. [15]. The timeline that was shown to the participants can be seen in Figure 3.4.

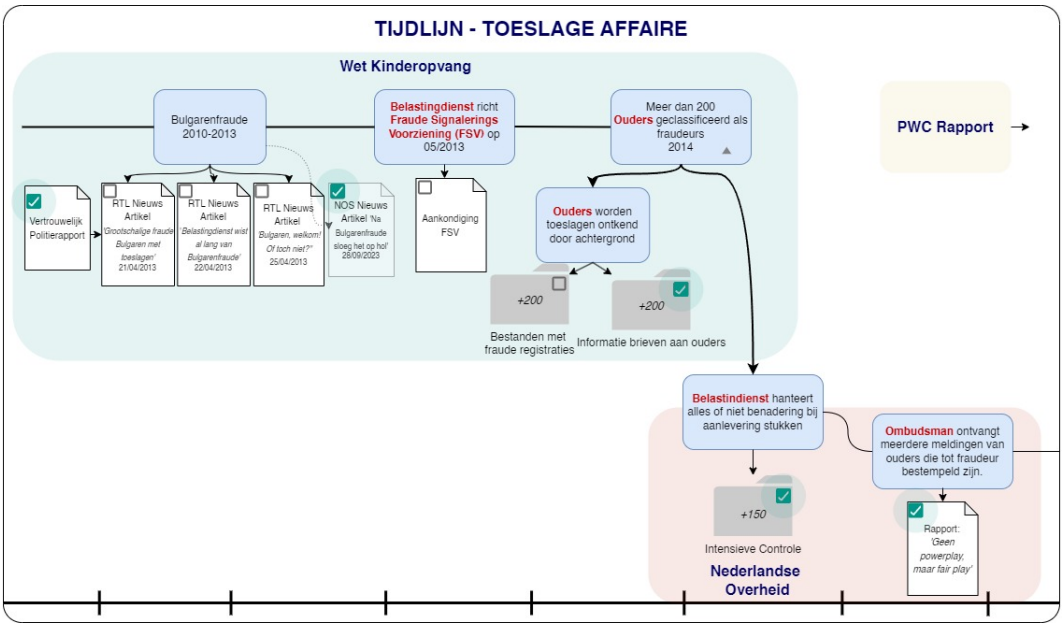


Figure 3.4: Example of Generated Timeline as shown to participants. Based on [15]

Subsequently, the 12 principles are presented. These principles are based on the different features and customisable aspects as found in Sections 2.2.2 and 2.2.3. For each principle, three conceptual implementations have been made which are either implementation possibilities or different additions to the principles. These three options are based on literature and information gathered in the preliminary research. Table 3.2 shows a list of all the principle names together with the names of options.

Table 3.2: Table of design principles with the names of the proposed possibilities.

Principle	Option 1	Option 2	Option 3
1. Topic Selection	Suggested Choices	Free Input	Important Concepts
2. Duration	Fill in Date	Pre-set Slider	Interval Slider
3. Actors, Place, Events	Filters	Selecting	As a Basis
4. Orientation	Choice	Switch	Checkbox
5. Level of Detail	Slider	Zoom	Pull-down Button
6. Filtering	Documents	Subjects	Entities/Actors
7. Time Units	Dropdown	Button	Detail Slider
8. Interactivity	Preview	Explanation	Opening Folders
9. Integration of Multimedia	Audio	Video	Images
10. Labels and Markers	Labels	Markers	-
11. Design Elements	Direct Adjustment	Menu	Themes
12. Manual Editing	Direct Adjustment	Text Only	Dragging

These options provide visualisations that can help the participant to get an idea of the concept. Figure 3.5 shows the first principle as it is shown to the participants. The other principles can be found in Chapter 4 and Appendix C.

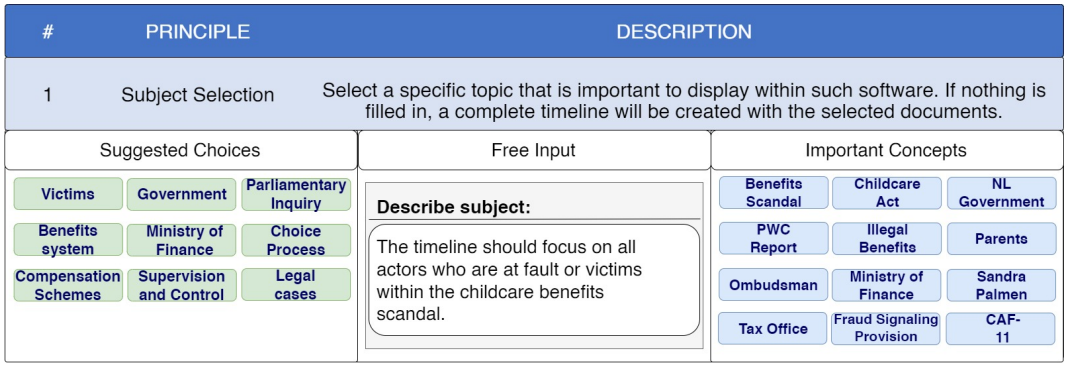


Figure 3.5: Principle 1 with the three propositions

Interview Structure To ensure the consistency of interviews, an interview structure is maintained. This document, containing the structure, can be found in the Interview Structure Sheets. Within the document, several questions regarding the experience of the participant in the Woo, the concept extraction and design principles are stated. However, since this is a semi-structured interview, follow-up questions are asked when deemed necessary. The Interview Structure sheet can be found in Appendix D. Furthermore Section 3.5 covers the full procedure of these interviews.

Another document that is used is the ranking sheet. This document contains a table where rankings during the interview can be easily saved for later analysis. The rankings are written down by the researcher so that the participant can focus on inspecting the artefacts sheet. This ranking document can be found in Appendix E. As mentioned, the interviews are recorded for later analysis. This is executed using an audio recording application on the phone of the researcher that stores the audio locally.

3.4.3 Materials Overview

Lastly, in two out of the ten cases, the interviews are preferred to be held online. Therefore the software UU Qualtrics together with Google Meet are used to perform the interviews. The survey contained the same information sheet, informed consent and images as seen in the Interview Artefacts Sheet. Therefore, there is no other difference between online and offline interviews other than the medium. Furthermore, an audio recorder software named Open Broadcaster Software is used to store the audio of those online meetings locally. The data gathered through UU Qualtrics is saved on the Utrecht University servers. This ensures that all the data gathered through these interviews are stored locally on a hard drive or a European server, according to the GDPR guidelines. Lastly, before conducting the experiment, the UU Ethics and Privacy Scan was fulfilled to ensure that the experiment was following the ethical guidelines of Utrecht University.

Table 3.3: Materials list used in research

Item	Material	Appendix
1	Informed Consent	A
2	Information Sheet	B
3	Interview Artefacts Sheets	C
4	Interview Structure Sheets	D
5	Ranking Sheet	E
6	Audio Recording Device	-
7	UU Qualtrics	-
8	Google Meet	-
9	Audio Recording Software	-
10	UU Ethics and Privacy Scan	-

3.5 Procedure

Interviews are held for the main research. These interviews will consist of three main parts: general questions, questions regarding the process concept extraction and design principles of timeline software. Furthermore, the interview will have a beginning and ending procedure for administrative tasks and final questions.

Part 1 will contain questions regarding the participant's view on the current state of the Woo process. These questions will mainly cover the flow of the current process and the problems within it. *Part 2* focuses on evaluating a proposal for concept extraction. Therefore, this part uses the proposed methodology as can be seen in Figure 3.3. The other two versions can be found in Appendix C. *Part 3* will gather information about general opinions as well as 12 design principles of a timeline-generating software. A list of these principles an example can be found in Section 3.4. If, during the interview, non-verbal communication is detected by the researcher, follow-up questions are asked to establish a verbal explanation.

Introduction

The interviews start by welcoming the participants and thanking them for participating in the study. After some small talk to make the participant feel at ease, the purpose of the interviews is explained and the *Information Sheet* (Appendix B) is presented to the participant. After reading the Information Sheet, the participant is asked if they have questions about it or the study itself. The Information Sheet, containing the researchers' contact information and their participant ID is given to the participant. Then the participants are presented with the *Informed Consent* (Appendix A). The statements on the document are mentioned by the researcher with an extra notice on the question if they allow an audio recording to be taken during the interview. After the participant agrees, they can sign the Informed Consent and several general questions will be asked.

Part 1 - General Questions

These general questions concerned the participants' job titles, tasks within that function and what their relation is with or within the Woo process. Af-

ter those questions, they are asked to tell about their experience with the Woo, and if there are any struggles or things that could be improved. If deemed necessary, follow-up questions are asked to gain further insight into their experience with the Woo process. Then the next part of the interview will be discussed: the proposal for concept extraction.

Part 2 - Concept Extraction
This part begins with the question of whether they have experienced any problems relating to interpreting or the interpretation of the Woo request. As a follow-up question, the participants are asked whether they have solutions to address this problem. Once answered, a description is given of the idea to extract topics, or as described in the theory, concepts out of a dataset to get an overview of topics that might be interesting for the Woo requester. A complete and detailed description of the proposal is given within the Artefacts Sheet. The participant can read through when further explanation is required.

Then three of the concrete process possibilities are shown to the participant. Figure 3.3 in the previous section, shows the first out of three proposals. The participants are explained that these examples are just broad opportunities, but any other ideas on certain parts of these proposals are encouraged to be shared. While presented with these three sheets, the participants are asked multiple questions regarding whether they think the concept works, if it would be beneficial within the Woo and in what part it should be implemented. This open method of presenting general process ideas was implemented to stimulate creative thinking and allow the participants to come up with their own interpretations and ideas.

After those questions, the participant is asked whether it would be necessary to give an official letter with the topics they have chosen with the help of the software. Finally, they are asked whether they think it is necessary to do this to be transparent and document the choices they have made. When deemed required by the researcher, follow-up questions are asked. Then the participant will be informed that the last part of the interview will be discussed.

Part 3 - Timeline Generation Software

Before giving an introduction to the software, the participants are notified that the concept extraction process is separately researched from the process of timeline automation. Therefore, the specification of the request could either be performed by a conversation with the requester or with the concept extraction.

Thereafter, an introduction is given to the concept of the timeline generation software. Then an example of what such a software would output is shown to the participant. This example is based on real software that has been created by the Delfts University team Epoch, on behalf of the National Archive. For further information see this link. After explaining the software, the participants are asked whether they have a clear image of the software and its purpose. When required, further explanation will be given. When a clear image is given, the participant is asked questions like what they think of the software, if it would fit within the Woo process, if they can think of situations where it would or would not work and if they have any other comments.

Once these questions are answered, the participant will be presented with a set of design principles. To minimise the time commitment required from our participants, only 6 out of 12 principles are answered. A participant could either get to see principles 1 till 3 and 7 till 9 or 4 till 6 and 10 till 12. These 12 principles can be seen in Table 3.2 from the Materials section. When showing the participants the design principles, they are asked to tell what they think about the principles, whether they think it is useful within the Woo process and if they could rank the three options per principle on usefulness. This is done for all principles that they have to cover.

Ending

After going through all the principles, they are asked whether they would think it should be possible to create multiple timelines for comparison reasons. After that, they are asked if they have any further additions to what has been discussed or about the research itself. If there are no further questions, the participant is thanked for their time, the audio recording is stopped and the interview ends.

The sequence of the procedure can be seen Figure 3.6. The full procedure and all questions of the study can be found in the *Information Structure* in Appendix D.

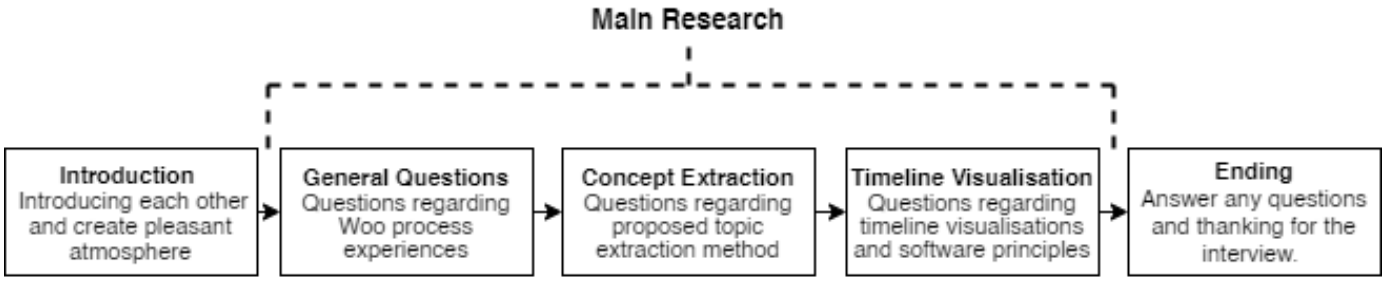


Figure 3.6: Sequence of procedure in interviews.

3.6 Data Analysis

A structured approach is employed to analyse the qualitative data gathered during the interviews. The Straussian Grounded Theory is not only used in the research to check its effectiveness during the Woo process but it will also be used to analyse the data gathered from the interviews. The following described steps will be executed four times in total. To find the opinion and view on the current Woo process, for the concept extraction part of the study, the general view on a timeline generation software and on the specific principles as proposed during the study. The first step of analysing the interviews is to transcribe the audio files.

Transcribing audio files The first step in performing this Grounded Theory is the transcription of the audio recordings. Therefore Utrecht University recommends the software Word Online, to automate a large part of the transcription process. Removing errors and language mistakes within the transcription is still necessary. The transcriptions will have time stamps to ensure that certain audio fragments can easily be accessed after transcription.

After the transcription of the audio files, the transcriptions will be coded. Coding is the process of categorising specific sentences within the transcriptions. Afterwards, these codings can be compared to see which participants agree with each other on certain topics. Therefore this research not only examines whether the Grounded Theory can be applied within the Woo process but will also use the Grounded Theory to analyse the transcriptions and retrieve relevant concepts. This method helps to objectively retrieve data out of qualitative data such as transcriptions which can later be used for further analysis.

Open Coding

After the transcriptions are created, they will be loaded into the qualitative data analysis software Nvivo. This software allows transcripts to be coded manually or automatically following the steps of the Grounded Theory. After loading the transcriptions into the software, the transcriptions are open-coded. This step involves highlighting text within the data and categorising

them into a code. During this process, questions such as who is involved in the process, what they think of the concept, how it can help or how it should be implemented are considered. Text over different transcriptions can be categorised under the same or different codes. By applying this technique an overview is created of particular codes across various participants.

To reduce the possibility of researcher bias, 2 of the 10 transcriptions are coded together with two other researchers with experience performing the Straussian Grounded Theory. This establishes an initial set of defined codes. After that, the found codes are discussed among researchers to reason why certain codes are created. Thereafter, the next stage of the Grounded Theory, *Axial Coding*, will be executed.

Axial Coding

The next step is to create a hierarchy between the established codes. Therefore all codes are read through to gain an overview. Considering all those codes, a hierarchy is established through the use of clusters with a bottom-up approach. Then all other codes are considered to either be categorised on their own or added to one of the other categories. This creates a hierarchy of codes. After that, it is important to determine if the categories can be further classified into main categories. That is the last step of the Grounded Theory called *Selective Coding*.

Selective Coding

Selective Coding can be executed after all hierarchies are established. This part specifically looks at generating core categories out of the hierarchies. These categories are established through the highest categories of the axial coding process.

Ranking

Another type of data that is gathered during the interviews is the given rankings of the proposed principles. Because of the scarce number of rankings given, the only analysis that will be executed is to calculate the average rankings per group and create new rankings given the average. This analysis aims to give numerical information about the rankings and can provide insight into the qualitative data analysis.

4. Results

The data collection took place in the period of 13-04-2024 up to and including 14-05-2024. This resulted in the total data of 10 participants (n = 10). Eight of the interviews were held physically in different locations in the Netherlands and two of the interviews were held online. Furthermore, the first two participants of the study were interviewed simultaneously. The participants were divided into different participant groups as can be seen in Section 3.3. The analysis performed on these interviews is primarily qualitative as the Straussian Grounded Theory was performed. However, a small quantitative analysis was executed on the rankings given by the participants. This chapter will cover all the results of the distinct parts of the interviews. The retrieved codes will be highlighted in italics throughout the results.

In Section 4.1, the current Woo process will be discussed. Then in Section 4.2, results regarding concept extraction will be examined. Finally, the results of the automated timeline-generating software are discussed in Section 4.3. This section will make a distinction between the general view of the software itself and the different design principles that are examined.

4.1 Current Woo Process

This part of the interview asked the participants about their thoughts on the current Woo process. Therefore, questions were asked on the process itself, pitfalls, and improvement points. These questions yielded a substantial amount of information. Figure 4.1 shows an overview of all the codes gathered during this part of the interview in relation to the different groups of participants. The participant groups are shown on the left and the lines show which code they have mentioned. Therefore, the lines show what participant groups have experienced or seen the specific types of problems. Some codes have been categorised further into smaller problems once deemed necessary. This visualises shared problems, concerns, and requirements regarding the Woo.

Information Management

One of the biggest problems currently faced within the Woo is *Information Management*. Participants 1, 2, 4, 9, and 10 mentioned that this is the biggest problem within the current process. It is important to know that these problems have occurred in some situations and with different governmental organisations and therefore do not apply to all. Participant 4 stated that the Woo handlers are *dependent on other colleagues* to deliver the right documents to them. Some cases don't allow the handlers to search for all the required documents themselves and therefore depend on other colleagues to hand them specific documents.

This includes another problem in information management: *difficulty in locating information*. Participants 1, 2, and 5 mentioned that there is no single database or software that is used all over the government organisations. Participant 9 has mentioned that even within government organisations, departments can use the same software but with a variety of settings that might create difficulties during collaborations between departments.

Lastly, the *disorganisation of information* artefacts even resulted in a Woo case where the *government lost indexed files* according to Participant 10. The participant mentioned that they knew the document existed since it was referenced in another public document.

Requester Concerns and Experience

On the requester's side, multiple concerns have been raised. These concerns were stated by all groups since they have experienced it themselves or heard from other requesters. Participant 9 discussed that *fishing for information* has taken place to get more documents than is necessary. Feeling the need to *limit the request* as much as possible because they feel pressured by the handler, as mentioned by Participant 3, which results in *distrust in the handler*. Participant 3, mentioned that among some requesters, there is a *distrust that all the documents have been retrieved*.

Handler Concerns and Experience

On the handlers' side, multiple concerns and experiences have been mentioned. One is the concern for a *change in mindset* as mentioned by Participant 1. Since some handlers are still a bit *hesitant to publish data* during a Woo request, a change of mindset needs to be enforced. Furthermore, the participant mentioned that some handlers or managers try to *transfer responsibility* to others to not be held accountable when wrong or privacy-sensitive information has been shared with requesters.

Requirements for Smooth Process

Another category of things mentioned during the interviews is some requirements that smoothen out the process of the Woo. Participants 3, 4, and 9 mentioned that *transparency is necessary* during the process and involving the requester in choices is necessary. Furthermore, Participant 9 mentioned that during the current process, *timelines are sometimes used as an overview* of how events happened to inform the requester.

Problems in Process

During the Woo process, multiple problems occur. One of the problems is the difference in process between *regional and national* governmental organisations. Participant 8 mentioned that some regional organisations struggle to give out information and work less structured. However, on a national level, there is a more professional and structured process. Additionally, it has been mentioned by participants that the *delivery of documents* has been a problem. This problem includes: *1. documents which are merged in one file 2. an overload of documents is delivered* or *3. too much information is redacted*.

The insufficient *availability of staff* to respond to requests has also been identified as a problem. This results in *external hiring* or *too much workload* that is given to an individual. This in turn results in a massive delay in information delivery. Another issue that has been mentioned is the fact that handlers *do not have knowledge* about the request itself. As Participant 3 has said, it is sometimes hard for the handler to perform the conversation, because the requester and handler do not know what information is available. Furthermore, the handler is not always the policy officer of that subject. Therefore, policy officers need to be asked to help with the case to deliver the right documents, which costs time.

Other problems that directly affect process efficiency is *slow decision-making*. Participant 2 argues that one of the biggest problems within the Woo process is the *signature line*, or in Dutch *Parafeerlijn*. This is the process of getting approval from management to continue within a process or to publish results. Since this can happen multiple times within the process, it costs time to wait for approval. This *sluggish process* causes deadlines to be exceeded.

Transparency has also been identified as an important problem within the Woo process. Since documents can be privacy-sensitive, they need to be handled carefully. However, that affects the transparency that is shown to the requester. One example that is mentioned by Participants 1, 2, and 10 is that after the information is gathered, the requester is presented with a *list containing the titles* of the artefacts. These titles are the only way to choose what documents they want. Though *titles do not always represent the content* of the file. Therefore, the participant stated that *extra documents are requested* to cover the possibility of the documents being important.

Another transparency issue that occurs is the problem of *not involving the requester* enough in the handling process. Participant 3 discussed that some requesters do not know what steps have been taken to search for certain documents, which can cause suspicion. Participant 9 also stated another issue where other people reading a Woo online after publicising might *lack the context of the search process* or documents themselves.

4.1.1 Summary Current Woo Process

To summarise, a variety of problems have occurred within the Woo process. These problems concern information management, staff requirements or transparency issues. Furthermore, have the transparency issues resulted in attitudes and actions that are not beneficial to the process. The trust that stakeholders have in each other is troubling. This shows the need for a transparent and efficient solution as will be discussed in the Discussion Chapter 5.

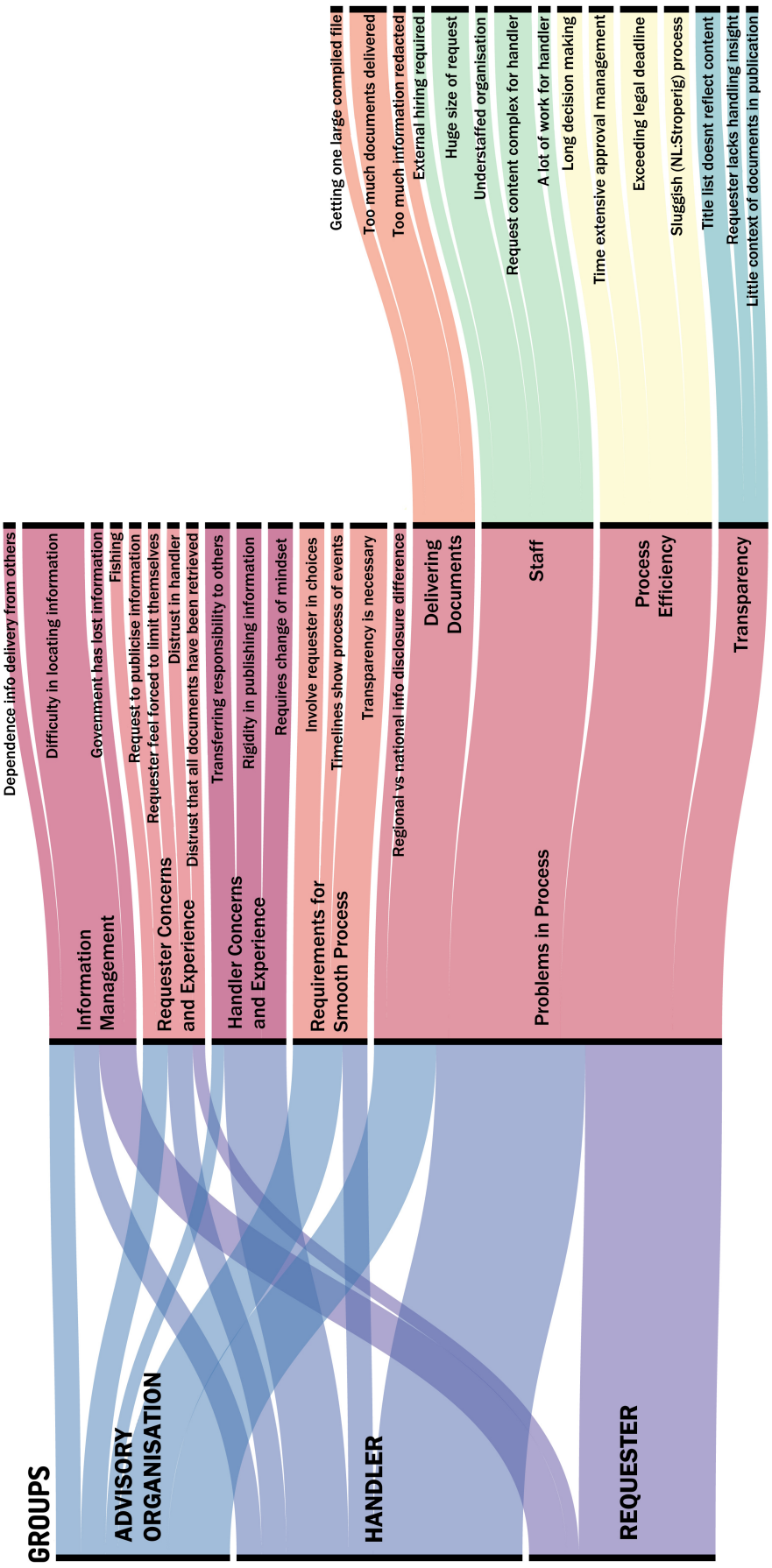


Figure 4.1: Codes found concerning current Woo Process

4.2 Concept Extraction

This section discusses the results that were gathered during the second part of the interview: concept extraction. Figure 4.5 shows all the codes and the hierarchical codes that were gathered during the Grounded Theory analysis. The colours within the diagram have been added to visually distinguish the different principles.

4.2.1 Software Characteristics

First, software characteristics that were gathered will be discussed.

Communication and Interaction

When showing the three possibilities for a concept extraction process, multiple characteristics regarding communication and interaction were mentioned. Eight out of ten participants mentioned that the proposed process of topic extraction can *help to specify* the Woo request. Participant 3 mentioned that the output of the proposed process could *help the process of locating documents*. Participant 9 also mentioned that this helps specify in a *structured method* instead of a conversation. Participant 3 mentioned that this could also serve as a *conversation starter* since you can go through the topics instead of trying to find them yourself. Furthermore, it allows the requester to *have more control* during the search process and provides an *overview of the search process* after publication. Moreover, it can help the handler *justify the choices* made in the search process with the help of the requested concepts. As discussed in more detail in Chapter 5, this implies that this method can help to improve a large problem within Woo: transparency.

User Needs and Challenges

The software also raised some concerns and challenges such as extra *context that could be desired* by the requester. Participant 4 stated that it remains *hard to know what type of documents* are behind a concept and what *risks you take when not selecting certain concepts*. Another challenge can be to gather an *initial set of documents* to extract the concepts from. Participant 3 discussed that the beginning of the proposed *concept extraction might not be the beginning of the Woo process* and first some initial conversations can be held.

Overview and Visualisation

Concerning the overview and visualisation, it was mentioned that it is useful to automatically generate an overview that helps to find topics discussed in the documents. Furthermore, Participant 5 mentioned that such a *process was also used* to gather documents on the Groningen gas request, which resulted in surprising topics that would not have been thought of. Finally, four participants mentioned that the last step of the proposed concept extraction would *help to give an overview of the correlation* between concepts as can be seen in Figure 4.2.

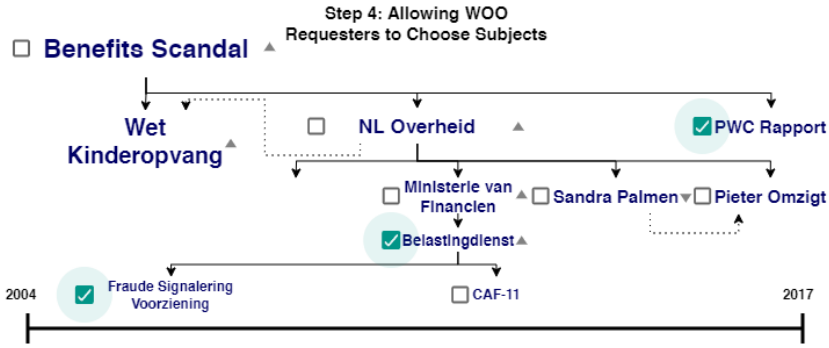


Figure 4.2: Visualisation of concept correlations translated to English.

Types of Request

The last discussed software characteristic is its beneficiality in different cases. Participants 1, 4, 6 and 7 mentioned that it *helps with large-scale requests* since specification within this process is beneficial. Furthermore, Participants 1, 4 and 7 mentioned that it would also be *useful for abstract requests*. Participants 2, 4, and 7 noted that some requests might be vague since the requester is unsure of their needs, and this helps to clarify. Furthermore, Participants 4 and 6 mentioned that *journalists* want to know the *steps of how policy choices* were made and the concept correlation can help with that. Participant 6 mentioned specifically that the match and count help to facilitate the choices.

Participants 2, 4, 6, 7, and 9 mentioned that this process is *unnecessary for concrete requests*. Furthermore, Participants 1 and 6 stated that requests concerning a couple of documents do not require this process. Participants 2 and 9 stated that it is *unnecessary for journalists* since they often know what they are looking for.

4.2.2 Features

During this section of the interviews, multiple features that would or would not benefit the software were discussed.

User settings and Preferences

One of the features that was mentioned to be useful by Participants 9 and 10, is the possibility to *switch between selecting important and unimportant* topics. The best functionality would depend on how many topics need to be discussed. On the other hand, Participants 9 and 10 argued that *only* allowing people to choose *what they want*, is better for specification since the requester thinks more about whether a concept found is necessary for them.

Overview and Summary

Concerning features for the overview and summary, one of the things that 9 out of 10 participants mentioned is the *necessity for a printout* of the agreed topics after the conversation. This creates transparency for the requester and helps to justify choices. Another feature that was proposed by Participant 4 is a *summary of the Woo request* to see whether the computer correctly understands the Woo request as it is imported. This summary can also be discussed with the requester to clarify. Participant 8 also mentioned that upon viewing the hierarchies of topics in the last part, *all concepts should be folded out* and not collapsed. The *wordweb*, as can be seen in Figure 4.3, was found to be a *chaotic overview* by Participants 6 and 9. However, Participant 3 mentioned that it might be *useful as an addition* to the concept list.

Figure 4.3: Visualisation of wordweb translated to English.

Search and Sorting Functions

Lastly, some search and sorting functions were discussed. One of the things that participants mentioned the software should have is the ability to *sort by document type*. Often, document types can already tell what kind of information the document contains, as mentioned by Participant 5. Furthermore, the ability to search for *documents behind concepts* was also requested by Participant 8. Since documents are already gathered, it would be an addition to *see the document or document titles*. Participant 9 mentioned another function: to *disable metrics* within the concept list to avoid the requester being confused by them.

4.2.3 Metrics

Within the proposed concept extractions, part 3 showed the metrics to be used for finding out whether a concept was relevant for the requester. Figure 4.4 shows the list of concepts with metrics as shown to the participants.

General

Participants 3, 6, 8, 9, and 10 mentioned that *descriptive metrics* about the concepts *can be useful* within this process to determine whether a concept is important for the request. Participant 9 also argued that these metrics *can be rather confusing* for the requester.

Count of Mentioned Concept

Participants 5, 6, 8, 9, and 10 specifically mentioned that they found the *number* of times that the concepts were mentioned over all documents, *useful* during the specification process. Participant 10 stated that it *provides insight* into the effects of selecting a specific concept. If it is not that relevant to a person, but is seen to be mentioned a lot over different documents can influence the person to select it or not.

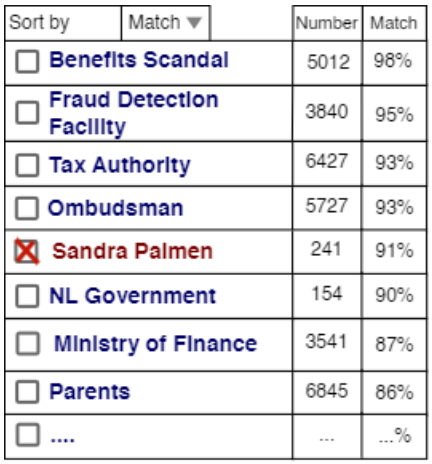


Figure 4.4: Visualisation of concepts with metrics.

Match with Request

Participants 3, 6, 8, 9, and 10 specifically mentioned that the *percentage of the match* with the Woo request is *useful*. Participant 6 suggested using it as an additional test to determine the request's significance, assuming the match is not adjustable. Participant 9 specifically stated that the match *can cause the most confusion*.

4.2.4 Requirements and Conditions

During the interviews, some requirements and conditions were mentioned for the process to work smoothly.

Communication

In the proposal, it was mentioned that the relevance coding was done by the handler only. However, Participants 3, 4, 5, 6, 8, and 10 mentioned that this *should be done together with the requester*. Participant 3 also states the *importance of using a handler* here to clarify things and help the requester through the process. The participant mentioned that this process should *not be used to replace a handler* but only to improve the process. Participants 5 and 9 mentioned that this process *should be done* together with the *policy officer and the information specialist* to clarify when requested by the requester.

User Experience

Other requirements and conditions concerning the user experience are the fact that the handler *must be able to* use the software properly and the requester and handler *must be willing to* use the software, as mentioned by Participants 3 and 7. Furthermore, Participants 1 and 2 mentioned that the *loading time should be short* to avoid irritation on either side.

Process Flow

One requirement mentioned by Participants 4 and 9 is that the process of concept extraction should be executed after an *initial conversation* with the requester to already clarify the unclarities and gain more information about the Woo request. As mentioned earlier, one of the conditions is that this software does *not replace human interaction*, as mentioned by Participants 3, 6, and 9. Participant 9 also mentioned that they assume there is a *risk that concepts are missing* since not all documents are used for this step and therefore a *disclaimer must be given* to the requester.

4.2.5 Summary Concept Extraction

To conclude, this section has discussed multiple ways in which the proposed methodology can support the current Woo process. It can help to create transparency as the requester is involved in the complete process of defining the search query. Furthermore, this defines accountability as there is an agreement on what information is required and what is looked for. The process also helps as a conversation topic to help specify the requesters' needs. Therefore it can be useful for certain types of requests. However, the results showed conditions that need to be met, such as a disclaimer, an initial conversation, and the necessity for a handler to ensure a successful implementation. These results will be further discussed in Chapter 5.

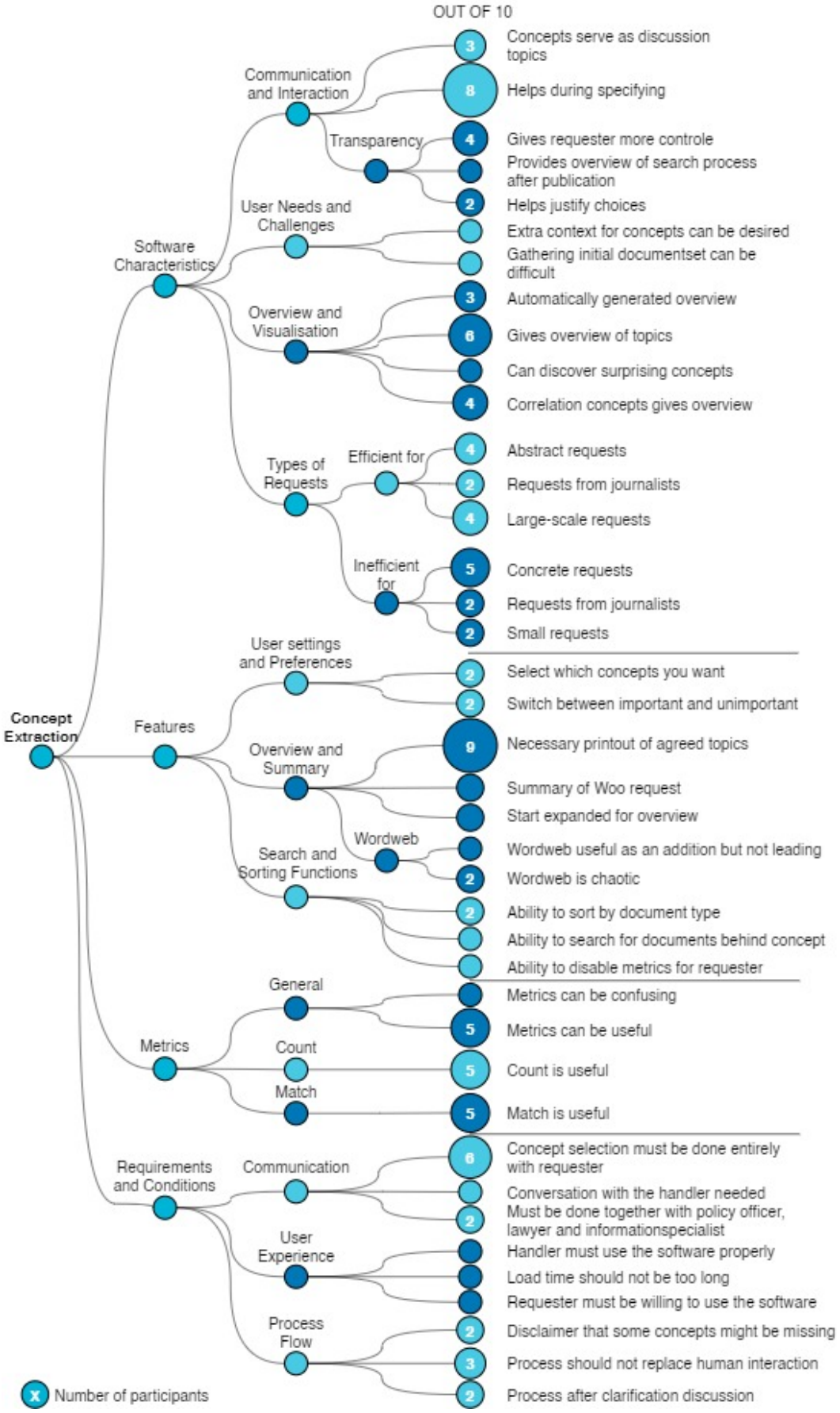


Figure 4.5: Codes found concerning the process of concept extraction

4.3 Timeline Visualisation Software

The third part of the interview concerned the automated timeline software where a separation was made between the concept of the software itself and specific design principles that could be implemented. First, the data gathered on the general view of the automated timeline generator will be discussed and then the specific design features will be discussed.

4.3.1 General

During the interviews, the participants were first introduced to the idea of automatic timeline-generating software. An image was shown to them to demonstrate the timeline. This image can be seen in Figure 2.3 in Section 3.4. Then the participants were asked to share their views on the software. The results from those questions are discussed in this section. Furthermore, Figure 4.6 shows an overview of all the codes gathered.

Communication

Concerning the communication, multiple comments were discussed. Participants 3 to 6, 8, and 9 mentioned that showing events in a timeline *gives an overview* of the process by which the documents are created. Furthermore, Participant 4 mentioned that a timeline like that *can be helpful as a conversation topic* since it shows a lot of information within visualisation and can therefore be used as a *structure for specifying* which documents are requested. Participant 2 mentioned that this visualisation of the timeline can also help to *simplify the signature line*. Giving context to the documents and showing the whole process can help management get an overview before signing the approval.

Document Management

Looking at document management, multiple codes are extracted. First of all, Participants 1, 3, 4, and 6 mentioned that the timeline *offers an overview of all the documents*. Participant 3 discussed that it can *help to find* important documents in a whole mass of documents. However, according to Participant 5, *journalists* often look at the *document level* rather than the process level. Participant 1 also stated that it is important to *indicate the document*

type within the timeline. Furthermore, Participants 5, 6, and 9 mentioned that an *extensive amount of documents can reduce overview*. Participant 5 argues that with some requests having thousands of documents, it seemed hard to keep oversight, even in a timeline. Participant 10 discussed that there is a *possibility for duplicates* and therefore could decrease the level of oversight. Participant 9 stated that names and other sensitive data in the *titles must also be redacted*. As discussed in more detail in Chapter 5, this implies that timelines could help within document management but multiple conditions have to be met.

Features

Two codes created during the analysis concerned creating multiple timelines for comparison. However, 5 out of 10 participants said that *multiple timelines are inconvenient and chaotic*. Participant 6 argued that when getting multiple timelines, they prefer to have them combined into one. However, Participants 1, 7, and 8 mentioned that, in some cases, *it might be useful* to have multiple timelines created for comparison.

Requirements and Conditions

Multiple requirements and conditions have been mentioned in relation to the automated timeline software. Participant 4 stated that when these timelines are *not used as a tool* but substantively, the *timelines have to be checked*. The participant adds that that is not what is desired. Participants 3 and 4 also mention that the software needs to *support and not replace*. Participant 2 argued that a prerequisite of the software working would be that the *information management works perfectly*. Furthermore, Participants 1, 2, and 4 mentioned that the software must *not be too hard to use* since it would reduce the adoption rate.

Software Properties

Multiple software properties were mentioned and coded. Participants 2, 6, and 10 noted that the software *helps to create a clear delineation*. Participant 6 stated that every attempt to streamline the documents required is encouraged. Furthermore, Participant 4 argued that this software would also be *good to use in other scenarios* like parliamentary inquiries, internal document overviews or other specific processes. Participant 9 mentioned that such software is a *good solution for document selection* and that in Miami Florida, United States, a similar method is already used. Lastly, Participant 4 stated that there might even be a possibility that *certain documents can emerge* that show potential turning points.

Transparency

Looking at codes concerning transparency, Participant 4 argued that this *software creates transparency* towards the requester since the computer creates the overview of documents. The participant adds that the computer acts as a *neutral actor* which is good, but there needs to be a disclaimer that this is a tool and there is *no guarantee for correctness*. Participants 3 and 4 also mentioned that they think it would be useful if the generated timeline is *added to the publication* documents. Participant 10 also argued that it would be good to *use the software after publication*. The participants argue that it would be useful to have the possibility to see the timeline and possibly *request the publication of documents* that haven't been released during that Woo request. As discussed in more detail in Chapter 5, this implies that both concept extraction and timeline visualisations help with creating transparency for the requester.

Types of Request

During the interviews, participants were also asked if they thought the software would be efficient or inefficient in certain cases. Their answers to those questions have been coded during the analysis. Participants 1 and 6 mention that it would be *good to use* in situations where requesters have submitted an *abstract request*. Participant 1 also noted it would be good to use in situations where requesters are *unsure of what they want*. Participant 6 mentioned that it would work well for larger-scale requests to streamline the process. Participant 2 suggested using it for smaller-scale requests due to doubt about the software's practical functionality.

Participant 9 was also a *bit sceptical* about using it for *larger requests*. Participant 8 mentioned that he thought it would be *good to use it in all cases* since a requester is better able to decide whether it is relevant to them or not.

4.3.2 Summary Timeline Visualisation Sofware

This section discussed the results that were gathered of the general views on implementing a timeline visualisation software. The results showed that participants found it a useful method as the timeline can provide an overview which helps in the specification process and provides accountability for what documents are agreed to publish. This results in transparency as well as showing all gathered documents in context. Additionally, different types of requests have been discussed as well as their usefulness after publication. Furthermore is it useful as a conversation topic as it supports the Woo Handler. However, the software must support and not replace the Woo handler. Furthermore, there must be a disclaimer of possible inaccuracies.

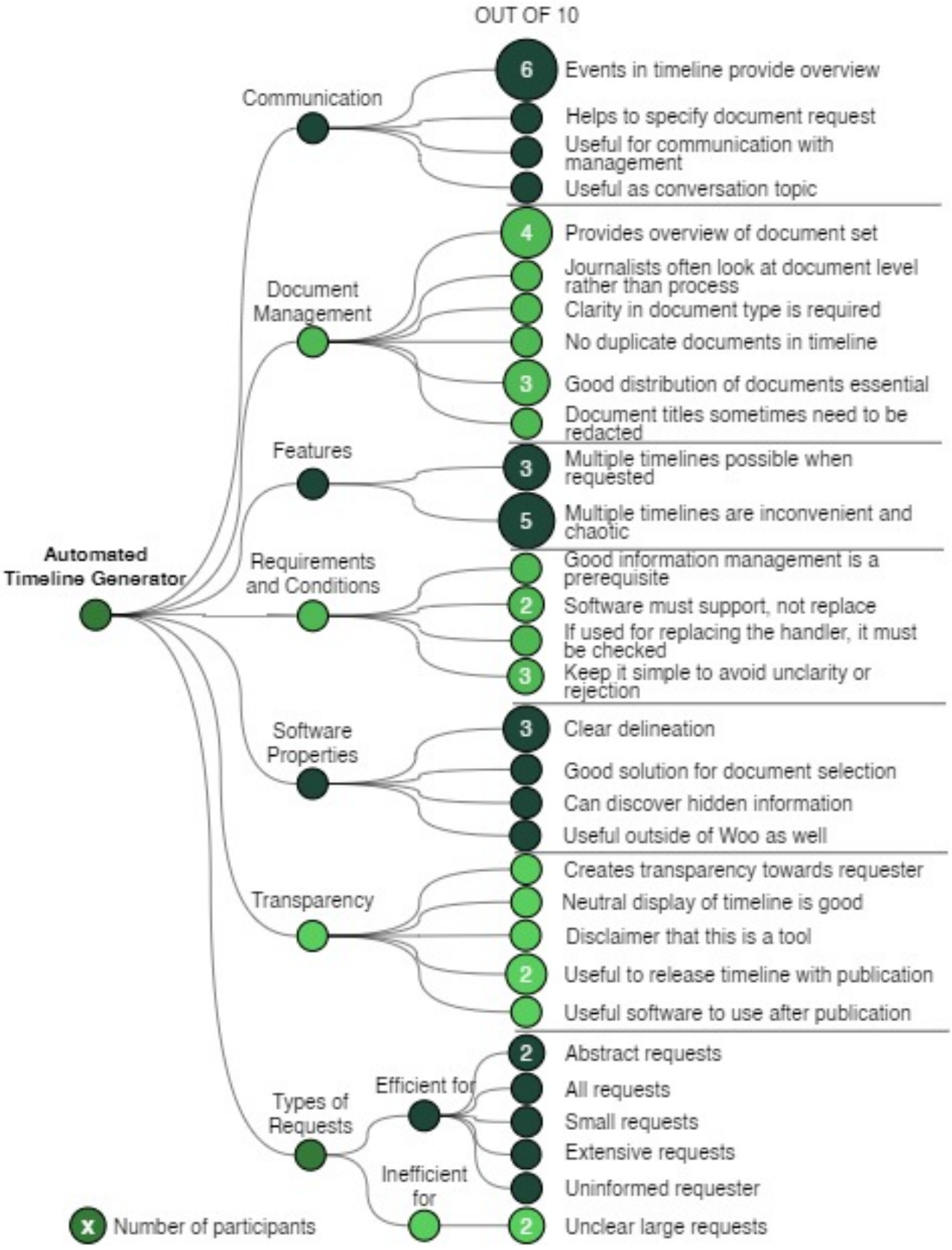


Figure 4.6: Codes found concerning the timeline visualisation software

4.3.3 Specific Principles

After getting information about the timeline software as a whole, further questions were asked about the specific principles that were proposed. Because of time constraints, all participants were shown six principles. Unfortunately, one participant refused to rank some of the principles and therefore, another participant was shown the same principles as the former participant. However, comments from that participant about the principles themselves are included in the analysis. That resulted in a dataset of six participants for principles 1-3 & 7-9 and four people for principles 4-6 & 10-12.

This division was used since the first six principles cover the implementation of basic features. In contrast to the first six principles, the last six cover extra features that might be useful. Therefore the division makes sure that all participants cover basic as well as extra principles. Figures 4.13 and 4.20, also indicate how many participants have mentioned a specific code, shown in the adjacent circle. To see the visualisations of the principles, please refer to Appendix C.

Principle 1

The first principle discussed was topic selection. This principle showed three ways of adjusting the timeline to show specific processes within the topic. Participant 6 mentioned that this would be a *good addition* to going through broad requests. To prevent seeing thousands of documents within one timeline, one can go through different sets one by one. Participant 2 mentioned that this would be *useful when submitting a Woo* request. The participant noted that if the information management was working properly, people could go through what information was available by means of the topics. Figure 4.7 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Suggested Choices. Participants 1 and 6 mentioned they thought the *Suggested Choices* would be *useful* to go through the *acquired documents* if the concept extraction process would be used beforehand.
2. Free Input. The *Free Input* option was claimed to be *useful* by Participants 2 and 10 with the expectation that it would give *a valid output.* Participant 10 argued that by using free input, certain *relations might be uncovered* that could require a confidentiality agreement. Relations between documents could be searched to gain a competitive advantage, for instance by looking at prices for medicines.
3. Important Concepts. Participant 2 mentioned that *Important Concepts* are not useful to use for this type of filtering. Participant 10 argued that this *should have been done a step before* and could raise distrust when asking about the specification of these concepts again.

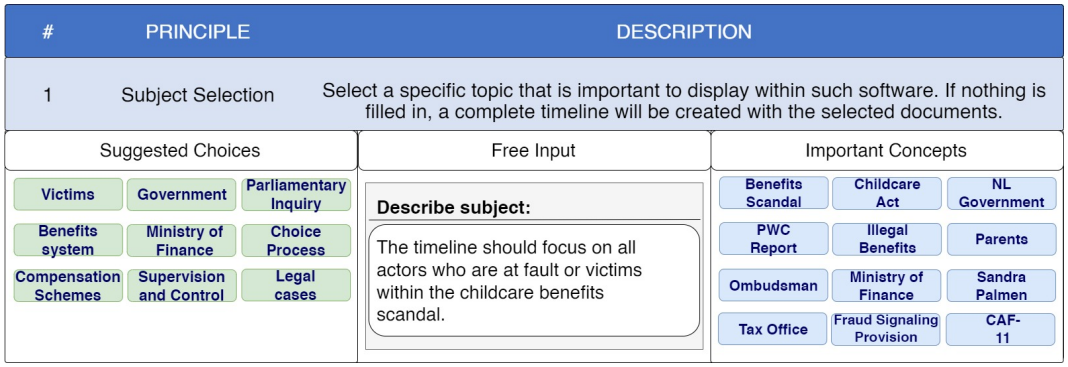


Figure 4.7: Principle 1, translated and enlarged as shown to the participants.

Principle 2

The second principle discussed, concerned selecting the beginning and end dates for the generated timeline. Participant 10 mentioned that it is important to make the decision possibility as simple as possible. Figure 4.8 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Fill in Date has been raised to be *very simple* to use by Participants 6 and 7.
2. Pre-set Slider has been indicated to be very visually pleasing to use by Participants 2 and 6. Participant 2 also mentioned that it would be useful to add the total duration time to the slider when moving.
3. Interval Slider was mentioned by Participant 1 to *look more like the output* of selecting the time rather than the input. Participant 6 mentioned that the interval slider is *not that useful* because it does not feel intuitive.

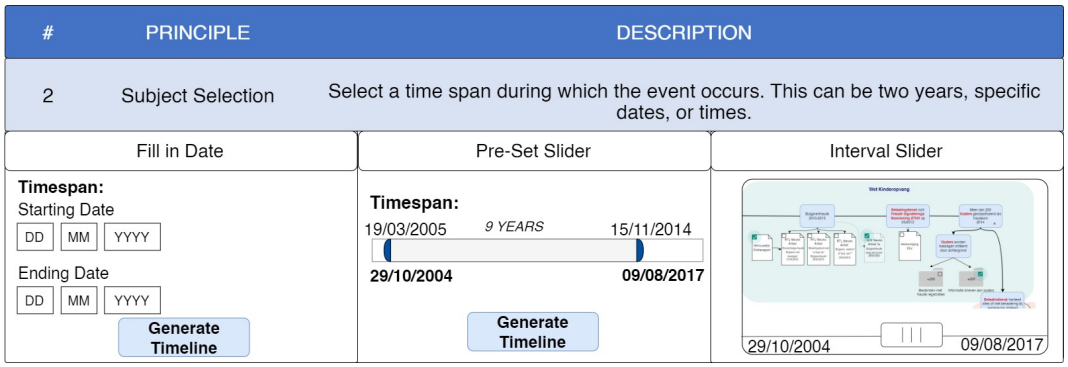


Figure 4.8: Principle 2, translated and enlarged as shown to the participants.

Principle 3

The third principle showed options to differentiate between actors, places and events. Therefore three possibilities were proposed to the participants to visualise actors, places, and events. Participants 7 and 10 stated that it would be *useful to look for actors*. Participants 2 and 10 mentioned that *events are also relevant* to search for and it would be useful to select them when requested. Participants 1, 5, 6, 7, and 10 discussed that they did *not find the place relevant* to be added. Participant 1 indicated that it would be *hard to retrieve* the information from documents and Participant 10 mentioned that rare cases might be *interested in the place*. Figure 4.9 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Filters was one of the *more pleasant* options to use according to Participant 10.
2. Selecting option was stated to be *unnecessary and a lot of work*, by Participants 1 and 10, since 200 events would require 200 actions. Participant 10 added that *it works as a filter* on a small scale.
3. As a Basis. According to Participant 1, it would be *pleasant to keep seeing* all the information within the timeline but highlighting some parts of the timeline could be useful.

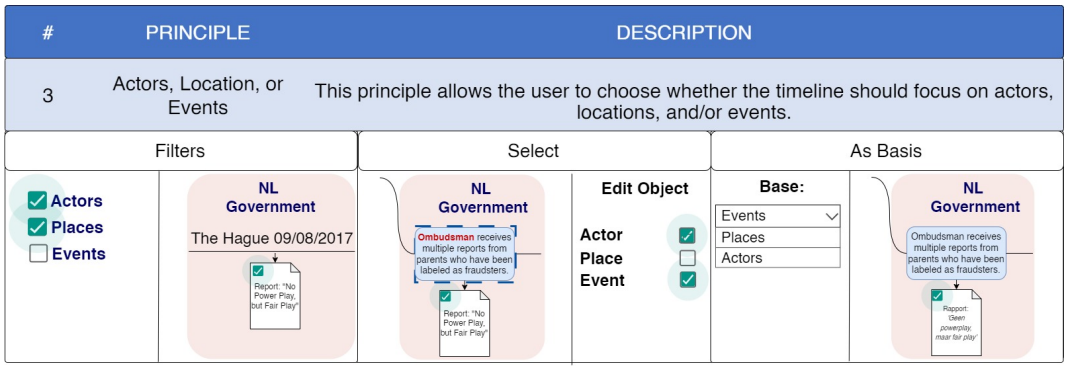


Figure 4.9: Principle 3, translated and enlarged as shown to the participants.

Principle 4

The fourth principle discussed options to change the orientation of the timeline, which could be horizontal or vertical. When presented with the idea, Participants 4 and 9 mentioned that scrolling vertically *feels very natural* to do. Participant 9 added that one of the advantages of having a vertical timeline is that it becomes *easier to print*. Participant 3 stated that since there are only two options possible, the possibility of change *should be kept simple*. Figure 4.10 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Choice, presented the idea of using a dropdown menu to change the orientation. Participant 3 mentioned that that was *way too many steps* for a small change and that a dropdown should be used *when there are multiple options*.
2. Switch, presented a switch that could be turned to the left or right side. Participant 8 mentioned that that looked *very simple to use*.
3. Checkbox showed the two options with two boxes next to it. Participants 3, 4, 8, and 9 stated that this option looked like you could *select both horizontally and vertically*. Furthermore, Participant 3 suggests that one *option could be blurred* to avoid confusion in selecting both options.

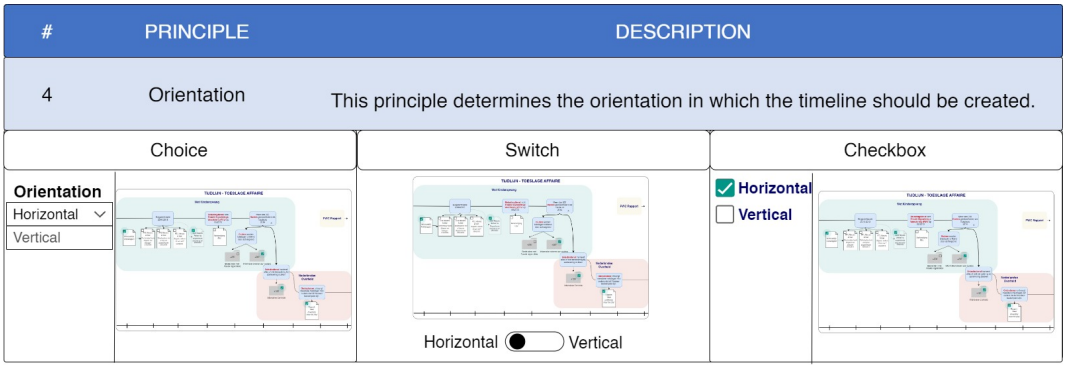


Figure 4.10: Principle 4, translated and enlarged as shown to the participants.

Principle 5

The fifth principle presented discussed the level of detail that was shown within the timeline. Participants 3, 8, and 9 mentioned that changing the level of detail was *important to keep the structure* and *frame parts* of the timeline. Participants 3 and 4 added that it is useful to have *multiple layers of event*s to have the ability to choose what layer you want to elaborate on. Figure 4.11 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Slider changed the level of detail by sliding. Participants 3 and 9 stated that this would *have to work well* and not be too sensitive. Participant 9 discussed a prior experience with annoying sensitivity. Participant 3 stated that *intervals would be useful* to avoid sensitivity. However, Participant 4 mentioned that a slider can *show too much information* at a time and that, in his experience, that is not how you read a timeline.
2. Zooming could become annoying when not working properly. The participant mentioned that when using zooming Google Maps, you *often don't end up where you want* to end up. The participant also mentioned that *zooming per section* can also be beneficial so that you are already within the right area to zoom.
3. Pull-Down Button was mentioned to be very useful by Participants 3, 4, 8, and 9 since the user can specifically *choose the area of expansion* and *level of detail*. Participant 3 also mentioned that *a combination* of the *Zoom* and *Pull-Down Button* would be a nice option to change the level of detail.

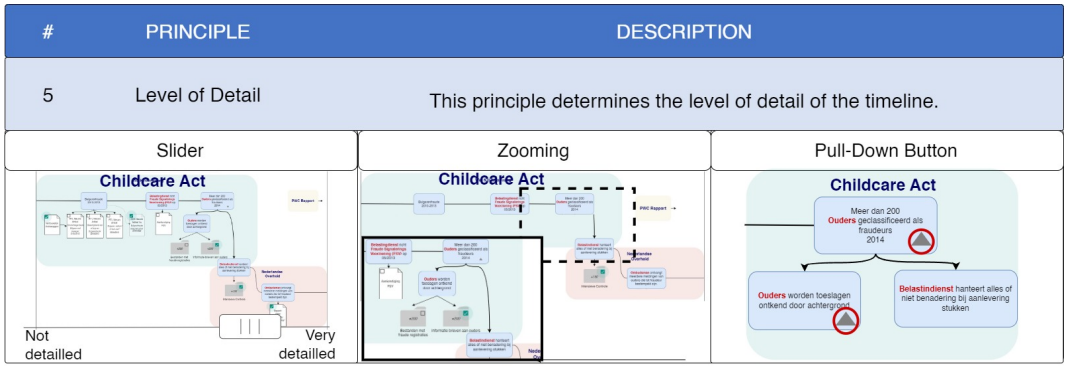


Figure 4.11: Principle 5, translated and enlarged as shown to the participants.

Principle 6

The sixth principle discussed was the principle of filtering. Participants 3, 4, 8, and 9 mentioned that filtering is a *very useful* function to have within this software since it helps to specify your search request. Participant 10, although not presented with this principle, has mentioned that filters are *essential to structure the search process*. Participants 3, 4, 8, and 9 added that it should be possible to *combine filtertypes*. Participants 3, 9, and 10 mentioned that the type of filter truly depends on the type of request. Figure 4.12 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Documents was mentioned to *already be used* extensively by Participants 4, 5, 8 and 9.
2. Subjects mentioned as *important* by Participant 9 but deemed *irrelevant* by Participant 4.
3. Entities/Actors was mentioned to be *hard to separate* by Participant 9.

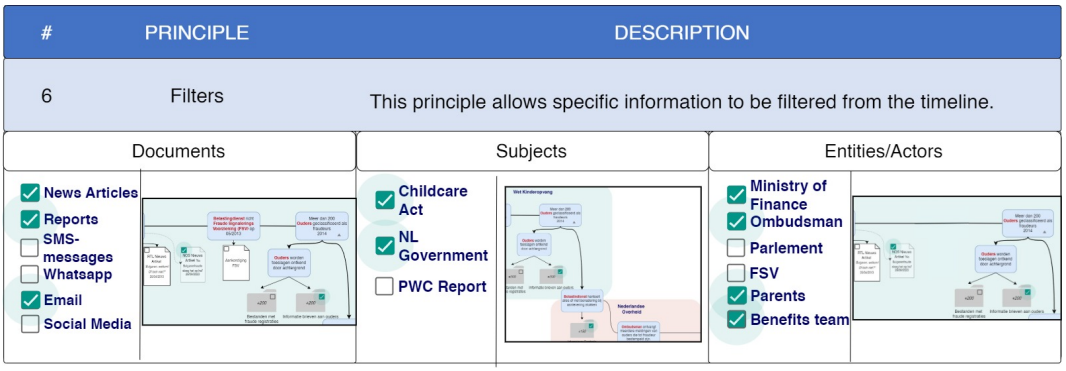


Figure 4.12: Principle 6, translated and enlarged as shown to the participants.

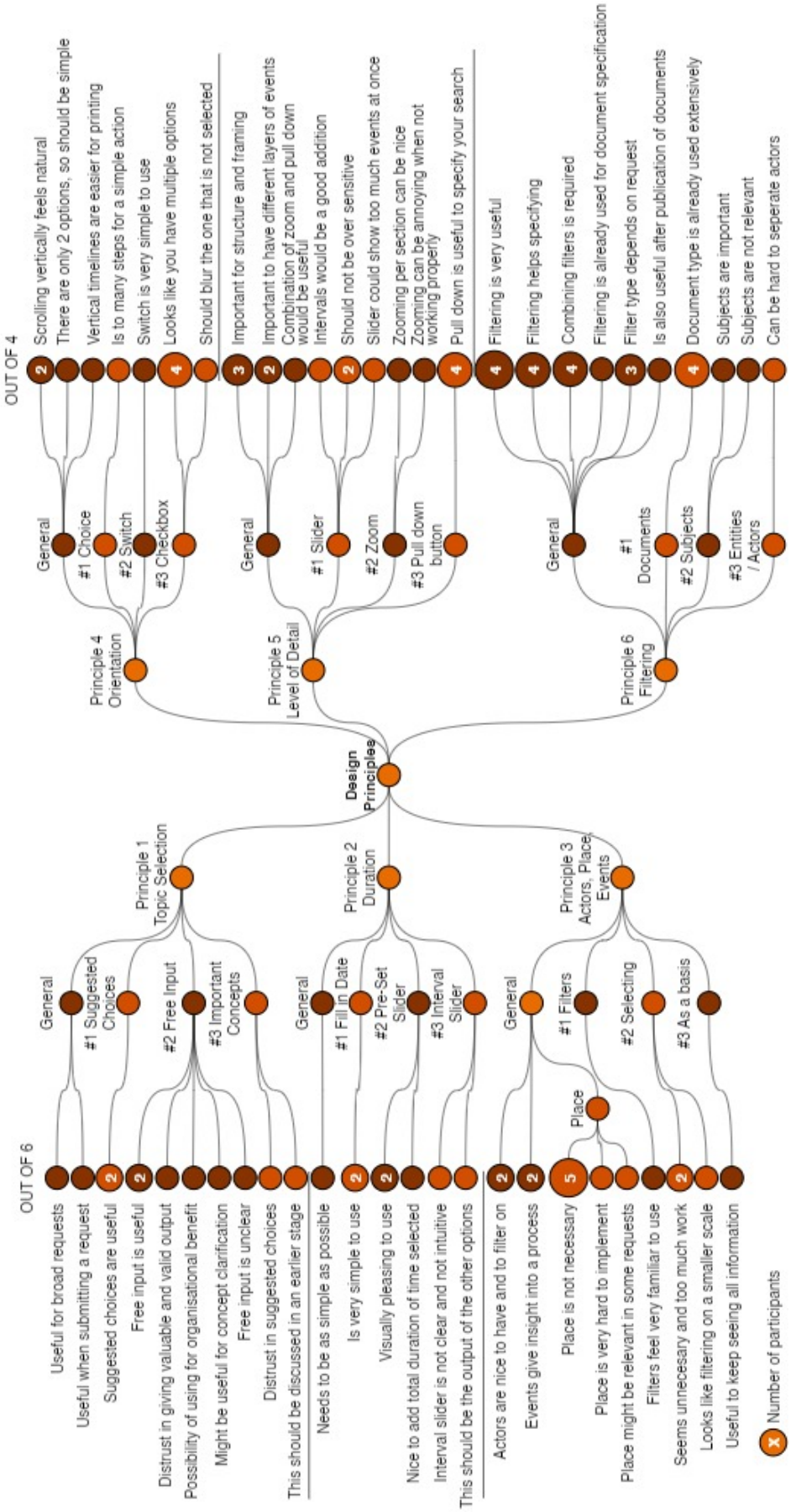


Figure 4.13: Overview of codes found concerning design principles 1-6

Principle 7

The seventh principle concerned time units. Time units refer to the unit in which the time is displayed within the timeline, for instance, days or hours. Participants 2 and 6 mentioned that the preferred unit *depends on the request* type. If you want to look at specific emails then hours or minutes would be a preferred unit. Participant 6 mentioned that *days are most preferred*. Participant 1 stated that it is *hard to filter* these timestamps from metadata since the data is not always correct. Participant 10 mentioned that it would also be good to add the *option to switch the order* from the newest to the oldest data within the timeline or the other way around. Figure 4.14 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Dropdown was mentioned to give the *best overview* according to Participant 2
2. Button was stated to be the *most intuitive* by Participant 6.
3. Detail Slider was mentioned to be the *least intuitive* by Participant 6.

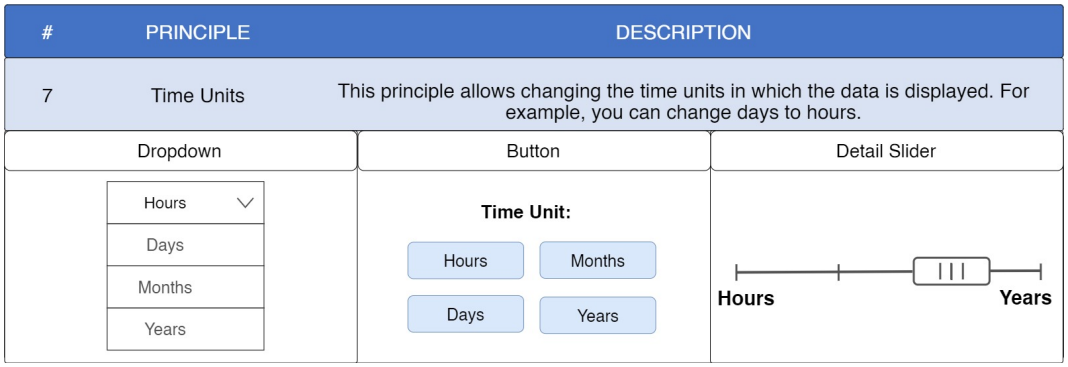


Figure 4.14: Principle 7, translated and enlarged as shown to the participants.

Principle 8

The eighth principle concerns the possibilities of interactivity with the timeline. Participant 10 mentioned that the type of interactivity can *depend on the request* size, requester or what the request is about. Figure 4.15 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Preview was considered to be a *useful addition* to Participants 6, 7, and 10. Participant 10 added that it can clear up unclarities on what type of document it is. Participant 10 also added that a preview should not be necessary for all documents to avoid having to go through all the previews.

2. Explanation was mentioned to be a *useful feature* by Participants 2 and 10. Participant 10 mentioned that it *could start a dialogue* since it can clear up unclarities. The participant added that it, therefore, should *maybe be added in the concept-defining phase* to avoid information asymmetry. Participant 2 added that the explanation *should be well-written* and pleasant to read.

3. Opening Folders was found to be *useful* by Participant 10 because it can specify the request and else you can still not see what the folder contains. Participants 2 and 6 did *not find the possibility useful* since they both were afraid that the document name still did not give away its contents

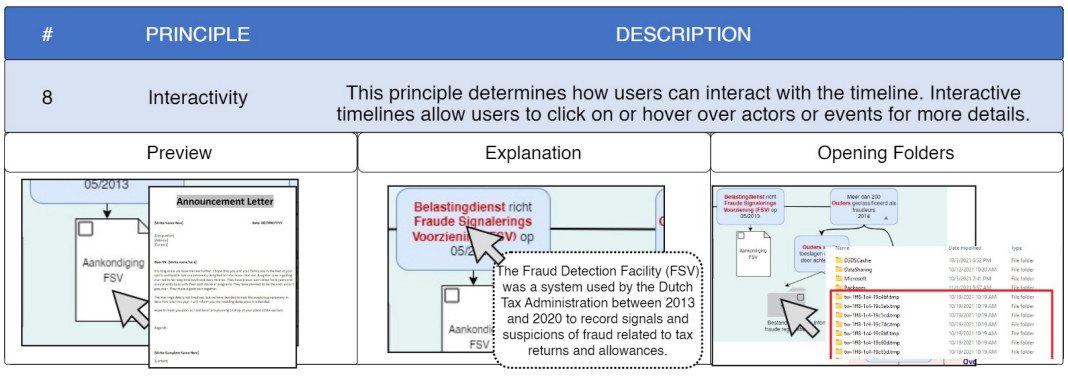


Figure 4.15: Principle 8, translated and enlarged as shown to the participants.

Principle 9

The ninth principle concerns the addition of multimedia. Participants 2, 7, and 10 thought this *could be useful* within the process. Participants 2, 6, and 7 stated that it currently is *not very common* because it is mostly textual but Participants 1, 2, 6, and 10 stated that this should be published more often. Participant 2 argued that *multimedia is a side issue* and should therefore maybe be a complete branch from the timeline. Participant 1 discussed that multimedia should *only be shared when requested* since it takes a lot more time to analyse those files compared to textual. Figure 4.16 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Audio was stated to be *useful* by Participant 10. Participant 6 discussed that a *summary* would be useful for audio files and Participants 2, 6, 7, and 10 mentioned that a *transcript* for the audio file should be appended. Participant 1 argued that it could be useful as long as you can *fast-forward* through it.

2. Videos have been stated to be *useful* by Participants 7 and 10. Participant 3 stated that *fast-forwarding* through videos is also essential. Participant 6 argued that a *summary of the video* would also be useful since it allows one to decide whether it is important without viewing it.

3. Images was mentioned to be the *most common* multimedia by Participant 1 and Participant 10. They stated that images *can be requested* specifically.

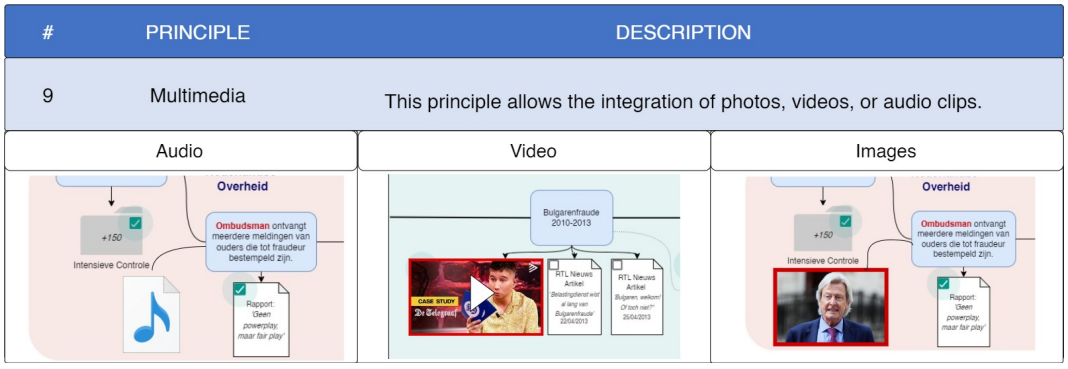


Figure 4.16: Principle 9, translated and enlarged as shown to the participants.

Principle 10

The tenth principle concerned the use of labels and markers within timelines. Participants 3, 8, and 9 mentioned that some *handlers might want to use* labels and markers, though adding these one by one was stated to be *too much work* by Participants 4 and 9. Participant 3 discussed that it might be good to add an *explanation on how to read* these labels and markers to the requester. Participant 3 also argued that it *can become quite unclear* when too many labels and markers are added. Figure 4.17 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Labels was argued by Participant 3 could help to *clear up* information without giving extra text, but *should be visualised better* than was presented, which Participant 8 also mentioned. When labels are added, it should truly be of *added value to the visualisation* as stated by Participants 3 and 4. visualisation.

2. Markers was stated by Participant 3, to only be visible when zoomed into the matter. The extra information should *only be an addition* and not shown primarily in the timeline.

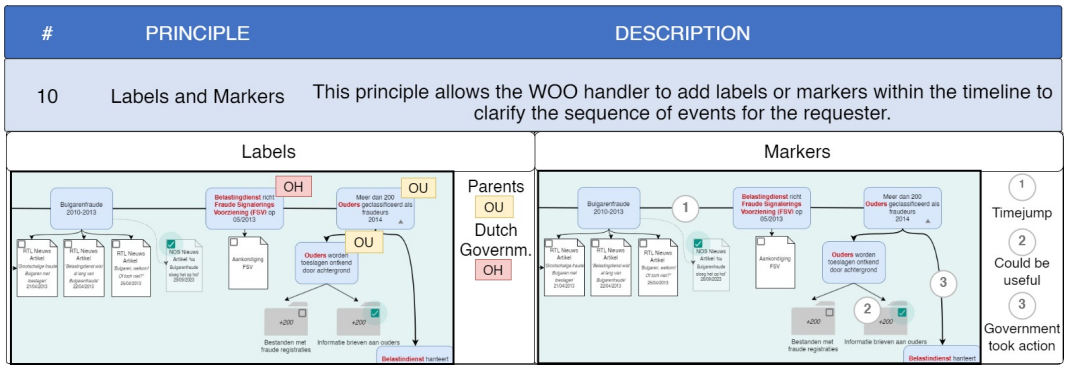


Figure 4.17: Principle 10, translated and enlarged as shown to the participants.

Principle 11

The eleventh principle discussed different types of *Design Elements*. Participant 4 mentioned that it is important to have a *visually attractive* timeline to avoid dismissal of the content. Participant 8 argued that it should be possible for the handler to *change the design* of the elements when requested. However, Participants 3 and 9 stated that too much difference in visualisations *can cause confusion*. Participant 9 added that handlers might create their *own colourful versions* of the timeline which can confuse requesters. Figure 4.18 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Direct Adjustment was stated to take too much time for the handler by Participants 4 and 9.
2. Menu was also stated to take too much time for the handler by Participants 4 and 9.
3. Themes could be useful but *standards for the visualisation* of the timeline were found to be the best by Participants 3, 4, 8, and 9. Participant 2 discussed that it would be useful for the requester if, the timelines across ministries have *similar aesthetics*. That allows the requesters to easily recognise the colours that represent, for instance, the government. Participant 9 mentioned that it would be useful to be implemented, but weighs the necessity to the fact that *departments at the ministries all work differently*.

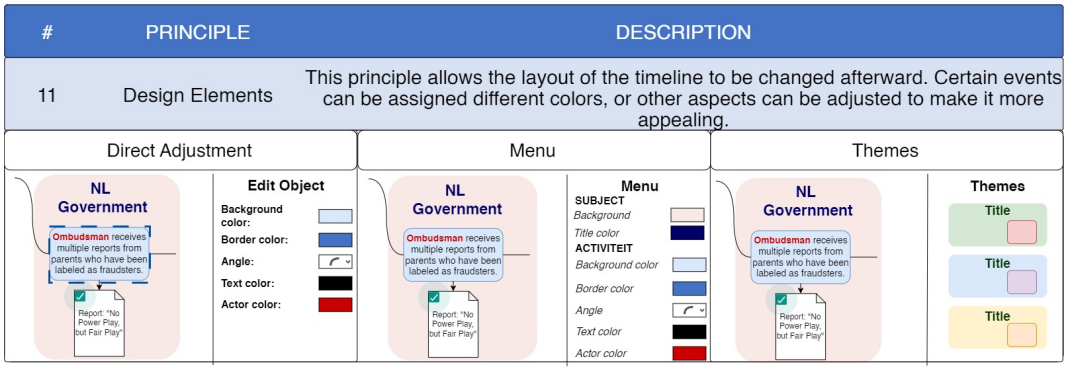


Figure 4.18: Principle 11, translated and enlarged as shown to the participants.

Principle 12

The twelfth and last principle discussed the possibilities for *Manual Editing*. Participants 3 and 4 mentioned that allowing the handlers to edit the content of the timeline, *decreases the value of automation and neutrality* of the system. Participant 4 added that if this software would be a replacement for the handler itself, the *possibility of editing* should be available. The participant adds that this software's *strength is to be used as an addition* to the current process and not as a replacement. Participant 3 added that if handlers are allowed to edit manually, the *edit history* should be available. Participant 9 added that it would be useful to *change small mistakes* such as weird text and Participant three mentioned that it would be useful for *optimisation*. However, Participant 4 stated that it is *okay* that the timelines *contain small mistakes*. Figure 4.19 shows the different propositions as presented to participants and below are the results concerning those propositions.

1. Direct Adjustment was argued to take too much time by Participant 3.
2. Text Only would be useful to change small mistakes. These mistakes were recognised to be possible because of spelling mistakes in titles or documents.
3. Dragging was stated to be not that useful by Participants 3, 4, and 9. Participant 9 also argued that the software should make sure that objects are placed correctly.

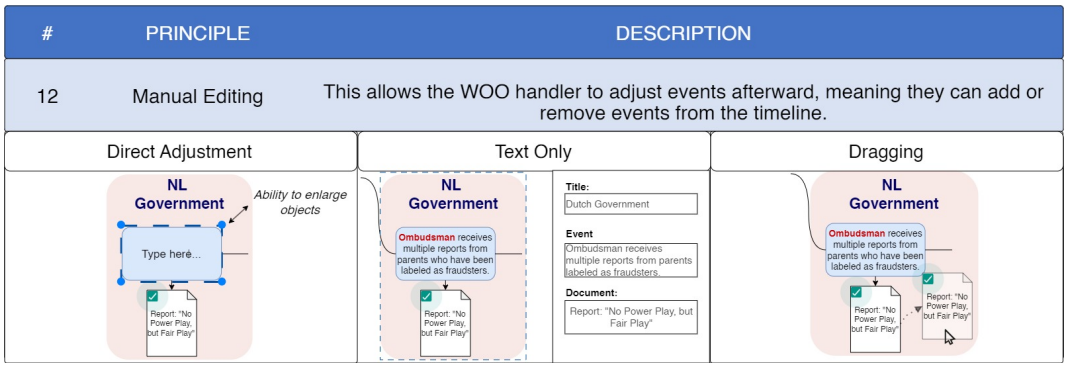


Figure 4.19: Principle 12, translated and enlarged as shown to the participants.

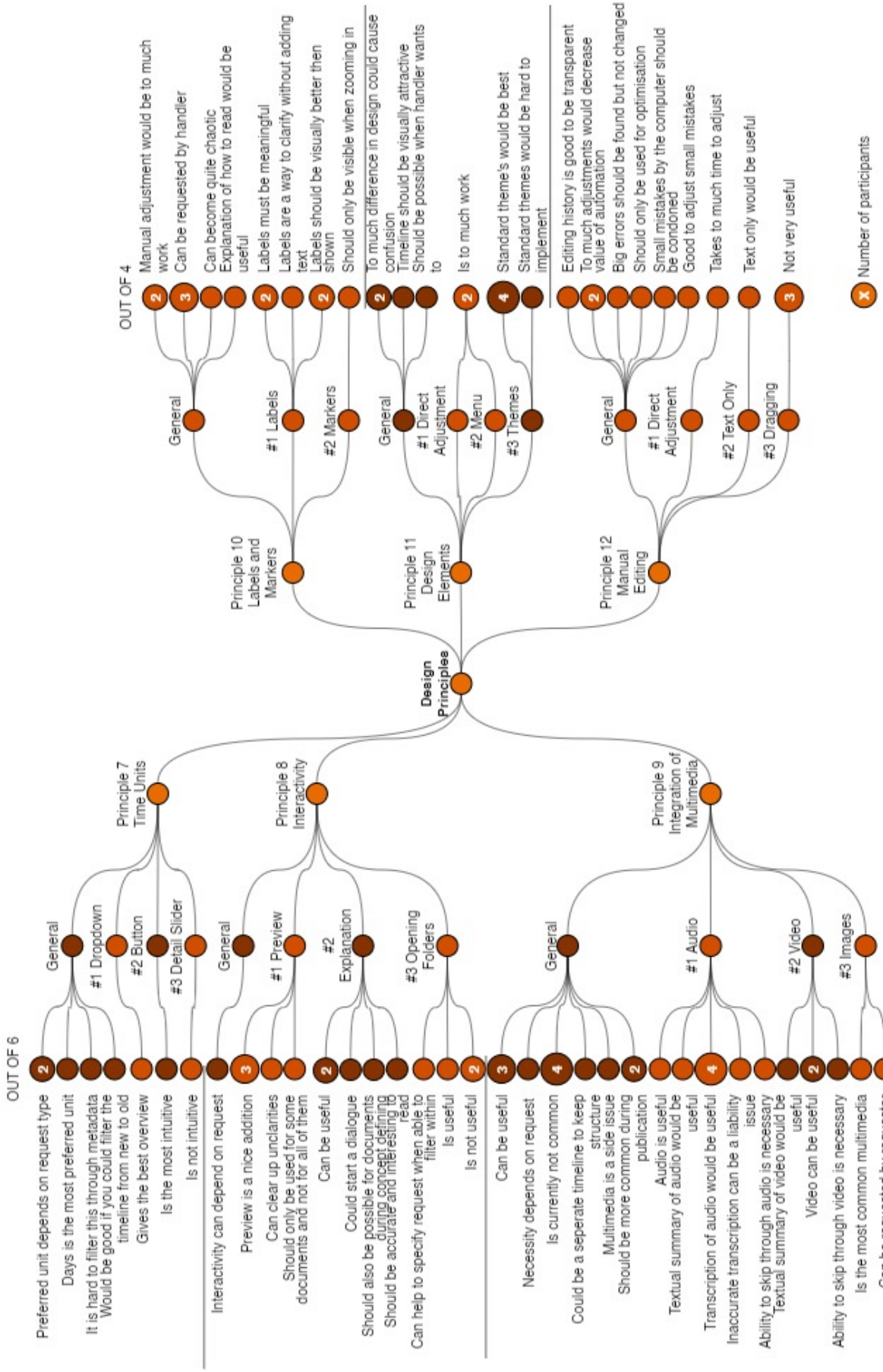


Figure 4.20: Overview of codes found concerning design principles 7-12

4.3.4 Ranking of Principles

An extra set of data was gathered during the interviews to get an insight into what ranking the participants gave to the principles. Unfortunately, one participant refused to rank some of the principles and was removed from the dataset for ranking analysis. That resulted in a dataset of 5 participants for the principles 1-3 & 7-9 and 4 people for the principles 4-6 & 10-12. Additionally, 2 participants from the Advisory Organisations have been presented with the same set of principles. Therefore, no data was gathered from that group for the principles 1-3 & 7-8. Table 4.1 shows a table with the names of the proposed possibilities as presented to the participants. This table was also shown in Chapter 3, but for convenience purposes, it has been chosen to show it again. A visualisation of the different options of each principle can be found in Appendix C. Table 4.2 shows the results from the ranking divided into three participant categories and as a total. Within the table, four groups are mentioned. The Advisory Organisation, Handlers, Requesters and a total group.

Table 4.1: Table of design principles with the names of the proposed possibilities.

Principle	**Option 1**	**Option 2**	**Option 3**
1. Topic Selection	Suggested Choices	Free Input	Important Concepts
2. Duration	Fill in Date	Pre-set Slider	Interval Slider
3. Actors, Place, Events	Filters	Selecting	As a Basis
4. Orientation	Choice	Switch	Checkbox
5. Level of Detail	Slider	Zoom	Pull-down Button
6. Filtering	Documents	Subjects	Entities/Actors
7. Time Units	Dropdown	Button	Detail Slider
8. Interactivity	Preview	Explanation	Opening Folders
9. Integration of Multimedia	Audio	Video	Images
10. Labels and Markers	Labels	Markers	-
11. Design Elements	Direct Adjustment	Menu	Themes
12. Manual Editing	Direct Adjustment	Text Only	Dragging

Table 4.2: Average rankings for three participant groups concerning 12 principles

Principle	Advisory Org.			Handler			Requester			Total		
	A1	**A2**	**A3**	**B1**	**B2**	**B3**	**C1**	**C2**	**C3**	**D1**	**D2**	**D3**
1	-	-	-	1.33	2.33	2.33	2.00	2.50	1.50	1.60	2.40	2.00
2	-	-	-	1.67	1.33	3.00	2.00	1.00	3.00	1.80	1.20	3.00
3	-	-	-	1.33	2.67	2.00	1.00	3.00	2.00	1.20	2.80	2.00
4	3.00	2.00	1.00	3.00	1.00	2.00	2.00	1.00	3.00	2.75	1.50	1.75
5	2.00	2.50	1.50	2.00	3.00	1.00	3.00	2.00	1.00	2.25	2.50	1.25
6	2.00	2.00	2.00	2.00	1.00	3.00	1.00	3.00	2.00	1.75	2.00	2.25
7	-	-	-	1.00	2.00	3.00	2.00	1.00	3.00	1.40	1.60	3.00
8	-	-	-	2.00	1.67	2.33	1.50	3.00	1.50	1.80	2.20	2.00
9	-	-	-	1.67	2.00	2.33	2.00	2.00	2.00	1.80	2.00	2.20
10	1.50	1.50	-	2.00	1.00	-	2.00	1.00	-	1.75	1.25	-
11	3.00	2.00	1.00	3.00	2.00	1.00	2.00	3.00	1.00	2.75	2.25	1.00
12	2.50	1.50	2.00	2.00	1.00	3.00	1.00	2.00	2.00	2.00	1.50	2.50

Looking at the table, it can be seen that some principles have a clear distinction of preference. These rankings either have a clear distinction between the group rankings or total rankings. Figure 4.21 shows the difference in the highest and lowest ranking. The difference between the highest and lowest total ranking of principles 2, 3, 7, and 11 is higher than 1.5. Knowing that the highest ranking can be 1 (most preferred), and the lowest ranking can be 3 (least preferred), there is a maximum difference of 2. Other principles where the total ranking difference is equal to or lower than 0.5 are 6, 8 and 9. This shows that these principle possibilities rankings are more distributed and therefore decrease the significance of the total rankings. This difference can be caused by heavy differences between groups or participants. For principles 6 and 8, there seems to be a large difference between the groups whereas principle 9 shows a clear difference between the participants within groups. The difference between the participants can be seen since the handler and requester groups of principle 9 have a maximum difference of 0.66.

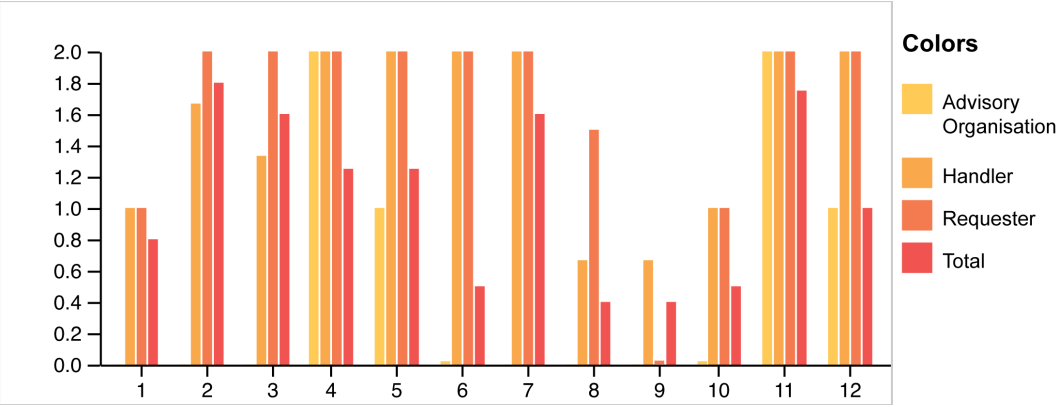


Figure 4.21: Difference in highest and lowest rankings

Using those averages a new rank table is created with the information on which proposed possibility is most preferred, ranked as 1, and least preferred, ranked as 3. If two possibilities score similarly, they will share the highest ranking. For instance, if you have the averages 2.5, 1.5 & 1.5, the rankings would be 1, 2, and 2 respectively. Table 4.3 shows the new table with the adjusted rankings.

Table 4.3: New rankings for three participant groups concerning 12 principles

Principle	Advisory Org.			Handler			Requester			Total		
	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3
1	-	-	-	1	2	2	2	3	1	1	3	2
2	-	-	-	2	1	3	2	1	3	2	1	3
3	-	-	-	1	3	2	1	3	2	1	3	2
4	3	2	1	3	1	2	2	1	3	3	1	2
5	2	3	1	2	3	1	3	2	1	2	3	1
6	1	1	1	2	1	3	1	3	2	1	2	3
7	-	-	-	1	2	3	2	1	3	1	2	3
8	-	-	-	2	1	3	1	3	1	1	3	2
9	-	-	-	1	2	3	1	1	1	1	2	3
10	1	1	-	2	1	-	2	1	-	2	1	-
11	3	2	1	3	2	1	2	3	1	3	2	1
12	3	1	2	2	1	3	1	2	2	2	1	3

4.3.5 Summary Timeline Visualisation Software

This section first discussed the general views on automatic timeline visualisation software. This resulted in a list of results as is discussed. Though here a general list of results is given to recap the results.

- **Effectiveness.** Helps create an overview, specify and provide conversation topic
- **Transparency.** Provides transparency as it gives more information on the documents through an arguably neutral medium.
- **Other uses.** Can be useful after publication as an overview or for other requesters. Furthermore, users outside Woo can also benefit from the software.
- **Conditions.** Several conditions have to be met to ensure successful implementation such as simplicity of software, clarity in document type, and a disclaimer that software can contain inaccuracies.
- **Process** Software must support and not replace the Woo Handlers

Afterwards, specific design principles were discussed for this timeline visualisation software. Below is a generalisation of the results discussed concerning the design principles.

- **Balance.** A balance has to be found between practicality and aesthetics.
- **Features.** Features must be carefully considered to not overwhelm users with functions.
- **Transparency.** If changes are made after the generation of the timeline, discuss them with the requester.
- **Limit Effort.** Limit the effort that needs to be put into the software for the Woo Handler. Manually changing specific objects costs too much time.
- **Standardising.** To ensure clarity, the software can implement standards in design.

An elaborate discussion of these results can be found in Chapter 5.

5. Discussion

Timelines have proven to be an effective tool for providing comprehensive overviews of events. By proposing two methods for request specification and document selection, this analysis tried to see whether automatically generated timelines can help when requesting government information through the Open Government Act (Woo). These methods were tested through ten interviews with participants representing various stakeholder groups. The data gathered was analysed using the qualitative research method of grounded theory. These results were mentioned in Chapter 4 and will be further discussed in Section 5.1. Subsequently Section 5.2 will give an example of how these findings can be used to implement within a software for both concept extraction and timeline generation. Afterwards, the Limitations and Future Work will be discussed in Sections 5.3 and 5.4 respectively. Thereafter, Section 5.5 provides a conclusion where the research questions will be answered.

5.1 Study Findings

During the analysis of the interview transcripts, multiple findings have been stated in Chapter 4. In this section, these results will be further discussed.

5.1.1 Preliminary Study

Process Mapping

During the first part of the preliminary study, the Woo-Request process is mapped. During this process, all the stakeholders that are directly involved have been extracted. Furthermore, the process is mapped into different parts finding multiple time-extensive points. Some of these time-extensive points have been verified by the participants in this research.

Signature Line: One time-extensive point mentioned was the signature line or 'parafeerlijn'. This concerns repeatedly having to ask the management for approval. Participant 2 stated that a timeline helps to speed things up since it gives the management an overview of the situation. This reduces the time of reading through all the reports. Section 2.1.3 also discusses the use of a timeline in the business world for analysis and overview. This helps to pinpoint why certain actions are taken and therefore helps to explain why requesters requested specific documents.

Topic and Document Specification: It was discussed that the proposed methods for request specification and document selection could help to reduce and specify what documents are requested. Therefore it can reduce the documents that need redaction in the process. Chapter 2 also describes this as timelines are used in various fields of application to pinpoint specific events, and turning points or to emphasise specific parts in the timeline.

Other Process Parts: Furthermore, multiple participants mentioned that the request specification and timelines could be added to the publication of the Woo to give context to the request and to justify why certain decisions were made. It was also mentioned that timelines could even be used after publication to show what documents were found. When some of these documents were not requested in the published requests, others could request them instead of going through the process themselves.

Grounded Theory

Another part of the preliminary study is the structured way of request specification using the Straussian Grounded Theory. During the mapping of the process, it was found that another possible time-extensive part of the process is the request specification. Some requests are rejected because it was too abstract or unclear. Furthermore, if accepted, there sometimes is a long process to exactly specify what the requester needs. During the interviews, Participants 3 and 9 added that this can be caused because the requester, and often the handler, do not know what information is available. Therefore, the handler and requester need to establish a common ground on what is available and what is needed.

To establish this common ground, this research aimed to find out whether the structured method of the Straussian Grounded Theory could improve the process. Therefore, extra research into the Straussian Grounded Theory was performed and an adjusted version of the methodology was created. This version contained 2 extra steps as can be seen in Figure 3.2.

1. Relevance Coding
2. Chronological Coding

Relevance Coding is the step of selecting which categories are useful, and *Chronological Coding* is the step of putting the mentioned categories in chronological order. The interviews showed that participants liked these added steps and helped with the specification process. Transparency was mentioned as an important characteristic of the system during this process. Therefore, it was necessary that *Relevance Coding* was performed with both the handler and requester.

Chronological Coding could also help to show transparency into when certain topics were discussed at the governmental organisation. However, participants mentioned that it would be more beneficial for abstract or large requests and in some cases also for journalists. Therefore, when correctly implemented, these extra coding steps can be beneficial to the process within certain types of requests.

5.1.2 Current Woo Process

Chapter 4 revealed multiple findings concerning the current Woo process. In this section, these findings will be discussed.

Information Management

One of the biggest problems within the process is information management. These results also correspond to the problems found in the *Ondraagelijk Traag* and *Matglas* reports as mentioned in the literature study. These results showed more specifically what the problems were within information management.

Locating Information: The difficulty in locating information was mentioned to be the biggest problem. One of the reasons for that is that, in some cases, officials are asked personally whether they have information on that topic. That means that the request handler is dependent on the official to deliver documents. This, of course, causes a huge time delay. Furthermore, looking for documents, scanning the documents and seeing if it is relevant is another problem which might be resolved by a detailed search software that is implemented over all government organisations. Although this study did not research solutions for these problems, these problems must be resolved to optimise the whole Woo process. It therefore also benefits the proposed methods discussed in this study

Concerns and Experiences

Other problems that were found during the interviews are some concerns and experiences of the requester and handlers.

Trust in Government: From the requesters' side, there is some distrust against the handler. Their distrust shows in the question if all documents are found and the feeling that the handler wants to limit the amount of information as much as possible. The requesters therefore can try to request as large and abstract as possible to make sure no information is withheld from them. However, requesting more information will also result in a longer redaction process. It was also mentioned that transparency is therefore important to increase trust and show that no information is withheld.

Handler Mindset: On the handlers' side, there is a rigidity in publishing information, and responsibility is transferred to others. Therefore it was mentioned that there needs to be a change of mindset on the handlers' side. Finding what causes this rigidity is also important to change the mindset.

Process Problems

Ensure Smooth Process: Because of these concerns and experiences, experience showed that there are some requirements for a smooth process. First of all is important to involve the requester in choices, resulting in a transparent process. This not only demonstrates the willingness from the handlers' side but also allows the requester to take some responsibility for the request. Furthermore, it was mentioned that, for a transparent workflow to be created, timelines were already made to demonstrate what information was known at what times within an organisation. It gives more context to the requester, which was stated to be helpful. Though these timelines were mostly abstract and time-consuming to create. This was also confirmed in the literature research as it is stated in Table 2.1 of Section 2.1.4. That shows the willingness to use a timeline and the usefulness of using it within the process.

Document Delivery: One of the problems concerning document delivery is the delivery of too many documents. This results in a long redaction process and the requester having too many documents to organise. Therefore, it is important to minimise the irrelevant documents. It was also discussed that documents have been merged into one file. This forces the requester to split the documents into separate ones before reviewing them. The reasons for the documents being merged are still unclear but should be avoided.

Staff Workload: Furthermore, the workload of the staff has been discussed to be a substantial problem. The problem is caused by the information management but also the size or complexity of the request. This results in organisations hiring externally to deal with the workload. However, the deadlines of the request are still exceeded. The methods discussed in this research therefore can reduce the workload as it increase the efficiency of the process and helps the handler to get an overview of the situation. This, in turn, helps to map large and complex requests.

So to conclude, there are multiple problems within the current Woo process. Some can be partially solved with the proposed methods, as will be discussed in the following subsections. However, this study will not solve all the problems as the biggest, information management, is a hard problem to solve.

5.1.3 Concept Extraction

Looking more specifically at the proposed concept extraction method using the Straussian Grounded Theory, multiple findings have been gathered. Overall, the proposed method was positively received and mentioned to be helpful.

Specific Features and Aspects

Document Types: First of all, selecting document types for a specific topic, was stated to be important. The interviews revealed that journalists are often searching for documents through the use of document types. The document types, like policy documents and monthly meetings, often contain the same type of information. Therefore specifying what document types are requested of the topics, can help to gather the right information and help set up a search query.

WordWeb: Another feature of the proposed methods was the word web. This is one of the visualisations that is more commonly used to demonstrate important concepts in the Grounded Theory. However, this was mostly found to be chaotic and unstructured for this process. However, it was mentioned that it is useful to quickly see the largest and most important topics. Nevertheless, simplicity was stated multiple times to be important since too many visualisations and features can be confusing. Consequently, it might be better to exclude this visualisation.

Metrics: A finding, concerning the concept list, was that the metrics helped decide what topics were relevant. The proposed metrics showed two things, the match that the topic has with the Woo request and how often it was mentioned in the initial set of documents. However, some participants interpreted the number of times mentioned as the number of documents in which it was mentioned. Therefore, these two types of numbers are useful to be included.

Moreover, it was mentioned that these metrics might also be confusing to people who are not submitting a Woo request that often and therefore, the feature should be able to be disabled. While these metrics may aid in making informed decisions, they also have the potential to create confusion. Therefore, the option to remove these metrics, could be beneficial.

Tick On or Off: Another feature that was different within the proposed methods, is the feature to select or cross off specific concepts. Although participants mentioned that a switch between these options would be useful, it might be better to select which is requested. This results in requesters knowingly choosing what topics they are interested in instead of irrelevant topics being overseen. A similar process is described in Section 2.1.1 as every document is carefully considered and checked to see whether it aligns with the drawn-up research question. This ensures that all the relevant information is retrieved, which is ultimately also the goal of this method.

Agreement Printout: Another feature that was not mentioned in the proposal, but was considered important is the ability to easily print out the agreed topics. This can help to reach an agreement on what the search query will be.

Method Benefits and Conditions

Structured Specification: Overall, the software helps specify and the results can serve as discussion topics. If certain topics are found that are surprising, it can be investigated why that is and result in a conversation. Furthermore, when going through all the steps of the method, can guide the requester and handler to a common ground. Participants found the list of concepts a useful and structured way of going through the topics.

Request Types: These request types are abstract, large-scale requests and some requests made by journalists. Journalists can find it helpful to get an overview of the process flow. However, they could figure this out themselves and therefore not require this step. Therefore, an initial conversation with the handler can point out whether this method would be required or if it falls into a category which does not require concept extraction.

Process Requirements: Concept selection must be entirely done with the requester since this is the most transparent way of selecting topics. The handler must know how to use the software properly so an explanation or intuitive interface is required. The requester must also be willing to use the software and should explain why and how that method works. If either the requester does not want to use the method or the handler does not know how to use it, concept extraction should be reconsidered.

Loading time should not be too long, since that can cause irritation which reduces the adoption rate and decreases the efficiency. Since the initial document set is not complete with all the documents, there must be a disclaimer that some concepts might be missing. Furthermore, as already discussed, an initial conversation before performing concept extraction should be performed to introduce each other, clarify the request, and see whether concept extraction is necessary.

5.1.4 Timeline Software - General

The next step after specifying a request is the selection of the documents. For this step, a proposed method was used to select documents by linking the documents to events and putting them in chronological order. The results showed that this way of creating a timeline can be helpful within the Woo process and different design principles of such a software have been investigated. this section will discuss the results gathered.

Process Benefits and Conditions

Benefits of Method First of all, this study gathered multiple views on an automatic timeline-generating software. The study revealed that linking the documents to events in a timeline created an overview of the situation. It helps to create a perspective on the process flow of document creation. This process flow can be used as a neutral conversation topic and could help to specify the document request. This means that the results show, that an automatic timeline software can (partially) solve multiple challenges and limitations of manually drawing up timelines as discussed in Section 2.1.4. These challenges are the subjectivity of the researchers, understanding relations between elements, the thorough analysis that is required and the time-consuming nature of drawing up manually.

Request Type: Lastly, during the study, participants were asked whether they thought using the timelines would work best in some cases or wouldn't work in others. This resulted in a mixed set of results. Large requests, for instance, can be specified using the timeline since it groups the documents into events. However, it can also lead to too many events causing an extensive and chaotic timeline.

Features

Document Type: As mentioned in subsection 5.1.3, the document type is especially important for journalists since they know what information the document types contain. When using the software, it needs to be checked whether document titles contain sensitive information that might need redaction. Furthermore, it is crucial to maintain the simplicity of the software to ensure ease of use and prevent it from being discarded.

Parallel Timelines: Another finding within this study is that multiple timelines to compare different perspectives were expected to be chaotic and inconvenient. Therefore one timeline containing all the information and labeled for different perspectives was mentioned to be better by one participant.

Transparency

Neutrality: The software can also help with discovering hidden or forgotten information. This means that documents which are not frequently accessed or viewed may end up being included in the automatically generated timeline. This in return helps with creating transparency, since these documents would maybe be less likely to be chosen by the requester through a list. Given that the timeline is generated by a computer, it has the potential to minimise the bias that might be present if a person were to create the timeline. Again referring to the subjectivity of researchers in 2.1.4.

Reliability However, there needs to be a disclaimer that this is a tool to help the process and mistakes might be made by the computer. These timelines can also be published with other documents for other people to see. The software itself could also be used after publication to allow others to see what documents were found. This allows them to choose other documents from that list instead of going through a whole Woo request process themselves.

5.1.5 Timeline Software - Design Principles

The results furthermore showed different findings concerning the specific design principles that were presented. These results will be further discussed in this section.

Balance Simplicity and Aesthetics

One of the findings during the study is that small actions should be easy to perform. Changing orientation, for instance, uses a switch or button for a simple, one-click action that displays all options. This can also be seen in the rankings since the average ranking shows that the switch is most preferred and the checkbox is second. However, when there are two easy actions, the most visually pleasing is preferred as the rankings of duration show. *Pre-set slider* can be considered as easy to use as the *Fill in Date* since it costs two clicks to select. However, the *Pre-set Slider* has been mentioned to be more visually pleasing.

Combinations in Principles

Level of Detail: Another finding of the study is the fact that multiple possibilities of the principles were preferred to be combined. Within the level of detail, for instance, a combination of the *Zoom* and *Pull-Down Button* was preferred by a participant. Zooming assures you focus on the correct part of the timeline, while pull-down buttons allow you to select your desired specifications. However, the rankings show that the *Pull-Down Button* is most preferred and the *Zoom* is least preferred. Other participants were not questioned on this combination and therefore remains to be investigated.

Filtering: All participants who were confronted with this principle mentioned that a combination is preferable. Examining the rankings of the various groups, it is evident that each group has ranked the filters differently. This variation might be explained by the fact that filter types can depend on the specific request. Another principle that was mentioned to depend on the request is the *Interactivity* principle. These rankings also differ greatly among the groups, and when looking at the total column. This shows that multiple options are preferred.

Uncommon Features

During the interviews, one principle, *Integration of Multimedia*, was noted as uncommon but potentially interesting. It was suggested that media should be more common in publications since the current society is a digital world. However, it was also deemed less important than other principles and could increase the software's complexity if implemented. The participant recommended keeping it simple. The rankings for this principle also varied greatly, with a difference of 0.4 between the highest and lowest average total ratings. This again suggests that the type of media needed might depend on the specific request.

Manual Actions

Principles 10, 11 and 12 concerned manually changing the design or text of the generated timeline. These three principles all share the common feedback that they were mentioned to be too time-consuming.

Neutrality: The labels and markers were stated to be a lot of work to adjust all the labels manually or add extra information through the use of markers. The markers were mentioned to be useful when extra context needs to be given. However, one could ask if this should be implemented in the timeline or maybe as an extra notification outside it. Adding it inside can maybe harm the neutrality and therefore transparency of the software. Adjusting the timeline through design or text could also cause this and therefore, should be dealt with, with caution.

Standardisation: The design elements were stated to be too much work when adjusting every event, document or group manually. Themes were mentioned to be easier to implement, as they can be applied with a few clicks. However, the use of different themes over timelines might lead to confusion. Therefore it was suggested to have a standardised visual layout where specific groups have a standardised colour. This can standardise the design of timelines for different governmental organisations. In turn, requesters who submit Woo requests over multiple organisations benefit from this as they understand how to read the timelines better. Furthermore, this benefits neutrality as it removes the option for personal adjustments.

Time-Consuming: The last principle was *Manual Editing* which covers the changing of the content within the timeline. The usefulness of themes was questioned because timelines should function as a tool rather than a guideline. For the sake of transparency, it was suggested that timelines should remain neutral. An editing history should accompany any changes made. Additionally, if the contents contain numerous mistakes, it should be considered whether the software requires modifications. These mistakes can occur in different fields but mistakes in timeline structure can be avoided through the quality assessments in Section 2.2.4. Most importantly of all, these manual adjustments are time-consuming. Having the option to change things can also lead to extra change. One of the benefits of creating a timeline automatically is the fact that it solves the problem of taking too much time to create as discussed in Section 2.1.4.

It is therefore probably best to not allow the handlers to make any adjustments, standardise the layouts and preserve the neutrality of the timeline.

5.2 Recommendation

This chapter covers the recommendations for implementing this system. This study discovered specific and structured methodologies that can improve the efficiency of the current Woo process. Therefore, the findings of this study are combined to create two layouts. The first layout that is created will cover the findings of the concept extraction part of the study. Figure 5.1 shows the created layout for the concept extraction. The second layout will cover the findings on the timeline visualisation software and the design principles. This layout can be seen in Figure 5.2. After each section, a use case is presented to show a possible usage of the software.

5.2.1 Concept Extraction

Figure 5.1 illustrates the proposed layout incorporating the findings of this study concerning concept extraction. First of all, the left side of the interface shows the menu.

Menu

The first tab on the menu is the *Upload Documents* tab where the initial set of documents can be uploaded. After that the concepts are automatically extracted from the set of documents and relevant concepts can be selected in the *Select Concepts* tab. Furthermore, does the menu contain two other tabs *Print Concepts* and *Check Woo Request*. The print concepts is a tab where the selected concepts can be printed to hand the requester the list with all the concepts that have been agreed to search for. The *Check Woo Request* allows the user to take another look at the Woo Request as it was submitted.

Possibility of Summarisation

Another possibility is a summarisation of the request by the computer. This shows the user how the computer processes the request to associate a match percentage with the concepts. This can easily be implemented as the automatic timeline creator from the Epoch from Delft University can already summarise documents. This summary, in combination with keyword extraction, is used for the timeline creation and can therefore also be used to summarise the Woo request. For further information see this link.

Transparency

As was suggested by a participant, a disclaimer is given in the bottom left corner that states this is a tool, and there is a possibility that some concepts are not found. Furthermore, you can see two places where the selected concepts can be printed out. This ensures that the agreed concepts are saved and accountability on what concepts are required and searched for, is recorded.

Visualisation

The right side shows two visualisations of the concepts that are found. The initial visualisation contains a list of concepts. For each concept, it displays the match percentage, the frequency of mentions across all documents, and the number of distinct documents in which the concept is mentioned. Every concept also contains an information button that can give further information about the concept itself. Given that document types are frequently utilised by journalists, this distinction is already incorporated within the various types of concepts to further specify the request. The user can access these document types by clicking the expand arrow.

Metrics

Furthermore, the user can sort the concepts based on one of the metrics, and the concept number displayed on the left will update accordingly when the sorting method is changed. When the search function is used to look for concepts, the number can indicate the relevance since it shows the rank of the concept based on the selected sorting method. On the bottom, a structure can be seen of the found concepts to illustrate the links between concepts. Another feature that the handler can use is to remove the metrics from the concept list. It was mentioned that the metrics can cause unclarity and therefore, there is a possibility to hide the metrics. Lastly, the loading time is mentioned at the bottom to indicate the speed of the software.

Use Case - Topic Selection of Child Benefits Scandal

A handler would receive a request for documents regarding organisations and teams responsible within the Child Benefits Scandal. *1.* First of all, the handler plans an initial conversation with the requester to discuss the Woo request and change the unclarities in the request itself. *2.* After the initial conversation, the handler collects a set of documents regarding various organisational processes during the scandal. *3.* After collecting these documents they load them into this software together with the Woo request itself. *4.* This outputs a list of topics discussed in these documents, the metrics of these topics and a hierarchy of these topics. *5.* The handler invites the requester for a meeting where they can discuss and select relevant topics and their document types. *6.* The requester selects all the topics relevant to them with or without the help of the metrics. After discussing it with the handler, *7.* the concepts are saved and printed out. This printout will be given to the requester as proof of the agreement on topics.

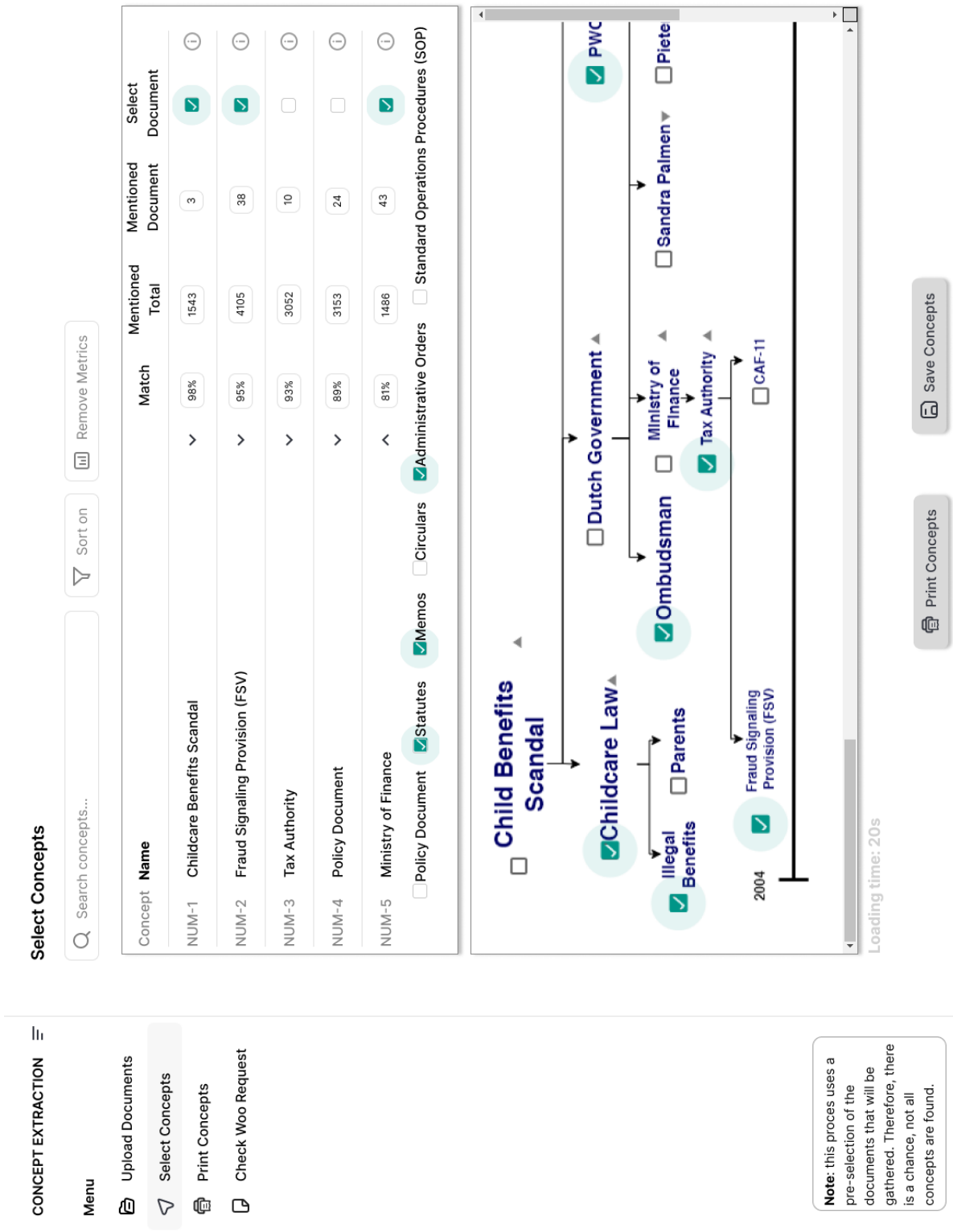


Figure 5.1: Example of User Interface (UI) for Concept Extraction.

5.2.2 Timeline Visualisation

Figure 5.2 illustrates a proposed layout incorporating the findings of the timeline visualisation part of the study. This illustration is purely to demonstrate the different possibilities that the software should have. Less attention was given to the content that was given within the timeline itself.

Menu

The left side of this visualisation shows different design principles implemented. These design principles concerned principles that can be considered a setting of the timeline.

Topic Selection: The first one is *Topic Selection* where can be chosen what specific things, the timeline should focus on. The selected topics can be seen with a green stroke around them. This principle shows the implementation of the suggested choices as was examined to be the best implementation.

Duration: The next item shows the implementation of the pre-set slider. This proposition was ranked the highest because of its simplicity and visual pleasantness. You can therefore select the beginning and end date of the timeline that is shown. Furthermore, it shows the duration between the beginning and end date as was suggested by a participant.

Orientation: To change the orientation, the highest total average ranking was the switch and therefore it is implemented here. It is a quick way to change the orientation and, as suggested by a participant, the selected orientation has a darker colour than the other option.

Time Units: can be simply selected using a dropdown menu since that was ranked the highest. Although a button might be a bit simpler, it becomes less visually aesthetic. This is caused by the fact that there are multiple options to choose from.

Markers: can be added through the button there which pops up a marker that can be dragged on the timeline. To avoid overlap, this marker cannot be dragged over events or text on the timeline. Section 5.1 discussed the separation of markers from the visualisation. However, since this option was not researched, it was implemented within the visualisation.

Actors and Events

The third principle: Actors, Places, and Events, was implemented using different tabs. Even though the *Filters* proposition was ranked the highest, it was mentioned that actors could be displayed differently to emphasise the differences. However, it was important to still show all the information. Therefore another visualisation seems to be a better solution. What the visualisation then looks like should be further investigated. Furthermore, there will still be an option to filter specific actors on the right side of the interface. Places were removed since this was not found relevant.

Visualisation

The visualisation itself shows different colours for the government (displayed red) and civilians (displayed purple). Furthermore, the timeline shows an audio recording of a meeting. The button on the top right of the audio recording also allows the user to look at the transcription. The timeline level of detail is determined using the expand buttons and the possibility of zooming. All events on the timeline begin to expand, as this was noted to enhance oversight.

Filtering

Furthermore, a search button was implemented to search specifically for documents within the timeline. Although this was not researched, it was mentioned multiple times to be helpful. Furthermore, all filter propositions are implemented since that was preferred.

Interactivity

The three propositions for interactivity are also implemented. Although only the preview of documents is shown, the user can also open folders to see the documents and get extra information on specific definitions.

Editing Design and Text

It was chosen to prohibit the user from editing the design and text of the timeline to preserve neutrality and avoid time-consuming editing. Instead, standards for the design are implemented and small errors within the timeline should be condoned.

Transparency

Lastly, three buttons are added to save the timeline, a list of documents, and to send the documents to the redaction department. The *'Save Timeline'* exports the timeline as multiple images and creates a link that can be shared to see the timeline with the documents selected. The *'Save Documents'* button will give a list of the documents and document titles that are selected. The *'Send to Redaction'* button will automatically send the documents to the redaction department. From there the documents can be finalised for publication.

Use Case - Document Selection of Child Benefits Scandal

This use case continues after a topic selection procedure has been completed.

After the handler has gathered all the documents that are relevant for the requester, *1.* they are loaded into the timeline software. After loading it into the software, a timeline is generated and displayed in the software. *2.* After the timeline is loaded in, the handler can add markers in the software to ensure clarity for the requester. *3.* Thereafter, the requester is invited to select documents. They can do that by changing all the variables in the software to select relevant documents. *4.* They can search, filter or change the visualisation to improve visibility. *5.* If the requester feels that some documents are missing, this will be discussed with the handler. The handler then asks the requester to select the documents they require and another meeting will be planned for additional documents. *6.* After a selection of documents has been made, the timeline is saved and documents will be sent off for redaction. *7.* A printout of a list with document titles and a link to the timeline with selected documents, will be given to the requester as an agreement. *8.* After the redaction process, the requester will be notified of the publication and a link will be shared with them to download the files.

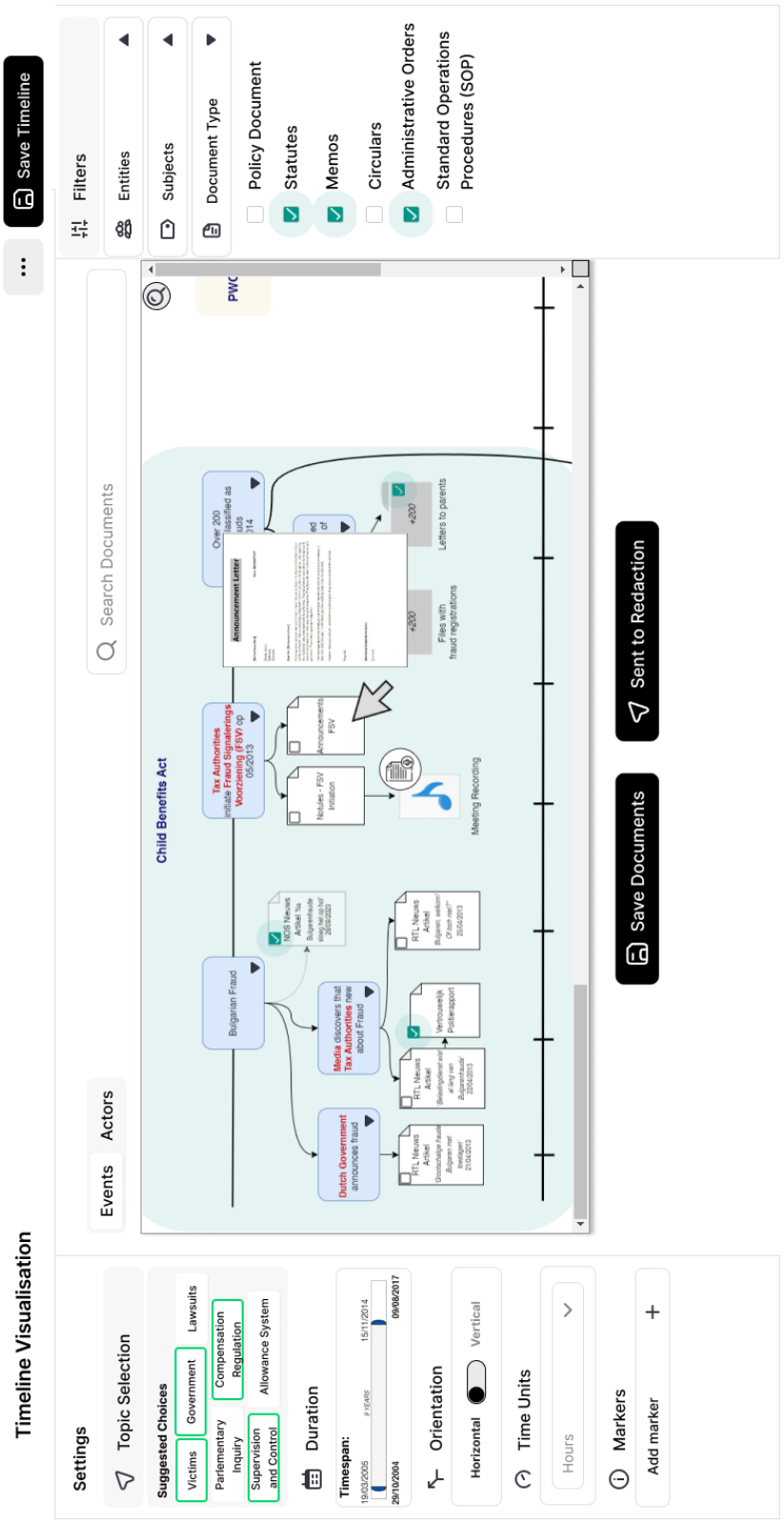


Figure 5.2: Example of User Interface (UI) for Timeline Generation Software.

5.3 Limitations

While this research aims to test methods for improving the Woo process, it is important to acknowledge the limitations that this research has.

Literature Study

Multiple factors have influenced not only the methodology used for this research but also the literature study that has been executed beforehand. This research was conducted as a thesis under the supervision of Max Muller at the National Archives. Although the goal was to perform a comprehensive and thorough literature review, it is important to acknowledge a potential bias. Mr. Muller was already engaged in the development of automatic timeline generation software, which helped in guiding this research but may have influenced the direction and focus of the literature research.

Research Method

Participants: The participant amount can also influence the outcome of the research. Although the participants represented various stakeholder groups involved in the Woo process, the sample size of ten participants might not be enough to draw accurate conclusions.

Group Division: Furthermore, the stakeholder groups were not equally represented during the part of the study questioning the design principles. Since all the participants only represented half of the design principles, the sample size covering the principles is limited. This division was made since the duration of an interview would have increased extensively which could have resulted in a smaller sample size.

Execution: Furthermore, the difference between online and physical methods could lead to different outcomes compared to conducting all of them physically. The same applies to the instance where two participants were interviewed simultaneously. These limitations need to be acknowledged because of their ability to influence the results.

Materials

Scenario: During the interviews, a scenario was given to the participants concerning a Woo. Even though this scenario was checked by multiple people who have had experience with Woo, one participant mentioned that it was an information request and not necessarily a Woo request. The scenario given requested information about a process instead of documents about a process. Therefore, this situation may have caused some confusion among the initial participants. Nonetheless, each scenario was also explained verbally, and after inquiring whether there were any questions and if they understood the scenario, all participants confirmed with a 'yes'.

Terminology: Another potential limitation of the research is that while discussing the concept extraction method, the term "concepts" might have confused the participants. In the Woo process, the term is also used to refer to concept documents, which are earlier versions of the published documents. There has been ongoing speculation about whether these concept documents should also be published. However, when explaining the idea behind the concept extraction and providing the definition of concepts within this context, all participants appeared to understand the definition. This definition was chosen since the Grounded Theory describes codes as concepts.

5.4 Future Work

This research opens opportunities that can be researched. Therefore these opportunities are discussed here.

Usability Testing

In future research, the usability of the design principles can be tested. Therefore, a prototype of the software can be created to check whether the principles are correctly implemented. These usability tests can validate whether the design principles are easy to use and not cause irritation because of unclarities. This test can be performed using, for instance, the System Usability Scale (SUS) to score the usability of individual features.

Implementation Study

Once the usability of implemented features has been tested, a study can be performed on the implementation of the software. Therefore, the software can be tested on numerous Woo requests to determine whether the handlers find it useful and if the requesters appreciate the methodologies used. This can be tested by an observational study, by gathering data afterwards or with the help of screen recording analysis. Furthermore, the quality assessment as discussed in Section 2.2.4 can be used to assess the quality. This part is essential to make sure that the software works properly.

Exploring Different Domains

Another research that can be performed is to map what other applications the software can be applied to. As mentioned during this study, parliamentary inquiries might benefit from an automated timeline generator. Other parts within or outside the government can benefit from these timelines. As was mentioned in Section 2.1.3, timelines are used for numerous amounts of applications. Therefore, the applications mentioned in this research can be tested or new applications can be tested.

5.5 Conclusion

This study aimed to find out how the implementation of timeline generation software could increase the understanding of document requests and improve the effectiveness of the Woo request process by providing a more comprehensive overview of information. To investigate this research question, three sub-questions were drawn up which will be discussed individually.

Subquestion 1

The first step in answering this research question was to find out which stakeholders were involved in the Woo process. After mapping the stakeholders, it was looked at how timeline visualisation software could benefit the process by facilitating the speed and clarity of helping with document management. This resulted in the following subquestion:

S1: *What stakeholders of the WOO request process can benefit from a timeline visualising software to facilitate the speed and clarity of information management?*

To answer this question, a preliminary study was conducted where initial conversations with stakeholders were held and the process was mapped. This resulted in a general process diagram of how the process could go. This process diagram shows the variety of stakeholders namely the requester, contact person, handling team, management, redactor, legal counsellors, and third parties.

Although the process mapping identified multiple time-consuming parts, it was chosen to specifically look at two parts: *1. The specification of the Woo request* and *2. the selection of the relevant documents*. The stakeholders involved in these parts are the *Woo Requester* and *Woo Handler*.

These parts showed a lack of a common ground that had to be established. For the request specification, a structural approach to establish relevant topics had to be undertaken as an initial step. This ensures a specification of the search query to reduce the number of objects seen within a timeline. To solve this issue, the Straussian Grounded Theory was adjusted and examined. After that, the following sub-question was drawn up.

Subquestion 2

To investigate this subquestion, three different and possible processes were illustrated. These illustrations were used during interview sessions where questions were asked about these processes. Initial thoughts on whether such a process would be beneficial, and how the process should be implemented, were asked. These questions were investigated during the interviews to answer the following subquestion:

S2: *To what extent can topics be identified to help define the required information and perspective(s) within the Woo-request process?*

The results of this study showed that the use of this structured method for finding concepts in datasets was expected to be beneficial as it could clear up the gap between what information is required and what is available. *Metrics* can support the choice of topics by giving extra context and helping to guide the Woo requester.

Additionally, it was found that the *transparency* of topics discussed could help to increase the trust between stakeholders. Furthermore, it is important to document the selected topics to facilitate the search for relevant documents. This should be recorded by creating a printout or using another method that can demonstrate what has been agreed upon. This agreement helps to establish *accountability* as the requester has indicated which documents are relevant and the handler has agreed which documents will be looked for.

Subquestion 3

The last subquestion concerns the requirements for a timeline visualisation software that is used during the selection of relevant documents. During the interviews, it was investigated how timelines can support the process of document selection. Furthermore, specific design principles were investigated that could benefit the process. To investigate this, the following subquestion was drawn up.

S3: *To what extent do general design features of software apply to timeline generation software that can be used within the WOO-request process?*

The results of this study showed that timelines can be an *effective tool* to *gain overview* of the documents available and help to *specify* which desired documents. Design principles were investigated and helpful features and guidelines were established.

First of all, creating an overview of documents helps the handler, requester as well as management to understand the situation. This overview supports the reasoning of why certain documents can be relevant to the requester. Additionally, transparency through overview helps to increase trust and avoid miscommunications. The decrease in miscommunications can, in turn, increase accountability as a clearer distinction is made in an agreement. Finally, this software can be beneficial in other domains.

Looking specifically at the software requirements, the software must be simple to use, visually pleasing and balance the amount of features. The use of multimedia, for instance, can be seen as an extra feature but can be helpful in certain requests. Furthermore is it important to have a disclaimer that this software can be used as a tool and not as the truth. Mistakes can easily be made by errors in documents or software and therefore, the human aspect of the Woo handler is absolutely necessary.

To conclude, this study discovered and investigated multiple methods that can be implemented to structurally improve the effectiveness and efficiency of the Woo request process once the conditions are met. Therefore, these solutions can enhance the Woo process, advancing the goal of establishing a more transparent government.

6. Epilogue

Dear reader,

As I bring this thesis to a close, I reflect on the journey that has brought me here, both academically and personally. This master´s thesis was written to fulfil the graduation requirements of the Human-Computer Interaction masters programme at Utrecht University. Hence, my years as a student have now reached their conclusion. Writing this thesis was not without obstacles in my research process as well as my personal life. However, I have managed to overcome these obstacles and am proud of the work that I have delivered.

During my years as a student, I discovered a passion for investigating software to enhance organisational processes. Helping the people within an organisation to improve their daily workload motivates me to perform at my absolute best.

In a gap year between my bachelor's and master's, I started working for a small profitable organisation. To gain experience in a contrasting organization, one that is large and non-profit, I sought out an assignment where I could both apply my passion and acquire valuable work experience. That is how I found this assignment at the Dutch National Archive. The support, enthusiasm and motivation of my colleagues were an inspiration throughout the process.

This research would not have been possible without the support of a group of people. First and foremost, I want to thank Max Muller and my other colleagues at the National Archive for the daily support that they have provided. Although the team was new when I joined, there was a welcoming feeling and a pleasant atmosphere. Furthermore, I would like to thank my supervisor Eelco Herder. Eelco has gladly helped and supported me throughout the process. His view was often different from mine which resulted in a balanced collaboration that has looked at topics from multiple perspectives. Finally, I would like to thank my friends, family and girlfriend Dana for their support throughout the process.

Finally, thanks to you, the reader, for reading this thesis and showing your interest in my research. I hope you have enjoyed reading.

Roy Schoonwater
28th of June 2024

Bibliography

[1] C. B. voor de Statistiek, *Minste vertrouwen in tweede kamer in 10 jaar tijd*, May 2023. [Online]. Available: https://www.cbs.nl/nl-nl/nieuws/2023/19/minste-vertrouwen-in-tweede-kamer-in-10-jaar-tijd#:~:text=Na%20een%20toename%20in%202020,voor%20grote%20bedrijven%20en%20ambtenaren..

[2] J. Kang and G. Hustvedt, "Building trust between consumers and corporations: The role of consumer perceptions of transparency and social responsibility," *Journal of business ethics*, vol. 125, pp. 253–265, 2014.

[3] V. Fanoy, C. Kaandorp, M. Roebroek, *et al.*, *Matglas - onderzoek naar de invoering van de wet open overheid*, Mar. 2023. [Online]. Available: https://openstate.eu/nl/2023/03/onderzoek-naar-wet-open-overheid-wijst-uit-veel-actie-maar-weinig-resultaat/.

[4] R. Kitchin, "Big data, new epistemologies and paradigm shifts," *Big data & society*, vol. 1, no. 1, p. 2 053 951 714 528 481, 2014.

[5] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, communication & society*, vol. 15, no. 5, pp. 662–679, 2012.

[6] C. M. Christensen, *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press, 2013.

[7] T. H. Davenport and J. Dyché, "Big data in big companies," *International Institute for Analytics*, vol. 3, no. 1-31, 2013.

[8] E. Karafiloski and A. Mishev, "Blockchain solutions for big data challenges: A literature review," in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*, IEEE, 2017, pp. 763–768.

[9] K. Jaseena, J. M. David, *et al.*, "Issues, challenges, and solutions: Big data mining," *CS & IT-CSCP*, vol. 4, no. 13, pp. 131–140, 2014.

[10] C. D. Hansen and C. R. Johnson, *Visualization handbook*. Elsevier, 2011.

[11] C.-h. Chen, W. K. Härdle, and A. Unwin, *Handbook of data visualization*. Springer Science & Business Media, 2007.

[12] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–236, 2017.

[13] P. von Stackelberg, "Footprints of the future: Timelines and exploratory forecasts in futures research," *Journal of Futures Studies*, vol. 13, no. 4, pp. 13–34, 2009.

[14] H. K. Adriansen, "Timeline interviews: A tool for conducting life history research," *Qualitative studies*, vol. 3, no. 1, pp. 40–55, 2012.

[15] M. L. Muller, E. Saaman, J. M. E. van der Werf, C. Jeurgens, and H. A. Reijers, "Timeflows: Visualizing process chronologies from vast collections of heterogeneous information objects," pp. 203–219, 2024.

[16] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu, "Storyflow: Tracking the evolution of stories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2436–2445, 2013.

[17] N. Gutehrlé, A. Doucet, and A. Jatowt, "Archive timeline summarization (atls): Conceptual framework for timeline generation over historical document collections," *arXiv preprint arXiv:2301.13479*, 2023.

[18] D. P. WATR2207, *Onderzoek memo toeslagen (memo-palmen)*, Oct. 2021. [Online]. Available: `https://www.tweedekamer.nl/debat_en_vergadering/uitgelicht/onderzoek-memo-toeslagen-memo-palmen`.

[19] Jun. 2024. [Online]. Available: `https://herstel.toeslagen.nl/dashboard-kinderopvangtoeslag/`.

[20] C. T. Street and K. W. Ward, "Improving validity and reliability in longitudinal case study timelines," *European journal of information systems*, vol. 21, pp. 160–175, 2012.

[21] J. Fulda, M. Brehmer, and T. Munzner, "Timelinecurator: Interactive authoring of visual timelines from unstructured text," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 300–309, 2015.

[22] A. Bagnoli, "Beyond the standard interview: The use of graphic elicitation and arts-based methods," *Qualitative research*, vol. 9, no. 5, pp. 547–570, 2009.

[23] P. Hu, M. Huang, P. Xu, W. Li, A. K. Usadi, and X. Zhu, "Generating breakpoint-based timeline overview for news topic retrospection," in *2011 IEEE 11th international conference on data mining*, IEEE, 2011, pp. 260–269.

[24] S. Elshout, M. Elshout, and R. Giesen, "An, verhei en, m., & drahmann,.(2023)," *from https://hdl. handle. net/1887/3665574 Version: Publisher's Version License: Leiden Uni ersit Non-exclusi e license Downloaded from: https://hdl. handle. net/1887/3665574*,

[25] D. De Vaus and D. de Vaus, *Surveys in social research*. Routledge, 2013.

[26] H. Colley and K. Diment, "Holistic research for holistic practice: Making sense of qualitative research data," 2001.

[27] N. K. Kayesa and M. Shung-King, "The role of document analysis in health policy analysis studies in low and middle-income countries: Lessons for hpa researchers from a qualitative systematic review," *Health Policy OPEN*, vol. 2, p. 100 024, 2021.

[28] G. A. Bowen, "Document analysis as a qualitative research method," *Qualitative research journal*, vol. 9, no. 2, pp. 27–40, 2009.

[29] M. D. White and E. E. Marsh, "Content analysis: A flexible methodology," *Library trends*, vol. 55, no. 1, pp. 22–45, 2006.

[30] K. Kolar, F. Ahmad, L. Chan, and P. G. Erickson, "Timeline mapping in qualitative interviews: A study of resilience with marginalized groups," *International journal of qualitative methods*, vol. 14, no. 3, pp. 13–32, 2015.

[31] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner, "Timelines revisited: A design space and considerations for expressive storytelling," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 9, pp. 2151–2164, 2016.

[32] J. Sheridan, K. Chamberlain, and A. Dupuis, "Timelining: Visualizing experience," *Qualitative research*, vol. 11, no. 5, pp. 552–569, 2011.

[33] B. F. Keith Norambuena and T. Mitra, "Narrative maps: An algorithmic approach to represent and extract information narratives," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW3, pp. 1–33, 2021.

[34] R. E. Roth, "Cartographic design as visual storytelling: Synthesis and review of map-based narratives, genres, and tropes," *The Cartographic Journal*, vol. 58, no. 1, pp. 83–114, 2021.

[35] D. Metz, S. Klassen, B. A. McMillan, M. P. Clough, and J. Olson, "Building a foundation for the use of historical narratives," *Science & Education*, vol. 16, pp. 313–334, 2007. DOI: 10.1007/S11191-006-9024-Z.

[36] S. Colby, "Energizing the history classroom: Historical narrative inquiry and historical empathy," *Social Studies Research and Practice*, 2008. DOI: 10.1108/ssrp-03-2008-b0005.

[37] G. Novack, "Understanding history; marxist essays," 1972.

[38] S. Lévesque and P. Clark, "Historical thinking: Definitions and educational applications," *The Wiley international handbook of history teaching and learning*, pp. 117–148, 2018.

[39] J. Banks-Wallace, "Storytelling as a tool for providing holistic care to women," *MCN: The American Journal of maternal/child nursing*, vol. 24, no. 1, pp. 20–24, 1999.

[40] N. Tahmasebi, L. Borin, G. Capannini, *et al.*, "Visions and open challenges for a knowledge-based culturomics," *International Journal on Digital Libraries*, vol. 15, pp. 169–187, 2015.

[41] A. S. Chan, "Storytelling, culture, and indigenous methodology," in *Discourses, dialogue and diversity in biographical research*, Brill, 2021, pp. 170–185.

[42] H. White, "The value of narrativity in the representation of reality," *Critical Inquiry*, vol. 7, pp. 5–27, 1980. DOI: 10.1086/448086.

[43] C. Atman, "Design timelines: Concrete and sticky representations of design process expertise," *Design Studies*, 2019. DOI: 10.1016/j.destud.2019.10.004.

[44] G. Karam, "Visualization using timelines," pp. 125–137, 1994. DOI: 10.1145/186258.187157.

[45] M. Brehmer, B. Lee, B. Bach, N. Riche, and T. Munzner, "Timelines revisited: A design space and considerations for expressive storytelling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 2151–2164, 2017. DOI: 10.1109/TVCG.2016.2614803.

[46] M. C. Clark, "Narrative learning: Its contours and its possibilities," *New directions for adult and continuing education*, vol. 126, no. 3, p. 3, 2010.

[47] K. G. Shaver, *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media, 2012.

[48] S. Konaté and B. Pali, "You have to talk with us, not about us": Exploring the harms of wrongful accusation on those affected in the case of the dutch 'childcare-benefit scandal," *Revista de Victimología/Journal of Victimology*, no. 16, pp. 139–164, 2023.

[49] D. Government, *Onderzoek memo toeslagen (memo-palmen)*, Oct. 2021. [Online]. Available: https://www.tweedekamer.nl/debat_en_vergadering/uitgelicht/onderzoek-memo-toeslagen-memo-palmen.

[50] D. Buchanan and P. Dawson, "Discourse and audience: Organizational change as multi-story process," *Journal of Management Studies*, vol. 44, no. 5, pp. 669–686, 2007.

[51] J. Phelan, *Narrative as rhetoric: Technique, audiences, ethics, ideology*. Ohio State University Press, 1996.

[52] O.-H. Ylijoki, "Academic nostalgia: A narrative approach to academic work," *Human Relations*, vol. 58, no. 5, pp. 555–576, 2005.

[53] D. Rosenberg and A. Grafton, *Cartographies of time: A history of the timeline*. Princeton Architectural Press, 2013.

[54] A. Kalizhanova, R. Zhussupova, and T. Maryshkina, "Using timeline in teaching the history of the english language to efl students,"

[55] D. Ovigli, N. Bossolan, M. Oliveira, and L. Beltramini, "The history of science in the interactive space of cbme," *Revista de Ensino de Bioquímica*, vol. 5, no. 2, p. 11, 2007.

[56] A. Cesta, A. Finzi, S. Fratini, A. Orlandini, and E. Tronci, "Validation and verification issues in a timeline-based planning system," *The Knowledge Engineering Review*, vol. 25, no. 3, pp. 299–318, 2010.

[57] M. C. Mayer, A. Orlandini, and A. Umbrico, "A formal account of planning with flexible timelines," in *2014 21st International Symposium on Temporal Representation and Reasoning*, IEEE, 2014, pp. 37–46.

[58] G. M. Chryssochoidis and V. Wong, "Rolling out new products across country markets: An empirical study of causes of delays," *Journal of Product Innovation Management: AN INTERNATIONAL*

PUBLICATION OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION, vol. 15, no. 1, pp. 16–41, 1998.

[59] A. A. Bui, D. R. Aberle, and H. Kangarloo, "Timeline: Visualizing integrated patient records," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 4, pp. 462–473, 2007.

[60] A. Ledesma, N. Bidargaddi, J. Strobel, *et al.*, "Health timeline: An insight-based study of a timeline visualization of clinical data," *BMC medical informatics and decision making*, vol. 19, pp. 1–14, 2019.

[61] S. Ozturk, M. Kayaalp, and C. J. McDonald, "Visualization of patient prescription history data in emergency care," in *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, vol. 2014, 2014, p. 963.

[62] J. Gill, T. Chearman, M. Carey, S. Nijjer, and F. Cross, "Presenting patient data in the electronic care record: The role of timelines," *JRSM short reports*, vol. 1, no. 4, pp. 1–10, 2010.

[63] A. Braga, F. Portela, M. F. Santos, *et al.*, "Pervasive patient timeline for intensive care units," in *New Advances in Information Systems and Technologies: Volume 2*, Springer, 2016, pp. 527–536.

[64] N. Momen, M. Kendall, S. Barclay, and S. Murray, "Using timelines to depict patient journeys: A development for research methods and clinical care review," *Primary Health Care Research & Development*, vol. 14, no. 4, pp. 403–408, 2013.

[65] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic, "Interpretable representation learning for healthcare via capturing disease progression through time," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 43–51.

[66] J. Danczyk, S. Kane, D. Housten, M. Voshell, R. M. Kilgore, and C. Hogan, "Using temporal representations for understanding complex interrelationships for mission planning," in *Advances in Ergonomics in Design: Proceedings of the AHFE 2016 International Conference on Ergonomics in Design, July 27-31, 2016, Walt Disney World®, Florida, USA*, Springer, 2016, pp. 183–190.

[67] M. Cialdea Mayer, A. Orlandini, and A. Umbrico, "Planning and execution with flexible timelines: A formal account," *Acta Informatica*, vol. 53, no. 6-8, pp. 649–680, 2016.

[68] A. Cimatti, A. Micheli, and M. Roveri, "Timelines with temporal uncertainty," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 27, 2013, pp. 195–201.

[69] N. Munier and N. Munier, "Project planning and scheduling," *Project Management for Environmental, Construction and Manufacturing Engineers: A Manual for Putting Theory into Practice*, pp. 119–150, 2013.

[70] E. Bjarnason, A. Hess, R. B. Svensson, B. Regnell, and J. Doerr, "Reflecting on evidence-based timelines," *IEEE software*, vol. 31, no. 4, pp. 37–43, 2014.

[71] J. Crow, E. Whitworth, A. Wongsa, L. Francisco-Revilla, and S. Pendyala, "Timeline interactive multimedia experience (time) on

location access to aggregate event information," in *Proceedings of the 10th annual joint conference on Digital libraries*, 2010, pp. 201–204.

[72] M. Champagne, "Diagrams of the past: How timelines can aid the growth of historical knowledge," *Cognitive Semiotics*, vol. 9, no. 1, pp. 11–44, 2016.

[73] P. Bennett, M. Fraser, and M. Balaam, "Chronotape: Tangible timelines for family history," in *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, 2012, pp. 49–56.

[74] A. Bezerianos, P. Dragicevic, J.-D. Fekete, J. Bae, and B. Watson, "Geneaquilts: A system for exploring large genealogies," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1073–1081, 2010.

[75] M. Ogawa and K.-L. Ma, "Software evolution storylines," in *Proceedings of the 5th international symposium on Software visualization*, 2010, pp. 35–42.

[76] M. Marti, J. Vieli, W. Witoń, *et al.*, "Cardinal: Computer assisted authoring of movie scripts," in *23rd International Conference on Intelligent User Interfaces*, 2018, pp. 509–519.

[77] J. A. Jones and R. Donmoyer, "Improving the trustworthiness/validity of interview data in qualitative nonprofit sector research: The formative influences timeline," *Nonprofit and Voluntary Sector Quarterly*, vol. 50, no. 4, pp. 889–904, 2021.

[78] R. B. Allen, "Interactive timelines as information system interfaces," in *Symposium on digital libraries*, vol. 175, 1995, p. 180.

[79] R. Swan and J. Allan, "Automatic generation of overview timelines," in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, 2000, pp. 49–56.

[80] R. Prabowo, M. Thelwall, and M. Alexandrov, "Generating overview timelines for major events in an rss corpus," *Journal of Informetrics*, vol. 1, no. 2, pp. 131–144, 2007.

[81] R. Yan, X. Wan, J. Otterbacher, L. Kong, X. Li, and Y. Zhang, "Evolutionary timeline summarization: A balanced optimization framework via iterative substitution," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 745–754.

[82] Y.-H. Tseng, C. C. Chen, S.-H. Wu, and L.-C. Yu, "Exploiting timelines to enhance multi-document summarization," in *ACL*, 2014.

[83] G. Pasquali, P. Cintia, F. Rossi, S. J. J. Saporito, A. Cossu, and F. Giannotti, "Tls-covid19: A new annotated corpus for timeline summarization," in *ECIR*, 2021.

[84] G. Tran, M. Alrifai, and E. Herder, "Timeline summarization from relevant headlines," in *Advances in Information Retrieval: 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29-April 2, 2015. Proceedings 37*, Springer, 2015, pp. 245–256.

[85] R. Campos, A. Pasquali, A. Jatowt, V. Mangaravite, and A. M. Jorge, "Automatic generation of timelines for past-web events," in *The Past Web: Exploring Web Archives*, Springer, 2021, pp. 225–242.

[86] M. La Quatra, L. Cagliero, E. Baralis, A. Messina, and M. Montagnuolo, "Summarize dates first: A paradigm shift in timeline summarization," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 418–427.

[87] Y. Zhou, J. Kietzmann, and F. Can, "Real-time timeline summarisation for high-impact events in twitter," in *ECAI*, IOS Press, 2016.

[88] Y. Yu, A. Jatowt, A. Doucet, K. Sugiyama, and M. Yoshikawa, "Multi-timeline summarization (mtls): Improving timeline summarization by generating multiple summaries," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 377–387.

[89] Y. Duan, A. Jatowt, and M. Yoshikawa, "Comparative timeline summarization via dynamic affinity-preserving random walk," in *ECAI 2020*, IOS Press, 2020, pp. 1778–1785.

[90] X. Wan and J. Yang, "Improved affinity graph based multi-document summarization," in *Proceedings of the human language technology conference of the NAACL, Companion volume: Short papers*, 2006, pp. 181–184.

[91] N. Ketui and T. Theeramunkong, "Inclusion-based and exclusion-based approaches in graph-based multiple news summarization," in *Knowledge, Information, and Creativity Support Systems: 5th International Conference, KICSS 2010, Chiang Mai, Thailand, November 25-27, 2010, Revised Selected Papers*, Springer, 2011, pp. 91–102.

[92] H. L. Chieu and Y. K. Lee, "Query based event extraction along a timeline," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 2004, pp. 425–432.

[93] K.-H. Nguyen, X. Tannier, and V. Moriceau, "Ranking multidocument event descriptions for building thematic timelines," in *COLING 2014, the 25th International Conference on Computational Linguistic*, 2014, pp. 1208–1217.

[94] A. Pasquali, V. Mangaravite, R. Campos, A. M. Jorge, and A. Jatowt, "Interactive system for automatically generating temporal narratives," in *Advances in Information Retrieval: 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part II 41*, Springer, 2019, pp. 251–255.

[95] T. Ge, W. Pei, H. Ji, S. Li, B. Chang, and Z. Sui, "Bring you to the past: Automatic generation of topically relevant event chronicles," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on*

Natural Language Processing (Volume 1: Long Papers), 2015, pp. 575–585.

[96] K. Masters, "Artificial intelligence in medical education," *Medical Teacher*, vol. 41, no. 9, pp. 976–980, 2019.

[97] A. Lee, D. Archambault, and M. Nacenta, "Dynamic network plaid: A tool for the analysis of dynamic networks," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.

[98] S. Di Bartolomeo, A. Pandey, A. Leventidis, *et al.*, "Evaluating the effect of timeline shape on visualization task performance," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.

[99] B. A. Price, R. M. Baecker, and I. S. Small, "A principled taxonomy of software visualization," *Journal of Visual Languages & Computing*, vol. 4, no. 3, pp. 211–266, 1993.

[100] C. V. Siang, F. B. Mohamed, F. M. Salleh, M. I. B. M. Isham, A. H. Basori, and A. B. Selamat, "An overview of immersive data visualisation methods using type by task taxonomy," in *2021 IEEE International Conference on Computing (ICOCO)*, IEEE, 2021, pp. 347–352.

[101] Monday.com, *What is a timeline creator and how they can help with project management*, Feb. 2024. [Online]. Available: `https://monday.com/blog/project-management/timeline-creator/`.

[102] D. Shahaf, C. Guestrin, and E. Horvitz, "Trains of thought: Generating information maps," in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 899–908.

[103] D. Shahaf, J. Yang, C. Suen, J. Jacobs, H. Wang, and J. Leskovec, "Information cartography: Creating zoomable, large-scale maps of information," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1097–1105.

[104] A. Aizawa, "An information-theoretic perspective of tf–idf measures," *Information Processing & Management*, vol. 39, no. 1, pp. 45–65, 2003.

[105] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang, "Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework," in *Proceedings of the 17th ACM international conference on Multimedia*, 2009, pp. 481–490.

[106] B. Liu, D. Niu, K. Lai, L. Kong, and Y. Xu, "Growing story forest online from massive breaking news," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 777–785.

[107] J. Ansah, L. Liu, W. Kang, S. Kwashie, J. Li, and J. Li, "A graph is worth a thousand words: Telling event stories using timeline

summarization graphs," in *The World Wide Web Conference*, 2019, pp. 2565–2571.

[108] L. Qiang, C. Bingjie, and Z. Haibo, "Storytelling by the storycake visualization," *The Visual Computer*, vol. 33, no. 10, pp. 1241–1252, 2017.

[109] T. J. Tanenbaum, K. Tanenbaum, M. S. El-Nasr, and M. Hatala, "Authoring tangible interactive narratives using cognitive hyperlinks," in *Proceedings of the Intelligent Narrative Technologies III Workshop*, 2010, pp. 1–8.

[110] D. Government, *Requesting information from the government (woo request)*, 2022. [Online]. Available: `https://business.gov.nl/regulation/freedom-of-information/`.

[111] G. Enthoven, S. Wiemers, S. d. Uijl, *et al.*, "Ondraaglijk traag - analyse afhandelen wob-verzoeken," *Instituut Maatschappelijke Innovatie*, Jan. 2022. [Online]. Available: `https://openstate.eu/wp-content/uploads/sites/14/2022/01/Ondraaglijk-traag-280122-def.pdf`.

[112] B. Ogunleye, T. Maswera, L. Hirsch, J. Gaudoin, and T. Brunsdon, "Comparison of topic modelling approaches in the banking context," *Applied Sciences*, vol. 13, no. 2, p. 797, 2023.

[113] B. Glaser and A. Strauss, *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 1967.

[114] D. Walker and F. Myrick, "Grounded theory: An exploration of process and procedure," *Qualitative health research*, vol. 16, no. 4, pp. 547–559, 2006.

Appendices

A. Informed Consent

Universiteit Utrecht,
Faculteit Informatiewetenschappen
Afdeling, Human Computer Interaction,
MsC. Student Roy Schoonwater
Postbus 80140,
3508 TC Utrecht,
r.a.g.schoonwater@students.nl

Universiteit Utrecht

Toestemmingsformulier

Onderzoeksproject:
Master Human-Computer Interaction
Onderzoeken van methodes voor informatie-uitwisseling binnen WOO-proces
Universiteit Utrecht

Naam onderzoeker:
Roy Schoonwater

Doel van Onderzoek:
Het onderzoeken van de toepasbaarheid van de Straussian Grounded Theory voor het versnellen van het WOO-proces en het verfijnen van zoektermen. Daarnaast wordt onderzocht welke ontwerpprincipes geschikt zijn voor gebruik in tijdlijnvisualisatie software om documenten overzichtelijk te presenteren.

Toestemming:
Voordat het onderzoek kan beginnen moet u akkoord gaan met de volgende verklaringen:

- ☐ Ik bevestig dat ik 18 jaar of ouder ben.
- ☐ Ik bevestig dat het onderzoeksproject "An Overseeable Overview: An Exploratory Study for Information Exchange using Timelines within Governmental Practices" aan mij is uitgelegd. Ik heb de gelegenheid gehad om vragen over het project te stellen en deze zijn naar tevredenheid beantwoord.
- ☐ Ik geef toestemming dat het gesprek wordt opgenomen in de vorm van een audio-opname.
- ☐ Ik geef toestemming voor het gebruik van het door mij bijgedragen materiaal om inzichten te genereren voor het onderzoeksproject.
- ☐ Ik begrijp dat mijn deelname volledig vrijwillig is.
- ☐ Ik begrijp dat ik mij op elk moment kan terugtrekken uit het onderzoek.
- ☐ Ik begrijp dat deelname aan het onderzoek niet vereist is voor mijn werk.
- ☐ Ik geef toestemming om de volledig geanonimiseerde gegevens te gebruiken voor toekomstige publicaties en andere wetenschappelijke middelen om de bevindingen van het onderzoeksproject te verspreiden.
- ☐ Ik begrijp dat de verzamelde informatie/gegevens veilig worden opgeslagen door onderzoekers, maar dat geanonimiseerde gegevens in de toekomst beschikbaar kunnen worden gesteld aan anderen voor onderzoeksdoeleinden.
- ☐ Ik begrijp dat ik kan verzoeken om verwijdering van alle door/van mij verzamelde gegevens.
- ☐ Ik ga akkoord om deel te nemen aan bovengenoemde studie

Naam: ……. Plaats: ……

Datum: …….. Identificatienummer:……

Handtekening:

B. Information Sheet

Universiteit Utrecht,
Faculteit Informatiewetenschappen
Afdeling, Human Computer Interaction,
MsC. Student Roy Schoonwater
Postbus 80140,
3508 TC Utrecht,
r.a.g.schoonwater@students.nl

Universiteit Utrecht

Informatieblad

Onderzoeksproject:
Master Human-Computer Interaction
Onderzoeken van methode om WOO-proces te verbeteren
Universiteit Utrecht

Naam onderzoeker:
Roy Schoonwater

Doel van Onderzoek:
Het onderzoeken van de toepasbaarheid van de Straussian Grounded Theory voor het versnellen van het WOO-proces en het verfijnen van zoektermen. Daarnaast wordt onderzocht welke ontwerpprincipes geschikt zijn voor gebruik in tijdlijnvisualisatie software om documenten overzichtelijk te presenteren.

Uitleg Onderzoek:
Bedankt voor je deelname aan deze wetenschappelijke studie over methoden om het WOO-proces te verbeteren.

Voor deze studie word je zo meteen geïnterviewd. Het interview is opgesplitst in twee delen. In het eerste deel worden er vragen gesteld over een objectieve methode om een *common ground** te creëren tussen welke informatie er beschikbaar is en welke er nodig is. Deze methode zou transparantie moeten bieden over alle beschikbare gegevens en moet helpen bij het verminderen van het aantal documenten. Dit kan later in het proces bijdragen aan het verminderen van het tijdrovende werk van documentreductie. Hiervoor test dit onderzoek de Straussian Grounded Theory, ontwikkeld door de sociologen Barney Glaser en Anselm Strauss. In deze studie wordt er onderzocht of deze benadering ook nuttig kan zijn binnen het WOO-proces om gemakkelijk te onderscheiden welke informatie beschikbaar is, zodat de aanvrager duidelijk kan aangeven welke aspecten van de data belangrijk zijn. Het tweede deel van het onderzoek richt zich vooral op een vervolgstap na identificeren van belangrijke concepten. Deze stap kijkt naar het opstellen van tijdlijnen om een duidelijk overzicht te geven tussen activiteiten, betrokkenen en de beschikbare documenten. Stel je dus voor dat je gemakkelijk een tijdlijn kan creëren van een bepaalde situatie met een geïntegreerd programma. Het tweede deel zal zich vooral focussen diverse ontwerpaspecten van zo'n software voor tijdlijnvisualisatie. We stellen vragen over hoe tijdlijnen kunnen worden ingezet voor meer duidelijkheid en welke functies noodzakelijk zijn voor de uiteindelijke software.

Mocht je tijdens of na de studie vragen hebben, neem dan gerust contact op via r.a.g.schoonwater@students.uu.nl. Je kunt op elk moment tijdens of na de studie verzoeken om je gegevens te laten verwijderen. Jouw gegevens worden tijdens deze studie veilig opgeslagen en zullen volledig worden geanonimiseerd. Tijdens dit onderzoek wordt het concept getoetst, niet de deelnemer. Alle antwoorden die je geeft, zijn waardevol.

Uw identificatienummer: _____

**Het bewerkstelligen van overeenstemmende informatie*

C. Interview Artefacts Sheets

Universiteit Utrecht

Interview Artefacten

Een document met afbeeldingen en uitleg

Deel 1 – Concept Extractie

Scenario:

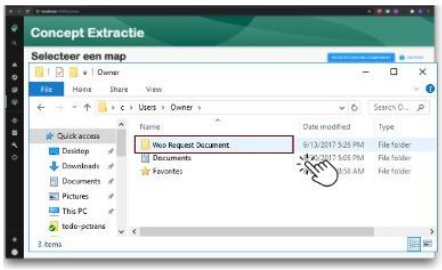
Stel je voor dat je een WOO-verzoek ontvangt met de aanvraag voor informatie van alle betrokkene bij de toeslagenaffaire. Jouw taak is om alle relevante documenten te lokaliseren. Aangezien het een verzoek is waarbij relatief veel documenten vereist zijn, zoek je een manier om uit te zoeken welke stukken informatie precies van belang zijn voor de Aanvrager. Overigens is het tijdrovende karakter van documentredactie nog een reden om precies die documenten te identificeren die noodzakelijk zijn voor de Aanvrager, waardoor het benodigde volume van documenten wordt verminderd. Om dit te bereiken, gebruik je een nieuwe functie in je huidige gebruikte programma om de belangrijkste concepten uit de documenten te halen. Presenteer deze concepten aan de Aanvrager om de relevantie te bepalen. Op de volgende pagina's zijn meerdere manieren waarop dit proces kan verlopen. Dit zijn slechts voorbeelden om de mogelijkheden te demonstreren en uw input over dit proces is het meest van belang. Vragen over deze voorbeelden kunt u altijd stellen.

Universiteit Utrecht

1e Concept Extractie

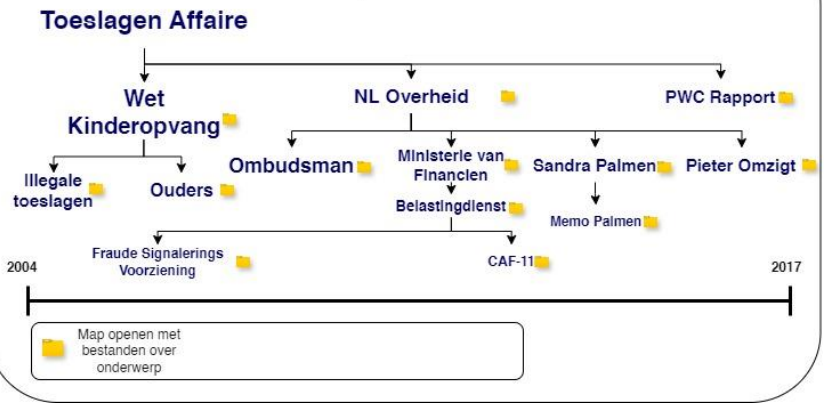
Deel 1: Documenten Verzamelen

Deel 2: Laad Folder in het Programma



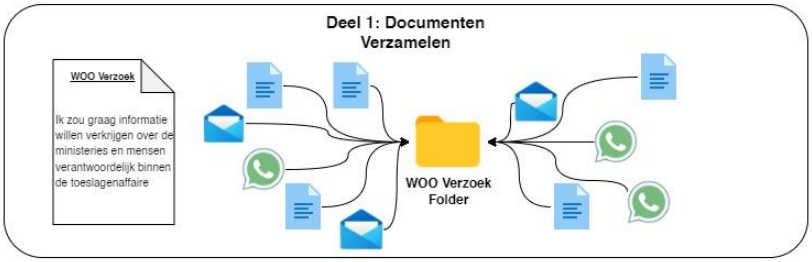
Deel 3: Verwijder Onbelangrijke Concepten

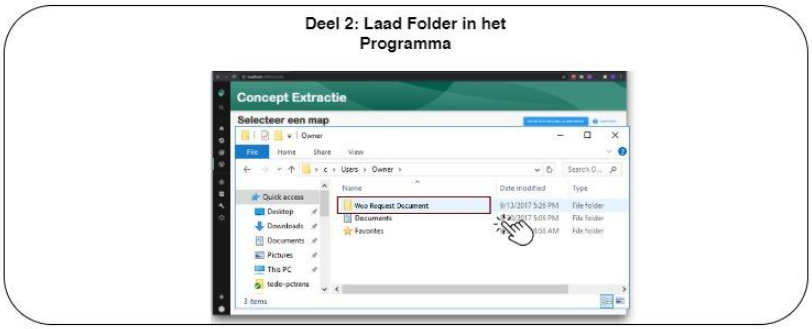
Deel 4: WOO Verzoeker Concepten Laten Kiezen

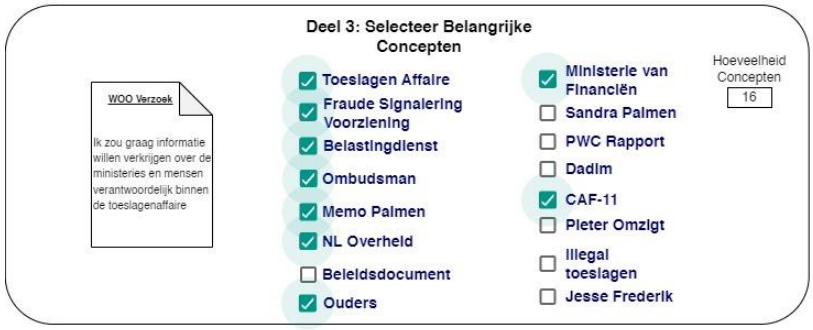


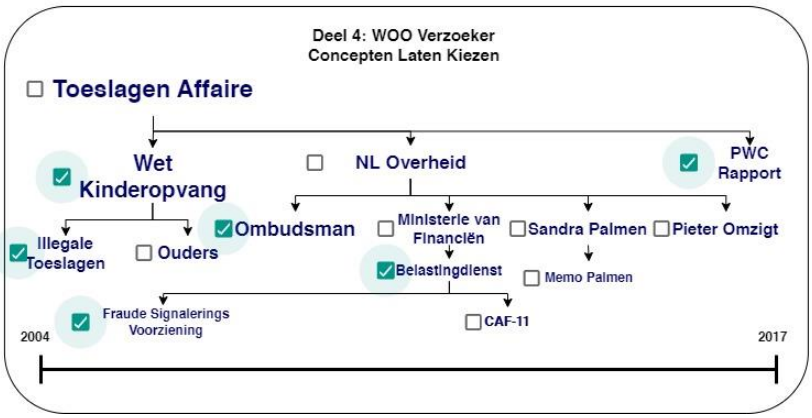
Universiteit Utrecht

2e Concept Extractie







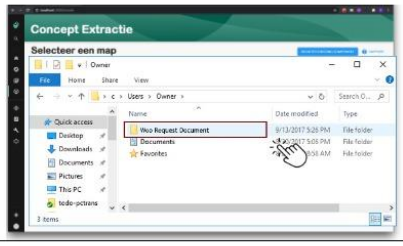


Universiteit Utrecht

3e Concept Extractie

Deel 1: Documenten Verzamelen

Deel 2: Laad Folder in het Programma

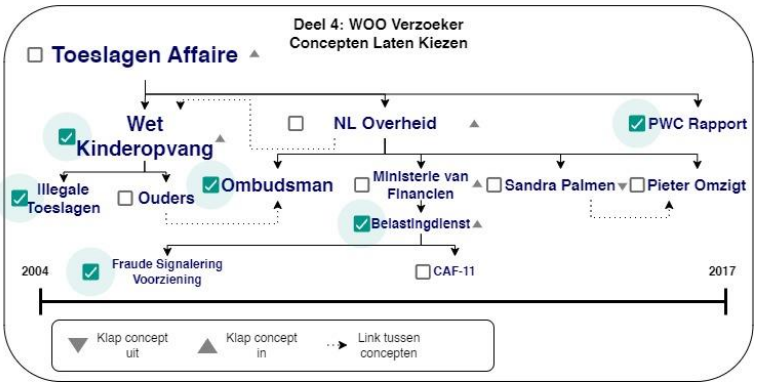


Deel 3: Verwijder Onbelangrijke Concepten

Sorteren op: Match ▼	Aantal	Match
☐ Toeslagen Affaire	5012	98%
☐ Fraude Signalerings Voorziening	3840	95%
☐ Belastingdienst	6427	93%
☐ Ombudsman	5727	93%
☒ Sandra Palmen	241	91%
☐ NL Overheid	154	90%
☐ Ministerie van Financien	3541	87%
☐ Ouders	6845	86%
☐	...	...%

	Aantal	Match
☒ Pieter Omzigt	24	84%
☒ Jesse Frederik	14	81%
☒ Dadim	1	76%
☐ Memo Palmen	59	62%
☐ CAF-11	148	57%
☐ PWC Rapport	35	55%
☐ Illegale Toeslagen	52	49%
☐ Beleidsdocument	7	36%
☐	...	...%

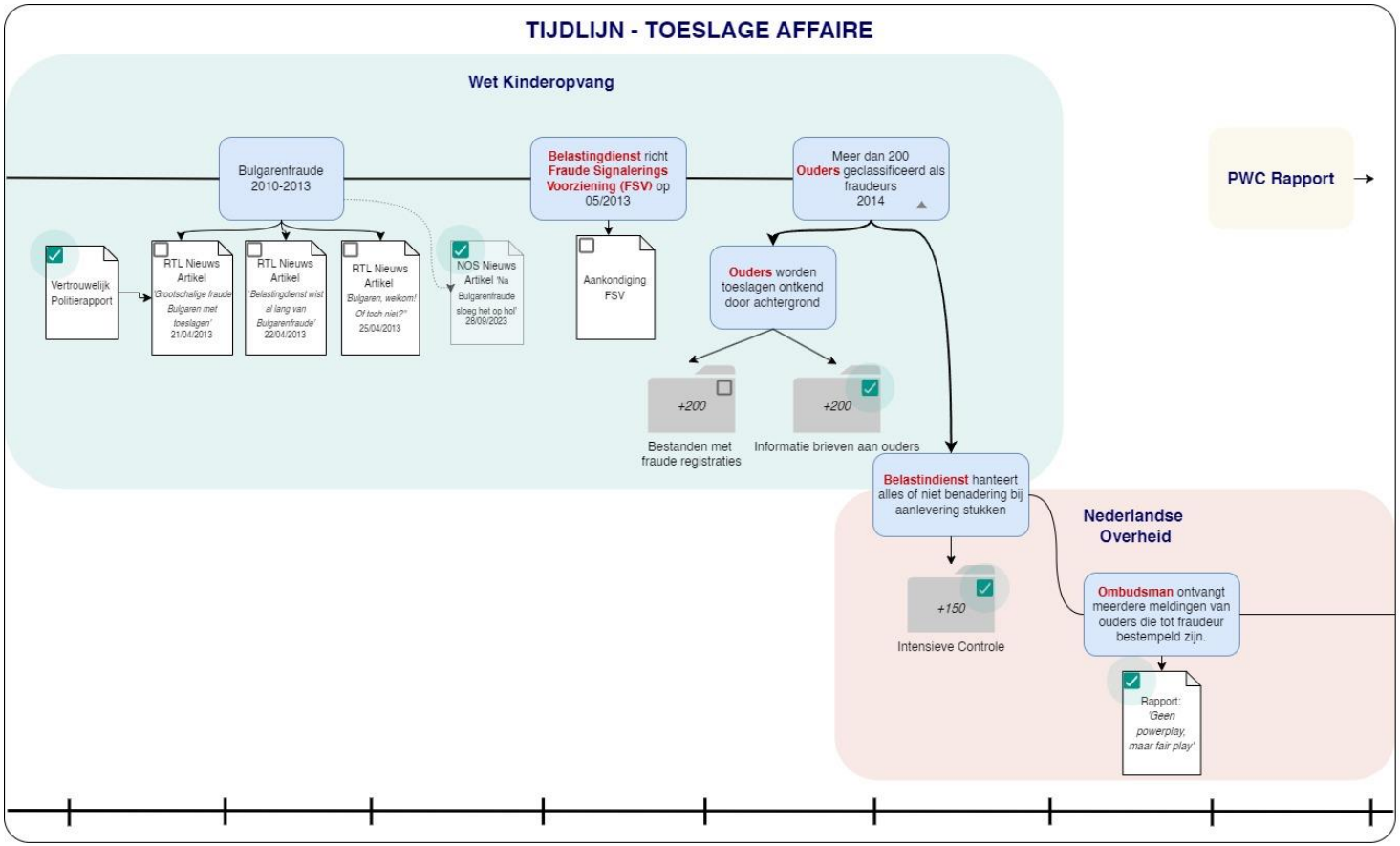
Deel 4: WOO Verzoeker Concepten Laten Kiezen



Universiteit Utrecht

Deel 2 – Design Principles

De voorgestelde software vergemakkelijkt de visualisatie van interactieve tijdlijnen, ook wel verhaallijnen genoemd, voor verschillende situaties. Deze tijdlijn/verhaallijn zal opgesteld worden nadat er een duidelijke common ground is gecreëerd over wat de aanvrager nodig heeft en welke informatie er beschikbaar is. Dit kan door middel van de methode zoals besproken in deel 1, of op andere manieren. Het primaire doel van deze software is om een holistisch overzicht te bieden van specifieke scenario's, waardoor gebruikers gemakkelijk sleutelcomponenten zoals gebeurtenissen, betrokkene en informatie-artefacten binnen deze scenario's kunnen identificeren. Bovendien biedt het de mogelijkheid voor gebruikers om binnen de tijdlijnvisualisatie te zoeken naar documenten die gekoppeld zijn aan gebeurtenissen of personen, waardoor een intuïtieve manier wordt geboden voor gebruikers om relevante documenten op te halen op basis van hun behoeften of interesses. In dit deel gaan we het hebben over verschillende design principes die geïmplementeerd kunnen worden.



Universiteit Utrecht

#	PRINCIPE	BESCHRIJVING
1	Onderwerp Selectie	Selecteer een specifiek onderwerp dat belangrijk is om binnen dergelijke software te tonen. Door niets in te vullen wordt een volledige tijdlijn gecreëerd met de geselecteerde documenten.

Voorgelegde Keuzes

Slachtoffers | Overheid | Parlementair Onderzoek
Toeslagen stelsel | Ministerie van Financiën | Verloop van Keuzes
Compensatie regelingen | Toezicht en Controle | Rechtzaken

Vrije Input

Omschrijf onderwerp:

De tijdlijn moet zich focussen op alle actoren die schuld hebben of slachtoffer zijn binnen de toeslagenaffaire.

Belangrijkste Concepten

Toeslagen Affaire | Wet Kinderopvang | NL Overheid
PWC Rapport | Illegale Toeslagen | Ouders
Ombudsman | Ministerie van Financien | Sandra Palmen
Belastingdienst | Fraude Signalering Voorziening | CAF-11

2	Tijdsduur	Selecteer een tijdspanne waarin het onderwerp plaatsvindt. Dit kan twee jaar zijn, specifieke data of tijdstippen.

Datum invullen

Tijdsspanne:

Begindatum:

DD | MM | YYYY

Einddatum

DD | MM | YYYY

Genereer Tijdlijn

Pre-Set Slider

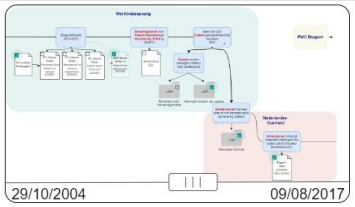
Tijdsspanne:

19/03/2005 — 15/11/2014

29/10/2004 — **09/08/2017**

Genereer Tijdlijn

Interval Slider



3	Actoren, Plaats of Gebeurtenissen	Dit principe stelt de gebruiker in staat te kiezen of de tijdlijn zich moet richten op actoren, plaatsen en/of gebeurtenissen.

Filters

Actoren
Plaats
Gebeurtenissen

Selecteren

Bewerk Object

Actor
Plaats
Gebeurtenis

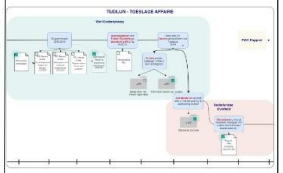
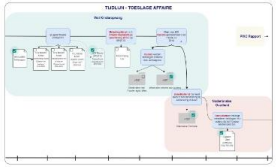
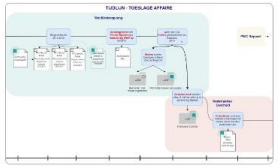
Als basis

Basis:

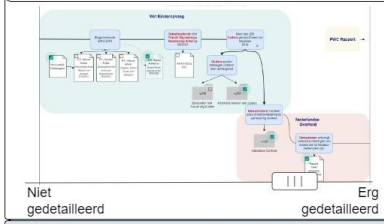
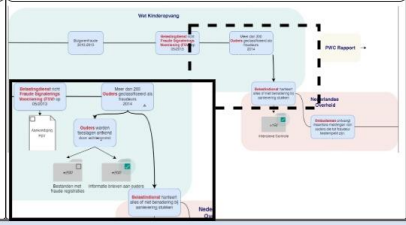
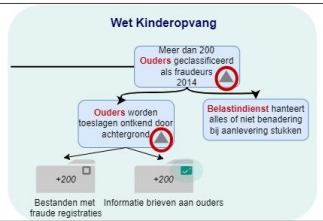
Gebeurtenissen
Plaatsen
Actoren

Universiteit Utrecht

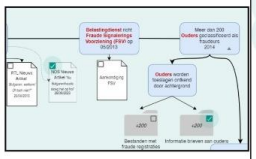
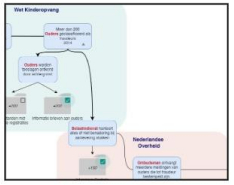
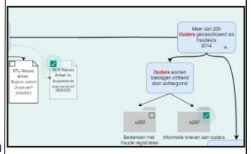
#	PRINCIPE	BESCHRIJVING
4	Oriëntatie	Dit principe bepaalt in welke oriëntatie de tijdlijn moet worden gecreëerd.

Keuze	Switch	Cirkelvormig
Oriëntatie: Horizontaal Verticaal 	 Horizontaal ⬤ Verticaal	☑ Horizontaal ☐ Verticaal 

#	PRINCIPE	BESCHRIJVING
5	Detail Niveau	Dit principe bepaalt hoe gedetailleerd de tijdlijn is.

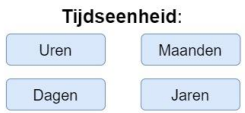
Slider	Zoom	Uitklapknop
 Niet gedetailleerd — Erg gedetailleerd		

#	PRINCIPE	BESCHRIJVING
6	Filteren	Dit principe maakt het mogelijk om specifieke informatie uit de tijdlijn te filteren.

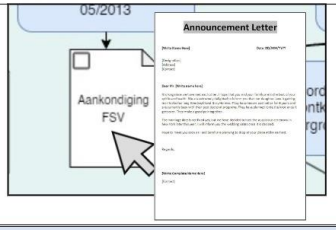
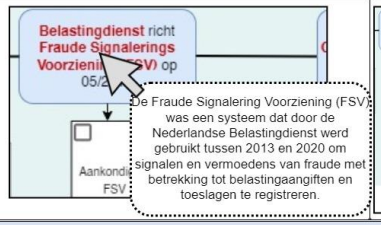
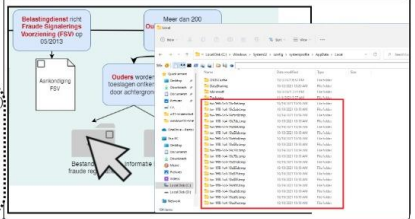
Documenten	Onderwerpen	Entiteit Filter
☑ Nieuws Artikelen ☑ Rapporten ☐ SMS-berichten ☐ Whatsapp 	☑ Wet Kindertoeslag ☑ NL Overheid ☐ PWC Rapport 	☑ Ministerie van Financiën ☑ Ombudsman ☐ Parlement ☐ FSV ☑ Ouders ☐ Toeslagenteam 

Universiteit Utrecht

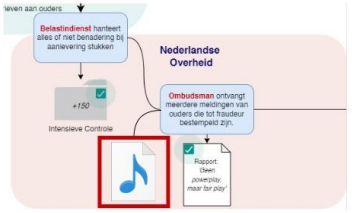
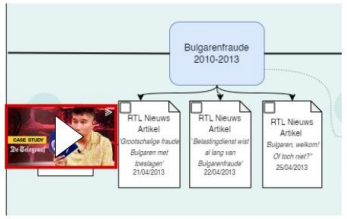
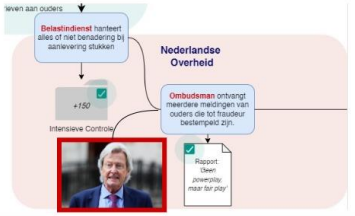
#	PRINCIPE	BESCHRIJVING
7	Tijdseenheid	Dit principe maakt het mogelijk om de tijdseenheden waarin de gegevens worden getoond, te veranderen. zo kunt u bijvoorbeeld dagen veranderen in uren.

Dropdown	Knop	Detail Slider
		

#	PRINCIPE	BESCHRIJVING
8	Interactiviteit	Dit principe bepaalt hoe gebruikers kunnen intergreren met de tijdlijn. Interactieve tijdlijnen stellen gebruikers in staat om op actoren of gebeurtenissen te klikken of eroverheen te zweven voor meer details.

Preview	Uitleg	Folders Openen
		

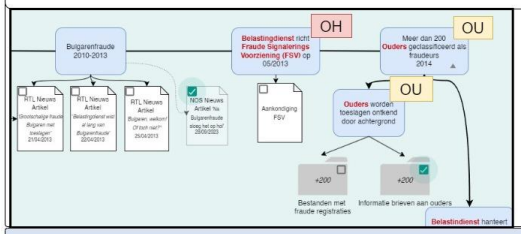
#	PRINCIPE	BESCHRIJVING
9	Integratie van Multimedia	Dit principe maakt integratie van foto's, video's of audioclips mogelijk.

Audiofragmenten	Videofragmenten	Afbeeldingen
		

Universiteit Utrecht

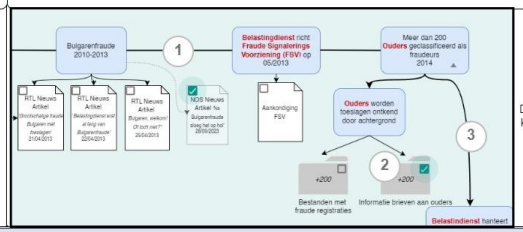
#	PRINCIPE	BESCHRIJVING
10	Labels en Markers	Dit principe stelt de WOO-behandelaar in staat om labels of markers toe te voegen binnen de tijdlijn die de volgorde van de gebeurtenissen kunnen verduidelijken voor de aanvrager.

Labels



Ouders
OU
Overheid
OH

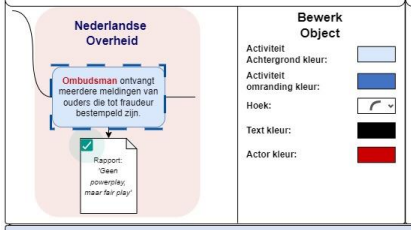
Markers



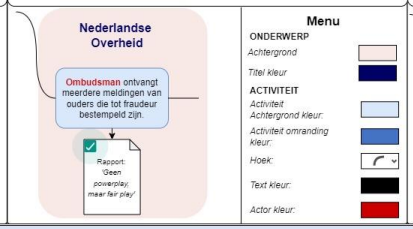
1 Tijdsprong in tijdlijn
2 Deze bestanden kunnen van pas komen
3 Stap waarbij overheid actie ondernam

11	Design Elementen	Dit principe maakt het mogelijk om de lay-out van de tijdlijn naderhand te veranderen. Bepaalde gebeurtenissen kunnen een andere kleur krijgen of andere aspecten kunnen worden aangepast om het aantrekkelijker te maken.

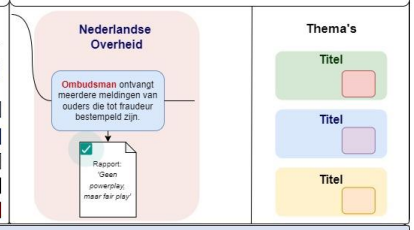
Directe Bewerking



Menu

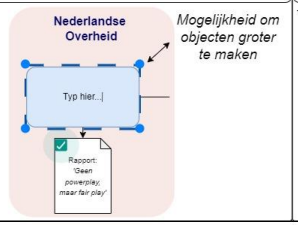


Thema's

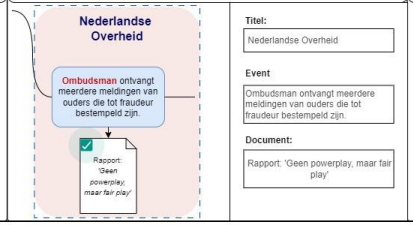


12	Handmatig Bewerken	Dit stelt de WOO-behandelaar in staat om gebeurtenissen naderhand aan te passen. Dit betekent dat ze gebeurtenissen aan de tijdlijn kunnen toevoegen of verwijderen.

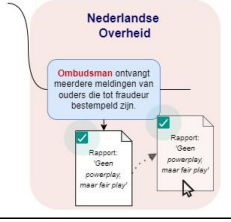
Directe Bewerking



Alleen tekst



Slepen



Universiteit Utrecht

D. Interview Structure Sheets

Universiteit Utrecht

Interview Structuur

Een gids voor een semi-gestructureerd interview

Begin van het Interview

- Allereerst, verwelkom de deelnemer en stel jezelf voor.
- Nodig hen vervolgens uit om plaats te nemen en zorg dat ze zich comfortabel voelen.
- Na een kort gesprekje, toon je de deelnemer het informatieblad en geef je de deelnemer tijd om die te lezen.
- Vraag daarna of de deelnemer vragen heeft.
- Verder, geef de deelnemer het toestemmingsformulier en geef de deelnemer tijd om ook dat te lezen.
- Als de deelnemer akkoord gaat, laat hen dan het contract ondertekenen en bewaar het document veilig.
- Zorg altijd dat er meerdere toestemmingsformulieren zijn, zodat ze altijd de mogelijkheid hebben om opnieuw te lezen wat ze hebben ondertekend.
- Geef hen ook het toestemmingsformulier of informatieblad mee, aangezien daar de contactgegevens op staan.
- Vertel de deelnemer dat het interview niet bedoeld is om informatie over hen te verkrijgen, maar om inzichten te verzamelen over een systeem, dus er zijn geen foute antwoorden.
- Noteer vervolgens welk type belanghebbende de deelnemer is: A. Afhandelaar B. Aanvrager.

Achtergrond

Aanvrager

De eerste vragen die gesteld zullen worden gaan over de demografische gegevens van de deelnemers.

1. Kunt u mij vertellen wat uw beroep is?
2. Hoe vaak heeft u een WOO-verzoek ingediend of geholpen met het indienen?
3. Op een schaal van één tot tien, wat is uw ervaring met het indienen van een WOO-verzoek?
4. Heeft u moeilijkheden ondervonden bij het definiëren van de informatie die de Aanvrager vraagt?

Afhandelaar

1. Kunt u mij vertellen wat uw beroep is?
2. Wat is uw specifieke taak binnen het WOO-proces?
3. Wat is uw ervaring met het WOO-proces?
4. Heeft u moeilijkheden ondervonden bij het definiëren van de informatie die de Aanvrager vraagt?

Deel 1 - Straussian Grounded Theory

Dit deel van het interview gaat over een aanpak voor belangrijke conceptextractie en documentreductie. Deze aanpak is gebaseerd op de Straussian Grounded Theory.

Laat de participant de voorbeelden van het proces op het blad *Interview Artefacten* bekijken.

1. Denkt u dat deze methode een goede manier zou zijn om concepten uit grote datasets te extraheren?
2. Denkt u dat deze methode voordelig zou kunnen zijn wanneer deze binnen het WOO-proces geïmplementeerd wordt?
3. In welk deel van het proces denkt u dat deze methode binnen het WOO-proces geïmplementeerd zou kunnen worden?
4. Heeft u verdere opmerkingen, tips of toevoegingen over het proces?

Stel nu voor dat de dit programma geïmplementeerd zou worden nadat de eerste set van informatieve artefacten is verzameld.

1. Denkt u dat deze methode de Aanvrager een overzicht kan geven van welke informatie beschikbaar is?
    a. Denkt u dat deze methode u een overzicht kan geven van welke informatie beschikbaar is?
2. Denkt u dat deze methode de Aanvrager kan helpen om aan te geven welke concepten hij relevant vindt?
    a. Denkt u dat deze methode u kan helpen om aan te geven welke gevonden concepten relevant voor u zijn?
3. Denkt u dat het verstanding zou zijn dat er een bevestingsmail gestuurd wordt na het selecteren naar zowel de aanvrager als behandelaar?
    a. Zou dit met een lijst kunnen of een QR-code als echtheidscertificaat?

Een ander project dat momenteel loopt bij het Nationaal Archief is de automatische creatie van verhaallijnen om aan WOO-aanvragers te tonen. Meerdere voorbeelden van verhaallijnen zijn te zien in op het blad *'Interview Artefacten'*. De methode die u zojuist heeft bekeken zou dan gebruikt kunnen worden als een filter om specifieke aspecten van een situatie binnen de verhaallijnsoftware te tonen.

1. Denkt u dat het implementeren van deze methode om informatie te verzamelen over wat de Aanvrager nodig heeft, nuttig is?

Algemene Ontwerpprincipes

Voordat we overgaan tot het laatste deel van het interview, geef ik een introductie over een voorgestelde software om te specificeren welke documenten een WOO belanghebbende wilt. Stel je voor dat een software wordt geïmplementeerd die van bepaalde situaties, automatisch interactieve tijdlijnen creëert, ook wel bekend als verhaallijnen. Deze tijdlijn wordt gecreëerd nadat duidelijk is geworden welke concepten belangrijk zijn voor de Aanvrager. Dit kan door middel van de methode zoals besproken in deel 1, of op andere manieren. Deze tijdlijn stelt de Aanvrager in staat om een volledig overzicht van een bepaalde situatie te zien, waar gebeurtenissen, betrokkene en informatieve data gemakkelijk te identificeren zijn. De aanvrager kan dan zien welke documenten gekoppeld zijn aan bepaalde gebeurtenissen en beslissen of zij deze documenten al dan niet willen ophalen. U kunt een voorbeeld zien op het ‘Extra Documenten’ blad.

- Heeft u een duidelijk beeld van de zojuist geschetste situatie?
- Wat vindt u van de software die zojuist omschreven is?
- Zou deze software passend werken binnen het WOO-proces?
- Kunt u voorbeelden geven van situaties waarin u denkt dat de software uitzonderlijk nuttig zou zijn?
- Kunt u voorbeelden geven van situaties waarin u denkt dat de software niet nuttig zou zijn?
- Hoe belangrijk is het dat deze software integreert met bestaande systemen of databases binnen uw organisatie?
- Zijn er andere opmerkingen die u kwijt wilt over deze voorgestelde software?

In het volgende deel van het interview zullen we verschillende ontwerpfuncties en aanpasbare aspecten bespreken die binnen dergelijke software geïmplementeerd kunnen worden. Deze vindt u ook op een blad papier voor u. We zullen elk van deze ontwerpprincipes doornemen en ik zou graag willen dat u deze drie vragen voor elk principe beantwoordt:

- Aan de hand van de voorbeelden die u bij elk principe ziet, wat vindt u goede of slechte eigenschappen aan elk voorbeeld?
- Hoe zou je deze principes ranken? Geef een 1 aan de beste en 3 aan de slechtste.
- Zijn er nog principes die toegevoegd moeten worden voor een software zoals omschreven?

PRINCIPLE	A.	B.	C
Onderwerp Selectie			
Tijdsduur			
Actoren, Plaats Geb			
Oriëntatie			
Detail Niveau			
Filteren			
Tijdseenheid			
Interactiviteit			
Integratie van Mult.			
Labels en Markers			
Design Elementen			
Handmatig Bew.			

- Denkt u dat het ook goed zou zijn als er meerdere tijdlijnen gemaakt kunnen worden in verschillende perspectieven om ze te vergelijken?
- In welke perspectieven zou het dan moeten kunnen verschillen?
    - Denkt u dat met de huidige principes, deze verschillende perspectieven gecreëerd kunnen worden?
- Denkt u dat er nog andere principes toegevoegd moeten worden?
- Heeft u nog andere opmerkingen?

Einde van het interview

Bedank de deelnemer voor hun tijd en inbreng

- Vraag de deelnemer of er nog onduidelijkheden of vragen zijn?

E. Ranking Sheet

IDENTIFICATIENUMMER:

PRINCIPLE	A.	B.	C
Onderwerp Selectie			
Tijdsduur			
Actoren, Plaats Geb			
Oriëntatie			
Detail Niveau			
Filteren			
Tijdseenheid			
Interactiviteit			
Integratie van Mult.			
Labels en Markers			
Design Elementen			
Handmatig Bew.			

IDENTIFICATIENUMMER:

PRINCIPLE	A.	B.	C
Onderwerp Selectie			
Tijdsduur			
Actoren, Plaats Geb			
Oriëntatie			
Detail Niveau			
Filteren			
Tijdseenheid			
Interactiviteit			
Integratie van Mult.			
Labels en Markers			
Design Elementen			
Handmatig Bew.			