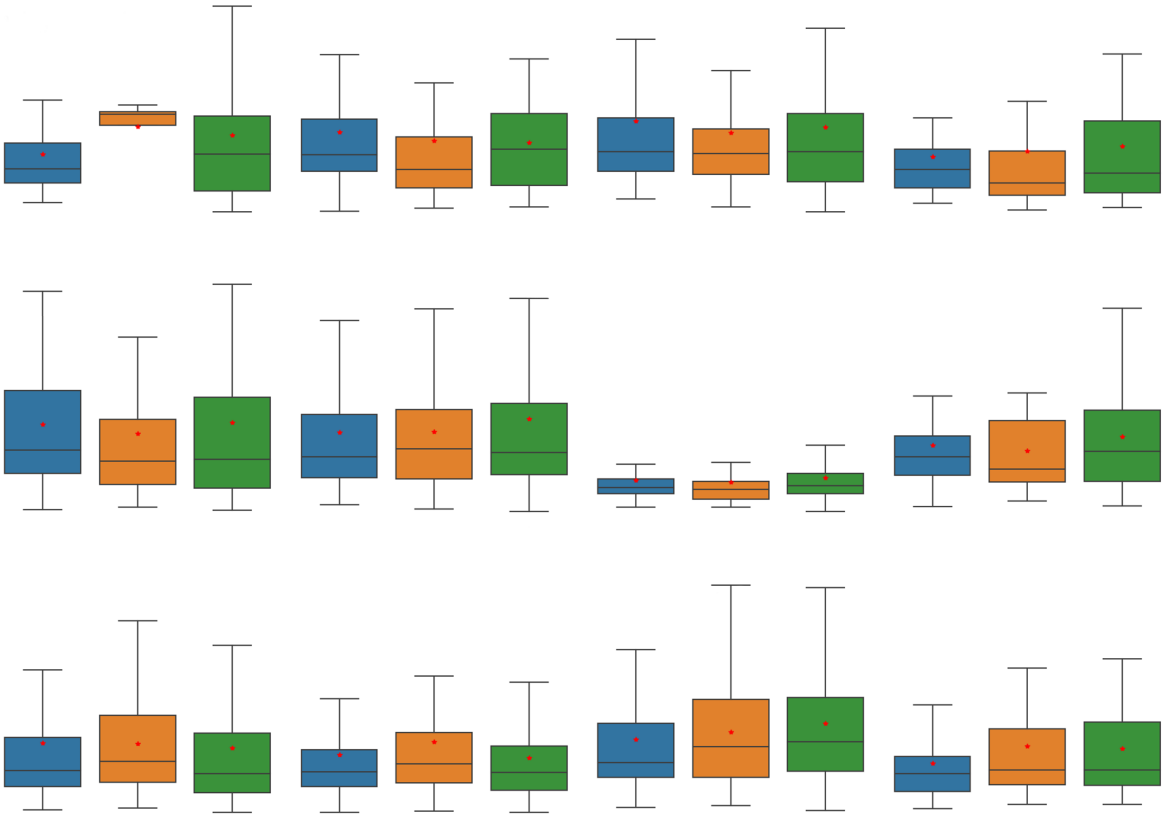
Data extraction from Woo-decisions



Joran Verweij

Layout: typeset by the author using LaTeX.
Cover illustration: Joran Verweij

Data extraction from Woo-decisions

Joran Verweij
12874132

Bachelor thesis
Credits: 18 EC

Bachelor *Informatiekunde*

University of Amsterdam
Faculty of Science
Science Park 900
1098 XH Amsterdam

Supervisor
dr. M.J. Marx

Informatics Institute
Faculty of Science
University of Amsterdam
Science Park 900
1098 XH Amsterdam

Semester 1, 2023-2024

Abstract

The report Matglas discusses amongst other things the lead times of Dutch freedom of information requests. Request- and decision dates have been manually gathered from thousands of Woo-decisions which took a considerable amount of time and effort. In this research a rule-based approach is used to extract the lead times automatically. After testing no significant difference has been shown to exist between distributions of the manual and automatic data, so a reliable estimation of lead times can be made automatically. The biggest obstacle in calculating the lead times is the use of non-machine readable methods for noting down decision dates which resulted in a recall of 60% when tested against decision dates found in the metadata. Final decisions have been extracted and a classifier has been trained for four categories with a micro F-score of 0.76, which is good enough for a very basic indication, but no conclusions can be drawn from the results. There is also potential for additional analysis of Woo-decisions as the automatic method can extract additional data with relative ease.

Contents

1. Introduction

In a democracy citizens can elect representatives to represent their opinions in parliament. To check if these representatives do their work well, the government needs to be transparent, partly for this reason freedom of information laws are made. The Dutch freedom of information law is called 'Wet open overheid', or 'Woo' for short. It describes the right to request information about anything the government does and states that decisions on requests have to be made with a maximum lead time of 42 days (*Wet Open Overheid*, 2023). However a report called 'Matglas' has shown that this time limit has been exceeded in the majority of Woo-decisions, and the average lead time even worsened compared to the previous year (Fanoy et al., 2023).

Plans are made to improve the process, the action plan 'Open Overheid 2023-2027' discusses previous and new methods to work towards an open and transparant government, and shorter lead times on Woo-decisions (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2023). In this action plan a point is made to promote the publishing of documents under the FAIR principles, making information in the documents findable, accessible, interoperable and reusable (Wilkinson et al., 2016). Elements of these principles can be found in the Dutch freedom of information law, article 2.4 states that documents produced as result of the procedure, should be in good, orderly and accessible shape, and where possible documents should be published in a machine-readable, electronic format together with metadata (*Wet Open Overheid*, 2023).

In the report Matglas and its preceding report Ondraaglijk Traag, information has been manually gathered for an estimated amount of 2100 documents, filling in around 22.000 cells in a spreadsheet (Fanoy et al., 2023; Enthoven et al., 2022). It can be assumed that this took a considerable amount of time and effort, and with documents published according to the FAIR principles it should be able to be automated. The method described in this paper will be used for the successor of the Matglas report, and so a considerable amount of time will be saved which can be used for other research concerning the Woo-process. The following research question has been formulated to test whether this is possible:

How well can texts from Woo-decisions obtained through optical character recognition be used for the analysis of Woo-decisions?

To answer the research question, the following sub questions will be answered:
1. How accurately can the lead time be automatically calculated?
2. What is the effect of non-machine readable decision days?
3. How accurately can decisions be classified?
4. What other data can be useful to analyse Woo-decisions?

2. Related work

Extraction of dates, and more specifically dates in combination with events has been attempted in earlier work. Woo-decisions contain descriptions of events, and often they are paired with a date. To connect these dates and events, a method has been developed by Bakker (2022) in which a graphical timeline is constructed after extracting dates using spaCy, and using ChatGPT to classify events and connecting these events to a date, however, too many mistakes were deemed to be made when ChatGPT had to decide if a date had an event or not. In this paper, ChatGPT will not be utilized, but the dates will be extracted directly from the text resulting from the OCR (optical character recognition)

using a more rule-based approach. Rule-based information extraction methods have high maintainability and transparency due to their declarative nature, and can go hand in hand with machine learning approaches by assisting in the creation of datasets used to train machine learning algorithms (Waltl et al., 2018).

Woo-dossiers did not comply with the FAIR principles as the data was not findable, not accessible, not interoperable and not reusable (Larooij, 2022). The current inclusion of text obtained by OCR in the Woogle dataset does open up new possibilities for data extraction, which is what will be attempted in this research.

3. Methodology

Before a comparison between the data from the reports and the automatic method can be made, the same periods as used in the reports will have to be defined. Comparing lead times requires two dates to be extracted first, the decision date and the request date. How these are extracted is discussed in this section, followed by a description of how the final decision of a Woo-decision is extracted and classified. Lastly, two additional potentially interesting points of information are extracted from the Woo-decisions. The data and code described in this section is available at *https://github.com/JoranIK/ondraaglijk-matglas*

3.1 Setting the periods for comparison

How are the limits of each period set?

The report Ondraaglijk Traag uses Wob-decisions (the Woo replaced the Wob, or Wet openbaarheid van bestuur on May 1, 2022) published on rijksoverheid.nl in the period of October 2020 until September 2021. For some ministries a different period is used, for example November 2020 until October 2021. For the Ministry of Health, Welfare and Sport a divergent period is used, April 2019 until March 2020 (Enthoven et al., 2022). The dataset used for the report has been granted access to by Maarten Marx and Serv Wiemers.

In the report Matglas Wob-decisions are used which have been published on rijksoverheid.nl until the 10th of January with a decision date in the period of the 1st of January until the 30th of April 2022, and Woo-decisions with a decision date from the 1st of May 2022 until the 31st of December 2022 (Fanoy et al., 2023). The dataset used in the report is granted access to by Maarten Marx and Veerle Fanoy.

The automatic method described in this paper uses a dataset from Woogle. Documents on Woogle are extracted from the Dutch government website https://open.overheid.nl/ through their API. At the time of writing the Woogle dataset contains 2.8 million documents from amongst others, municipalities, provinces, universities and ministries (Marx et al., n.d.). The Woogle dataset has been granted access to by Maarten Marx.

The period of 2023 will be defined as documents published from the 1st of January 2023 until the 10th of January 2024, with a decision date from the 1st of January 2023 until the 31st of December 2023. The choice for the 10th of January 2024 has been made because documents can be published at a later time than their decision. The dataset containing 2.8 million documents has to be pre-processed and reduced to only the data that is necessary to analyse the Wob- and Woo-decisions from the periods used in the reports Ondraaglijk Traag, Matglas, and the period of 2023. After setting the correct datatypes, combining necessary dataframes and filtering, it contains only Wob- and Woo-documents marked

as decision in Woogle with a document name that does not contain the words 'bijlage', 'inventaris' or 'verzoek' with a decision date from the 1st of April 2019 until the 10th of January 2024, published by ministries. For comparison between the previously manually obtained data used in the reports Ondraaglijk Traag and Matglas, the dataset has been limited as follows. For Ondraaglijk Traag, documents published between October 2020 and September 2021 are used, and for the Ministry of Health, Welfare and Sport documents published between April 2019 and March 2020 are used. For Matglas documents the same limits as used in the report have been applied.

3.2 Extracting the decision date and request date

What data is needed to calculate the lead time?

Lead time is defined as the time needed from the moment an information request has been received until the moment a decision is published. For the calculation of the lead time two dates are needed, the date of receival, and the decision date. The texts used to extract data from, are texts obtained by scanning the PDF-files of Woo-decisions using the open source Tesseract OCR (Marx, 2023).

Decision date

The decision date appears in different forms throughout the texts as interpreted by the OCR. Four different forms were distinguished and separate regular expressions were developed for the extraction of these forms, with a fifth method as last resort.

1. 'Datum 18 maart 2023' or 'Datum 18 03 2023'
   The date is noted after 'Datum' without other characters following the word 'Datum'.
2. 'Datum: 18 maart 2023' or 'Datum: 18 03 2023'
   The date is noted after the word 'Datum' with a colon following the word 'Datum'.
3. 'Datum 18-maart-2023' or 'Datum 18-03-2023'
   The date is noted after the word 'Datum' with dashes separating the individual date components.
4. '18 maart 2023 Datum' or '18 03 2023 Datum'
   In the documents, stamped dates, handwritten dates or poorly made scans can make recognizing the decision date more difficult for the OCR, this can result in the decision date being seen as appearing before the word 'Datum'.
5. Other forms
   As a last method to extract the decision date, the first occurrence of the word 'Datum' has been searched. If a match has been found, eight words before, and six words after are selected and given to a function that looks for, and extracts a date. If multiple dates have been found, the most recent date will be given priority to prevent the request date to be extracted instead.

The first four forms result in a date in text form and are transformed to a date-type using the Python library dateparser that utilizes NLP to recognize written months in Dutch. These five results are combined into one end-result, for which the first form is taken as base case. If form one yielded no result, the next form is used in the listed order. If no date can be found, it will be marked as no decision date found. In some cases the date can be incomplete, dates will only be saved if they contain a day, month and year.

Extracting the decision date proved more difficult than initially expected. The metadata of the documents contained information called 'document datum' on the open.overheid.nl

website, and 'foi_decisionDate' in the Woogle dataset. This date was equal to the decision date in the majority of the cases, but sometimes deviated a few days. For further comparisons this date is used as the decision date unless otherwise mentioned.

Request date

The structure in which the request date is mentioned can vary, five regular expressions have been developed to extract it.

1. The sentence containing the request date is the first sentence after the word 'Geachte'.
   If the word 'Geachte' is found, the following 15 words are selected, and the earliest date is extracted.
2. The sentence containing the request date starts with 'In uw'.
   After the words 'In uw', the following 15 words are selected, and the earliest date is extracted.
3. The first sentence containing the word 'verzoek'.
4. The first sentence containing the word 'verzocht'.
5. The first sentence containing the word 'ontving'.
   The first sentence containing one of the above three words is selected, and the earliest date is extracted.

The results are processed and combined in the same manner as the decision date. As with the decision date, incomplete dates are not saved. If the decision date and the request date are found, a lead time can be calculated by subtracting the request date from the decision date and taking the difference rounded to whole days.

3.3 Classifying the decision

What types of decision can be distinguished?

The decisions made in Woo- and Wob-decisions fall in four categories.
1. Rejected (afgewezen)
The request has been rejected in full, or the requested date has already been made public.
2. Not present (niet aanwezig)
The requested information is not present at the ministry the request has been made to.
3. Partially made public (deels openbaar)
The requested documents will be partially made public, or the requested documents will all be made public, with the exception of information concerning personal information.
4. Public (openbaar)
The requested information will be made public in full.
The British government makes a comparable distinction (*Freedom of Information Statistics: Annual 2022 Bulletin*, 2023), and this distinction between decision types has been confirmed by the lead author of the report Matglas, V. Fanoy (Personal communication, 29/11/2023).

How has the decision classifying model been trained?

To classify the final decisions, the sentences containing the decision have first been extracted from decisions made in 2021 to 2023. In most cases the decision is in a separate section, marked by the header 'Besluit'. The first four sentences of the section have been extracted and saved. After manual classification using Doccano, it became clear that the ratio in

which the decision types appeared was strongly imbalanced. the decision type 'Partially made public' occurred almost twice as many times as the other three decision types combined.

To prevent majority class classification the ratio has been manually adjusted by applying the imbalanced model to the extracted decision sentences of 2021 to 2023 and thus having a basic distinction between decision types. These decision types were divided and manually re-classified, with an emphasis on the minority classes. The final dataset contained 900 manually labelled decisions, consisting of 293 partially public, 237 not present, 190 rejected and 180 public decisions. The decisions have been shuffled and split into a training set of 75%, a test set of 10% and a validation set of 15% of the decisions. This dataset is then trained with the Python library spaCy, using the textcat preset with the large Dutch language model 'nl_core_news_lg'.

3.4 What other data can be useful to analyse Woo-decisions?

The process from request until decision can contain numerous different obstacles, the following datapoints have been chosen because they can have an impact on the time taken to reach a decision and therefore might be interesting for analysis, although these do not encompass all possible obstacles.

Determining if a Woo-decision is adjourned

The theoretical maximum lead time of a document is 42 days. Article 4.4 of the Wet Open Overheid states that a maximum of four weeks is given to make a decision starting from the day the request has been confirmed as received. The decision date can be adjourned with two weeks if the request is of considerable size, or if the request is complicated. An additional two weeks can be taken if the information contains information of which a third party could have objections against, in those two weeks the third party is given the chance to review the information, however these last two weeks only start after the decision has been made public. To see if a decision has been adjourned, the document is searched to see if it contains the word 'verdaag'. Sentences containing this word have been selected and saved. If a sentence was found, the document will be marked as 'adjourned'.

Determining if a Woo-decision is a partial decision

Decisions can be divided into multiple smaller decisions. To determine if a document is part of a partial decision, the document has been searched for the word 'deelbesluit'. If this word is found, the sentence containing it is saved. If a sentence was found, the document is marked as partial decision.

4. Results

To answer the main research question, the sub questions will be answered in this section. Firstly, the lead times of each separate ministry are compared using four statistical measures which will point out if a significant difference exists. Secondly, an error analysis will describe with a precision and recall metric why the decision date is often the biggest obstacle in determining the lead time of a document, causing the decision date found in the metadata to be a better option. Thirdly, the decision classifying model is tested against its test set, and last, two results will show that additional data can be extracted from the texts.

4.1 How accurately can the lead time be calculated?

The lead times obtained through calculating the difference between the decision date and the request date can be compared to manually obtained data used for the reports Ondraaglijk Traag and Matglas. The comparison is made by calculating if the lead times show a significant difference compared to the manually obtained data, using four statistical tests from the Python package SciPy.

The lead times of the automatic method have been filtered by removing outliers. Outliers have been removed by calculating the 99.5th percentile for each separate ministry. This limit is chosen because it rejects a minimal amount of (considerably high) lead times while staying close to the maximum of the manually obtained data. Lead times exceeding their corresponding limit have been removed for each ministry. Negative lead times have also been removed. As previously mentioned, the decision date found in the metadata of the documents is used instead of the decision date obtained through extraction from the document text.

Statistical comparison between the lead times of the report and the automatic method

Four statistical tests have been conducted to test whether the automatically extracted lead times are comparable with the manually obtained lead times found in the previous reports. The chosen tests are used to evaluate if a significant difference between the distributions of lead times can be found. The Cramér-von Mises two-sample test is used to determine if the distribution of two samples comes from the same continuous distribution. The Mann-Whitney U test is used to evaluate if the underlying distribution of two samples is the same. The Kolmogorov-Smirnov two-sample test is used to determine if two distributions are identical, and the Anderson-Darling k-sample test is used to determine if k-samples are drawn from the same distribution. The chosen tests are non-parametric because the data for the separate ministries did not always follow a (log)normal distribution, and the sample size can be small (as low as 6).

These tests have been applied to the distributions of all separate ministries from the manually obtained data paired with the automatically extracted data from the same period and ministry. The tables with exact results can be found in Appendix A.

Concluding this, except for the Ministry of Foreign Affairs in the report Matglas, no significant difference can be found within the lead times of both reports and the automatic method as all p-values exceeded the chosen significance level at 0.05. Noting the values of the statistics, indications of a difference can still be found, but they are not significant.

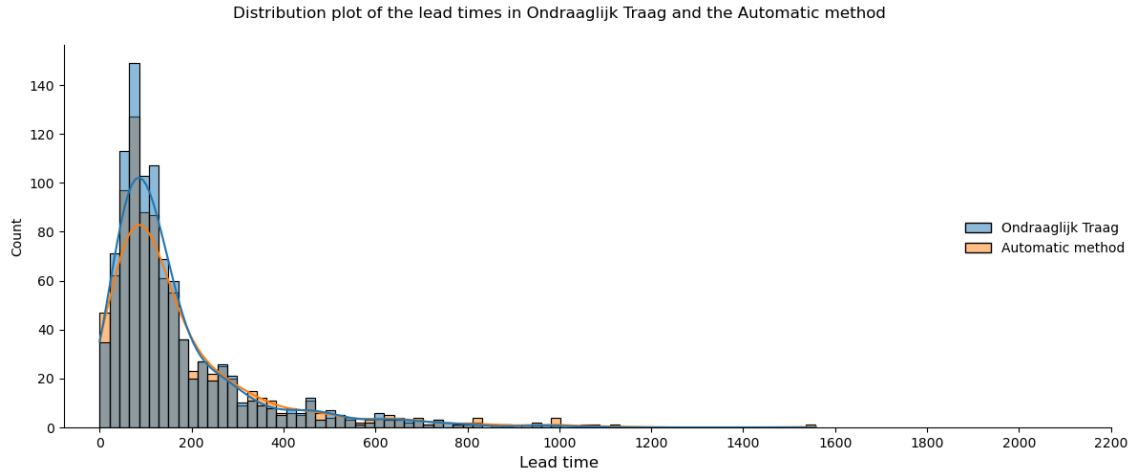


Figure 1: Distribution plot of the lead times in Ondraaglijk Traag and the Automatic method

As there is no significant difference in lead times for both previously made reports tested against the automatic method, it can be expected to give a similar result for the lead times calculated for the year 2023. Having calculated the lead times, a comparison can be made between each report and ministry to give an overview of the lead times through the years. To make this comparison, the data used to make the previous reports has been used, with the addition of the automatically gathered data for the year 2023.

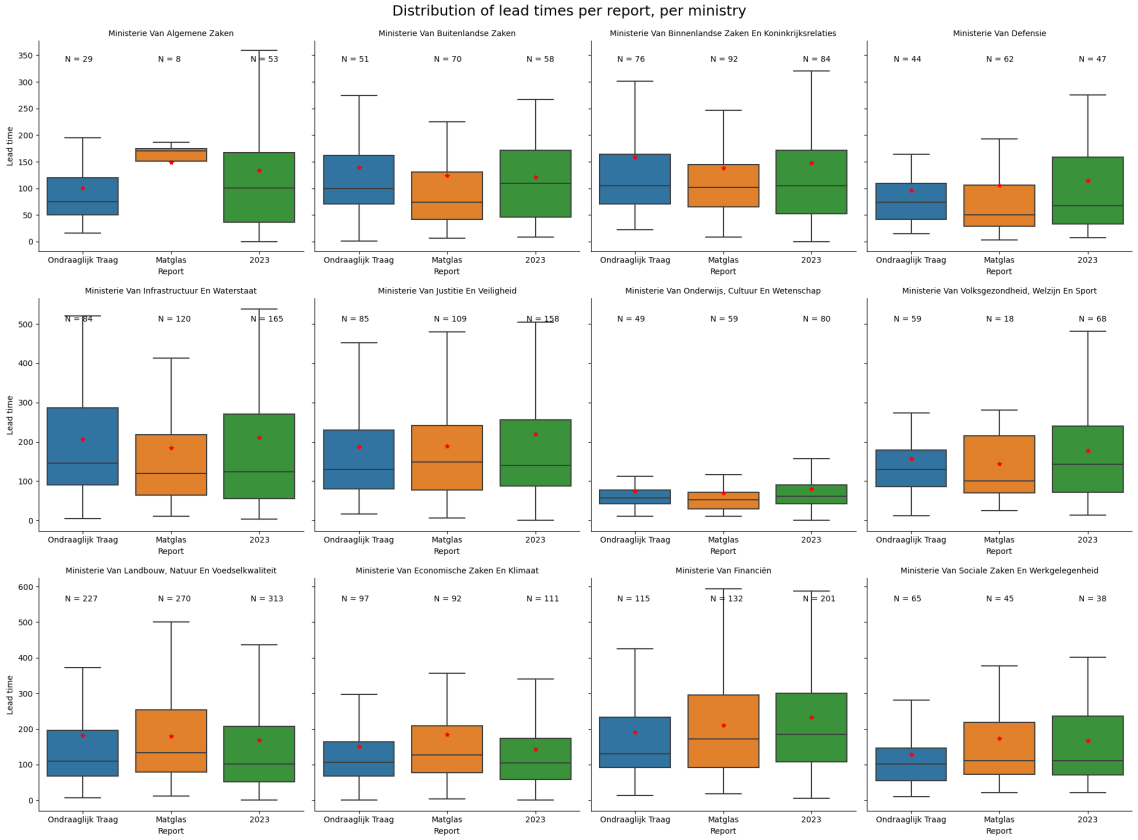


Figure 2: Distribution of lead times for each ministry and report, with a red star depicting the mean lead time.

Not all ministries show improvements in their lead times. Two are apparent, the Ministry of Agriculture, Nature and Food Quality and the Ministry of Economic Affairs and Climate

Policy show declining quartiles as well as a declining median and mean.

4.2 What is the effect of non-machine readable decision days?

The technique used to extract decision dates from the text obtained using OCR did not always succeed to find a decision date. This can often be attributed to the use of stamps, hand-written dates, or dates where it is unclear what the day or month is.
When taking documents used in the period of 2023 as example, a manual examination shows that in the 197 documents where the automatic method failed to retrieve a decision date, 126 stamps were used, 17 documents were poorly scanned from a print, 15 documents had no decision date, 11 dates were handwritten, and 28 had other issues which can be a broken link, misspelling, incomplete date or unusual document structure (Table 1, the abbreviations used for the ministries are defined in appendix C).

Table 1: Comparison between the decision date listed in the metadata and the automatically extracted decision date

Ministry	Same decision date	Different decision date	Missing automatic decision date	Total documents
AZ	52	1	1	54
BZK	42	9	37	88
BuZa	56	2	2	60
Def	18	25	11	54
EZK	81	14	21	116
Fin	173	25	11	209
I&W	126	12	45	183
J&V	158	7	6	171
LNV	270	24	36	330
OCW	78	7	4	89
SZW	38	1	1	40
VWS	44	7	22	73
Total	**1136**	**134**	**197**	**1467**

The automatically extracted decision date has a precision of 0.72, a recall of 0.60 and an F-score of 0.64 (N=1467) when tested against the metadata decision date. The mean absolute error of different decision dates is considerably high at 1001.89 days because of extremely high outliers, the median absolute error lies at 146 days. Missing or wrongly extracting 40% of decision dates would impact the sample size of lead times considerably. Which is why the decision date found in the metadata has been used to acquire a larger and more accurate number of lead times.

The Ministry of General Affairs (AZ) had no stamps in their 54 Woo-decisions, while the Ministry of the Interior and Kingdom Relations (BZK) used 72 stamps in their 88 Woo-decisions, which can be expected given the amount of missing automatically extracted dates. Of 197 decision dates that could not be automatically extracted, the stamps, handwritten, poor scans and absent decision dates attributed to 85.8% of the cases. To improve the machine-readability of Woo-decisions, the use of stamps, handwriting and scanned documents should be minimized, which in turn will improve the extent in which

the Woo-documents comply with the FAIR principles and the Dutch freedom of information law.

Figure 3: Examples of decision dates as found in the documents of 2023

4.3 Classification of the decision

The end result has been classified using the model trained on 810 decisions. Unfortunately there is no gold standard to test it against. The model has been measured against the test set, the metrics and their results are as described below in Table 2.

Table 2: Results of the decision classifier tested against the test set (N=90)

	Precision	Recall	F-score
Openbaar	0.59	0.84	0.70
Deels openbaar	0.96	0.68	0.79
Afgewezen	0.6	0.82	0.69
Niet aanwezig	0.84	0.84	0.84
Micro scores	**0.76**	**0.76**	**0.76**

The results show that the class deels openbaar has a low recall score (0.68), not all decisions with this decision type are classified as such. The classes openbaar and afgewezen have relatively low precision scores, decisions which do not belong to these classes, are often classified as such. A possible solution would be to gather more samples belonging to the classes openbaar and afgewezen to improve the accuracy of the model, these classes have the least samples in the training data. Applying the model to the data of the year 2023 results in the following graph (Figure 4).

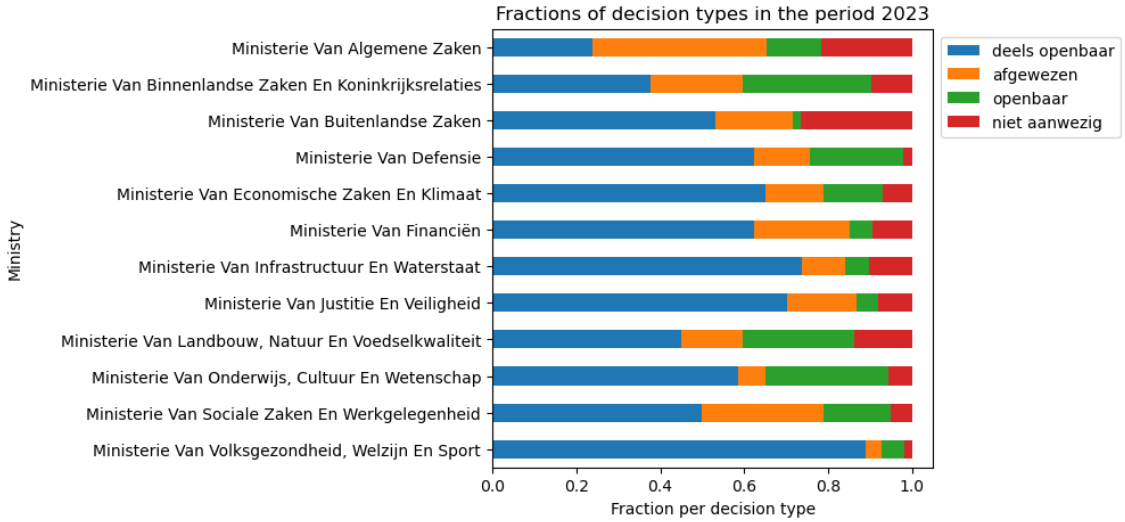


Figure 4: Distribution of decision types using the decision classifier model

The low recall score of the class deels openbaar indicates that it is underrepresented in the results. As the classes openbaar and afgewezen have low precision scores, it is likely that these are overrepresented at the cost of the class deels openbaar.

4.4 What other data can be useful to analyse Woo-decisions?

What fraction of Woo-decisions has been adjourned?

Adjourning Woo-decisions can be an indication that a ministry receives complicated or sizable requests for information, and thus might have a higher median lead time. Figure 5 shows the fraction of Woo-decisions which mentioned the word used to describe an adjournment.

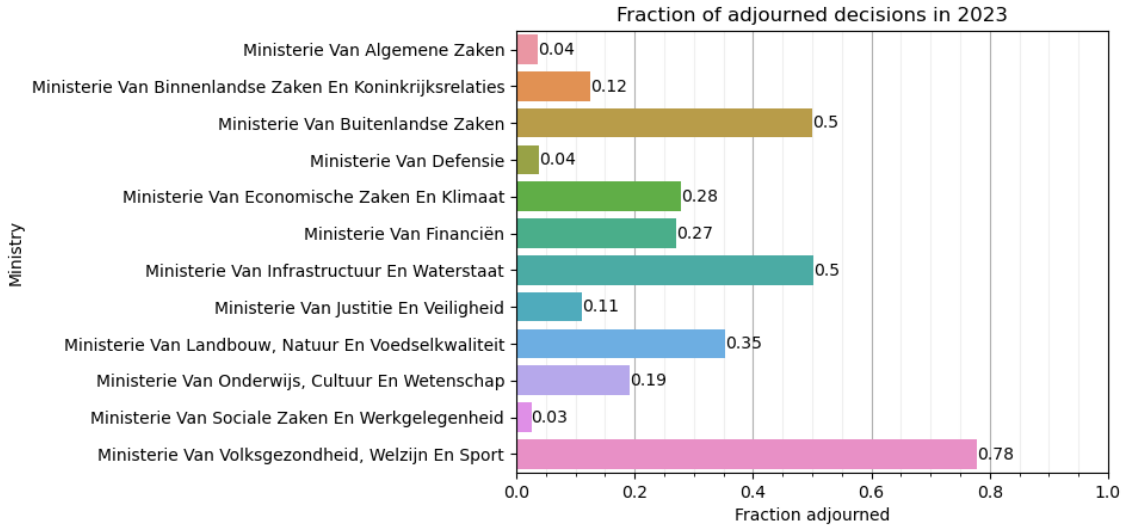


Figure 5: Fraction of adjourned decisions in 2023

What fraction of Woo-decisions is a partial decision?

Making partial decisions can give the requester of information updates throughout the process if it is expected to take a considerably long time to decide. Figure 6 depicts the

amount of Woo-decisions which mentioned the word used to describe a partial decision as a fraction of the total amount of Woo-decisions in the concerning ministry.

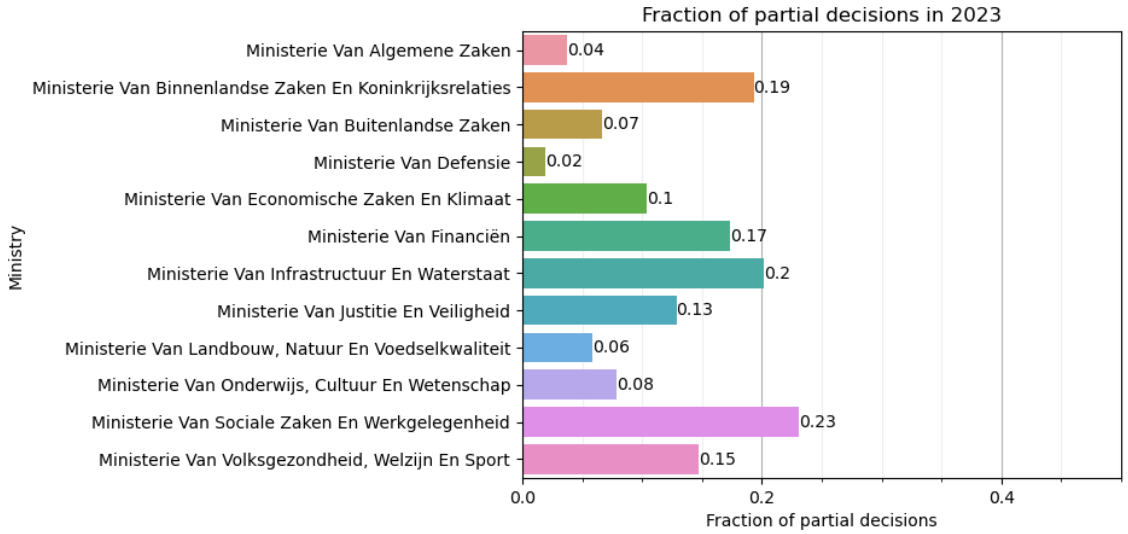


Figure 6: Fraction of partial decisions in 2023

5 Discussion

During the research, three limitations were noticed regarding partial decisions, objections and the formatting of text. The generality of the automatic method will also be discussed. Two possibilities regarding partial decisions and objections against Woo-decisions will be pointed out for further research.

5.1 Limitations of the study

Partial decisions are treated as regular decisions in this study. A document is published, with a request date and decision date, and the automatic method calculates the lead time. However it is not optimal to calculate it this way, as it is part of a single request with a single final decision date, only it is split in parts. It is better to see these partial decisions as a whole, with the intermediate decisions as additional data. Combining these partial decisions was not an achievable goal seen the time of this study and the state of the data.

No distinction was made between decisions and objections to decisions. The law concerning the 'Wet Open Overheid' does not explicitly state whether the same maximum lead time is true to make a decision on an objection. Objections to Woo-decisions fall under the rules of general objections against the government.

Request dates are sometimes written in a specific abbreviated format. The year is written as the abbreviation 'jl.'. The method used to extract dates can not convert this to the corresponding year, and the dates are dismissed as result.

As shown in the action plan 'Actieplan Open overheid 2023-2027', the Dutch freedom of information law might undergo changes in the future to (amongst others) increase the machine readability and alignment with the FAIR principles. While developing the automatic method, only Woo-decisions from ministries ranging in the timespan of April 2019 to January 10, 2024 are taken into consideration. If the structure of future Woo-decisions changes significantly, the automatic method might have to be reworked. As an

additional remark, the method used is not tested on other instances than the 12 ministries. The method is still quite general, it can be tested on other governmental bodies such as municipalities or provinces if they adhere to a similar document structure and wording as the ministries.

5.2 Further research

Apart from the further development needed on the decision classifier, two more subjects that can be interesting for further research are listed below.

In the case of partial decisions, connecting identifiers which referred to identifiers of the previous partial decision were sometimes present in the text, and other clues could connect them to each other. If no metadata is added which allows these documents to be connected, it would be interesting to see if it is possible to link these using for example titles or the documents content. If this appears to be impossible, it would not be in line with the FAIR principles as stated by Wilkinson et al. (2016), due to the lack of findability and reusability.

Objections against Woo-decisions can quickly escalate to a hearing in court and with these hearings, money is involved. It can be useful to examine how many objections end in a revision of the previous decision, and what it takes to reach a new conclusion.

6 Conclusion

To answer the main research question, the four sub questions will be answered first, and then the answers will be combined into a final conclusion.

1. *How accurately can the lead time be automatically calculated?* The lead times resulting from automatic extraction in combination with the metadata decision date are not significantly different than earlier manually obtained data, except for the ministry BZK in the Matglas period. While they do not have a significant difference, it can not be concluded that the distributions are the same, therefore the resulting distributions can be used to make a reliable estimation, in particular for the median and quantiles as they are less prone to distortion because of outliers.

2. *What is the effect of non-machine readable decision days?* The decision date is the biggest obstacle in automatically determining lead times because of the use of stamps, handwritten dates, poorly scanned documents and absent decision dates which can cause up to 40% of decision dates to be extracted incorrectly or not at all. Non-machine readable decision days have a considerable effect on the amount of lead times that can be calculated, which in turn gives an indication that the content of the published Woo-decisions still do not follow the FAIR principles well enough to be machine-readable, as also found by Larooij (2022).

3. *How accurately can decisions be classified?* The decisions made can be extracted and classified, however the trained model does not perform well enough to draw conclusions, more data has to be gathered to improve the performance.

4. *What other data can be useful to analyse Woo-decisions?* More data can be extracted from Woo-decisions than dates and decisions, sentences stating if documents are adjourned or part of a partial decision can be used for further analysis, showing that with automatic extraction, it is relatively easy to retrieve additional information.

Texts from Woo-decisions obtained through OCR can be used for the analysis of Woo-decisions and additional information can be extracted with relative ease. However the Woo-decisions are not yet fully machine-readable, which is why currently only a reliable estimation can be made by using this method. If exact numbers of the lead times are required, further development is necessary. This is also the case for the decision classifier, which in its current state can give a basic indication of the distribution of decisions, but needs further development for precise results.

References

Bakker, F. T. (2022). *Timeline extraction from decision letters using ChatGPT.*

Enthoven, G., Wiemers, S., Den Uijl, S., Nouwen, A., Kuilman, E., Jorissen, R., & Vos-Goedhart, T. (2022). *Ondraaglijk traag.*

Fanoy, V., Kaandorp, C., Roebroek, M., Tazelaar, E., Den Uijl, S., Schuil, J., Van De Beek, A., Wiemers, S., & Enthoven, G. (2023). *Matglas.*

Freedom of Information statistics: annual 2022 bulletin. (2023, May 15). GOV.UK. https://www.gov.uk/government/statistics/freedom-of-information-statistics-annual-2022/freedom-of-information-statistics-annual-2022-bulletin#outcomes

Larooij, M. (2022). *De FAIRificatie van Woo-dossiers.*

Marx, M. (2023, May 12). *Woogle data nu vrij beschikbaar.* Wooverheid. Retrieved January 22, 2024, from https://wooverheid.nl/2023/05/12/woogle-data-nu-vrij-beschikbaar/

Marx, M., Kamps, J., & Larooij, M. (n.d.). *Woogle overview.* Woogle. Retrieved January 20, 2024, from https://woogle.wooverheid.nl/overview

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2023). *Actieplan Open Overheid 2023-2027.*

Waltl, B., Bonczek, G., & Matthes, F. (Eds.). (2018). Rule-based Information Extraction - Advantages, Limitations, and Perspectives. *Jusletter IT, 22.* https://wwwmatthes.in.tum.de/pages/1w12fy78ghug5/Rule-based-Information-Extraction-Advantages-Limitations-and-Perspectives

Wet open overheid. (2023, April 1). https://overheid.nl. Retrieved January 25, 2024, from https://wetten.overheid.nl/BWBR0045754/2023-04-01

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Da Silva Santos, L. O. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T. W., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18

Appendix

A. Statistics comparing lead times of Ondraaglijk Traag and the automatic method

A.1 Resulting p-values comparing Ondraaglijk Traag and the automatic method

Table 3: Statistical comparison of the lead times of Ondraaglijk Traag and the automatic method for a similar period. This table contains the p-values, for the statistic, see the next table.

	Cramer von Mises 2 sample p-value	Mann-Whitney U p-value	Kolmogorov-Smirnov 2 sample p-value	Anderson-Darling k sample p-value	N (Automatic method)	N (Matglas)
AZ	1.00	0.86	1.00	0.25	26	29
BZK	0.61	0.38	0.81	0.25	73	76
BuZa	0.46	0.32	0.41	0.25	51	51
DEF	0.99	0.99	1.00	0.25	30	44
EZK	0.59	0.33	0.88	0.25	91	97
FIN	0.91	0.69	0.98	0.25	102	115
I&W	0.69	0.37	0.87	0.25	91	84
J&V	0.66	0.51	0.67	0.25	86	85
LNV	0.89	0.80	0.98	0.25	235	227
OCW	0.64	0.44	0.75	0.25	41	49
SZW	0.79	0.56	0.94	0.25	69	65
VWS	0.35	0.94	0.35	0.24	21	59

A.2 Resulting statistics comparing the lead times of Ondraaglijk Traag and the automatic method

Table 4: Statistical comparison of the lead times of Ondraaglijk Traag and the automatic method for a similar period. This table contains the statistics, for the p-value, see the previous table.

	Cramer von Mises 2 sample statistic	Mann-Whitney U statistic	Kolmogorov-Smirnov 2 sample statistic	Anderson-Darling k sample statistic	N (Automatic method)	N (Matglas)
AZ	0.02	366	0.07	-1.10	26	29
BuZa	0.10	2544	0.10	-0.67	73	76
BZK	0.13	1150	0.18	-0.23	51	51
DEF	0.03	658	0.08	-1.09	30	44
EZK	0.10	4049.5	0.08	-0.62	91	97
FIN	0.05	5681.5	0.06	-0.92	102	115
I&W	0.08	3521.5	0.09	-0.60	91	84
J&V	0.09	3869.5	0.10	-0.05	86	85
LNV	0.05	26307	0.04	-0.94	235	227
OCW	0.09	1100.5	0.13	-0.02	41	49
SZW	0.07	2109.5	0.09	-0.86	69	65
VWS	0.17	612	0.23	0.34	21	59

B. Statistics comparing lead times of Matglas and the automatic method

B.1 Resulting p-values comparing Matglas and the automatic method

Table 5: Statistical comparison of the lead times of Matglas and the automatic method for a similar period. This table contains the p-values, for the statistics, see the next table.

	Cramer von Mises 2 sample p-value	Mann-Whitney U p-value	Kolmogorov-Smirnov 2 sample p-value	Anderson-Darling k sample p-value	N (Automatic method)	N (Matglas)
AZ	1.00	1	0.95	0.25	6	8
BuZa	0.94	0.63	0.95	0.25	90	101
BZK	0.03	0.03	0.05	0.04	72	70
DEF	0.40	0.27	0.52	0.25	56	64
EZK	0.63	0.37	0.87	0.25	88	96
FIN	0.46	0.26	0.73	0.25	133	139
I&W	0.12	0.12	0.19	0.12	126	128
J&V	0.89	0.62	0.95	0.25	133	111
LNV	0.73	0.47	0.82	0.25	251	280
OCW	0.98	0.88	0.99	0.25	46	60
SZW	0.42	0.26	0.58	0.25	45	48
VWS	0.83	0.53	0.90	0.25	39	18

B.2 Resulting statistics comparing Matglas and the automatic method

Table 6: Statistical comparison of the lead times of Matglas and the automatic method for a similar period. This table contains the statistics, for the p-value, see the previous table.

	Cramer von Mises 2 sample statistic	Mann-Whitney U statistic	Kolmogorov-Smirnov 2 sample statistic	Anderson-Darling k sample statistic	N (Automatic method)	N (Matglas)
AZ	0.03	23.5	0.25	-0.99	6	8
BZK	0.04	4358.5	0.07	-1.05	90	101
BuZa	0.54	1984	0.22	2.31	72	70
DEF	0.15	1583.5	0.14	-0.40	56	64
EZK	0.09	3898.5	0.08	-0.66	88	96
FIN	0.13	8516	0.08	-0.38	133	139
I&W	0.31	7163	0.13	1.02	126	128
J&V	0.05	7108.5	0.06	-0.89	133	111
LNV	0.07	33859	0.05	-0.55	251	280
OCW	0.03	1356	0.08	-0.84	46	60
SZW	0.14	931.5	0.16	-0.31	45	48
VWS	0.06	313.5	0.15	-0.77	39	18

C. Abbreviations

Table 7: Abbreviations used for the 12 ministries

Fully written out ministry name	Abbreviation
Ministerie Van Algemene Zaken	AZ
Ministerie Van Binnenlandse Zaken En Koninkrijksrelaties	BZK
Ministerie Van Buitenlandse Zaken	BuZa
Ministerie Van Defensie	DEF
Ministerie Van Economische Zaken En Klimaat	EZK
Ministerie Van Financiën	FIN
Ministerie Van Infrastructuur En Waterstaat	I&W
Ministerie Van Justitie En Veiligheid	J&V
Ministerie Van Landbouw, Natuur En Voedselkwaliteit	LNV
Ministerie Van Onderwijs, Cultuur En Wetenschap	OCW
Ministerie Van Sociale Zaken En Werkgelegenheid	SZW
Ministerie Van Volksgezondheid, Welzijn En Sport	VWS