Text2Storyline: Generating Enriched Storylines From Text

Francisco Manuel Pires Gonçalves
Mestrado em Ciência de Computadores
Departamento de Ciência de Computadores
2022

Orientador
Prof. Ricardo Campos, Instituto Politécnico de Tomar

Coorientador
Prof. Alípio Mário Guedes Jorge, Faculdade de Ciências da Universidade do Porto
Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_________
Sworn Statement

I, Francisco Manuel Pires Gonçalves, enrolled in the Master Degree Computer Science at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offense.

Francisco Manuel Pires Gonçalves
17th November 2022
Text2Storyline: Generating Enriched Storylines From Text

Author: Francisco GONÇALVES
Supervisor: Ricardo CAMPOS
Co-supervisor: Alípio Mário Guedes JORGE

A thesis submitted in fulfilment of the requirements for the degree of MSc. Computer Science at the Faculdade de Ciências da Universidade do Porto Departamento de Ciência de Computadores

November 17, 2022
“Why do we fall, sir? So that we can learn to pick ourselves up”

Michael Caine as Alfred Pennyworth, in *Batman Begins*
Acknowledgements

First of all, I would like to thank Professor Ricardo Campos for all the hours of guidance and patience and for motivating to push through any adversity that arose throughout this project. To Professor Alípio Jorge for providing his expertise on more sensible matters and for his aid in moments of need. I would also like to thank Jorge Duque, who was responsible for developing the Time-Matters online demo, for his availability and help, which were fundamental during the initial stages of this work. Thank you to everyone who took the time to participate in either or both of our surveys.

To my family, my biggest thanks to my parents in particular, for the unconditional love and support and for always providing me with everything I could ever need. I could not have gotten this far without them.

To all my friends, for showing interest in my work and helping me accomplish this task, but above all for being what a friend should be throughout the years.

A special thank you to my girlfriend Yasmin, for always supporting me, especially in the toughest moments and for listening to my endless ramblings about this project.
Abstract

Faculdade de Ciências da Universidade do Porto
Departamento de Ciência de Computadores

MSc. Computer Science

Text2Storyline: Generating Enriched Storylines From Text

by Francisco GONÇALVES

In recent years, the amount of information generated, consumed and stored has grown at an astonishing rate, making it difficult for those seeking information to extract knowledge in good time. This has become even more important, as the average reader is not as willing to spare more time out of their already busy schedule as in the past, thus prioritizing news in a summarized format, which are faster to digest. On top of that, people tend to increasingly rely on strong visual components to help them understand the focal point of news articles in a less tiresome manner. Information retrieval and data visualization are currently very important tools that enable information summarization. However, despite development and implementation of new algorithms in this domain, the problem of building and presenting consistent narrative structures in the spectrum of web articles is yet to be completely resolved. This growing demand, focused on exploring information through visual aspects, urge the need for the emergence of alternative approaches concerned with text understanding and narrative exploration. This motivated us to propose Text2Storyline, a platform for generating and exploring enriched storylines from an input text, a URL or a user query. The latter is to be issued on the Portuguese Web Archive (Arquivo.pt), therefore giving users the chance to expand their knowledge and build up on information collected from web sources of the past. To fulfill this objective, we propose a system that makes use of the Time-Matters algorithm to filter out non-relevant dates and organize relevant content by means of different displays: ‘Annotated Text’, ‘Entities’, ‘Storyline’, ‘Temporal Clustering’ and ‘Word Cloud’. To extend the users’ knowledge, we rely on entity linking to connect persons, events, locations and concepts found in the text to Wikipedia pages, a process also known as Wikification. Each of the entities is then illustrated by means of an image collected from the Arquivo.pt. Our proposal was evaluated
by conducting two surveys as a means to assess Text2Storyline’s usability and effectiveness, respectively. The results obtained showed that potential users value this kind of platform expressing particular interest in the ‘Storyline’ and ‘Annotated Text’ features. However, other components, such as the ‘Entities’ feature also show potential due to its appealing yet simple and extremely informative design. Such results show that one such system may configure an important tool for generating automatic narratives for large scale stories from the current and the past web. Notwithstanding, there is still room for improvement, as results showed, before users can have a complete and flawless experience. As future work, we plan to improve our image selection method, by developing and implementing an algorithm capable of extracting more accurately event-related images. On top of that, our system is also limited when handling events and temporal expressions together. Currently, the highest scored set of keywords is used as the occurrence that the respective date depicts. This may not always work correctly and a more precise event detection mechanism must be implemented in the future.

**Keywords:** Narrative Visualization, Temporal Information Extraction, Named Entity Recognition, Wikification, Web Archiving
Nos últimos anos, a quantidade de informação gerada, consumida e armazenada tem crescido a um ritmo surpreendente, dificultando a quem procura informação extrair conhecimento em tempo útil. Isso se tornou ainda mais importante, pois o leitor comum não está tão disposto a perder tanto tempo da sua agenda já lotada como no passado, priorizando as notícias em formato resumido, que são mais rápidas de digerir. Além disso, as pessoas tendem a confiar cada vez mais em fortes componentes visuais para ajudá-las a entender o ponto focal das notícias de uma maneira menos cansativa. A recuperação da informação e a visualização de dados são atualmente ferramentas muito importantes que habilitam a sumarização da informação. No entanto, apesar do desenvolvimento e implementação de novos algoritmos neste domínio, o problema de construir e apresentar estruturas narrativas consistentes no espectro de artigos da web ainda não foi completamente resolvido. Esta crescente demanda, voltada para a exploração da informação por meio de aspetos visuais, urge a necessidade do surgimento de abordagens alternativas voltadas à compreensão do texto e à exploração narrativa. Isso motivou-nos a propor o Text2Storyline, uma plataforma para gerar e explorar histórias enriquecidas a partir de um texto ou de um termo de pesquisa. Este último será emitido no Arquivo da Web Portuguesa (Arquivo.pt), dando assim aos utilizadores a possibilidade de expandirem os seus conhecimentos e acumularem informações recolhidas em fontes web do passado. Para cumprir este objetivo, propomos um sistema que utiliza o algoritmo Time-Matters para filtrar datas não relevantes e organizar conteúdos relevantes por meio de diferentes exposições: ‘Texto Anotado’, ‘Entidades’, ‘Storyline’, ‘Temporal Clustering’ e ‘Word Cloud’. Para expandir o conhecimento dos usuários, recorremos à vinculação de entidades para conectar pessoas, eventos, locais e conceitos encontrados no texto e páginas
da Wikipédia, processo também conhecido como Wikificação. Cada uma das entidades é então ilustrada através de uma imagem recolhida do Arquivo.pt. A nossa proposta foi avaliada através da realização de dois inquéritos como forma de avaliar a usabilidade e eficiência do Text2Storyline, respectivamente. Os resultados obtidos mostraram que os potenciais utilizadores valorizam este tipo de plataforma, manifestando interesse específico nas funcionalidades ‘Storyline’ e ‘Texto Anotado’. No entanto, outros componentes, como a funcionalidade ‘Entidades’ também apresentam potencial devido ao seu design apelativo mas simples e extremamente informativo. Tais resultados mostram que um sistema destes pode ser uma ferramenta importante para gerar narrativas automáticas para histórias em larga escala da web atual e passada. Não obstante, ainda há espaço para melhorias, como os resultados mostraram, para que os usuários possam ter uma experiência completa e sem falhas. Como trabalho futuro, planeamos melhorar o nosso método de seleção de imagens, desenvolvendo e implementando um algoritmo capaz de extrair imagens relacionadas a eventos com mais precisão. Além disso, o nosso sistema também é limitado ao manipular eventos e expressões temporais juntos. Atualmente, o conjunto de palavras-chave com o valor mais alto é usado como a ocorrência que a respectiva data retrata. Isso pode nem sempre funcionar corretamente e um mecanismo de detecção de eventos mais preciso deve ser implementado no futuro.

**Palavras-chave:** Visualização de Narrativas, Extração de Informação Temporal, Reconhecimento de Entidades Nomeadas, Wikificação, Arquivamento da Web
Contents

Acknowledgements v
Abstract vii
Resumo ix
Contents xi
List of Figures xv
Glossary xix

1 Introduction 1
1.1 Motivation ....................................................... 1
1.2 Objective ..................................................... 2
1.3 Context ......................................................... 3
1.4 Contributions .................................................. 3
1.5 Time Frame .................................................... 4
1.6 Thesis Structure ............................................... 5

2 Related Work 7
2.1 Narratives ..................................................... 7
2.2 Temporal Information Extraction ............................... 13
2.3 Keyword Extraction ........................................... 15
2.4 Named Entity Recognition .................................... 17
  2.4.1 Wikification ............................................... 18
2.5 Text Summarization .......................................... 21
2.6 Data Collection and Web Archiving .......................... 22
2.7 Summary ....................................................... 24

3 Text2Storyline Architecture 25
3.1 Technologies .................................................. 26
  3.1.1 Angular Framework ........................................ 26
  3.1.2 TypeScript ................................................ 26
  3.1.3 Python .................................................... 26
  3.1.4 Flask ...................................................... 27
  3.1.5 Docker ..................................................... 27
3.2 Foundational Algorithms ........................................ 27
  3.2.1 Time-Matters .............................................. 28
  3.2.2 YAKE! ...................................................... 32
  3.2.3 Wikifier .................................................... 34
  3.2.4 MediaWiki ................................................... 35
3.3 Features .......................................................... 36
  3.3.1 Annotated Text .............................................. 36
  3.3.2 Entities ...................................................... 38
  3.3.3 Storyline ..................................................... 39
  3.3.4 Temporal Clustering .................................... 40
  3.3.5 Word Cloud .................................................. 40
  3.3.6 Search Results ............................................ 41
3.4 Summary .......................................................... 42

4 Text2Storyline Showcase ........................................ 43
  4.1 Home Page ...................................................... 43
  4.2 Interface Exploration ...................................... 46
    4.2.1 Creating Narratives from a URL ...................... 47
      4.2.1.1 Queen Elizabeth II ................................. 48
      4.2.1.2 Dante Alighieri ................................... 54
    4.2.2 Creating Narratives from a Query .................... 58
    4.2.3 Creating Narratives from Free Text .................. 66
  4.3 Summary ........................................................ 72

5 Results and Discussion .......................................... 73
  5.1 User Preference Evaluation ................................. 74
    5.1.1 Users’ Preferences on Text2Storyline components .. 75
    5.1.2 Storyline Visual Component Evaluation .............. 80
    5.1.3 Entities Visual Component Evaluation ............... 83
  5.2 Comparability Evaluation ................................ 85
    5.2.1 Storyline Feature’s Effectiveness ................... 86
      5.2.1.1 Queen Elizabeth II ................................. 86
      5.2.1.2 Mário Soares ....................................... 87
      5.2.1.3 Dante Alighieri ................................... 89
      5.2.1.4 Middle East Respiratory Syndrome ............... 90
    5.2.2 Storyline Relevance Date Effectiveness ............. 93
      5.2.2.1 Queen Elizabeth II ................................. 94
      5.2.2.2 Mário Soares ....................................... 95
      5.2.2.3 Dante Alighieri ................................... 97
      5.2.2.4 Middle East Respiratory Syndrome ............... 98
    5.2.3 Storyline Image Effectiveness ........................ 100
    5.2.4 Entities’ Effectiveness ................................. 101
      5.2.4.1 Queen Elizabeth II ................................. 102
      5.2.4.2 Mário Soares ....................................... 103
      5.2.4.3 Dante Alighieri ................................... 104
  5.3 Summary ........................................................ 106
6 Conclusions 107
A Survey: Text2Storyline Usability 111
B Survey: Text2Storyline Evaluation 115

Bibliography 131
List of Figures

1.1 Jornal Público timeline on TAP’s privatization ........................................ 2
1.2 Time Frame of tasks performed throughout the year ................................. 4

2.1 Wikipedia Beta Cluster’s Wikistories for Dog ........................................ 10
2.2 Digital Libraries’ Narrative on Dante Alighieri Source: https://dlnarratives.eu/project.html ................................................................. 11
2.3 Digital Libraries’ Main Hub for the query “Mário Soares” ......................... 12
2.4 Temporal Expressions obtained for the running example using Heideltime’s online demo ................................................................. 14
2.5 Time-Matters’ Timeline visualization for the running example .................... 15
2.6 Keywords obtained using YAKE! for the running example ........................ 16
2.7 NER and Entity Linking processes exemplified. Source: https://en.wikipedia.org/wiki/Entity_linking .................................................. 17
2.8 Dependency of entities in a sentence of the running example using spaCy .... 18
2.9 Named Entity Recognition of the running example using spaCy ............... 19
2.10 Illustrative example from the Wikifier online demo, showing the annota-
tions (a) and the full text marked with the same annotations as hyperlinks to the respective Wikipedia pages (b) ........................................... 20
2.11 Text Summarization of the running example using the BERT-extractive-summarizer approach Source: SMRZR.io (https://smrzr.io/) ............... 22

3.1 Heideltime approach with Time-Matters for temporal expression recognition 29
3.2 Rule-based approach with Time-Matters for temporal expression recognition 30
3.3 Keywords obtained with YAKE! .............................................................. 32
3.4 Entity behaviour ...................................................................................... 35
3.5 Text2Storyline diagram ........................................................................... 37
3.6 Annotated Text feature ............................................................................ 38
3.7 Entities feature ....................................................................................... 38
3.8 Storyline feature .................................................................................... 39
3.9 Single event of the Storyline for a query ................................................ 39
3.10 Temporal Clustering feature .................................................................. 40
3.11 Word cloud feature ............................................................................... 41
3.12 Search Results feature ........................................................................... 42

4.1 Home page .............................................................................................. 44
4.2 Second part of the home page ................................................................... 45
4.3 Free Text section of Text2Storyline ....................................................... 46
4.4 News article page on Queen Elizabeth’s life ............................................ 47
5.6 Distribution of votes on each score for every option and respective average rank from a question in the first survey to order different ‘Search Results’ options .............................................. 80
5.7 Question from the first survey on the preferable description for a storyline event .................................................. 81
5.8 Two questions from the first survey to inquire feedback on usage of illustrations for Storyline events ................................. 82
5.9 Two questions from the first survey on a comparison between Text2Storyline’s Entities and Dandelion’s Entities ............................. 84
5.10 Storyline events with the most positive (a) and most negative (b) results ................................................................. 87
5.11 Full distribution of votes in percentage (%) for all questions for the example on Queen Elizabeth II ................................. 87
5.12 Storyline events with the most positive (a) and most negative (b) results ................................................................. 89
5.13 Full distribution of votes in percentage (%) for all chronological moments for the example on Mário Soares ................................. 89
5.14 Storyline events with the most positive (a) and most negative (b) results ................................................................. 91
5.15 Full distribution of votes in percentage (%) for all chronological moments for the example on Dante Alighieri ................................. 91
5.16 Storyline events with the most positive (a) and most negative (b) results ................................................................. 92
5.17 Full distribution of votes for all questions for the example on the Middle East Respiratory Syndrome (MERS) ......................... 93
5.18 Average score for each element across all examples .................. 93
5.19 Storyline events with a high score (0.429) (a) and a low score (0.228) (b) ................................................................. 95
5.20 Full distribution of results in percentage (%) for all questions on the relevance of events on Queen Elizabeth II ......................... 95
5.21 Storyline events with a high score (0.642) (a) and a lower score (0.416) (b) ................................................................. 96
5.22 Full distribution of results in percentage (%) for all questions on the relevance of events on Mário Soares ................................. 97
5.23 Storyline events with a low score (a/b) .................................................. 98
5.24 Full distribution of results in percentage (%) for all questions on the relevance of events on Dante Alighieri ................................. 98
5.25 Storyline events with a low score (a/b) .................................................. 99
5.26 Full distribution of results in percentage (%) for all questions on the relevance of events on MERS ................................. 100
5.27 Questions from the second survey comparing Text2Storyline’s automatically selected illustration (Option A) to a manually selected image (Option B) ........................................................................ 102
5.28 Comparison of entities in general (a/b) and considering a single entity (c/d) between Text2Storyline and Dandelion ................. 103
5.29 Comparison of entities in general (a/b) and considering a single entity (c/d) between Text2Storyline and Dandelion ................. 104
5.30 Comparison of entities in general (a/b) and considering a single entity (c/d) between Text2Storyline and Dandelion .................. 105
A.1 Complete survey on Text2Storyline’s usability ................................. 111
A.1 Complete survey on Text2Storyline’s usability ................................. 112
A.1 Complete survey on Text2Storyline’s usability ................................. 113
A.1 Complete survey on Text2Storyline’s usability ................................. 114
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 115
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 116
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 117
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 118
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 119
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 120
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 121
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 122
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 123
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 124
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 125
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 126
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 127
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 128
B.1 Complete survey on Text2Storyline’s evaluation ........................................ 129
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>BART</td>
<td>Bidirectional and Auto-Regressive Transformers</td>
</tr>
<tr>
<td>BERT</td>
<td>Bidirectional Encoder Representations from Transformers</td>
</tr>
<tr>
<td>EL</td>
<td>Entity Linking</td>
</tr>
<tr>
<td>HTML</td>
<td>HyperText Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>IE</td>
<td>Information Extraction</td>
</tr>
<tr>
<td>IE</td>
<td>Information Retrieval</td>
</tr>
<tr>
<td>JS</td>
<td>JavaScript</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>KE</td>
<td>Keyword Extraction</td>
</tr>
<tr>
<td>NBVT</td>
<td>Narrative Building and Visualising Tool</td>
</tr>
<tr>
<td>NER</td>
<td>Named Entity Recognition</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>NLTK</td>
<td>Natural Language Toolkit</td>
</tr>
<tr>
<td>PaaS</td>
<td>Platform as a Service</td>
</tr>
<tr>
<td>PWA</td>
<td>Portuguese Web Archive</td>
</tr>
<tr>
<td>RAKE</td>
<td>Rapid Automatic Keyword Extraction</td>
</tr>
<tr>
<td>REST</td>
<td>Representational State Transfer</td>
</tr>
<tr>
<td>TIE</td>
<td>Temporal Information Extraction</td>
</tr>
<tr>
<td>TS</td>
<td>Text Summarization</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>WWW</strong></td>
<td>World Wide Web</td>
</tr>
<tr>
<td><strong>YAKE!</strong></td>
<td>Yet Another Keyword Extractor</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Visual narratives are advantageous as they are able to condense the useful information in a chronological order and serve it to the user in a light, yet appealing and informative manner. They are often used as a means to extend the readers’ knowledge with elements that are not easily interpreted when reading the text at first hand. Our main goal in this thesis, was to achieve that very same idea, by providing features and tools that allow for more innovative and creative ways of experimenting with the narrative as a whole.

1.1 Motivation

Recent years have shown a clear trend, especially in the younger generations, towards the consumption of information from different formats [1]. Driven by this new paradigm, different stakeholders have made an effort in an attempt to adapt their content to the consumption habits of an increasingly digital audience. In this context, the representation of texts from timelines appears as an alternative to the presentation of data made solely from textual structures, offering users the possibility to become familiar with a given event in a short space of time. Several news outlets have been making efforts in this regard. One illustrative example of this is the storyline (see Figure 1.1) of a Portuguese reference newspaper (Jornal Público1), which documents the privatization of the Portuguese national flag carrier - TAP.

Despite the growing importance of timelines in the context of summarizing data from multiple documents, little is known about their use and application in the context of individual documents and visual narratives. On the other hand, the immense volume of data

1https://www.publico.pt/
existent in web documents also makes it prohibitive to manually build and make this type of interface available. This work is motivated by the concept of exploring more appealing and innovative ways to represent narratives and provide creative tools and features to enhance the user experience.

1.2 Objective

In this thesis, we intend to propose an alternative to the availability of purely textual structures, offering users the possibility to create automatic visual narratives with a temporal focus [2]. Each visual narrative is complemented with a set of related images automatically obtained from a collection of 584 million images preserved by the Arquivo.pt\(^2\), the Portuguese web archive infrastructure. As a way of complementing the information obtained from text documents, we propose to identify a set of relevant keywords and named entities. The detailed information about each entity is obtained by connecting to an external database, Wikipedia\(^3\), thus expanding the information initially obtained from the text. The ultimate objective of this thesis is to understand how users perceive and value this kind of structures.

\(^2\)https://arquivo.pt/
\(^3\)https://en.wikipedia.org/wiki/Main_Page
1.3 Context

The starting point of this thesis was not a blank canvas. This concept of a visual narrative has already been explored and developed to some extent in the past within our team [3], and although more primitive, in comparison with the current version, it served its initial goal of showcasing the Time-Matters\textsuperscript{4} [3] algorithm. However, just as technology keeps advancing, Text2Storyline\textsuperscript{5} was born. Using the foundations from its predecessor, which only created a rather limited number of visual narratives when a URL of an article or a single text was provided, we recreated Text2Storyline to consider new features and visual elements.

In comparison to Time-Matters, Text2Storyline allows users to search for queries which are issued on the Arquivo.pt, and from there create and present visual narratives that are not restricted to a single text, but to a topic instead. In addition to this, it expands the user’s knowledge by presenting automatically identified potentially relevant entities such as persons, locations, objects or concepts and linking them to external knowledge databases, such as Wikipedia, that the user can be redirected to. Multiple other components are also on display to summarize the events or supply a general sense of the text or of the topic provided, by means of temporal clusters or word clouds.

This thesis was carried out as part of the Text2Story\textsuperscript{6} project, which is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within the project UIDB/50014/2020 and LA/P/0063/2020.

1.4 Contributions

With the completion of this work, we developed a pipeline that allows users of the Text2Storyline framework to automatically transform a text or a group of texts into temporal representations that can be explored through a number of different visual elements. Our contributions are as follows:

1. An online website\textsuperscript{7} that makes use of the collection of documents and images preserved by the Arquivo.pt, with the ability to automatically create visual narratives from both individual documents and search queries;

\textsuperscript{4}http://time-matters.inesctec.pt/
\textsuperscript{5}http://text2storyline.inesctec.pt/
\textsuperscript{6}https://text2story.inesctec.pt/
\textsuperscript{7}http://text2storyline.inesctec.pt/
2. A solution that enables showcasing the Time-Matters\(^8\) algorithm on top of queries, rather than just on text or urls \([3]\);

3. Expand the user’s knowledge by identifying narrative elements of a text and connecting them to external knowledge databases, such as Wikipedia;

4. A study that, not only evaluates the developed pipeline, but it also studies how the proposed narrative elements are perceived and valued by users;

5. Participation in the 2022 Arquivo.pt Award.\(^9\)

1.5 Time Frame

Throughout the entire project, multiple tasks were performed from 5th October 2021 until 5th September 2022. Figure 1.2 displays a time frame of the different phases of work to design and develop this project as well as assess it and write on it. Initially, full focus was given to the development of Text2Storyline, in the sense of both back-end and front-end. Then, in March, we deployed the first version of our project and shifted our focus towards fixing minor issues. Once the final version met our standards and expectations, all that remained was writing this thesis, which had started in the middle of April. A period for final reviews and considerations of this dissertation was reserved to make sure all aspects were addressed before submission.

---

\(^8\)https://github.com/LIAAD/Time-Matters

1.6 Thesis Structure

This thesis is organized as follows. Chapter 2 will introduce and explain concepts related to the topic of this thesis as well as provide simple, yet illustrative examples of said concepts. The main focus is the definition of a narrative and multiple tools that explore its content to the fullest. This is followed by Chapter 3, that explores the methods previously addressed and provides an in-depth explanatory and illustrative review of each tool that Text2Storyline makes use of. This chapter is subdivided according to the multiple components available when creating a narrative. Next, comes Chapter 4 which guides the reader through the platform, showcasing its entire array of features by using several examples. Strengths and flaws are displayed and justified. Furthermore, Chapter 5 assesses the entire platform’s components based on feedback from the public. Observations are made for every result obtained. Lastly, Chapter 6 sums up every outcome and respective conclusions as well as note areas to improve and possible ideas for future work.

As our running example, that will be used to demonstrate various tools and features throughout this thesis, consider a text extracted from a Portuguese media news outlet, Dinheiro Vivo\textsuperscript{10}, which reports the death of Mário Soares, former President of Portugal. The content, which is a preserved text dated from 7th January 2017, was obtained from the Arquivo.pt infrastructure\textsuperscript{11} and can be read below\textsuperscript{1}:


\textsuperscript{10} https://www.dinheirovivo.pt/
da ditadura. Soares acabaria por ser um preso político na altura. O antigo chefe de Estado concorreu ainda ao cargo de Presidente da República, em 2006, tendo ficado em terceiro lugar; Cavaco Silva foi, na altura, eleito Chefe de Estado. (Notícia atualizada às 16h09)”

An English version of this text can be found below:

(1) “Mário Soares died this Saturday. The former President of the Republic was 92 years old and had been hospitalized at the Hospital da Cruz Vermelha since December 13, 2016. In addition to being Head of State between 1986 and 1996, Mário Soares was Prime Minister twice and Member of the European Parliament. Founder of the Socialist Party in 1973, Mário Soares is considered one of the “fathers” of Portuguese democracy, a designation that he himself has always rejected, saying only that he is “the father of two children, but father of democracy, I’m no such thing. I contributed in some way to the triumph of democracy”. As prime minister, he was one of the main responsible for Portugal’s accession to the then European Economic Community (EEC), whose accession treaty was signed in June 1985. During his 92 years of life, Mário Soares was also a lawyer, having defended dozens of political prisoners during the dictatorship period. Soares would end up being a political prisoner at the time. The former head of state also ran for the position of President of the Republic, in 2006, having been in third place; Cavaco Silva was, at the time, elected Head of State. (News updated at 4:09 pm)”
Chapter 2

Related Work

This chapter focuses on providing information on the most central notions of the Text2Storyline project. This includes multiple definitions of specific terms and methods necessary for the proper functioning of the platform as a whole. Examples of specific research tools regarded as state-of-the-art will also be shown for each procedure discussed, in order to provide a greater understanding on the value of their use in this thesis.

The remainder of this chapter is organized as follows. Section 2.1 addresses the concept of narratives and its importance to this thesis. The following five sections discuss different methods and are complemented by state-of-the-art tools and respective exemplifications. In particular, Sections 2.2, 2.3 and 2.4 introduce the concept of temporal expressions, keyphrases and entities, respectively, and how to identify and score them in texts. Section 2.5 mentions techniques to summarize the content of a text. Section 2.6 addresses the way texts and media resources are extracted from web sources to be used. Lastly, Section 2.7 recapitulates the most important issues discussed in this chapter.

2.1 Narratives

A narrative, according to Chatman [4], can be defined as a structure with two parts: the story and the discourse. The story is the “what” that is portrayed in a narrative, while the discourse is the “how”. It is also defined by Riedl et al. [5], as a cognitive structure designed to represent life events to better understand the world around us. In short, what makes up a narrative are the entities, the temporal data, and the events found in it, thus giving it a meaning.
In today’s society, with an ever growing technological range of resources, narratives can be a powerful method of showing news or information on a topic, event or personality in a more appealing manner, in order to attempt to match the ever demanding levels of satisfaction of the average Internet user when it comes to the design and functionality of the platform. Despite the main struggles regarding the extraction of knowledge from news archives’ platforms and subsequent data visualization, some media platforms started adopting timelines as an alternative to the presentation of data made solely from textual structures, offering users the possibility to familiarize themselves with a certain event in a short time lapse. Despite the growing importance of timelines in the context of summarizing data from various documents, their use is still quite rare. In this work, we go a step forward, when compared to the Time-Matters online demo, by considering both single and multiple documents. Taking all of this into account, Text2Storyline combines the importance of narratives that structure the text of a given news article or text by summarizing and displaying it in a chronological order of events, with images representative of each event to offer users the ability to automatically create visual narratives from a given text.

Multiple initiatives have been taking place in this regard. The Text2Story workshop\(^1\), the Financial Narrative Processing Workshop\(^2\) or the Storytelling workshop\(^3\) are just some examples. Research Labs such as BBC\(^4\) have also been at the forefront by continuously analysing what type of information would be consumed by audiences, especially the younger ones. Their research indicated a lighter approach would be favored, focusing on telling stories through image and little text, similar to what was already being done in some social media platforms like Instagram\(^5\) or Snapchat\(^6\). Secondly, the subject of those news pieces is also important, with younger audiences taking greater interest in stories around mental health and pop culture.

Multiple groups of people, institutions or even companies focused on assuring their content of information or, in the very least, part of it, could be consumed in a more appealing manner by telling stories in a more visual or illustrative way, which can be achieved by making use of images, videos or recordings. For instance, the latter was specifically explored by Gillick et al.\(^7\), shaping an idea that more components than simply text can be used in a creative and intuitive manner to enrich a piece of information. Soundtracks were

---

\(^1\)https://text2story22.inesctec.pt/
\(^2\)https://bbcnewslabs.co.uk/projects/graphical-storytelling/
\(^3\)https://www.instagram.com/
\(^4\)https://www.snapchat.com/pt-BR
used to enhance storytelling centered around films and television shows accompanied by their respective music pieces in an audio format.

These processes can always be done manually but, taking into account today’s technological power and in order to reduce the workload of writers and journalists, automation of storytelling keeps gaining more traction over the recent years [10]. Many different approaches have been considered and developed in this area, such as a growing forest of stories that automatically clusters documents from different breaking news into events and connecting them to form a cohesive story [11] or THEaiTRE⁵, which emerged with the original idea of having a robot create a theatre play. It premiered on 26 February 2021 with the title “AI: When a Robot Writes a Play” and was watched online by 18,450 devices [12].

An emphasis on using visual components also followed this automation trend, by considering relevant sentences or expressions from the source text and gathering and displaying images that enhance the story [13] in a lighter yet meaningful way. Advances in this field made it possible to generate a picture book from a story by analyzing its text and extracting key information of characters, actions and scenes that are then organized and visualized to form a coherent story from images alone [14]. This process is comparable to Text2Storyline’s ability to automatically generate narratives.

The Wikimedia Research⁶ team has also been developing a rather interesting project called Wikistories⁷, that aligns itself with the idea of visually summarized content. It is still in an early phase and only available in Indonesian Wikipedia pages as of the writing of this thesis, therefore an accurate review and comparison to Text2Storyline could not be performed, despite showing great promise. Wikistories allows Wikipedia editors to create short stories with a primary visual component suitable for quick consumption and easy sharing, similar to Instagram’s stories feature. It is attached to the respective wikipedia page and presents a brief summary of the most important parts of information from the full text. Figure 2.1 illustrate this behaviour for one of the Wikipedia Beta Cluster pages that has this feature available, Dog⁸. As can be seen, 7 Wikistories have been created by editors. This initiative focuses on images as its central point with small portions of text to provide context. This behaviour is similar in Text2Storyline’s Storyline feature (see

---

⁵https://theaitre.com/
⁶https://research.wikimedia.org/
⁷https://www.mediawiki.org/wiki/Wikistories
⁸https://en.m.wikipedia.beta.wmflabs.org/wiki/Dog
Section 3.3.3), as various moments are described with a sentence and accompanied by an appropriate illustration.

![Wikipedia Beta Cluster’s page for Dog](image1)

![Wikipedia Beta Cluster’s page for Dog](image2)

![Wikipedia Beta Cluster’s page for Dog](image3)

**Figure 2.1:** Wikipedia Beta Cluster’s Wikistories for Dog

All of these initiatives provide great insight on what can be achieved when it comes to creative storytelling and makes way for future endeavours. However, the most similar project to ours is Digital Libraries’ Narratives\(^9\) [15–17] which can be observed in Figure 2.2. In this project, the authors offer a similar approach to ours by providing an informative timeline capable of accurately ordering events in a chronological fashion. If an accurate enough illustration of an event is found, it is also displayed along with the text. A key point of this tool is that it focuses on creating semantic relations between different events to form a meaningful semantic web. The core of Digital Libraries’ project is their Narrative Building and Visualising Tool (NBVT), a semi-automatic tool to construct and visualise narratives [18]. It obeys an ontology\(^10\) for narratives they created themselves [19, 20] and

---

\(^9\)https://dlnarratives.eu/

\(^10\)Field of study that seeks the classification and explanation of entities
makes use of Wikidata\textsuperscript{11} as a reference knowledge-base for importing entities [21–23] and Wikimedia Commons\textsuperscript{12} as an external source of images. Their information is mainly, yet not completely, extracted from Europeana\textsuperscript{13}, a digital library that holds different types of content from various kinds of heritage institutions, such as museums, archives, libraries and audiovisual collections.

To gain access to other stories, users are redirected to the NBVT\textsuperscript{14} online demo. This requires contacting the authors to create a user account\textsuperscript{15}.

When entering the system, users are given the chance to query the tool. Once a query is submitted, the main hub of Digital Narratives displays the various entities related to the specified query, colored according to their main class (people, organizations, objects, concepts, places, works). Events are also displayed in this hub if any exist, as Figure 2.3 illustrates for the query “Mário Soares”\textsuperscript{16}. Any narrative generated is saved to the user’s dashboard so it can be easily accessed at any time. Changes can be made to said narratives, such as removing events or entities as the user deems appropriate. The entire narrative can be exported in a JSON format.

\begin{table}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Figure 2.2: Digital Libraries’ Narrative on Dante Alighieri}
\end{table}

\begin{verbatim}
To gain access to other stories, users are redirected to the NBVT\textsuperscript{14} online demo. This
requires contacting the authors to create a user account\textsuperscript{15}.

When entering the system, users are given the chance to query the tool. Once a query
is submitted, the main hub of Digital Narratives displays the various entities related to
the specified query, colored according to their main class (people, organizations, objects,
concepts, places, works). Events are also displayed in this hub if any exist, as Figure
2.3 illustrates for the query “Mário Soares”\textsuperscript{16}. Any narrative generated is saved to the
user’s dashboard so it can be easily accessed at any time. Changes can be made to said
narratives, such as removing events or entities as the user deems appropriate. The entire
narrative can be exported in a JSON format.
\end{verbatim}
The authors of Digital Narratives keep working on implementing new features, such as a story map, which will allow to formally represent geospatial knowledge. However, giving users the chance to enter a free text or a URL doesn’t seem to be under consideration. In contrast to Digital Libraries, Text2Storyline is also not restricted to accept queries of well-known entities, being indeed able to produce results for any subject that has been covered in the past by media outlets, thus making it an open and versatile system, though also more subject or prone to errors when it comes to automate the narrative process. Further to this, we make use of the Time-Matters \cite{3, 24} algorithm as a means to determine the relevance of the temporal expressions found. Based on this, we can then filter out parts of the story, which despite being informative, may not be relevant or of interest to the reader. Finally, we also made sure that our platform was open and available for anyone to have complete access to it, as opposed to Digital Libraries, which requires permission from its creators. Although their process isn’t entirely automated, especially for their default narratives, the strong point of Digital Libraries is the existence of a semantic web that links distinct events as they share one or multiple aspects and are obviously related to the main theme or query (which is, however, expected to be of a well-known entity). A minor difference is the aesthetic of both platforms. Even though they share the same goal, they were structured and designed differently, from the color scheme to the positioning of the various features each has. In our work, entities are displayed all at once in a specific area, whereas Digital Libraries displays entities grouped with the specific event where they were identified. Just as Text2Storyline extracts its content from Arquivo.pt,
a Portuguese archive, which may not contain news relative to regional news from other location in the globe, the same is also verified with Digital Libraries, as it works best with Italian personalities or events (e.g., for the query “Mário Soares”, Text2Storyline gathers more events from local articles whereas for the query “Dante Alighieri”, Digital Libraries generates a more complete and concise narrative).

All things considered, both platforms aim to generate narratives by using slightly distinct methods and tools and, consequently, create unique features and displays.

### 2.2 Temporal Information Extraction

Temporal Information Extraction (TIE) constitutes an important part of many Natural Language Processing (NLP) applications. This process, which mainly consists of recognizing temporal relations between time and events to chronological order them [25], has been the subject of several studies over the years [26, 27]. Temporal expressions present in a text can be expressed in many different ways and can often be: implicit [24], as they usually refer to less specific periods of time that require normalization (e.g., “Christmas 2020”); explicit by being clearly marked in the text as an absolute date (e.g., “June 2016” or “16th March 1998”); relative, which require more context from the rest of the text to be normalized (e.g., “today”) [28].

Temporal tagging or detection [29] is the task of finding phrases with temporal meaning within the context of a document. The whole process of temporal tagging can be divided in two distinct tasks: extraction and normalization. The extraction task simply consists of identifying temporal expressions in the text. Temporal normalization [29] is the task of mapping from a textual phrase describing a potentially complex time, date, or duration to a context-independent, easy-to-use temporal representation. It is worth noting that not all temporal expressions need to undergo this normalization process.

Tools that perform these actions are called temporal taggers. Currently, multiple methods are considered state-of-the-art and commonly used for NER. Heideltime\(^{17}\), for example, is a multilingual, domain-sensitive temporal tagger developed at the Database Systems Research Group\(^{18}\) at Heidelberg University\(^{19}\). It extracts temporal expressions from documents and normalizes them according to the TIMEX3 annotation standard. Since

---

17 https://github.com/HeidelTime/heideltime
18 https://dbs.ifi.uni-heidelberg.de/
19 https://www.uni-heidelberg.de/en
being introduced in 2010, strategies for temporal tagging in different domains were implemented [30]. Currently, it contains hand-crafted resources for 13 languages [31], such as English and Portuguese among others, as well as automatically created resources for more than 200 languages [32], although these resources are of lower quality than the manually created ones. A python package wrapper was developed to facilitate its usage but also to make it more versatile and responsive to the user’s needs, while obviously allowing it to be run in a Python-based program. Consider the example displayed in Figure 2.4 using the running example 1. The temporal expressions are identified and marked in the text using Heideltime’s online demo.

Another popular temporal tagger is SUTime, a Java library for recognizing and normalizing time expressions. It is a rule-based system particularly devoted to the English language and can be used to annotate documents with temporal information [33]. It currently does not support Portuguese. Certain algorithms exist with the aim of making use of the identification properties of temporal taggers, such as the ones previously addressed, and provide relevance scores to each temporal expression. That is the case of

---

20 https://github.com/JMendes1995/py_heideltime
21 https://heideltime.ifi.uni-heidelberg.de/heideltime/
22 https://nlp.stanford.edu/software/sutime.shtml
Time-Matters\textsuperscript{23}, which requires a Heideltime or a rule-based approach as a temporal tagger to function properly and is used in the context of the ‘Annotated Text’ component (see Section 3.3.1). The Time-Matters online demo\textsuperscript{24} offers a timeline visualization of the scores assigned to each temporal expression. Figure 2.5 shows the results of that timeline for the running example 1.

![Figure 2.5: Time-Matters’ Timeline visualization for the running example](image)

### 2.3 Keyword Extraction

Keywords are used to mark the most important parts of a text. They may also be called keyphrases as they may consist of groupings with more than one word. It is a widely used method when it comes to analysis, indexing and retrieval of information. The method of obtaining the keywords of a document is called Keyword Extraction and it can be defined as the process that automatically identifies a set of elements that best describes the subject of a text by providing metadata that summarizes and characterizes the document. With the extremely large number of documents available online, manual KE is entirely unachievable. Taking this into account, various methods are used to perform such tasks with moderate ease, which are backed by research in this field [34–36]. Various tools have been developed to challenge the efficiency of its predecessor or of its competition. Consider Yet Another Keyword Extractor (YAKE!)\textsuperscript{25}, a light-weight unsupervised automatic keyword extraction method which rests on text statistical features extracted from single documents to select the most important keywords of a text. Since it is an unsupervised

---

\textsuperscript{23}https://github.com/LIAAD/Time-Matters

\textsuperscript{24}http://time-matters.inesctec.pt/

\textsuperscript{25}http://yake.inesctec.pt/
method, it does not need to be trained on a particular set of documents, neither does it depend on dictionaries, external-corpus, size of the text, language or domain. This system can be beneficial when handling a large number of tasks and a plethora of situations where the access to training corpora is either limited or restricted [37, 38]. It supports a multitude of languages, among them English and Portuguese. One such tool will be used in the context of the ‘Annotated Text’ component (see Section 3.3.1). For the sake of continuity, the same text of our running example 1 is used to exemplify YAKE!’s behaviour when tasked with identifying relevant keywords. The results can be seen in Figure 2.6.

![Figure 2.6: Keywords obtained using YAKE! for the running example](image)

In a quite similar manner, Rapid Automatic Keyword Extraction (RAKE) [39, 40] also exists as an extremely efficient, domain and language-independent method to extract keywords from individual documents to enable application to dynamic collections. It can be easily applied to new domains, and performs well on several types of documents, specially those that do not follow specific grammar conventions. Similarly to YAKE!, it also supports English and Portuguese, among various other languages.

Taking into account what the two previous keyword extractors can already accomplish, KeyBERT\(^\text{26}\) was created as a powerful method for extracting keywords and keyphrases [41] supported by the latest advances of neural networks. It supports English and Portuguese, among many others. This tool extracts document embeddings with BERT in order to get a document-level representation. Then, word embeddings are extracted for relevant words or phrases. Lastly, cosine similarity\(^\text{27}\) is used in this algorithm to find the most similar expressions to the document [42].

---

26 [https://github.com/MaartenGr/KeyBERT](https://github.com/MaartenGr/KeyBERT)
27 measure of similarity between two sequences of numbers
It is worth noting the existence of pke\textsuperscript{28}, an open source python-based keyphrase extraction toolkit \cite{pke}. It currently implements over ten keyphrase extraction models and it allows for easy benchmarking of state-of-the-art keyphrase extraction models \cite{keyphrasification}. On the topic of keyphrases, it is also relevant to consider a recent tutorial on Keyphrasification\textsuperscript{29}, which took place at the 44th European Conference on Information Retrieval (ECIR 2022)\textsuperscript{30}. Keyphrasification \cite{keyphrasification} is defined as the task of summarizing texts with keyphrases, which this tutorial covered along with a hands-on approach on several popular tools and models.

## 2.4 Named Entity Recognition

Named Entity Recognition (NER) is a subtask of Information Extraction with the goal of locating and classifying named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc. This process helps answer pertinent questions regarding the text \cite{ner}. Entity Linking is another NLP task commonly used together with Entity Recognition as it assigns a unique identity to previously discovered entities (e.g., in “Paris is the capital of France” the entity “Paris” refers to the city and not “Paris Hilton” or any similar entity). Figure 2.7 shows this behaviour in a clearer manner.

![Figure 2.7: NER and Entity Linking processes exemplified. Source: https://en.wikipedia.org/wiki/Entity_linking](https://en.wikipedia.org/wiki/Entity_linking)

Many different implementations have been developed and released that can achieve these NLP tasks. They all have unique advantages and consequent disadvantages to their usage, but they are all considered state-of-the-art when it comes to performing Named Entity Recognition.

\textsuperscript{28}https://github.com/boudinfl/pke
\textsuperscript{29}https://keyphrasification.github.io/
\textsuperscript{30}https://ecir2022.org/
Entity Recognition. SpaCy\textsuperscript{31} is an open-source software library for advanced natural language processing, written in Python and Cython. It currently has models for 15 languages and a multi-language model with 10 languages, which include English and Portuguese. Figures 2.8 and 2.9 show a tiny preview of what spaCy’s features have to offer. Figure 2.8 shows the entity dependency when provided with text. Only the first sentence of the running example 1 was considered for display purposes, although this tool is capable of handling the entire text. Figure 2.9 shows the actual NER process applied to the entirety of the running example 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{dependency.png}
\caption{Dependency of entities in a sentence of the running example using spaCy}
\end{figure}

Another example of an NER tool is Apache OpenNLP\textsuperscript{32}. This machine learning based toolkit is commonly used for the processing of natural language text. It supports the most common NLP tasks, such as language detection, tokenization, sentence segmentation, part-of-speech tagging, named entity extraction, chunking, parsing and coreference resolution. These tasks are usually required to build more advanced text processing services.

It is also noteworthy to mention Natural Language Toolkit (NLTK)\textsuperscript{33}, a suite of libraries and programs in Python that is widely used for NLP tasks \cite{47} such as classification, tokenization, stemming, tagging, parsing, and semantic reasoning. It also provides interfaces to over fifty corpora and lexical resources.

\subsection{Wikification}

This process \cite{48, 49} takes both Entity Recognition and Linking a step further, by associating the already identified entities with their corresponding Wikipedia page, as the name

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{31}https://spacy.io/
\item \textsuperscript{32}https://opennlp.apache.org/
\item \textsuperscript{33}https://www.nltk.org/
\end{itemize}
\end{footnotesize}
of the method suggests.

The prime example of this process is Wikifier\(^{34}\), a web service that takes a text document as input and annotates it with links to relevant Wikipedia concepts \(^{50}\). It was implemented by Janez Brank, the Artificial Intelligence Laboratory\(^{35}\) and the Jožef Stefan Institute\(^{36}\). Its parameters are highly customizable, allowing the user to decide how strictly filtered the results are. As of the writing of this thesis, it supports texts in English, Portuguese, Italian, French, Spanish, Dutch and German. An illustrative example of Wikifier using the running example \(^{1}\) can be observed in Figure 2.10. This tool is used in the context of the ‘Annotated Text’ and the ‘Entities’ components (see Sections 3.3.1 and 3.3.2 respectively).

While Wikification as a whole is quite a restrictive method in the sense that it specifies that entities found are linked to a Wikipedia page, several similar tools fit this field, even though they may link the identified entities to other external knowledge databases. Such is the case of DBpedia Spotlight\(^{37}\), a tool capable of automatically annotating mentions of DBpedia resources in text, providing a solution for linking unstructured information sources to the Linked Open Data cloud through DBpedia \(^{51}\). It supports 12 languages thus far, with English and Portuguese among them. It consists of a four step approach:

\(^{34}\)https://wikifier.org/
\(^{35}\)https://ailab.ijs.si/
\(^{36}\)https://www.ijs.si/ijsw
\(^{37}\)https://www.dbpedia-spotlight.org/demo/
spotting, candidate selection, disambiguation and filtering. These steps combined lead to a text of identified entities marked with hyperlinks to the respective DBpedia\(^{38}\) pages.

Since DBpedia is a very similar platform to Wikipedia, web services such as this one can be considered a Wikification tool.

Lastly, the ability to provide access to more than one external knowledge database is quite trivial. Therefore, APIs such as Dandelion\(^{39}\) exist. This web service combines a multitude of semantic text analytics features such as: text similarity, language detection,

\(^{38}\)https://www.dbpedia.org/

\(^{39}\)https://dandelion.eu/
sentiment analysis, entity extraction and wikification. Therefore, given a text it is able to, among other things, extract and correctly identify the appropriate entities and match them with both the respective Wikipedia and DBpedia pages.

2.5 Text Summarization

Text Summarization (TS) aims to summarize relevant information so it can be easily read by the user. Automatic Text Summarization adds layers of complexity to this definition as it must focus on identifying the important content. This is usually done at sentence level, therefore extracting the relevant sentences of a text. TS methods can be either extractive or abstractive [52–54].

The former, as the term suggests, identifies the most appropriate expressions or sentences from the source text, extracts them and joins them to form a summary. Extractive methods are usually subdivided into a scoring step and a selection step. First, sentences are classified according to their importance to the context of the text. Then, the ones with the highest scores are selected and extracted to form a summary. The latter is quite the opposite, as it produces a summary from scratch by paraphrasing the information from the source text.

Taking this into account, summarization models may then differ according to the architecture chosen as they can use an encoder-decoder approach which means that a given neural network method reads a text, encodes it and then generates the target text. Afterwards, the decoder will extract information from the encoder, based on the scores of the source text. The encoder can be a Convolutional Neural Network (CNN) [55], a Recurrent Neural Network (RNN) [56], Long Short-Term Memory (LSTM) [57] or Transformers [58]. The decoding can be handled in either an auto-regressive or a non auto-regressive approach.

Other factors can influence how a TS model works: learning schematics (supervised or reinforced learning approaches) [59] or the existence of external transferable knowledge, which is a category algorithms like GloVe [60] or Bidirectional Encoder Representations from Transformers (BERT) [61] fall into.

Different Text Summarization techniques should be considered to better understand the importance of implementing this process in the Text2Storyline project. GSum⁴⁰, which stands for Guided Summarization, is an extensible guided summarization framework that

⁴⁰https://github.com/neulab/guided_summarization
can take different kinds of external guidance as input [62]. The model is based on neural encoder-decoders with pre-trained language models, like BERT [61] and Bidirectional and Auto-Regressive Transformers (BART) [63].

On the opposite side in terms of procedure, NeuSum, which is also known as Neural Extractive Document Summarization, is a framework that learns to score and select sentences, integrating them into an end-to-end trainable model. As a neural network model, it learns to identify the relative importance of sentences, predicting the relative gain given the sentence extraction and the partial output summary. This model has two parts: the document encoder and the sentence extractor [64].

In Figure 2.11, the running example 1 is presented after being summarized using SMRZR.io, an online demo of the BERT-extractive-summarizer tool [65]. All of these tools support English and Portuguese, among many other languages.

Figure 2.11: Text Summarization of the running example using the BERT-extractive-summarizer approach
Source: SMRZR.io (https://smrZR.io/)

2.6 Data Collection and Web Archiving

In order to properly generate a storyline of events, it is necessary to first acquire a text or gather a portion of texts. A common denomination for this process is also Data Acquisition, which consists of fetching and then extracting data from websites. Common web pages are built using text-based mark-up languages such as HTML and are designed for human users and not for ease of automated use. As a result, specialized tools and software have been developed to facilitate the scraping of web pages. A prime example of
such a tool is Newspaper3k\(^{43}\), the most popular Python package when it comes to extracting texts from a given URL. It works on over 10 languages and is particularly efficient on news articles.

Custom and specifically built APIs or extensions are the most common method to extract the specific data needed from a website. Major news or information sources have developed their very own web services that, once given permission can access and extract the desired content. Such is the case for one of the most popular online encyclopedias, Wikipedia\(^{44}\). Its API called MediaWiki\(^{45}\), is entirely free to use and allows users to access any Wikipedia page and extract its content, images included. It supports any language used in Wikipedia pages. Another example comes from a very well-known newspaper, The New York Times\(^{46}\), which allows users to operate their array of web services. Similarly to MediaWiki, each of the NY Times’ APIs\(^{47}\) perform a specific action on the newspaper’s online database going from retrieving any relevant articles based on a query to obtaining the most popular stories either in general or by topic, among other features.

Another way of collecting data is through web archives, a process that consists on collecting portions of the World Wide Web (WWW) to ensure the information is preserved in an archive for future researchers, historians, and the public \([66]\). One such process is of the utmost importance grounded on the fact that content referenced by most URLs in the World Wide Web has a half-life of just a few years, which is even more aggravated with social media content that hardly lasts more than a week. Web archiving is thus an essential process that allows temporary pieces of information to be preserved forever \([67]\).

Several web archiving techniques have been created around the world ever since 1996. Currently, there are close to 100 web archiving initiatives around the globe, summing up to a total of 625 billion archived web pages \([68]\). The most well-known domain for this task is the Internet Archive\(^{48}\), a non-profit library of millions of free books, movies, music, software and websites, among other content. It exists with the main goal of providing universal access to all knowledge. It digitizes about 1,000 books a day for a total of more than 2 million books, financially supported by libraries and foundations. It is worth acknowledging the Wayback Machine\(^{49}\), a digital archive of the World Wide Web founded

\(^{43}\)https://github.com/codelucas/newspaper

\(^{44}\)https://en.wikipedia.org/wiki/Main_Page

\(^{45}\)https://www.mediawiki.org/wiki/API:Main_page

\(^{46}\)https://www.nytimes.com/international/

\(^{47}\)https://developer.nytimes.com/apis

\(^{48}\)https://archive.org/

\(^{49}\)https://web.archive.org/
by the Internet Archive. It allows users to see exactly how websites looked in the past by preserving archived copies of defunct websites. A Portuguese equivalent is the Portuguese Web Archive (PWA) [69], more commonly known as Arquivo.pt⁵⁰, which holds preserved pages with the .PT domain from 1996 onwards. Until 2007, these contents were mainly extracted from the Internet Archive. Afterwards, Arquivo.pt began to make its own collections from the web. Even though the majority of the archived content is in Portuguese, it also preserves content in other languages such as English, Spanish or French. As of January 2022, it has over 13 million archived web files, 28 million websites and 852 TeraBytes of compressed data.

The Arquivo.pt Text Search API⁵¹ is how Text2Storyline gains access to the text content of any web page, given any query issued by a user. Images are also extracted from the Arquivo.pt, although this process is exclusively performed through a different API, namely the Image Search API⁵², which we use to extract the most accurate images according to related keywords found in the narrative. This process is used in the context of the ‘Entities’ and ‘Storyline’ components (see Sections 3.3.2 and 3.3.3 respectively).

2.7 Summary

In this chapter, an overview of all the tools used or discussed in the Text2Storyline project was presented. Basic concepts, such as the definition of narratives, temporal information extraction, keyword extraction, named entity recognition, text summarization, data collection and web archiving were established, among other related denominations, to better understand the aim and importance of this work. Examples were also provided to easily understand the expected results of each process addressed in this chapter. Taking into account the basis laid out, it is important to further expand the knowledge on each of the tools that were introduced and how their respective algorithms work.

---

⁵⁰https://arquivo.pt/
⁵¹https://github.com/arquivo/pwa-technologies/wiki/Arquivo.pt-API
Chapter 3

Text2Storyline Architecture

Various concepts were learned during the development of this project and, consequently, multiple tools and features were implemented, which resulted in the usage of different programming languages, packages and APIs. This chapter focuses on an in-depth look into the entirety of Text2Storyline’s features and discusses how specific tools and their usage affect each of these features. Some other information on technologies and trivial concepts that weren’t addressed in the previous chapter will also be tackled.

This work was built to consider two distinct iterations: a single-document approach and a multi-document one. The former corresponds to applying the Time-Matters algorithm [3, 24] to a single text. This text can be provided to the Text2Storyline platform either directly or through a URL, where a web extraction tool will first extract the content of the link provided. The latter is used for multiple documents and is used when a search term, also known as a query, is provided instead. In this case, all relevant documents to the query will be fetched and considered by the Time-Matters algorithm, which, will then identify and score temporal expression with regards to its relevance in the corresponding text [3, 24]. In this thesis, documents are to be retrieved from the Arquivo.pt web infrastructure. In the future, other data sources may also be included.

Along this chapter, multiple comparisons and comments will be made regarding both iterations, as they essentially use the same tools despite evident differences in the outcome display.
3.1 Technologies

The following technologies discussed in this chapter were fundamental in defining and developing the entire architecture of this project. They consist of the most basic, yet essential and necessary tools to bring this project into fruition. We refer to programming languages and frameworks, among others.

3.1.1 Angular Framework

The whole project was built in Angular\(^1\) or, as it is commonly referred to, Angular 2+ or Angular CLI which is a TypeScript-based web application framework and development platform for creating efficient and sophisticated single page apps. The original project had already been developed in Angular. Therefore, it only made sense to continue using this tool for this work. The learning curve at the start is relatively steep, although previous knowledge and experience on web systems is quite valuable.

3.1.2 TypeScript

TypeScript\(^2\) is the main programming language used when developing Angular applications, although very similar to JavaScript as it is a strict syntactical superset of JS and adds optional static typing to the language. It is developed and maintained by Microsoft and is designed for the development of large applications and transpiles\(^3\) to JavaScript. As it is a superset of JavaScript, existing JS programs are also valid TypeScript programs.

3.1.3 Python

Python\(^4\) is a high-level, interpreted, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically-typed, garbage-collected and supports multiple programming paradigms, including structured, object-oriented and functional programming. All of the APIs that were developed and used in this work were coded in Python.

---

\(^1\)https://angular.io/
\(^2\)https://www.typescriptlang.org/
\(^3\)Type of translator that takes the source code of a program written in a programming language as its input and produces an equivalent source code in the same or a different programming language
\(^4\)https://www.python.org/
3.1.4 Flask

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

This tool was used for RESTful API development, which is particularly useful in giving the user complete control over how they access their data. A RESTful API is essentially an architectural style that uses HTTP requests to access and use data and is based on the REpresentational State Transfer (REST). REST is usually preferred over similar technologies as it uses less bandwidth, making it more efficient for internet usage tasks.

3.1.5 Docker

Docker is a set of platform as a service (PaaS) products that use OS-level virtualization to deliver software in packages called containers which are isolated from one another and bundle their own software, libraries and configuration files. They can communicate with each other through well-defined channels. Because all of the containers share the services of a single operating system kernel, they use fewer resources than virtual machines. All the APIs developed for this project were deployed using Docker for easier accessibility and usage.

3.2 Foundational Algorithms

Specific operations happen throughout the entire Text2Storyline platform and are not directly attached or related to one specific feature. This section focuses on those operations and addresses the corresponding algorithms, distinguished by the goals each tool aims to achieve. Subsection 3.2.1 addresses the Time-Matters algorithm whose aim is to identify...
and score temporal expressions in the text. Subsection 3.2.2 discusses the usage of YAKE! to identify keywords, which are subsequently used in other components of the architecture. Lastly, Subsection 3.2.3 and Subsection 3.2.4 tackle the usage of the tools needed to identify relevant entities and display them accordingly, that is Wikifier and WikiMedia.

### 3.2.1 Time-Matters

As it was discussed in the previous chapter, the process of Temporal Information Extraction allows for a text to have its temporal expressions identified and scored according to its importance. For the Text2Storyline project, this is achieved using Time-Matters [3] together with a temporal tagger. A key aspect of this algorithm is its ability to value or score the identified temporal expressions in the text. Other characteristics involve the fact that it does not require any training stage and builds upon local text statistical features extracted from documents, making it an unsupervised, domain corpus and mostly language-independent solution. As for the temporal taggers it works together with two solutions. By default, that temporal tagger would be Heideltime\(^9\) [30–32]. However, a rule-based approach can also be used, instead. Both taggers used, allow for a specification of a granularity value [70], which defines whether the algorithm should gather dates in the most detailed manner or focus on more specific formats (e.g., only years (YYYY), only months (MM-YYYY) or days (DD-MM-YYYY)). The first (Heideltime), is a more efficient, though time-consuming solution. The latter (rule-based) can be considered when the target texts are rather extensive and when the need to return results faster is more important than the effectiveness of the results themselves.

In detail, Heideltime\(^10\) is a multilingual, domain-sensitive temporal tagger developed at the Database Systems Research Group\(^11\) at Heidelberg University\(^12\). It extracts temporal expressions from documents and normalizes them according to the TIMEX3 annotation standard. Since being introduced in 2010, strategies for temporal tagging in different domains were implemented [30]. Currently, it contains hand-crafted resources for 13 languages [31] as well as automatically created resources for more than 200 languages [32], although these resources are of lower quality than the manually created ones. A python package wrapper named py_heideltime\(^13\) was developed to facilitate its usage but also to

\(^9\)https://github.com/JMendes1995/py_heideltime
\(^10\)https://github.com/HeidelTime/heideltime
\(^11\)https://dbs.ifi.uni-heidelberg.de/
\(^12\)https://www.uni-heidelberg.de/en
\(^13\)https://github.com/JMendes1995/py_heideltime
make it more versatile and responsive to the user’s needs, while obviously allowing it to be run in a Python-based program. Figure 3.1 shows the results of applying py_heideltime to our running example 1. In the figure, one can observe that temporal expressions such as 1973, the date when the Socialist Party was founded, or 13th December 2016 (normalized to 2016-12-13) when Mário Soares was hospitalized, were identified by py_heideltime. In addition to this, users can also have access to a temporal-annotated version of the text, together with the heideltime_processing field which gives users information about how much time it took to process the given text, which in the case of our running example was approximately 3.6s.

![Figure 3.1: Heideltime approach with Time-Matters for temporal expression recognition](image)

In addition to py_heideltime, Time-Matters also makes use of a py_rule-based approach, a self-defined regex solution filled in with pre-defined rules that enable detecting various date formats in the text. Figure 3.2 exemplifies the outcome of this simple

---

temporal tagger when applied to our running example 1. Similarly to py_heideltime, py_rule-based was built to give users a list of the identified temporal expressions (e.g., a YYYY year-format, such as “2016”), a temporal-annotated version of the text and information about the time it took to process the given text, which was approximately 0.0002s in the case of our running example 1.

```
, 
{‘rule_based_processing’: 0.0002040863037109375, 
’rule_based_text_normalization’: 0}]
```

**Figure 3.2:** Rule-based approach with Time-Matters for temporal expression recognition

Considering the results obtained and shown in Figures 3.1 and 3.2, there are significant differences that can be spotted right away. As previously noted, the rule-based approach is faster as it only took 0.0002 seconds as opposed to the 3.598 seconds that were required by Heideltime. On the other hand, py rule-based was only able to able to annotate 6 temporal instances as opposed to the 9 temporal expressions identified by py_heideltime. This has to do with the fact that the rule-based tagger only considered dates in the YYYY format and disregarded everything else, whereas py_heideltime identified all possible temporal expressions with full granularity. While Heideltime is effectively a superior temporal tagger, it may pay-off to use the rule-based solution in longer documents and in cases where the text is simple in terms of the temporal expressions it contains.

Each of the identified temporal expressions can then be scored using a ByDoc or a By-Sentence approach if a single text is taken into account. With the former (ByDoc), Time-Matters retrieves a unique single score for multiple occurrences of a temporal expression in different sentences (e.g., two instances of “2016” in a text will always have the same
score such as “0.92”), thus considering the context of the entire text, regarding all the relevant keyword that the temporal expression in cause co-occurs with. The latter (By-Sentence) retrieves multiple and consequently, different scores for multiple occurrences of a temporal expression in different sentences (e.g., two instances of “2016” in a text may have different scores such as “0.92” for the first instance and “0.77” for the second), which means it only considers the keywords in the same sentence as the temporal expressions being scored.

When it comes to scoring temporal expressions in a set of documents, rather than just a single text, more options are available: ByCorpus, ByDoc and ByDocSentence. The first option retrieves a unique single score for each temporal expression found in the corpus of documents. The remaining two work similarly to the ones previously described in the context of a single text.

For this project, the pros and cons for each parameter and their options were considered. To simplify the accessibility and usage for the average user, such parameters were kept fixed and cannot be changed in the website itself. Therefore, the option adopted follows a ByDoc principle, for both single and multiple documents. This means that the score for multiple occurrences of a temporal expression in different sentences is the same. Also, the granularity value is set to be as precise as possible, in order to identify every possible temporal expression. This setting, as previously explained, defines whether the algorithm gathers dates in specific formats or in any possible format detected. Taking into account the Figures 3.1 and 3.2, dates that follow a YYYY (e.g., “1986”) or similar formats (e.g., “junho de 1985” or “13 de dezembro de 2016”) are detected as well as other dates that don’t follow such kinds of patterns (e.g., “92 anos” or “este sábado”).

The assumption for the logic of the algorithm [24] is that the relevance of a candidate date \(d_j\) may be determined with regards to the relevant terms \(W_j^*\) that it co-occurs within a given context, which can be a window of \(n\) terms in a sentence, the sentence itself, or even a corpus of documents in case we are talking about a collection of multiple documents. Therefore, the more a given candidate date \(d_j\) is correlated with the most relevant keywords \(W_j^*\) of a document or documents \(t_i\), the more relevant the candidate date is.

In order to model such temporal relevance, Generic Temporal Similarity Measure [71] is relied on, which makes use of co-occurrences of keywords and temporal expressions as a means to identify relevant dates within a text. GTE is formalized in Equation 3.1, whose
values range from 0 (irrelevant) to 1 (relevant). The InfoSimba similarity measure [72] is represented by IS.

\[
GTE(t_i, d_j) = \text{median}(IS(W_{t_i}, d_j)), \; w_{t_i} \in W_i^*
\]  

(3.1)

The algorithm of Time-Matters is used in multiple components of Text2Storyline, namely in the 'Annotated Text' (see Subsection 3.3.1), the ‘Storyline’ (see Subsection 3.3.3), the ‘Temporal Clustering’ (see Subsection 3.3.4), the ‘Word Cloud’ (see Subsection 3.3.5) and the ‘Results’ (see Subsection 3.3.6) components.

### 3.2.2 YAKE!

The process of obtaining relevant keywords or phrases from the target document or documents is performed along with the Time-Matters algorithm by using YAKE!. In the context of the Time-Matters algorithm, YAKE! is set to consider the extraction of 1-gram, which means that the algorithm will consider single words in a text. On the other hand, when extracting keywords for the ‘Word Cloud’ component (see Subsection 3.3.5), an n-gram value of 3 is considered instead, meaning that YAKE! will consider groupings of up to 3 words. Figure 3.3 below shows the annotated keywords when running YAKE! on top of our running example 1 with a 1-gram setting. It is worth noting that only the keywords identified in the same sentence as each temporal expression are considered for its co-occurrence calculation (e.g., “13 de dezembro de 2016” takes into account the keywords “antigo”, “Presidente”, “República”, “Hospital”, “Cruz” and “Vermelha”).

![Figure 3.3: Keywords obtained with YAKE!](image-url)
When it comes to Keyword Extraction, multiple techniques exist, all of which are capable of performing such as task. YAKE! is considered a statistical approach [38] and, as such, it aims to find the score of the terms present in the document using different types of statistics calculated over a single document. Then, those terms are ordered based on their scores and displayed as important keywords.

YAKE! extracts features by calculating five unique features: TCase (Casing), TPos (Term Position), TFNorm (Term Frequency Normalization), TRel (Term Related To Context) and TSent (Term Different Sentence) [37]. The first one (TCase) considers that an uppercase term, except one that starts sentences, is more valuable than a lowercase term. Acronyms are also considered. The second feature (TPos) follows the intuition that words in the beginning of the document should be more important than the ones placed in the middle or in the end. YAKE! distinguishes itself in how the position of a word is obtained. Rather than directly using the word’s position in the text, it instead uses the position of the sentence in which the word occurs. The third feature (TFNorm) refers to the frequency of a word in the text. The fourth feature (TRel) is used to quantify the significance of a word based on its context. This means that the higher the number of different terms co-occurring with a word, the lesser is the significance of that word. The fifth and last feature (TSent) considers that a word that appears in many different terms has a greater chance of being important.

All of these features previously described are combined to calculate the score of a given word $t$, formalized in Equation 3.2.

$$S(t) = \frac{T_{Rel} \times T_{Pos}}{T_{Case} + T_{TFNorm} + T_{Sent}}$$

Lastly, to take into consideration the $n$-gram value, the final score $S(kw)$ of a candidate keyword $kw$ is given by Equation 3.3.

$$S(kw) = \frac{\prod_{t \in kw} S(t)}{KF(kw) \times (1 + \sum_{t \in kw} S(t))}$$

The YAKE! extraction tool is used in multiple components of Text2Storyline, namely in the ‘Annotated Text’ (see Subsection 3.3.1), the ‘Storyline’ (see Subsection 3.3.3), the ‘Temporal Clustering’ (see Subsection 3.3.4) and the ‘Word Cloud’ (see Subsection 3.3.5) components.
3.2.3 Wikifier

Entity Recognition aims to locate and classify entities (persons, organizations, countries, etc) in a text. Such entities can then be linked to knowledge-based sources, such as Wikipedia, a process known as Wikification. In this project, we apply Named Entity Recognition and Entity Linking to allow users expanding his/her knowledge regarding a given entity. Among a few already available solutions, upon consideration of their respective pros and cons, Wikifier\(^{15}\) [50] became our NER and Entity Linking choice. Upon some tweaking to remove irrelevant annotations, the final array of results is usually quite appropriate, vast, quick and informative to the user. Consequently, Wikipedia is the source to where the users are redirected to, whenever they would be interested in knowing more about any entity.

This entire process’s main idea consists of recognizing Wikipedia concepts by performing a disambiguation step when a word or phrase might refer to more than one concept [50]. Firstly, an annotation step occurs, by identifying concept titles in the text and gathering related concepts to those titles (e.g., the phrase “skeletal muscle” in the text is bound to both [Skeletal], [Muscle] and [Skeletal Muscle]) to be scored. The scoring formula for a candidate concept \(c\), which is presented in Equation 3.4, considers the following criteria: the length, in words, of the longest title of the concept anywhere in the target document \(T_c\); how deeply the concept is integrated into the target document \(I_c\); the average degree of similarity between the official title and surface form of the concept \(H_c\).

\[
S(c) = T_c I_c H_c \quad (3.4)
\]

At each text location, the concepts are ranked according to their score as only the highest ranked concept is kept. The top three concepts can be kept as well if their title is longer than the top concept, which ensures that words in the document are never labeled by more than three Wikipedia articles. Regardless, multiple final annotations are a rare occurrence. Wikifier excels at its task by tracing the author’s train of thought phrase by phrase through the text it analyzes.

\(^{15}\)https://wikifier.org/
3.2.4 MediaWiki

Even though Wikifier proved to be quite an indisposable tool, it still lacked an aesthetic, more detailed and illustrative approach. The solution became not only convenient but ideal. Using the already identified entities and respective links to the corresponding Wikipedia page, web extraction tools could be used to automatically obtain the image and the text from each of those pages. Wikipedia’s very own specialized API, MediaWiki\(^\text{16}\), allows the user to access any Wikipedia page and extract its content, which includes all text and images. This tool was particularly useful in extracting the first few words of a Wikipedia page corresponding to an entity previously detected and identified in a text. The method to extract the main picture was also considered as a backup in case the main method, using the Arquivo Image Search API couldn’t return any image. Since the focus of this project is also to showcase Arquivo.pt’s archive which includes all news articles and respective images, if there was an available image then it is granted priority to that extraction process, regardless of the quality of the image.

The combination of these tools, Wikifier, MediaWiki and Arquivo Image Search, is displayed in Figure 3.4. The entity “Mário Soares”, among others, is identified and classified as “human” by Wikifier. This tool also provides a link to the corresponding Wikipedia page that the user can access by clicking the respective card. MediaWiki and Arquivo Image Search handle the description and image, respectively.

\(^{16}\text{https://www.mediawiki.org/wiki/API:Main_page}\)
3.3 Features

This section addresses the many features that Text2Storyline contains and analyzes its behaviour through the tools used for each. Taking into account the information established in the previous chapter on each tool and method, it is as important to understand how they are used and what can be done when multiple techniques are combined to shape a component into something user-friendly, but also useful. As Figure 3.5 illustrates, users are given the choice to interact with the system by either providing a text or URL (light-blue box on the left side) or a search term (light-purple on the right side). The first option allows Time-Matters to focus on a single document and, consequently, it can provide the user with its contents annotated through the ‘Annotated Text’ feature (first light-green box on the left). On the other hand, using a query as input, will instead show the ‘Search Results’ feature (first light-green box on the right) due to the fact that multiple articles are fetched. Even though Time-Matters iterates over all of them, there is no suitable or appealing way of displaying the annotated content for every document. However, this component allows the user to analyze the annotated text for each article individually. The remaining four components remain the same regardless of the type of input supplied.

The following subsections address a component each: ‘Annotated Text’ (Subsection 3.3.1); ‘Entities’ (Subsection 3.3.2); ‘Storyline’ (Subsection 3.3.3); ‘Temporal Clustering’ (Subsection 3.3.4); ‘Word Cloud’ (Subsection 3.3.5); ‘Search Results’ (Subsection 3.3.6). Each of these features offer a distinct look into the content of the text supplied or texts obtained from a given query. It is important to mention that this chapter follows an overall discussion on each component rather than a divisive approach into the components that are made available to the user when a text or URL is provided versus when a query is supplied. The chosen structured method works best as most features addressed in this chapter are similar, regardless of the user’s input choice. The only deviation is the ‘Search Results’ component that is only available when a search term is provided. This specific outlier will be addressed in the Subsection 3.3.6.

3.3.1 Annotated Text

The ‘Annotated Text’ feature uses the text extracted from a user-provided link or text and applies Temporal Expression Recognition, Keyword Extraction and Entity Recognition to identify temporal expressions, keywords and entities, respectively. They are then
all marked in the text in distinctive ways: temporal expressions identified and scored by Time-Matters are marked in a range of colors according to their score, going from red to green; keywords obtained by YAKE! are presented in bold; entities determined by WikiFier, and their linkage to Wikipedia are marked in dark blue, the characteristic accent color of the Text2Storyline platform. These features can be shown all at once or toggled to show only the ones the user desires. Hovering over any keyword will reveal its score given by YAKE!. The same can be done with temporal expressions which show the score assigned to them by Time-Matters. This behaviour is exemplified in Figure 3.6.

It is worth noting that the ‘Annotated Text’ component is not available if a search query is provided, as multiple news articles and texts will be returned in this case, instead of just one. However, as it will be further analyzed in the subsection that corresponds to the ‘Search Results’ component (see Subsection 3.3.6), an option will be available to create
a narrative for each of the obtained texts that were deemed relevant to the query provided. That process will show the respective annotated text, among the other components that will be addressed in the following subsections.

3.3.2 Entities

The ‘Entities’ component gathers any person, location or concept identified from a text or texts and displays them in a card format, that redirects to the corresponding Wikipedia page. Each entity has a primary class, which is identified through the Wikifier web service. A brief description of said entity is also provided by making use of MediaWiki’s API, which extracts the text from the target Wikipedia page, displaying the first 50 words. The illustration present for each entity is obtained in a similar fashion. Figure 3.7 illustrates the tab that contains the entities found for the running example 1.
3.3.3 Storyline

The ‘Storyline’ component displays events extracted from the main text or query source, by making use of text summarization, as expressions or sentences surrounded by relevant temporal expressions are pulled from the main text or texts. It offers a interactive and visual horizontal approach with the resource of images. Consider Figure 3.8, which shows the storyline generated for the running example 1.

The ‘Storyline’ component, being the main attraction, also makes use of web and keyword extraction tools. For each event sourced from applying text summarization, YAKE! is applied to obtain the most relevant expression or word, which is then used as that event’s title. Images are obtained through the Arquivo.pt Image Search API, by providing the title previously acquired through the keyword extraction tool, and using the most relevant image that the system returns. For storylines that are generated from a query, every description contains a mark (‘[+]’) that allows the user to be redirected to the article that event was extracted from. This particular distinction can be confirmed with Figure 3.9 as it shows one event from the storyline obtained for the following query: “Mário Soares”.

![Figure 3.8: Storyline feature](image)

![Figure 3.9: Single event of the Storyline for a query](image)
3.3.4 Temporal Clustering

The ‘Temporal Clustering’ component is rather similar to the ‘Storyline’ one in terms of the content that is presented. It shows the same number of events in a more compact and easy-to-read manner, as well as the option to include the least relevant dates. This feature is a vertical presentation of events, for a quicker overall assessment of the narrative created. The main divergence lies in the usage of illustrative elements. To keep this feature simple yet practical, no images are associated with each event. The date is displayed in a prominent manner, as is the score which is shown in green if the temporal expression in a given event is considered relevant, which happens if the score displayed is higher or equal to 0.35. Otherwise, the score is shown in red. The phrase where the temporal expression was extracted from is displayed right below. The date is marked in bold. Similarly to the ‘Storyline’ component, each event’s description is obtained by using the same extractive method of text summarization. Figure 3.10 shows this component’s behaviour for the narrative created using the running example 1.

![Figure 3.10: Temporal Clustering feature](image)

3.3.5 Word Cloud

A word cloud or tag cloud is a visual representation of relevant words in a text or extract. It is commonly used to provide the reader with a quick idea of what a document or concept is about, by displaying words or expressions in different sizes that are representative of their importance.

This was achieved by using a sub-tool from YAKE!. By extracting the most relevant keywords in a text or set of texts, YAKE! is also capable of creating a word cloud based on
those results. Consider the Figure 3.11 that exemplifies this tool for the running example 1.

By analyzing Figure 3.11, it can be observed that the words that were considered the most relevant are “Soares”, “Mário” and “Morreu”, which depict an accurate summary of the text from the running example 1.

3.3.6 Search Results

The ‘Search Results’ component is only available when a query is provided, as this feature’s purpose is to present the user with all the articles considered relevant to the search term provided.

Consider the following search query (2) as a running example to demonstrate Text2Storyline’s behaviour when handling multiple documents:

(2) “Síndrome Respiratória do Médio Oriente”

This query translates to Middle East Respiratory Syndrome or MERS, a viral respiratory illness caused by Coronavirus (MERS-CoV). Figure 3.12 shows the appearance of the ‘Search Results’ feature for the running example (2).
The user can then consult any of said articles or generate a narrative for each of them individually. Doing the latter will open a new window and show a display of five new components based on the contents of the article in question as if the link of the chosen article had been provided initially. Those five components are: ‘Annotated Text’, ‘Entities’, ‘Storyline’, ‘Temporal Clustering’ and ‘Word Cloud’, the components that would exist whenever a text or a URL is provided.

3.4 Summary

In this chapter, information was provided on diverse technologies that were utilized such as the Angular framework, Docker and Flask or programming languages like Python and TypeScript. Additional tools and respective algorithms that were considered vital and, therefore, foundational to this project’s development were addressed, introduced or further discussed, providing meaningful examples. Furthermore, the various components that constitute this work were shown and their behaviours were explained, all of which were accompanied by illustrative examples.
Chapter 4

Text2Storyline Showcase

The main aspects that Text2Storyline has to offer, as well as the tools and algorithms it makes use of have been addressed and shown in the previous chapters. From this stage onward, it is important to take a guided look at the platform’s home page (see Section 4.1), for the many distinct, although optional, starting points the user can make use of to create narratives as mere suggestions of potential topics of interest.

Subsequently, detailed discussions of multiple experiments are made in order to showcase Text2Storyline’s behaviour for different types of inputs or sources, such as URLs, queries and free text (see Section 4.2).

4.1 Home Page

The home page of Text2Storyline is the first point of contact between the user and all of the platforms’ features. Thus, an appealing yet clear design is key. The primary color is a dark blue tone. As Figure 4.1 illustrates, the first part of this scrollable home page contains the project’s logo and a search bar at the top that can be used to input queries or links.

Right below, at the center, there is an image carrousel with pre-defined queries that can be clicked to create narratives accordingly. “António Costa”, the first picture, will create a narrative for the Prime Minister of Portugal (as of the date of writing this thesis) based on the news articles that were archived in the past by the Arquivo.pt infrastructure. By its sides, are two tables that display recent suggestions of queries in an attempt to link the present interest to the past [2]. These are filled with the hourly trending search
queries from Google, using the Google Trends\textsuperscript{1} data. The main engine for this process relies on pytrends\textsuperscript{2}, a Python package responsible for obtaining the necessary data, which was implemented during the development of the API capable of providing such data for the tables in question from Google Trends. Other methods include fetching data for: a country’s most searched queries during the previous year; interest on a specific query by region which would return a list of the countries who searched that term the most; related queries of a specific query, which led users to search other terms related in some way to the original one.

By scrolling down through the home page, a set of manually selected narratives from relevant news articles are on display, as Figure 4.2 illustrates, and they can all be used to create the corresponding narratives.

Furthermore, the users are invited to input their own text through the ‘Texto Livre’ or ‘Free Text’ option that can be found on the top banner of the home page. Figure 4.3 exemplifies this behaviour. Note that the text area contains four different examples distinguishable by the language they are written in: Portuguese, English and French. Two of them are in Portuguese. The user is, however, free to write their own text as long as it doesn’t surpass 5000 characters. The resulting narrative will be accompanied by all the features that were previously addressed, similarly to when a URL is provided. This specific type of input also makes use of a language detection tool web service, developed

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{home_page.png}
\caption{Home page}
\end{figure}

\textsuperscript{1}https://trends.google.com/trends
\textsuperscript{2}https://github.com/GeneralMills/pytrends
Figure 4.2: Second part of the home page
from the langdetect\textsuperscript{3} Python package, which is capable of identifying up to 183 different languages and return its specific two-letter ISO 639\textsuperscript{4} code. Regardless of all these possibilities, note that Text2Storyline can only accept 7 distinct languages, namely Portuguese, English, Spanish, German, French, Italian and Dutch. This is directly related to py\_heideltime’s default installation, although support for more languages can be added as we see fit. This language identification is also crucial to ensure the Wikipedia pages linked to detected entities are correct, as well as the extracted text from each one.

The other option present in the banner allows access to an ‘About’ section, which discloses information on the team that contributed to this project as well as special thanks to vital organizations and a walk-through video, highlighting everything that can be achieved with Text2Storyline.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.3.png}
\caption{Free Text section of Text2Storyline}
\end{figure}

\section{4.2 Interface Exploration}

This section intends to show Text2Storyline’s behaviour, emphasizing its strong points while acknowledging its weaknesses and errors. In order to form a comprehensive report on this project’s effectiveness of its multiple available features, multiple inputs are considered, namely using a URL (see Subsection 4.2.1), a query (see Subsection 4.2.2) and a free text extract (see Subsection 4.2.3).

\textsuperscript{3}https://pypi.org/project/langdetect/

\textsuperscript{4}A standardized nomenclature used to classify languages
4.2.1 Creating Narratives from a URL

This subsection addresses experiments performed by using links (URLs) to generate the corresponding narratives and analyze the outcome accordingly. In order to correctly explore this project’s behaviour, two distinct examples are used. The first one is a news article from Jornal Público (a Portuguese daily newspaper) centered on the Queen of the United Kingdom, Elizabeth II, as she recently (2022) celebrated the Platinum Jubilee, which marks the 70th anniversary of her accession to the throne (see Figure 4.4).

![Image of news article page on Queen Elizabeth’s life](image)

**Figure 4.4:** News article page on Queen Elizabeth’s life

With this example we aim to gauge Text2Storyline’s potential at extracting the relevant content and form accurate and concise events from a news source, ranging from appropriate titles and descriptions to illustrations of said events. This particular article was written by Público’s journalists in a chronological fashion, detailing pertinent occurrences throughout Queen Elizabeth’s life or the Royal Family’s. Such format can be used to compare directly to the events generated in the ‘Storyline’ component. It also contains images for some of its events, which is useful to evaluate this project’s effectiveness at acquiring and displaying appropriate illustrations by comparing the ones extracted automatically through a web service with images manually selected by the author.

The second example is the Wikipedia page of Dante Alighieri, an Italian poet, writer and philosopher better known for his “Divine Comedy”, a narrative poem (see Figure 4.5).

---


4.2.1.1 Queen Elizabeth II

For the first example, as with all of the following test subjects, their various features will be discussed. Given the input URL, the system applies the newspaper3k library\textsuperscript{7} to extract the corresponding text. Once the text is extracted, two APIs are applied: (1) Time-Matters, for temporal scoring and keyword identification - through YAKE!, and (2) Wikifier, for entity detection. In what follows, we start with the 'Annotated Text' component. In this example, we will only consider a small part of the original article, as taking into account the entire text would be impractical due to its extensive nature. Figure 4.6 illustrates this feature. An obvious divergence from the original news article is the formatting of the text, which is displayed as a simple running text marked with the annotations.

As can be observed, the annotated text provides an immense amount of information

\textsuperscript{7}https://newspaper.readthedocs.io/en/latest/
4. Text2Storyline Showcase

FIGURE 4.6: Extract of the annotated text

attached to the text. The marks on temporal expressions allow the reader to easily analyze the most important parts of the text associated with the most relevant dates, according to the Time-Matters algorithm. In the figure, two dates (“2 de Junho de 1953” and “1981”) are marked with yellow color. By mouse hovering them, we can observe a score of 0.555 for the first and 0.543 for the second. Two other dates (“1977” and “25 anos”) marked in bright red are still considered relevant (0.39 for both). The remaining four dates (“1533”, “1603”, “24 de Novembro de 1953” and “1970”) are deemed irrelevant as their scores are lower than 0.35 (0.227 for the first two, 0.211 for the third and 0.327 for the last). Looking at the small excerpt, one can also observe, marked in bold, a list of 1-gram keywords, mostly focused on personalities of the Royal Family such as “Isabel” (Elizabeth), “Carlos” (Charles) and “Diana”, or events, namely “Jubileu”, “visita” (visits) and “cerimónia” (ceremony). The entities identified in blue color follow a similar pattern as locations (“Abadia de Westminster” (Westminster Abbey), “Nova Zelândia” (New Zealand) or “Grã-Bretanha” (Great Britain)) or royal titles (“Rainha” (Queen), “Príncipe” (Prince) or “Monarca” (Monarch)).

It must be noted that out of the 81 total dates identified in the text, 50 are deemed relevant. Keywords will also stand out for the remainder of the text, enabling the user to grasp the main idea of the text quite quickly. Entities are marked as well, with a unique color making them distinguishable from the remainder of the text. It is important to acknowledge that not all entities identified are marked in the text, due to their names, at times, not appearing in the text in the same manner as their Wikipedia page titles. Consequently, only 26 out of the imposed limit of 50 entities are marked in the text.

When addressing the ‘Entities’ component, all entities are correctly displayed in the sense that no person, concept, location or organization feels like it should not be there, although some images obtained from Arquivo.pt through their Image Search API may be suboptimal. As Figure 4.7 suggests, these entities are incorrectly illustrated, even though everything else such as their class and description are correct. It is important to note
that this phenomenon is rather uncommon and the amount of entities with correct and accurate illustrations far outweigh the entities with poor ones, which is evidenced with Figure 4.8.

The ‘Storyline’ feature may suffer from the same issue as well, as images are obtained with the same method as Figure 4.9 exemplifies. The illustration of a radio host is in no sense connected to the information of Queen Elizabeth’s children. In the context of this component, it may be a more severe matter as having an event accompanied by the
wrong illustration can confuse the user and diminish its appeal. Although not a certainty, it happens more frequently than in the ‘Entities’ feature.

**Figure 4.9:** Example of storyline event with poorly displayed elements

It must be noted that responsibility is not completely assigned to the tool responsible for fetching illustrations. On occasion, the title chosen by YAKE! for an event can also be incomplete or incorrect. That same title is used to search for an adequate image. This behaviour is clear in Figure 4.10, where the selected title does not reflect the important part of the description of the event, thus misleading the image extraction tool. This should be tackled in future work. Note, however, that the image provided is quite accurate to the title in question.

**Figure 4.10:** Example of storyline event with poorly displayed elements

On the other hand, plenty of events in the storyline do not show signs of these problems as Figure 4.11 proves. Still, further alternative titles could be considered instead. For instance, in Figure 4.11a, a better title would be “Princesa Isabel casa-se” (Princess Elizabeth weds). However, such title would be slightly more ambiguous than the current one, which could have a poor influence on the consequent illustration. Therefore, it is important to establish a balance between title and image. In contrast, the second example given by Figure 4.11b, has an accurate title and illustration and leaves little to no room for improvement. Other elements that compose an event of a storyline, namely the description and the date cannot be discussed in the same manner as the previous two. The date, being identified by Time-Matters and being used is always correct regardless of its format. The description, being the location of the corresponding date cannot be changed as its part of
the source document. However, it is important to note that, by default, seemingly irrelevant events with a Time-Matters score of less than 0.35 are hidden. In some cases, this is an acceptable behaviour as the user can still choose to reveal those events. Nonetheless, it must be acknowledged that such events may not always be correctly scored, as Figure 4.11c confirms. An event such as the birth of the current Queen of the United Kingdom should be relevant, as opposed to having a score of 0.228. This is not entirely an issue of Time-Matters, as the low score was justified by being an old date, only mentioned once throughout the entire text.

As for the ‘Temporal Clustering’ component, since it makes no use of illustrations it does not suffer from these issues. However, just as it was observed when addressing storylines, events with low scores are hidden by default, therefore contracting the same problem. Not much else can be noted from this feature, as it shares a lot of traits with the ‘Storyline’ feature, which were already covered before. Figure 4.12 shows the behaviour of this feature for a select range of four dates, as it would be impractical to display the 63 events that were formed with this example.
Lastly, the ‘WordCloud’ component simply achieves what it promises, a quick summary of the context of the source text by showing keywords where their size is directly proportional to their importance, namely “Isabel” (Elizabeth) and “rainha” (queen), which are displayed as the primary terms from this document, as Figure 4.13 shows. Note that some words that appear are not entirely appropriate but they appear a lot throughout the text. Such is the case of “Foto” (Photo), “Arquivo” (Archive), “REUTERS” and “The Print Collector”. All of these are constantly mentioned in captions of the illustrations used in the news article, making them important keywords yet rather useless for the reader.
4.2.1.2 Dante Alighieri

In this second example, we aim to create a narrative from Dante’s Wikipedia page. Similarly to the previous example, as excerpt from the main source document is considered for a more readable exploration and respective illustration of the ‘Annotated Text’ feature, available in Figure 4.14. Looking at the figure, one can observe that only four dates exist in this excerpt, which accurately represents the amount of dates throughout the entire document. They are all considered irrelevant (0.259 for the first, 0.133 for the second and third and 0.334 for the last one). Keywords are also scarce, marking mostly personalities such as “Dante” and “Beatriz”. This is due to having an extensive source text, from which keywords are extracted, while only a single excerpt is shown. Entities capture, for this excerpt, concepts, namely “Pintura” (Painting), “Amor” (Love), “Escritor” (Writer), “Política” (Politics) and “Filosofía” (Philosophy) but fail to mark “Dante” and “Beatriz”, even though they were identified as being in the text. This is due to those entities being identified as “Dante Alighieri” and “Beatriz Portinari”, respectively, which is not how they appear in this excerpt.

![Figure 4.14: Extract of the annotated text](image)

Given the extensive nature of this document compared to the article of the previous example that featured a list of chronological (thus likely relevant) events, less dates are included in this text. Differently to the previous test subject, the majority of dates identified have low scores as a considerable amount are deemed by Time-Matters algorithm as irrelevant (36 dates out of a total of 65). This goes in line with the expected summarization aspect of this feature, which aims to hide from the reader less relevant parts of the story likely to happen in larger articles such as wikipedia ones. The behaviour of identified keywords is consistently correct once again and the entities, although correctly identified fail to be marked in their entirety once again, even if to a much lesser extent. Out of the total limit of 50 entities, 45 are marked in the text.
The ‘Entities’ feature confirms that the illustration issues noted in the previous example are more common. It is, however, to be expected that such issues become more apparent for this text due to the fact that, despite being a somewhat well-known personality, Portuguese sources would have very little information or pieces written on such a figure that lived in Italy between the 13th and 14th centuries, which is evidenced by Figure 4.15. Even though some entities are indeed mentioned in the source text, their importance and contribution towards the user’s knowledge is questionable, as Figure 4.16 illustrates. These entities fall in a grey area, as some users may find it valuable to have definitions of such concepts whereas others will find it trivial and therefore, unnecessary.

![Figure 4.15: Example of entities with poor illustrations](image)

Proceeding to the ‘Storyline’ component, illustrations are once more rather poor, which is due to having a poor title (see Figure 4.17a), very little data on the specific topic (see Figure 4.17b) or personality (see Figure 4.17c). Note that, despite the increase in the frequency of illustrative errors in this example’s storyline, there are positive cases, as Figure 4.18 proves. The lack of effectiveness registered in this example can be associated once more to images being pulled from an archive that primarily focuses on saving Portuguese news articles since 1996 and images fetched come from that array of archived write-ups. Therefore, it is believable that a considerably small amount of said sources, if not none at all, contain media assets adequate to illustrate these events from the past, added to Dante
Alighieri not being an important figure in Portugal’s culture overall, in today’s society or in the past.

Just as it was remarked for the news article on Queen Elizabeth and in all possible inputs, ‘Temporal Clustering’ will share most of the positive and negative aspects with the exception being the illustrative features. This example has a considerable amount of events rated with a score of 0. This is also verified in the storyline component. It can be confirmed that, even though some are correctly graded and, therefore, completely irrelevant, this is not the case for all of them as Figure 4.19 illustrates. The events with the dates 1289, 1301 and 1318 do refer to moments in the life of Dante and are incorrectly rated. For the date 1301, this score could be justified as it does not directly impact the narrative centered on the Italian poet. However, the event dated from 2002 is correctly scored as it merely refers to the date of the publication of a bibliographic reference used to write the content of the Wikipedia page.

Lastly, the ‘Word Cloud’ feature presents even more effective results than the one displayed in the previous example, as no words or phrases are incorrectly shown. This happens because, contrarily to a news article which often names sources whenever any third party content (in particular media assets) are used, that behaviour is not mirrored in other informative text pages such as Wikipedia. Even though references are included when extracting the source text, they do not have a meaningful impact, especially in this example.
1295
Cargo público devia

Score: 0.412

Foi, também, médico e farmacêutico; não pretendia exercer essas profissões mas, segundo uma lei de 1295, todo nobre que pretendesse tomar um cargo público devia pertencer a uma das guildas (Corporazioni di Arti e Mestieri - ou seja, "Corporação de Artes e Ofícios").

(A) Storyline event with poorly selected title and illustration

1290
Dante procurou refúgio

Score: 0.334

Quando Beatriz morreu, em 1290, Dante procurou refúgio espiritual na filosofia da Literatura latina.

(B) Storyline event with poorly selected illustration

01 DE NOVEMBRO DE 1301
Carlos de Valois

Score: 0.788

Entretanto, a 1 de novembro de 1301, Carlos de Valois entrou em Florença com os guelfos negros que, por seis dias, devastaram a cidade e massacram grande número de partidários da facção branca.

(C) Storyline event with poorly selected illustration

FIGURE 4.17: Examples of storyline events with poorly displayed elements

14 DE SETEMBRO DE 1321
Dante Alighieri

Score: 0.335

Dante Alighieri (Florença, entre 21 de maio e 20 de junho de 1265 d.C. — Ravena, 13 ou 14 de setembro de 1321 d.C.[1] foi um escritor, poeta e político florentino, nascido na atual Itália.

(A) Storyline event with adequate elements

1802
Francês Victor Hugo

Score: 0.123

Disse o escritor e poeta francês Victor Hugo (1802-1885) que o pensamento humano atinge em certos homens a sua completa intensidade, e cita Dante como um dos que “marcam os cem graus de génio”.

(b) Storyline event with adequate elements

FIGURE 4.18: Examples of storyline events with adequate displayed elements
that contains only three references. As Figure 4.20 shows, the most used words throughout the text are “Dante” and “Florença” (Florence).

4.2.2 Creating Narratives from a Query

This subsection addresses experiments performed by using queries to generate the corresponding narratives and analyze the outcome accordingly. In contrast to the examples
that use URLs, there is minimal distinction between any queries, therefore a single example will suffice. However, it is important to note that some search terms may extract more articles from the archive than others. This can happen for a multitude of reasons: having a particular topic or event get little coverage or interest in Portugal at the time; being a rather new subject and not many news sources have written any pieces on it. The query chosen that manages to depict this issue to an extent is “Síndrome Respiratória do Médio Oriente”, which translates to “Middle East Respiratory Syndrome” (MERS-CoV), a viral respiratory illness first reported in 2012 in Saudi Arabia. Symptoms include fever, cough and shortness of breath, as many infected have died.

Similarly to the previous subsection, where URLs were used as input, narratives generated from queries will be explored as to the quality of the corresponding storylines. It is taken particular awareness to the overall cohesion of multiple events that follow one another. Since multiple documents are used, there may be evident dissimilarities between events.

Differently than the previous two test subjects, this example created from a query does not possess an ‘Annotated Text’ component given the multitude of source texts gathered. Therefore, the principal point of contact to this narrative once it is created becomes the ‘Storyline’ component. A new and rather obvious issue, added to the issue with titles and illustrations, is the quality of the descriptions provided by Arquivo.pt. The extraction methods currently used by Arquivo.pt for a large amount of texts do not properly filter the content, resulting in descriptions filled with advertisements (see Figure 4.21a). In other cases, some of the pages returned are merely newspaper’s hubs where the news and its snippet is referenced (see Figure 4.22). Overall, no single description is as clean as the ones shown in previous examples, even if a minor issue in some events (see Figure 4.21b). This happens somewhat on purpose as snippets were analyzed instead of the full content of each article. Doing the opposite would increase the time necessary to create such narratives exponentially. Our hopes are that, upon our feedback, these issues will be fixed by Arquivo’s team in the near future.

Despite these negative issues, this ‘Storyline’ feature includes the option for the user to check the source for each event, which slightly remedies the issue previously addressed. It must be noted that, although a rare phenomenon, some archived pages take a long time to fully load and others seem to be inaccessible. This is evidently an issue on the side of Arquivo.pt. Another difference from this specific example to the previous two is the
(A) Storyline event with poor title, illustration and description

(B) Storyline event with poor description

FIGURE 4.21: Examples of storyline events with poorly displayed elements

FIGURE 4.22: Example of storyline whose source is a hub for multiple articles
number of events that form the storyline, as this only has 9 events. Once more, there is a limit as to how many events can be used in this ‘Storyline’ feature. However, in this case, the main reason is that simple there are not that many articles that fit the query used as input. The ‘Temporal Clustering’ feature, just as in previous examples, mirrors the ‘Storyline’ component in a more simplified display and, consequently shares most of its successes as well as failures.

Entities can suffer from the issues previously detected and addressed in previous examples, such as poor illustrations or redundant entities. However, taking into account the problems noted in the description of storyline events where multiple unrelated content or mentions to other articles can be attached, this can severely impact the identification of correct entities and increase significantly the amount of those that are in no way related to the query provided (see Figure 4.23). Nevertheless, some entities are wrongly identified due to the cluttered snippets with advertisements or the source’s signature (see Figure 4.24). Similar to other examples, plenty of entities still capture the main idea of what the user intends to see from the query used (see Figure 4.25).

Subsequently, a unique component only available for narratives created through queries, ‘Search Results’ displays all 12 sources obtained for the query “Síndrome Respiratória do Médio Oriente”, ordered according to the their score. This score is based on the highest
TEXT2STORYLINE: GENERATING ENRICHED STORYLINES FROM TEXT

**Figure 4.24:** Example of incorrectly identified entities from cluttered text

**Figure 4.25:** Example of adequate entities
score originating from one of the dates in the text of the respective source. Although this sorting is the default, the search results can be, instead, ordered according to their date. This method uses the earliest date found in each document. Each search result shown has its title, description and dates found displayed in the form of a list similar to Google’s search results. The score assigned to each source is only visible when using the score-based sorting. A simple yet intuitive and organized way of letting the user consult any sources they desire, as Figure 4.26 accurately describes.

Figure 4.26: Examples of the Search Results feature

In this feature, and through each article’s title and description, one can confirm that almost all sources are related to the query used for this example. The remaining sources are ones that, as already mentioned, redirect to hubs where a specific piece of information pertinent to this query exists and should not be ruled out. It is also possible to create a narrative for one of these sources at a time, in order to get a more in-depth look at specific
moments or parts of the overall narrative. This process is similar to the ones described for the first two examples as the URL of the selected source is used, therefore it will only be briefly discussed. Figure 4.27 describes this behaviour. The Google Trends API is also used exclusively in this feature to obtain up to five related queries to main one provided by the user. Those related search terms are “virus”, “sars mers virus”, “sars virus”, “mers virus” and “virus mers” as Figure 4.28 suggests.

Unlike what happened in the previous observations of the ‘Word Cloud’ feature, results for a query feel cluttered, primarily noting the repetition of several words that imply or mean the same such as “Médio Oriente” (Middle East), “Médio Oriente Menu” (Middle East Menu), “Médio Oriente Internacional” (Middle East International) “Médio Oriente causou” (Middle East caused), “Médio” (Middle) and “Oriente” (East). Two of these managed to also collect the name of the source as in the snippet they were formatted as
such. The remaining are parts of the whole which is “Médio Oriente” (Middle East) and is more than enough. This behaviour could also be due to not having more important words or phrases and some concepts end up getting repeated in order to reach the 20 total keywords for the word cloud. It must also be taken into account the fact that only snippets of up to the top ten highest scored source articles were used to form this word cloud, which has an impact on its effectiveness in exchange for faster results. Still, some important information can be gained besides knowing it is connected to the Middle East (“Médio Oriente”), but also to South Korea (“Coreia do Sul”), that is a Severe Acute Respiratory Syndrome (“Síndrome Respiratória” and “Respiratória Aguda Grave”), known as “MERS” as the Figure 4.29 illustrates.
4.2.3 Creating Narratives from Free Text

This subsection addresses experimentations performed by using the 'Free Text' input feature discussed previously in this chapter (see Section 4.1) to generate the corresponding narratives and analyze the outcome accordingly. Similarly to the previous type of input discussed, a single example will be taken into account, a text in English, available in Text2Storyline as an option. The text reports on the infamous Boston Marathon bombings that occurred in April 2013. Approximately 264 people were injured and 3 died. The following extract depicts the entirety of the text used:

1. “The Boston Marathon bombing was a terrorist attack, followed by subsequent related shootings, that occurred when two pressure cooker bombs exploded during the Boston Marathon on April 15, 2013. The bombs exploded about 12 seconds and 210 yards (190 m) apart at 2:49 pm EDT, near the marathon’s finish line on Boylston Street. The explosion killed 3 civilians and injured an estimated 264 others. The Federal Bureau of Investigation (FBI) took over the investigation and, on April 18, released photographs and a surveillance video of two suspects. The suspects were identified later that day as Chechen brothers Dzhokhar Tsarnaev and Tamerlan Tsarnaev. Shortly after the FBI released identifying images publicly, the suspects killed an MIT policeman, carjacked a civilian SUV, and initiated an exchange of gunfire with the police in nearby Watertown. During the firefight, a Massachusetts
Bay Transportation Authority Police officer was injured but survived with severe blood loss. A Boston Police Department officer was also injured and died from his wounds nearly a year later. Tamerlan Tsarnaev was shot several times in the firefight and his brother subsequently ran him over with the stolen SUV in his escape. Tamerlan died shortly after arriving at Boston’s Beth Israel Hospital. An unprecedented manhunt for Dzhokhar Tsarnaev ensued on April 19, with thousands of law enforcement officers searching a 20-block area of Watertown. During the manhunt, authorities asked residents of Watertown and surrounding areas, including Boston, to stay indoors. The public transportation system and most businesses and public institutions were shut down, creating a deserted urban environment of historic size and duration. Around 6:00 p.m., shortly after the “shelter-in-place” advisory was rescinded, a Watertown resident discovered Dzhokhar hiding in a boat in his back yard. Reports conflict as to whether or not he was armed. Located within the boat by thermal imaging, he was shot while in the boat, arrested, and then taken to a hospital shortly thereafter. During an initial interrogation in the hospital, Dzhokhar alleged that Tamerlan was the mastermind. He said they were motivated by extremist Islamist beliefs and the wars in Iraq and Afghanistan, and that they were self-radicalized and unconnected to any outside terrorist groups. According to him, they learned to build explosive devices from an online magazine of the al-Qaeda affiliate in Yemen. He said that he and his brother had decided after the Boston bombing to travel to New York City to bomb Times Square. Dzhokhar was indicted on April 22, while still in the hospital, on 30 charges relating to homegrown terrorism, including use of a weapon of mass destruction and malicious destruction of property resulting in death. He was found guilty on all charges on April 8, 2015, and the following month was sentenced to death.”

This final example’s narrative follows the same structure of the first two. Therefore, the ‘Annotated Text’ feature is the first one to be shown to the user. This component has very little to note in addition to previous observations from other test subjects, as temporal expression and keyword identification and consequent classification performed as expected. As for entities, out of the total 33 identified, only 17 are marked in the text. Part of this behaviour lies on the fact that multiple entities share words. In that case, only the first and most relevant entity is attached to the corresponding text. This happens with
the following entities, among others: "Boston Marathon", "Boston Marathon bombing", "Boston" and "Marathon". Some other entities may not be in the text explicitly which is the case for "Islamism", "September 11 attacks" and "Domestic terrorism" which are related to the topic of this example and can provide greater insight. Figure 4.30 illustrates the outcome of the behaviour of all elements of this component.

This ‘Entities’ feature has a main distinction in comparison to all the others previously addressed. Since the source text is written in English, some Wikipedia pages that link to identified entities have titles that do not exist or have different titles for Portuguese Wikipedia. For example, the page "Boston Marathon bombing"\(^8\) is equivalent to “Atentado à Maratona de Boston de 2013”\(^9\). Therefore, the Language Detection API was used in order to detect the language of the source text and then extract the initial text from the Wikipedia page of each entity in English as Figure 4.31 that contains multiple entities confirms. Considering the results, it has very few illustrative issues, oppositely to those that were noticed in previous examples. The fact this is a properly formatted text and is used as a default example must be taken into account, as not every free inputted text will have such positive results. Regardless, Figure 4.33a shows the entities that still display the least accurate and correct illustrations, despite still being somewhat related to the entity in question. One other aspect to address is that some entities do not have any classes assigned to them. Even though Wikifier is directly and completely responsible for this

---

\(^8\)https://en.wikipedia.org/wiki/Boston_Marathon_bombing
\(^9\)https://pt.wikipedia.org/wiki/Atentado_N%C3%A0%20Maratona%20de%20Boston%20de%202013
4. TEXT2STORYLINE SHOWCASE

process it still must be noted as a potential flaw in the display. Figure 4.33b illustrates how entities that suffer from this problem are shown.

The ‘Storyline’ component for this text contains 9 events of which 2 are irrelevant with a score of 0. Once again, some events are poorly illustrated, although they are a minority
and the main cause of this issue comes from an incorrect or incomplete title as Figure 4.32 shows. It must be noted that the last five events are incorrectly dated, due to the fact that the year is not being specifically written in the source text. Therefore, Time-Matters assumed the current year of 2022, instead of the year it was written, which would be 2013. Then, when the following year is mentioned it is referring to 2014 instead of 2023, which would be impossible. This behaviour is displayed by Figure 4.33.

**Figure 4.32: Examples of poorly displayed storyline events**

(A) Poorly illustrated storyline event

(B) Poorly illustrated storyline event

**Figure 4.33: Examples of incorrectly dated storyline events**

(A) Incorrectly dated storyline event

(B) Incorrectly dated storyline event
The ‘Temporal Clustering’ feature suffers from the same date identification issues. Otherwise, there is nothing else to note, except mentioning that 2 events are based on a specific time of the day (“Around 6:00 p.m.” and “2:49 pm”) and that is also included in the date, as Figure 4.34 illustrates. This behaviour is also existent in the ‘Storyline’ feature, even though it is being addressed here.

![Figure 4.34: Temporal Clustering events dated with hours](image)

Lastly, the ‘Word Cloud’ component correctly identifies the relevant words or phrases in the source text as Figure 4.35 shows. From the figure, it can be observed that the most relevant expressions are “subsequent related shootings”, “pressure cooker bomb”, “cooker bombs exploded” and “Boston Marathon”. These are appropriate, although the second and third are rather similar and can be understood as a sequence of two keywords. In addition to this, small-sized keywords can also be observed to get some more context and grasp the main idea behind the entire text. about this event. These are: “Boston Police Department”; “brother Dzhokhar Tsarnaev” or “marathon finish line”.

4.3 Summary

This chapter served as a guide to properly showcase and explore the vast array of functionalities that the Text2Storyline has to offer depending on the type of input the user provides. Given the nature of automation of generating narratives, other rather smaller aspects may have remained undisclosed as no example perfectly makes use of every feature or exposes every issue. Regardless, this discussion was consistently thorough and unbiased as a critical mindset was taken for every example. In conclusion, some features still require more improvement than others, although, for some issues, some time may be needed for web services used to be improved or for a improved version of those tools to come to fruition. The aesthetic parts of Text2Storyline were discussed as well, with a particular focus on the platform’s home page and all its possible paths from it, more specifically the ‘Free Text’ page that allow the user to generate narratives from any text they desire.
Chapter 5

Results and Discussion

This chapter makes use of all knowledge acquired throughout this thesis and focuses on properly assessing its components and respective effectiveness. For this, the previous chapter plays an important role as some of the examples already provided will be target of several points of discussion to ensure this project is capable of achieving what it promises. These results already explored will now be complemented with the feedback from the U.Porto\(^1\) academic community obtained through two distinct surveys. The first\(^2\) (see Section 5.1) aims to understand the preferences of users regarding the presentation of information through different visual components (see Appendix A). The second\(^3\) (see Section 5.2) intends to evaluate in a more formal and objective way some of the results obtained by comparing Text2Storyline with reference systems (see Appendix B). Responses for both surveys were collected in the span of a week, starting on 25th July 2022 throughout 1st of August 2022. The following excerpt is the English translation of the original message sent to the academic community of U.Porto in order to appeal for their collaboration in an unorthodox and informal manner:

1. “Hi,

Do you mind if I address you as “you” throughout this email? I hope the answer was “yes” because an email is not yet such an interactive process but here it goes!

I come to ask for your help in completing two surveys for my master’s thesis. If you choose to contribute and complete only the first questionnaire, it would be a

\(^1\)https://sigarra.up.pt/up/pt
\(^2\)https://forms.gle/dubfHnqNkMnk5778
\(^3\)https://forms.gle/ujj6k6SMaUMsEid7f9
huge help. I know filling out questionnaires is tedious, especially in July, but it only takes 5 minutes. If the topic interests you, the second survey, although a little more extensive, serves to assess my project in more depth. Both questionnaires contain all the necessary information about the platform created as part of my thesis, Text2Storyline. The main idea of this tool is to create visual narratives in order to enrich/complement the narrative elements found in the original text. You can test this tool at will (this is even more optional than surveys). Text2Storyline accepts text or news links and also search terms such as Google.

And if you think it’s a boring task, it’s okay to close the quiz halfway and pretend you never clicked on the link. Come on, we’ve all done this at least once!

You have a week to fill in any of them in case you are busy when you read this email for the first time.

Thanks for reading this far and thank you so much if you decide to help!”

We challenge the reader to also participate and fill out these surveys in order to be more familiarized with Text2Storyline and its features: Survey 1; Survey 2.

5.1 User Preference Evaluation

The first survey focuses on gathering the users’ feedback regarding the components available in Text2Storyline at first (see Subsection 5.1.1), and then, more specifically assessing key elements of prime features, namely the ‘Storyline’ (see Subsection 5.1.2) and the ‘Entities’ (see Subsection 5.1.3) components. The main focus is visual elements rather than their correctness, which is reserved for the second survey (see Section 5.2). A total of 267 submissions were registered. As expected, the majority of the participants are between 18 to 24 years of age (65%), followed by ages 25 to 39 (24%), older than 40 (10%), and lastly, 17 or younger (1%), as only 2 persons in this last age group participated in this survey. As for distribution based on academic qualifications, the majority of participants has a Bachelor’s degree (44%), followed by people with Secondary Education (30%), those with a Master’s degree (22%) and lastly, those with a Doctorate degree (4%).
5.1.1 Users’ Preferences on Text2Storyline components

The first set of questions focus on understanding what features from Text2Storyline do users consider the best in tackling text narratives. To this regard, we asked the participants to classify from 1 (least important) to 5 (most important) each of the following three components: ‘Annotated Text’, ‘Storyline’ and ‘Entities’. We focused on obtaining feedback on these features as they are not as well known as other components such as the ‘Word Cloud’ or ‘Temporal Clustering’, thus requiring more attention. This emphasis is evident throughout both surveys. Figure 5.1 displays these three questions. The results obtained show that the majority of the participants in the survey answered with 4 for the ‘Storyline’ (41%), the ‘Annotated Text’ (38%) and the ‘Entities’ (38%) features. The result distribution for these three components was rather similar between them. This shows that, although people are not quite familiarized with these features, they value its potential.

Then, we asked the participants to rank all the components from 1 (least important) to 6 (most important) (see Figure A.1f from Appendix A). A component called ‘Simple Text’, which consisted of text without any annotations, was included as a control variable, thus totaling 6 components, namely ‘Simple Text’, ‘Annotated Text’, ‘Entities’, ‘Storyline’, ‘Temporal Clustering’ and ‘Word Cloud’. Considering the variability of these results, in order to properly establish a final ranking order of all components, the average ranking score of each component was calculated. This was achieved by summing the number of votes (0 to 267) for each score multiplied by that score (1 to 6), which was then divided by the total number of respondents (267). Figure 5.2 shows the full distribution of results for this question, in the form of a chart, whereas Figure 5.3 shows a table with the amount of votes for each score for every component, as well as its average ranking score. The most popular score for each component is also marked accordingly. Analyzing the results from both illustrations, the two highest scored components are ‘Annotated Text’ and ‘Storyline’, registering an average rank of 4.29 and 4.33, respectively, making the latter the most important component to respondents, by a very small margin. Note that the ‘Annotated Text’ feature registers the highest amount of votes for the score of 6. However, a significant amount of participants also ranked it with only a 2 or 3. On the opposite side of the spectrum, the ‘Word Cloud’ component was undoubtedly chosen as the least important feature with an average score of only 1.89. For comparison, the next least important component, ‘Temporal Clustering’ registered an average score of 3.39. This feature shared
TEXT2STORYLINE: GENERATING ENRICHED STORYLINES FROM TEXT

Figure 5.1: Questions from the first survey to rate three components: (a) ‘Annotated Text’, (b) ‘Storyline’ and (c) ‘Entities’
5. RESULTS AND DISCUSSION

A very close average rank with the remaining two components, ‘Simple Text’ (3.42) and ‘Entities’ (3.67). Obviously, having the opportunity to read the original text without any annotations or formatting is essential to the majority, if not all participants. However, this is already a possibility with the ‘Annotated Text’, as all annotations can be disabled at will. Regardless of the reasoning, this preference towards simplicity of a text must be noted. We can assume that components that display content or information in the most direct way are still widely preferred. Still, the presence of illustrations is a plus to many users, which is evidenced by having the ‘Storyline’ component top over the others. The fact that the ‘Entities’ feature takes the last place of the podium also contributes to this observation, as it also relies on illustrations. Overall, users value, as they always have, information in its most pure and absolute form. However, these results also suggest they see potential in other methods that rely on visual elements to display the same information.

![Figure 5.2: Distribution of results for a question in the first survey to order all components from the least important to the most important](image)

![Figure 5.3: Distribution of votes on each score for every component and respective average rank from a question in the first survey to order all components from least important to most important](image)

Next, we aim to compare the users’ preferences when it comes to issuing a query into the system. To this regard, we put the ‘Search Results’ feature side-by-side with other platforms’ search results pages (see Figure A.1g from Appendix A). Five options (presented as letters from A to E) were made available to the users. Each letter refers to the results...
obtained by different platforms for the running query “Sindrome Respiratória do Médio Oriente” (Middle East Respiratory Syndrome (MERS)): A: Google; B: Wikipedia; C: Arquivo.pt; D: Text2Storyline ‘Search Results’ component; and E: Text2Storyline ‘Storyline’ component. It must be noted that the respondents were never explicitly told the source of any platform names, although some such as Google or Wikipedia are quite obvious. Figure 5.4 illustrates options A, D and E. The remaining figures can consulted directly in the survey or here. The participants were asked to rank, from 1 to 5 the five components, with 1 being the least important and 5 the most important. Considering the variability of these results, in order to properly establish a final ranking order of all options, the average ranking score of each search results system was calculated. This was achieved by summing the number of votes (0 to 267) for each score multiplied by that score (1 to 5), which was then divided by the total number of respondents (267). Figure 5.5 shows the full distribution of results for this question, in the form of a chart, whereas Figure 5.6 shows a table with the amount of votes for each score for every component, as well as its average ranking score. The most popular score for each component is also marked accordingly. By interpreting both representations of results, two polar opposites are rather noticeable: the most important search results platform, option E - Text2Storyline ‘Storyline’ component with an average score of 3.91 and a total of 147 votes that ranked it with a score of 6; the least important search results system, option A - Google with an average rank of only 2.24 as well as a total of 128 respondents rating it with a score of 1. Wikipedia’s search results were considered the second most important (3.38) with Text2Storyline’s ‘Search Results’ feature taking the last spot of the podium (2.81). Both showed a dispersal of votes, implying indecision from participants when sorting these options. Overall, this is a rather surprising outcome, implying that people prefer having results displayed in a chronological fashion, in the very least when it comes to queries that possess a historical component. Having people prefer Text2Storyline’s ‘Search Results’ over Google’s is equally unexpected and a few reasons can justify this outcome. Although only minimally distinct from Google’s interface, it may look more appealing by simply not being what most people are constantly used to, which would be Google or a similar search engine. On top of this, the ability to toggle between sorting search results by date or their relevance scores may be important to users, as well as the ability to create a narrative from one of them.
5. RESULTS AND DISCUSSION

**Figure 5.4:** Three of the five options for a question from the first survey to rank Search Results pages: (a) Option A, (b) Option E and (c) Option D
5.1.2 Storyline Visual Component Evaluation

The next two sets of questions focus on the storyline visual component, whose results and feedback from previous questions have shown to be one of the prime components of the Text2Storyline interface. In particular, participants were asked their preference regarding two distinct elements in events: textual preferences and media preferences.

For the first scenario, four different cases were considered (see Figures A.1h, A.1i, A.1j and A.1k from Appendix A) as participants were asked to choose between one of the following options: the first, which is used by Text2Storyline, where only the sentence that contains the temporal expression is used; the second, which contains more context in the form of an additional sentence that follows the one where the temporal expression lies in. Figure 5.7 shows an illustrative example of one of those instances. Four distinct questions were posed to participants in order to draw proper conclusions on the matter, as two questions indeed add useful information and the remaining do not. These questions follow a similar distribution of responses where less text is vastly preferred, recording
65% of submissions for the first event, 69% for the second and 75% for both the third and fourth. This means that most people would rather have as little text as possible, even if it adds more information to the event it describes. This is directly reflected by the first and second questions having descriptions where the additional text is rather useful to the event it depicts, thus showing less of a preference by respondents towards opting for less text. Regardless, the majority may have found that the overall quality of the storyline events would suffer by having multiple events with extensive texts, even if the reader would benefit from the additional content.

Which of the following options do you prefer to be used as description for an event in the Storyline?

- A (More text)
- B (Only the sentence that contains the date for the event)

Figure 5.7: Question from the first survey on the preferable description for a storyline event

For the second scenario, we aim to understand how much images (along with text) are valued by readers. For this, participants were faced with two events and were asked to decide if the presence of the illustration was beneficial to the event or not. Figure 5.8 shows both questions. Unsurprisingly, for both the first (73%) and the second (87%) questions, the vast majority decided images are important to improve the understanding of a text. This implies that, regardless of the respondent’s age, people almost always want an image to accompany a text. This point of study will be deepened when analyzing the results of the second survey, where the correctness of the illustrations will play an important role in people’s responses (see Section 5.2).
Which of the following options regarding how information is presented (A or B) do you prefer?

**OPÇÃO A**

![Image](image.png)

1977
Jubileu de Prata

A rainha marca o seu Jubileu de prata (25 anos como monarca) com uma digressão pelos países da Commonwealth e celebrações faustosas na Grã-Bretanha.

**OPÇÃO B**

A rainha marca o seu Jubileu de Prata (25 anos como monarca) com uma digressão pelos países da Commonwealth e celebrações faustosas na Grã-Bretanha.

○ A (with image)

○ B (without image)

(A) First question on presence of illustration in event of Storyline

Which of the following options regarding how information is presented (A or B) do you prefer?

**OPÇÃO A**

![Image](image.png)

JANEIRO DE 2022
Harry e Meghan

Logo em Janeiro, o anúncio de Harry e Meghan de abandonarem os deveres reais e passarem a morar longe constitui um duro golpe para a monarca.

**OPÇÃO B**

Logo em Janeiro, o anúncio de *Harry e Meghan* de abandonarem os deveres reais e passarem a morar longe constitui um duro golpe para a monarca. O casal, que pretendia manter-se no núcleo duro da família, mas com menos deveres, acabou por perder os títulos e mudar-se para Los Angeles, depois de uma tentativa de se instalar no Canadá. A avalanche causada pela decisão do segundo filho de Carlos acabou por ser relegada para segundo plano depois da declaração da pandemia de covid-19. A 5 de Abril, a rainha falou à nação sobre o efeito da doença nas vidas de todos.

○ A (with image)

○ B (without image)

(B) Second question on presence of illustration in event of Storyline

**Figure 5.8:** Two questions from the first survey to inquire feedback on usage of illustrations for Storyline events
5.1.3 Entities Visual Component Evaluation

The Entities feature was also tested as we aim to compare Text2Storyline’s entities with another service to flesh out potential flaws or positive visual aspects. In this set of questions, it was asked to choose between Text2Storyline and a platform capable of similar entity detection, Dandelion\(^4\). Both questions, which are illustrated in Figure 5.9, are centered on the news article on Queen Elizabeth II, as to which had the most appealing way of displaying entities. The first question (see Figure 5.9a) considered the overall look and accessibility of all entities detected. Out of the 267 participants, 244 (91\%) preferred Text2Storyline’s. However, when a single example is considered (see Figure 5.9b), the opinion shifted as 151 (57\%) found Dandelion’s approach more attractive, which has more information and is easy to read. This suggests that even if respondents may have found Text2Storyline’s interface more attractive when multiple entities are displayed, they valued the amount of information available much more. This is important feedback in order to help us understand what to improve.

This finishes the analysis of the questions from the first survey. From this, we can conclude that all of Text2Storyline’s features are broadly viewed as important but, from having respondents order them according to their relevance, people still prefer the components that display the most amount of information in a direct manner such as the ‘Annotated Text’. Then again, people still value the presence of visual elements such as illustrations in the ‘Storyline’ alongside as little text as possible. This may be an obvious disparity, although it can be explained by people preferring to have both components display information in distinct ways: ‘Annotated Text’ includes as much information as possible and ‘Storyline’ chronologically shows little bits of text for each each illustrated event for a lighter and more visual-oriented approach.

\(^4\)https://dandelion.eu/semantic-text/entity-extraction-demo/
Which of the following options (A or B) do you prefer as a means to display multiple entities regarding a news related to Queen Elizabeth II.

(A) First question on the overall look of the interface with entities
Which of the following options (A or B) do you prefer as a means to display one entity (Queen Elizabeth II).

(b) Second question on the specific look of the interface for one entity

Figure 5.9: Two questions from the first survey on a comparison between Text2Storyline’s Entities and Dandelion’s Entities
5.2 Comparability Evaluation

The second survey focuses on the effectiveness of Text2Storyline’s prime components, namely ‘Storyline’ and ‘Entities’. The remaining components, although just as important, were excluded from this assessment as their potential flaws and strong points were already evident and well-known. Therefore, all elements from the previously mentioned components are assessed and compared to the original sources, similar systems or manually selected alternatives. Being substantially more extensive than the first survey, it only gathered 36 responses. As expected, the majority of the participants are between 18 to 24 years of age (50%), followed by ages 25 to 39 (33%), older than 40 (16%), and lastly, 17 or younger (1%). As for distribution based on academic qualifications, the majority of participants has a Bachelor’s degree (50%), followed by people with Secondary Education (14%), those with a Master’s degree (28%) and lastly, those with a Doctorate degree (8%).

This questionnaire addresses the ‘Storyline’ and ‘Entities’ components for four distinct examples, which were already mentioned throughout this thesis and are the following: news article on Queen Elizabeth II; Wikipedia page on Italian poet Dante Alighieri; a news article on the death of the former Portuguese President of the Republic, Mário Soares, which is also our running example 1; and the query “Síndrome Respiratória do Médio Oriente” (Middle East Respiratory Syndrome).

The first part of the survey focuses on evaluating the ‘Storyline’ feature, as the first set of questions (see Subsection 5.2.1) focus on all three elements present in an event of the ‘Storyline’ feature, namely the related image, date and title. Relevance of events is also an important point of study (see Subsection 5.2.2), in order to note whether automatically scored temporal expressions which were classified as irrelevant are indeed irrelevant, thus worth hiding the respective events by default, and vice-versa. Then, focus is shifted to comparing illustrations from all four events to manually selected ones (see Subsection 5.2.3), in order to better understand how accurate the image selection tool is.

The second part of this survey (see Subsection 5.2.4) focuses on entities, expanding on what was inquired in the previous survey. However, the main focus is determining how effective Text2Storyline is at providing the most accurate or precise information for entities. This includes the images that represent each entity as well as a description and possible classes, among other elements. These questions begin by considering a set of multiple entities overall and then, focus on a single entity.
5.2.1 Storyline Feature’s Effectiveness

In this section, we continue what was previously discussed in the first survey (see Section 5.1.2). In addition to this previous experiment, participants are also asked to rate the title, image and date of 7 randomly chosen chronological moments for each event, from 1 (least) to 3 (most), according to how adequate each element is. Given the extensive number of questions, a summary will be presented and the most notorious or unexpected results will be shown and discussed.

5.2.1.1 Queen Elizabeth II

The first test subject to be analyzed is the news article on Queen Elizabeth II. Results vary but the general consensus is that dates are almost always accurate or correct (2 or 3). Titles are sometimes incomplete (2) but more frequently correct (3). This may happen because YAKE! is used with a 3-gram value to obtain the best keyphrase of the description and in some cases titles with more than 3 words are necessary. However, the correctness of images is uncertain, oscillating between completely wrong (1) to acceptable but flawed (2) yet almost never perfect (3). It is important to note that these observations were performed considering the example of the news article centered on Queen Elizabeth II, therefore, some participants may not have been totally aware of the topic despite some contextual information provided so, the margin for error in these submissions, although minor, must be taken into account. Figure 5.10 shows, for this example, the storyline events that achieved the most positive and negative feedback. The distribution of votes is also displayed. For the first event, (see Figure 5.10a), the majority believe the image and the date are accurate and adequate as it properly shows a photo of Queen Elizabeth and Prince Philip’s wedding. However, the title is simply “Marinha Filipe Mountbatten” (Navy Philip Mountbatten) which is clearly incomplete as it should mention the wedding as well as Elizabeth. As for the negatively rated event (see Figure 5.10b), all elements with the exception of the date and the description are severely flawed. The title is “Harry torna-se pública” (Harry becomes public) which by itself is nonsensical. Consequently, the illustration follows the same erratic behaviour as it is in no way related to the event in question.

Additionally, Figure 5.11 displays a table of the complete distribution of results in percentage (%). The most voted score for every element (title, image and date) of 7 randomly chosen chronological moments is marked accordingly for an easier analysis. The average
score that takes into account all assessed moments for each element is also shown. This was achieved by summing the number of votes for each score multiplied by that score (1, 2 or 3), which was then divided by the total number of respondents (36). This was performed for each chronological moment and its outcome represents the average score for that specific moment in the event. Then, those averages were summed and divided by the total number of moments in question (7). This process is done separately for each element, namely the title, the image and the date.

![Figure 5.10: Storyline events with the most positive (a) and most negative (b) results](image1.png)

![Figure 5.11: Full distribution of votes in percentage (%) for all questions for the example on Queen Elizabeth II](image2.png)

<table>
<thead>
<tr>
<th>Moment</th>
<th>Title</th>
<th>Image</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Moment 1</td>
<td>0.25</td>
<td>0.33</td>
<td>0.42</td>
</tr>
<tr>
<td>Moment 2</td>
<td>0.44</td>
<td>0.36</td>
<td>0.2</td>
</tr>
<tr>
<td>Moment 3</td>
<td>0.58</td>
<td>0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>Moment 4</td>
<td>0.31</td>
<td>0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Moment 5</td>
<td>0.39</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td>Moment 6</td>
<td>0.5</td>
<td>0.31</td>
<td>0.19</td>
</tr>
<tr>
<td>Moment 7</td>
<td>0.47</td>
<td>0.22</td>
<td>0.72</td>
</tr>
</tbody>
</table>

5.2.1.2 Márcio Soares

The next example, referent to events on the life of Márcio Soares, shows significantly better results, overall. However, the titles seem to have become worse taking on the role of the poorest displayed element. Dates are, once more, rarely rated with a score of 1 and
in cases when that happens, the majority still decides that it is acceptable, yet flawed (2). This comes down to some participants being more critical than others. As for illustrations, images are accurate and correct considering the title that is provided, but as the majority of participants is correlating the image to the entire context provided by the description of said event, the rating of both title and image decrease significantly. We may assume that the main issue lies in the title being incomplete (2). Consequently, the image suffers from this behaviour and is considered completely wrong (1). Similar to the previous example, two distinct events and respective results are displayed in Figure 5.12. Additionally, The positively rated event (see Figure 5.12a) is slightly confusing as the main occurrence is Mário Soares founding the Socialist Party. However, since the entire sentence where the date is found is used, it also mentions the notion of the former President of the Republic of Portugal being considered the father of Democracy towards the end of the sentence. This may have confused some participants resulting in the divisive decision in rating the title and image accordingly. The date is simply correct as the majority of the votes confirms. As for the second event (see Figure 5.12b), the date was still considered correct by most people. However, the title “Presidente da República” (President of the Republic) is incredibly incomplete as it fails to mention the actual occurrence for this date, which is Mário Soares being hospitalized. The image is also too generic, only showing four pictures of Mário Soares throughout his life. Note that, if the title had been correct then so would the illustration.

Additionally, Figure 5.13 displays a table of the complete distribution of results in percentage (%). The most voted score for every element (title, image and date) of 7 randomly chosen chronological moments is marked accordingly for an easier analysis. The average score that takes into account all assessed moments for each element is also shown. This was achieved by summing the number of votes for each score multiplied by that score (1, 2 or 3), which was then divided by the total number of respondents (36). This was performed for each chronological moment and its outcome represents the average score for that specific moment in the event. Then, those averages were summed and divided by the total number of moments in question (7). This process is done separately for each element, namely the title, the image and the date.
5. RESULTS AND DISCUSSION

Figure 5.12: Storyline events with the most positive (a) and most negative (b) results

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Image</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment 1</td>
<td>0.19</td>
<td>0.42</td>
<td>0.08</td>
</tr>
<tr>
<td>Moment 2</td>
<td>0.19</td>
<td>0.39</td>
<td>0.11</td>
</tr>
<tr>
<td>Moment 3</td>
<td>0.08</td>
<td>0.47</td>
<td>0.06</td>
</tr>
<tr>
<td>Moment 4</td>
<td>0.14</td>
<td>0.53</td>
<td>0.19</td>
</tr>
<tr>
<td>Moment 5</td>
<td>0.36</td>
<td>0.39</td>
<td>0.06</td>
</tr>
<tr>
<td>Moment 6</td>
<td>0.5</td>
<td>0.28</td>
<td>0.17</td>
</tr>
<tr>
<td>Moment 7</td>
<td>0.5</td>
<td>0.31</td>
<td>0.06</td>
</tr>
<tr>
<td>Avg Score</td>
<td>2.04</td>
<td>2.07</td>
<td>2.59</td>
</tr>
</tbody>
</table>

Figure 5.13: Full distribution of votes in percentage (%) for all chronological moments for the example on Mário Soares

5.2.1.3 Dante Alighieri

The next example, which is centered on Dante Alighieri, the famous Italian writer, displays an increasing trend on poorly displayed illustrations for events. Excluding 2 of the total 7 random chronological moments, the overwhelming majority decided the images were extremely incorrect or unrelated (1) to the events they try to depict. In contrast, dates are rated as very adequate (3) throughout, with one exception where votes were more spread out but still somewhat neutral (2), which shows participants were critical, yet did not dismiss this element as totally flawed. The rating of titles falls in-between the other two elements of a storyline, receiving a divisive feedback (1 or 2) from participants in almost all chronological events that were tested. This may be due to titles reflecting the most important part of the description but striking users as still not being good enough. Other, more critical participants may have decided extracting the main idea, location or
personality from the description is not enough of a condition to have the title be considered incomplete. Figure 5.14 illustrates the best and worst rated events for this example. The first (see Figure 5.14a), and overall positive event throughout all elements is merely a final annotation on a painting that depicts Dante and his lover. Despite this, the event is still properly dated and the illustration shows that painting which is proven by the distribution of results for those elements. However, the title is somewhat lackluster as it only mentions the author of the painting as “italiano Cesare Sacaggi” (Italian Cesare Sacaggi). The participants were rather benevolent in rating this title and the tie between poorly chosen (1) and somewhat acceptable (2) is understandable. As for the second chronological moment (see Figure 5.14b), which refers to the arranged marriage between Dante and Gemma Donati, is rather poorly rated. The only exception is the date, as it remains consistently well reviewed in all examples thus far. The title only refers to Gemma’s father, “Messe Manetto Donati” and that should not be the focal point of this occurrence. The title is in no way related to either the description or the title as it only shows a building. It must be noted that no images can be found of Gemma or her father, so illustrating this event would be incredibly difficult if not impossible. However, the image chosen is far from an acceptable solution.

Additionally, Figure 5.15 displays a table of the complete distribution of results in percentage (%). The most voted score for every element (title, image and date) of 7 randomly chosen chronological moments is marked accordingly for an easier analysis. The average score that takes into account all assessed moments for each element is also shown. This was achieved by summing the number of votes for each score multiplied by that score (1, 2 or 3), which was then divided by the total number of respondents (36). This was performed for each chronological moment and its outcome represents the average score for that specific moment in the event. Then, those averages were summed and divided by the total number of moments in question (7). This process is done separately for each element, namely the title, the image and the date.

5.2.1.4 Middle East Respiratory Syndrome

The last example focuses on the Middle East Respiratory Syndrome or MERS. Similarly to the previous examples, the dates maintain a consistent positive rating (3) throughout all events. The title suffers from the opposite kind of feedback, as all but one event have been considered by the majority to be inadequate (1). The illustrations can have two common
5. RESULTS AND DISCUSSION

(B) Storyline event and results

Figure 5.14: Storyline events with the most positive (a) and most negative (b) results

<table>
<thead>
<tr>
<th></th>
<th>Title 1</th>
<th>Title 2</th>
<th>Title 3</th>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
<th>Date 1</th>
<th>Date 2</th>
<th>Date 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment 1</td>
<td>0.44</td>
<td>0.36</td>
<td>0.19</td>
<td>0.86</td>
<td>0.11</td>
<td>0.03</td>
<td>0.06</td>
<td>0.19</td>
<td>0.75</td>
</tr>
<tr>
<td>Moment 2</td>
<td>0.25</td>
<td>0.39</td>
<td>0.36</td>
<td>0.94</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.11</td>
<td>0.83</td>
</tr>
<tr>
<td>Moment 3</td>
<td>0.08</td>
<td>0.61</td>
<td>0.31</td>
<td>0.89</td>
<td>0.08</td>
<td>0.03</td>
<td>0.25</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Moment 4</td>
<td>0.11</td>
<td>0.56</td>
<td>0.33</td>
<td>0.92</td>
<td>0.05</td>
<td>0.03</td>
<td>0</td>
<td>0.11</td>
<td>0.88</td>
</tr>
<tr>
<td>Moment 5</td>
<td>0.31</td>
<td>0.42</td>
<td>0.27</td>
<td>0.94</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.14</td>
<td>0.83</td>
</tr>
<tr>
<td>Moment 6</td>
<td>0.42</td>
<td>0.36</td>
<td>0.22</td>
<td>0.25</td>
<td>0.53</td>
<td>0.22</td>
<td>0.08</td>
<td>0.17</td>
<td>0.75</td>
</tr>
<tr>
<td>Moment 7</td>
<td>0.28</td>
<td>0.47</td>
<td>0.25</td>
<td>0.17</td>
<td>0.44</td>
<td>0.39</td>
<td>0.14</td>
<td>0.17</td>
<td>0.69</td>
</tr>
<tr>
<td>Avg Score</td>
<td>2.00</td>
<td>1.40</td>
<td>2.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.15: Full distribution of votes in percentage (%) for all chronological moments for the example on Dante Alighieri

outcomes, the first being an evident decision towards improper or incorrect selection. The second result is more positive as it registers a disparity between somewhat appropriate (2) or incorrect (1) as a small minority consider it is as correct and adequate as possible (3). Figure 5.16 shows the participant’s behaviour for two distinctly rated events. The event that gathered more positive results (see Figure 5.16a) is centered on the discovery of the Severe Acute Respiratory Syndrome (SARS) disease. The description contains two dates: “2003”, the one this event focuses on, and “2012”, the date the MERS disease was discovered. Therefore, even though it would be more suitable to focus on MERS rather than SARS, the date is correct, as the majority of participants decided. From this thought, then the title “Síndrome Respiratória Aguda Grave” (Acute Severe Respiratory Syndrome) is also mostly correct, although a better alternative should mention its discovery. This may have steered the majority of users to rate it as only being adequate enough (2). As for the image, it correctly illustrates SARS, especially to anyone ignorant on the matter and, as
such, no participant rated this illustration with a score lower than 2 and the vast majority decided it is a good image selection (3). When it comes to the more negatively assessed storyline event (see Figure 5.16b), the date element still registers a majority of votes towards being accurate and adequate (3). However, both the title and illustration have been rated as very poor choices (1) by the overwhelming majority. The title chose a part from the description that is merely advertisement, in part being in capitalized letters. This negatively impacted the whole event as the focal point should have been the death of the last known MERS infected patient in South Korea. This also influenced the quality of the image for this event. Even though it is a correct choice for this title, it is completely unrelated to the event and to MERS. Therefore, these negative results are expected.

Additionally, Figure 5.17 displays a table of the complete distribution of results in percentage (%). The most voted score for every element (title, image and date) of 7 randomly chosen chronological moments is marked accordingly for an easier analysis. The average score that takes into account all assessed moments for each element is also shown. This was achieved by summing the number of votes for each score multiplied by that score (1, 2 or 3), which was then divided by the total number of respondents (36). This was performed for each chronological moment and its outcome represents the average score for that specific moment in the event. Then, those averages were summed and divided by the total number of moments in question (7). This process is done separately for each element, namely the title, the image and the date.

Lastly, Figure 5.18 displays the average rating score of each element, considering all four examples (Queen Elizabeth II, Mário Soares, Dante Alighieri, MERS). These averages were obtained by summing the previous average scores for each instance and then
5. RESULTS AND DISCUSSION

Figure 5.17: Full distribution of votes for all questions for the example on the Middle East Respiratory Syndrome (MERS)

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Image</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment 1</td>
<td>0.14</td>
<td>0.58</td>
<td>0.28</td>
</tr>
<tr>
<td>Moment 2</td>
<td>0.39</td>
<td>0.44</td>
<td>0.17</td>
</tr>
<tr>
<td>Moment 3</td>
<td>0.44</td>
<td>0.44</td>
<td>0.11</td>
</tr>
<tr>
<td>Moment 4</td>
<td>0.5</td>
<td>0.33</td>
<td>0.17</td>
</tr>
<tr>
<td>Moment 5</td>
<td>0.67</td>
<td>0.14</td>
<td>0.19</td>
</tr>
<tr>
<td>Moment 6</td>
<td>0.69</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>Moment 7</td>
<td>0.39</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td>Avg Score</td>
<td>1.72</td>
<td>1.70</td>
<td>2.57</td>
</tr>
</tbody>
</table>

This process is done separately for each element, namely the title, the image and the date. These simple yet incredibly informative values show that dates were the highest rated element, having almost no issues throughout the total 28 (7 × 4) chronological moments that were assessed. The other two elements, title and image, received rather underwhelming ratings in comparison, 1.89 and 1.69, respectively. This shows us the aspects of the storyline that need improvement the most.

Figure 5.18: Average score for each element across all examples

5.2.2 Storyline Relevance Date Effectiveness

One other aspect of the ‘Storyline’ is once more assessed with the help of the survey for completely unbiased feedback. We aim to verify that low scores (below 0.35), which are assigned to dates by Time-Matters, are indeed not relevant to be shown to the users. Note that non-relevant dates are not shown in the ‘Storyline’ component by default. Therefore, 6 questions were asked for each of the four total examples, thus resulting in a total of 24 questions. For each example, 3 relevant events with scores higher than 0.35 and 3 with scores lower than 0.35 were used as test subjects and the participants of the survey were asked if the event in question was relevant to them, when taking into account the general sense and focus of each text or query. There are three possible choices for each question:
Yes, it is important; Yes it is important, but some elements such as title, date or image are incorrect; No, it is not important. Note that scores of each event were kept hidden from respondents to promote an unbiased behavior.

5.2.2.1  Queen Elizabeth II

The first set of questions corresponding to the narrative generated on the life events of Queen Elizabeth II presented divergent opinions for the so-called relevant chronological moments, although the small majority did agree all 3 randomly chosen moments are important while acknowledging there are some flaws, as Figure 5.19a confirms for one of those occurrences. The event is indeed important as it describes a key moment in the life of Elizabeth, however, the title and image are incorrect. One exception, however, steered all participants into agreeing it was a relevant event, although the votes were still split between the existence of errors or not. These are sensible questions, as the minimal issue in any element will tip the decisions towards that option. Regarding the irrelevant events, the same conflicting behaviour occurred in only one of the questions, whereas there was a unanimous agreement that the other two events are important, despite having a low score. Figure 5.19b proves these results by showing one of those events. This behaviour is entirely justifiable given the context of this event, which focuses on Queen Elizabeth’s 90th birthday. The image is acceptable and, while the title is incomplete, it is understandable how some people decided it was good enough to select the first option. This outcome may suggest that low scores do not necessarily mean that the subsequent occurrence it depicts is unimportant. Additionally, Figure 5.20 displays a table with the full distribution of results for all six chronological moments on the life of Queen Elizabeth II. The most voted option is marked accordingly and each question is automatically identified as relevant (score $\geq 0.35$) or irrelevant (score $< 0.35$). The average score distribution is also calculated for a more generic understanding on this subject. For this, we assumed that option “No” carries a value equivalent to a score of 1, “Yes, but has errors” means a score of 2 and “Yes” represents a score of 3. Then, the amount of votes (0 to 36) for each option were multiplied by their respective score (1, 2 or 3), such that the number of votes for “No” is multiplied by 1, the amount of votes for “Yes, but with errors” is multiplied by 2 and the number of votes for “Yes” is multiplied by 3. Then that value is divided by the total number of moments that are being assessed (6), which results in the final average score. This score has a value of 2.09, which implies that, being closer to the score of 2,
means most, if not all, chronological moments in question were voted by participants as being relevant to the event but having evident errors.

![Figure 5.19: Storyline events with a high score (0.429) (a) and a low score (0.228) (b)](image)

<table>
<thead>
<tr>
<th>Moment 1</th>
<th>Yes, but has errors</th>
<th>Yes</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>0.72</td>
<td>0.25</td>
<td>Irrelevant</td>
</tr>
<tr>
<td>Moment 2</td>
<td>0.31</td>
<td>0.50</td>
<td>0.19</td>
</tr>
<tr>
<td>Moment 3</td>
<td>0.14</td>
<td>0.47</td>
<td>0.39</td>
</tr>
<tr>
<td>Moment 4</td>
<td>0.36</td>
<td>0.39</td>
<td>0.25</td>
</tr>
<tr>
<td>Moment 5</td>
<td>0</td>
<td>0.46</td>
<td>0.54</td>
</tr>
<tr>
<td>Moment 6</td>
<td>0.46</td>
<td>0.50</td>
<td>0.04</td>
</tr>
<tr>
<td>Avg Score</td>
<td></td>
<td></td>
<td>2.09</td>
</tr>
</tbody>
</table>

![Figure 5.20: Full distribution of results in percentage (%) for all questions on the relevance of events on Queen Elizabeth II](image)

### 5.2.2.2 Mário Soares

For the next example, centered on the former President of Portugal, Mário Soares, all events are inherently relevant. The event with the lowest score (0.416) refers to his death, and is arguably the most important event from this storyline. The participants agree as 100% decided it was relevant, although a small majority (59%) decided some elements were flawed. As Figure 5.21a shows, the title was definitely the main motivation for this choice as it fails to mention the death of Mário Soares. The image is also not too illustrative but still shows Mário Soares from a moment not too distant from his passing. Similarly, all others events share this overwhelming majority, as only one event, which
is illustrated in Figure 5.21b, had 2 (6%) negative votes out of the total 36. The majority of people considered there were flaws in this event’s elements. Both title and image are not evidently wrong, but the title fails to convey the main occurrence, which is that he failed to be elected President of Portugal in 2006. Overall, even though the 6 events were already considered relevant, they are confidently confirmed to be important to the narrative. Additionally, Figure 5.22 displays a table with the full distribution of results for all six chronological moments on the life of Queen Elizabeth II. The most voted option is marked accordingly and each question is automatically identified as relevant (score $\geq 0.35$) or irrelevant (score $\leq 0.35$). The average score distribution is also calculated for a more generic understanding on this subject. For this, we assumed that option “No” carries a value equivalent to a score of 1, “Yes, but has errors” means a score of 2 and “Yes” represents a score of 3. Then, the amount of votes (0 to 36) for each option were multiplied by their respective score (1, 2 or 3), such that the number of votes for “No” is multiplied by 1, the amount of votes for “Yes, but with errors” is multiplied by 2 and the number of votes for “Yes” is multiplied by 3. Then that value is divided by the total number of moments that are being assessed (6), which results in the final average score. This score has a value of 2.59, which implies that respondents decidedly agreed all chronological moments are relevant. However, there appears to be less certainty when deciding on the existence of flawed elements, as the first three moments were quite convincingly voted as not having any issues, but the last three were not as discernible.
5. Results and Discussion

5.2.2.3 Dante Alighieri

The third example is centered on the Italian poet Dante Alighieri, as three irrelevant and three relevant events on his life were proposed to participants for evaluation. Even though, the events with a higher score were still considered relevant by at least 75% of people with a massive emphasis on the errors of their elements, it was not as clear of a decision as it was noted in the previous example. The events with low scores that were deemed irrelevant showed a polar opposite distribution of results, with one event having 100% of votes agree it is important and with only 6 votes (17%) being critical and considering at least one element is flawed, as Figure 5.23a illustrates. This question benefited from the benevolence of participants mixed with lack of knowledge on the topic as the image is indeed rather poorly selected, even though the title is optimal. On the other hand, another irrelevant event, shown in Figure 5.23b, has a majority of 75% of the participants deciding it is not relevant for this narrative, which is justifiable as it should not be an actual event since it is merely using the date of publication of a book used as a bibliographic reference during the writing of Dante’s Wikipedia page, where all texts for storyline events was extracted from. Even then, 25% still thought it was worth including and, taking into account that the image and title are appropriate justifies the majority of those votes identifying no flaws. Additionally, Figure 5.24 displays a table with the full distribution of results for all six questions of events on the life of Dante Alighieri. The most voted option is marked accordingly and each question is identified as relevant (score $\geq 0.35$) or irrelevant (score $\leq 0.35$). The average score distribution is also calculated for a more generic understanding on this subject. For this, we assumed that option “No” carries a value equivalent to a score of 1, “Yes, but has errors” means a score of 2 and “Yes” represents a score of 3. Then,
the amount of votes (0 to 36) for each option were multiplied by their respective score (1, 2 or 3), such that the number of votes for “No” is multiplied by 1, the amount of votes for “Yes, but with errors” is multiplied by 2 and the number of votes for “Yes” is multiplied by 3. Then that value is divided by the total number of moments that are being assessed (6), which results in the final average score. This score has a value of 2.02, implying that, by being closer to the score of 2, most, if not all, chronological moments were voted by respondents as being relevant to the event while having perceptible errors.

<table>
<thead>
<tr>
<th>Moment 1</th>
<th>Yes, but has errors</th>
<th>Yes</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment 2</td>
<td>0.31</td>
<td>0.61</td>
<td>0.08</td>
</tr>
<tr>
<td>Moment 3</td>
<td>0.08</td>
<td>0.58</td>
<td>0.33</td>
</tr>
<tr>
<td>Moment 4</td>
<td>0.17</td>
<td>0.72</td>
<td>0.83</td>
</tr>
<tr>
<td>Moment 5</td>
<td>0.75</td>
<td>0.04</td>
<td>0.21</td>
</tr>
<tr>
<td>Moment 6</td>
<td>0.28</td>
<td>0.61</td>
<td>0.11</td>
</tr>
<tr>
<td>Avg Score</td>
<td></td>
<td></td>
<td>2.02</td>
</tr>
</tbody>
</table>

**Figure 5.24:** Full distribution of results in percentage (%) for all questions on the relevance of events on Dante Alighieri

5.2.2.4 Middle East Respiratory Syndrome

The final example, a narrative generated from the query “Síndrome Respiratória do Médio Oriente” (MERS) only has one irrelevant event with a score of 0.224. A total of 75% of the participants decided it is worth being included in the storyline, despite having terribly flawed elements in the title and illustration, as it was previously observed (see Figure
The main information the description conveys is rather important, which justifies this distribution of results shown by Figure 5.25a. The remaining relevant events are positively rated overall and are all considered important to the narrative. Two of those events received unanimous approval, as in one, displayed by Figure 5.25b, the majority (75%) voted that no errors were found in the occurrence’s elements. In the other, the opposite was noted as most people (71%) considered there was an issue with at least one of the elements. This implies that although the events that are being formed are correct and related to the context of the narrative, some of its elements need improvement as they may confuse users or be a nuisance. Additionally, Figure 5.26 displays a table with the full distribution of results for all six questions of events on the Middle East Respiratory Syndrome (MERS). The most voted option is marked accordingly and each question is identified as relevant (score ≥ 0.35) or irrelevant (score < 0.35). The average score distribution is also calculated for a more generic understanding on this subject. For this, we assumed that option “No” carries a value equivalent to a score of 1, “Yes, but has errors” means a score of 2 and “Yes” represents a score of 3. Then, the amount of votes (0 to 36) for each option were multiplied by their respective score (1, 2 or 3), such that the number of votes for “No” is multiplied by 1, the amount of votes for “Yes, but with errors” is multiplied by 2 and the number of votes for “Yes” is multiplied by 3. Then that value is divided by the total number of moments that are being assessed (6), which results in the final average score. This score has a value of 2.22, meaning that all chronological moments were voted by respondents as being relevant to the event. This certainty was repeated when identifying potential flaws in the elements of the moments in question, as all but one were voted as having errors.
Lastly, the final average rating score considering all four examples combined (Queen Elizabeth II, Mário Soares, Dante Alighieri, MERS) sums up the ratings of all 24 assessed chronological moments. This value was obtained by summing the previous average scores for each instance and then dividing it by the total number of said instances (4). This score is 2.23, which can imply multiple conclusions. Overall, most respondents agreed that the majority of both non-relevant and relevant moments, save some exceptions, belong in the storyline. However, this is not entirely a positive result for Text2Storyline, as some initially considered lowly scored moments should be scored higher, thus identifying a potential flaw in the system.

### 5.2.3 Storyline Image Effectiveness

Lastly, we take a different approach that allows us to assess the degree of correctness of images in events by putting them against manually selected images of the same event. Given a random chronological event in the ‘Storyline’ component, accompanied by an illustration, we fetched an image that would be appropriate for said event and let respondents state their preference between either of them. This different methodology helps us grasp the notion of how effective the illustration selection tool is when compared to other images that participants should naturally favor. Figure 5.27 illustrates two questions that obtained the most interesting outcome. This is important to help us understand how much having somewhat incorrect or unrelated illustrations negatively impacts the reader’s perception of an event, taking into account that most respondents showed their preference towards having images in storyline events in the first survey. The results were somewhat expected in most questions as the overwhelming majority preferred the manually selected images, even when Text2Storyline’s illustrations were considered correct.
In one of these cases, all 36 (100%) participants choose that option (see Figure 5.27a). Interestingly, two questions managed to sway participants into voting for the automatically selected image. Both belong to Dante’s example and achieved 53% and 56% of the votes. Figure 5.27b displays the latter of these two instances. These distinct distribution in preference implies that images that may be arguably unrelated to the event will always be disregarded as it would rarely occur with a manually selected image. On the other hand, automatically chosen illustrations have the ability to be just as appropriate as manually selected ones, or even beat them in some instances. This can prove that, even though it is still flawed, elements such as images being automatically chosen can be a massive advantage when generating visual narratives on a larger scale, for which there are not enough human resources, or it would be costly. These results also imply that the image selection algorithm used in this project, which uses the title of a given moment in a storyline event and then uses it as a query to extract the best possible image from Arquivo.pt’s image archive, may not be the most appropriate. As such, other approaches should be considered in the future. Note that, prior knowledge in the topic at hand, or lack of it, can have an impact on participants preferences, which is slightly more evident in the question for the event of Dante’s face reconstruction. There was only a slight margin of votes that preferred Text2Storyline’s image which shows a clear hesitation from participants. However, the manually selected image poses that rare occasion where it is actually incorrect as it illustrates Dante’s popular death mask, instead of what this chronological moment is alluding to. This undoubtedly swayed some distracted respondents into opting for this choice.

Overall, almost all non-relevant dates were considered important to the event it was inserted in. The one exception was a fixture that referred to the date of publication of a book, used as a bibliographical reference, thus, it was never relevant to the event. As for all other irrelevant events, respondents decidedly agreed they belonged in the storyline, even if they added little information. However, it must be noted that the majority believe these events displayed at least one flawed element, be it the title, the image or the date.

5.2.4 Entities’ Effectiveness

The final set of questions in this second survey focus on the ‘Entities’ feature, in a similar way to how questions in the previous survey were handled, but with a different
motivation. The aesthetic aspect was disregarded to pivot onto the correctness of elements, such as images, classes and descriptions. Just as the main intent of these questions changed, so did our intentions, as we now aim to identify potential weaknesses and flaws in Text2Storyline’s entities. All examples that have been previously mentioned, with the exception of the “Síndrome Respiratória do Médio Oriente” (MERS) query example, were used for this set of questions. This exception comes from the fact that there would not be an accurate way to measure the two platforms as these entities were identified from multiple documents rather than just one. It should also be noted that Dandelion only supports texts up to 700 characters whereas Text2Storyline’s limit is 10,000 characters.

5.2.4.1 Queen Elizabeth II

The distribution of results followed a more extreme inclination than the votes registered in the first survey. The questions based on Queen Elizabeth had the most positive results, with 86% votes in favor for the entities overall, and 83% for a specific entity. Figure 5.28 shows the options given to participants in these two questions for this example. These results are somewhat expected as more entities seem accurate and correctly identified. Images are also essential as it is the element people first look at, and having a high amount of correct illustrations ensures this preference. Considering a single entity in particular
makes the decision even more obvious as Dandelion’s image for Queen Elizabeth II is incorrect, as can be seen in Figure 5.30d.

![Figure 5.28: Comparison of entities in general (a/b) and considering a single entity (c/d) between Text2Storyline and Dandelion](image)

5.2.4.2 Mário Soares

The entities obtained for the text on the death of Mário Soares show a similar distribution of results, as 86% participants preferred Text2Storyline over Dandelion, when considering all entities identified and displayed. Taking particular focus into one specific entity, Mário Soares, the majority (67%) still preferred our approach, but a drop was noted in comparison to the first question.
In contrast, for the example on Dante, 56% preferred Dandelion’s approach for the entities overall. However, that behaviour was not repeated for a single entity as only 36% chose Dandelion. Figure 5.30 shows the options given to participants in these two questions for this example. For the first result, some entities were incorrectly illustrated and it was an issue that stood out quite easily. As for the specific entity of Dante, Dandelion’s description was too short and referred to a painting that depicted the Italian writer instead of the person.

From these results, one can confidently confirm that the overall look and behaviour of
5. RESULTS AND DISCUSSION

(A) Overall look of Text2Storyline’s entities

(B) Overall look of Dandelion’s entities

(C) Particular look of one of Text2Storyline’s entities

(D) Particular look of one of Dandelion’s entities

Figure 5.30: Comparison of entities in general (a/b) and considering a single entity (c/d) between Text2Storyline and Dandelion

multiple entities being displayed is preferable to Dandelion’s approach. However, Dandelion manages to provide more focus to a single entity, showing more information and in a more centered way, by zooming in on the selected entity to distinguish from the remaining. This may be a behaviour to take into consideration for the future, and present more options to the user rather than simply redirecting to the entity’s Wikipedia page on click.
5.3 Summary

This chapter provided an increased level of analysis to the examples provided in the previous chapter. The surveys provided extremely valuable feedback which helped understand what features would be appreciated and used the most as well as what needs to be improved the most. The results from both surveys were extensive, thus very informative for every question and every test subject, even if not every question and respective outcome was analyzed in this thesis, the summary provided can be applied to the majority of questions. The exceptions and minorities were covered in detail to better understand what separated them from the rest, whether they were positive or negative.

As such, the most prominent results allowed us to conclude that, as planned, participants would consider the ‘Storyline’ feature a central piece of Text2Storyline, along with the ‘Annotated Text’ and ‘Entities’. The ‘Search Results’ component also received extremely positive feedback for the way results can be sorted and used to generate individual narratives. Focusing on the ‘Entities’ feature, the majority of results heavily favored Text2Storyline when compared with a different platform, Dandelion. It is worth noting that, our implementation of ‘Entities’ was only less favored twice due to two distinct reasons: the quality of the entities’ images; the amount of information available to a single entity. On the other hand, some flaws were also made evident, notably in elements of the ‘Storyline’ and ‘Entities’ features. Images and titles of moments received poor reviews throughout 4 distinct narratives. For the ‘Storyline’ this flaw may come from incorrect titles obtained with YAKE! and a better solution may be required. This may also solve the accuracy of automatically selected images, as it uses said title to search a relevant illustration from Arquivo.pt. However, this does not occur for images extracted for entities, implying that a more accurate image selection method may be needed as well. Lastly, although temporal expressions were correctly identified in texts, the fact that the majority of the respondents disagreed with the Time-Matters scoring on the non-relevant moments, thus being considered in fact important to the narrative, show that this behaviour may need to be inspected.
In current times, the increasing amount of information that is generated, consumed and stored has made it difficult for those seeking information to extract knowledge in reasonable time. Additionally, information has slightly shifted towards types of content that are easier and faster to digest, with a strong visual component. Despite several advancements in the field of information retrieval and data visualization, the issue of building and presenting consistent narrative structures in the domain of web articles is yet to be completely resolved. Motivated by this, we propose Text2Storyline, a platform for generating and exploring enriched storylines from an input text, a URL or a user query. This project was born by adapting a previous implementation of the Time-Matters online demo, which was limited by only accepting text as input and it exclusively annotated the temporal expressions and keywords in the text. Our efforts, produced a new online platform that distinguished itself from its predecessor in two main ways. Firstly, by allowing other types of inputs, namely URLs and queries, in addition to single texts. Secondly, by enabling the user to expand their knowledge with the identification of potentially relevant persons, events, locations, objects or concepts, which are called entities. For the former, we rely on a Python package (newspaper3k) which is able to get a URL and obtain the corresponding text, and rely on the Arquivo.pt infrastructure to search for documents that are deemed relevant to the search term provided and, from there, generate a concise temporal narrative. For the latter, we use an API (Wikifier) that performs Entity Linking, more specifically a task called Wikification, as it assigns a Wikipedia page to the entities found in the text.

Overall, Text2Storyline consists of five components which are key to understand the story of a text. They are the ‘Annotated Text’, the ‘Entities’, the ‘Storyline’, the ‘Temporal
Clustering’ and the ‘Word Cloud’ components. Few projects exist with similar objectives regarding temporal narratives for proper comparisons. The closest example is Digital Libraries’ Narratives, which shares some similarities with Text2Storyline. Their Narrative Building and Visualising Tool (NBVT) rivals Time-Matters. However, their system extracts content from Europeana while illustrations are obtained from Wikimedia Commons. In Text2Storyline both information is collected from the Arquivo.pt infrastructure which gives us the chance of collecting historical information, though it also raises questions as to the quality of the data collected, when compared to Europeana. These distinctions result in different storylines in each system, even when centered on the same subject.

Two surveys were created to assess both the usability of the platform as well as the quality of results displayed from the generated visual narrative. Responses from 267 and 36 participants respectively, were collected and analyzed extensively. When giving their feedback on the ‘Storyline’ component, respondents were adamant that the images and title could be improved in order to increase the value of this feature, even though results also showed it was the feature respondents considered the most important, followed by the ‘Annotated Text’ and the ‘Entities’ component. Similarly, our ‘Search Results’ component received positive feedback, being favored when compared to other rather popular systems and their search results pages. As for the effectiveness of our system, when scoring dates, results showed that most non-relevant moments were deemed relevant to be shown by default, even those that contain redundant or repetitive information. As such, these results are not particularly enlightening, as our scoring algorithm may be working accordingly, but the majority of people favor more content, even if minimally informative. The ‘Entities’ feature was also thoroughly analyzed, as our entities were compared to the ones obtained when using the Dandelion system, both in terms of usability and performance. Results clearly imply that Text2Storyline’s approach is more appealing and effective when displaying the list of resulting entities. However, Dandelion was slightly more favored when focusing on a single entity as it displayed more information on the subject.

Our solution, is a first step towards understanding and creating narratives automatically in a large scale environment, which may be beneficial for displaying many visually outdated articles from the past, stored in Arquivo.pt. This can be particularly useful for
journalists that may require to fact-check information from past stories. Notwithstanding, Text2Storyline still has room for improvement. As future work, we plan to improve certain features, more notably the ‘Storyline’ and ‘Entities’ components, by implementing a more accurate image selection method.
Appendix A

Survey: Text2Storyline Usability

(a) Survey introduction

(b) Personal questions on age and education

Figure A.1: Complete survey on Text2Storyline’s usability
(c) Question 1

(d) Question 2

(e) Question 3

(f) Question 4

FIGURE A.1: Complete survey on Text2Storyline’s usability
A. SURVEY: TEXT2STORYLINE USABILITY

FIGURE A.1: Complete survey on Text2Storyline’s usability
Figure A.1: Complete survey on Text2Storyline’s usability
Appendix B

Survey: Text2Storyline Evaluation

Figure B.1: Complete survey on Text2Storyline’s evaluation
Figure B.1: Complete survey on Text2Storyline’s evaluation
(I) Question 5

Classifique os seguintes elementos de 1 a 3, onde 1 significa nada adequado e 3 significa muito adequado para a descrição do momento cronológico datado de 2003:

- 2003
  - 1
  - 2
  - 3

(Título)

- Imagem
- Data

(SUGERENCIA) Para sugestões/comentários relativos a qualquer um dos elementos anteriormente analisados, use a seguinte caixa de texto:

Short answer text

(I) Question 6

Classifique os seguintes elementos de 1 a 3, onde 1 significa nada adequado e 3 significa muito adequado para a descrição do momento cronológico datado de 19 de maio de 2018:

- 19 de maio de 2018
  - 1
  - 2
  - 3

(Título)

- Imagem
- Data

(SUGERENCIA) Para sugestões/comentários relativos a qualquer um dos elementos anteriormente analisados, use a seguinte caixa de texto:

Short answer text

(K) Question 7

Nos nossos debates, pretende-se que avalie os elementos da storyline relacionados a uma notícia sobre a morte do antigo Presidente da República, Mário Soares.

(OUTUBRO DE 2019)

(Título)

- Imagem
- Data

(SUGERENCIA) Para sugestões/comentários relativos a qualquer um dos elementos anteriormente analisados, use a seguinte caixa de texto:

Short answer text

(L) Separator for questions on Mário Soares

(M) Question 8

(N) Question 9

FIGURE B.1: Complete survey on Text2Storyline's evaluation
FIGURE B.1: Complete survey on Text2Storyline’s evaluation
B. SURVEY: TEXT2STORYLINE EVALUATION

FIGURE B.1: Complete survey on Text2Storyline’s evaluation
TEXT2STORYLINE: GENERATING ENRICHED STORYLINES FROM TEXT

Figure B.1: Complete survey on Text2Storyline’s evaluation
B. SURVEY: TEXT2STORYLINE EVALUATION

Figure B.1: Complete survey on Text2Storyline’s evaluation
No âmbito de uma notícia centrada na família Isabel II do Reino Unido (que recentemente celebrou o Jubileu de Prata), indique se achou importante, ou não, de visualizar na storyline o momento cronológico destacado de 1991.

1991
Estados Unidos
1991, A princesa Isabel do Reino Unido e o seu marido, o infante Juan, visitaram os Estados Unidos e tornou-se a primeira monarca e rainha em 67 anos a visitar uma Conferência da NACE.

☐ Sim, é importante
☐ Sim, é importante mas alguns componentes (título, data ou imagem) não estão corretos
☐ Não, não é importante

[OPCIONAL] Para questões/comentários relativos a qualquer um dos elementos anteriores, use a seguinte caixa de texto:
Short answer text

No âmbito de uma notícia centrada na família Isabel II do Reino Unido (que recentemente celebrou o Jubileu de Prata), indique se achou importante, ou não, de visualizar na storyline o momento cronológico destacado de 31 de agosto de 1997.

31 DE AGOSTO DE 1997
Miliciano matado em Paris

☐ Sim, é importante
☐ Sim, é importante mas alguns componentes (título, data ou imagem) não estão corretos
☐ Não, não é importante

[OPCIONAL] Para questões/comentários relativos a qualquer um dos elementos anteriores, use a seguinte caixa de texto:
Short answer text

No âmbito de uma notícia centrada na família Isabel II do Reino Unido (que recentemente celebrou o Jubileu de Prata), indique se achou importante, ou não, de visualizar na storyline o momento cronológico destacado de 21 de abril de 2016.

21 DE ABRIL DE 2016
Isabel celebra
21 de abril de 2016, Isabel celebrou o seu 60º aniversário.

☐ Sim, é importante
☐ Sim, é importante mas alguns componentes (título, data ou imagem) não estão corretos
☐ Não, não é importante

[OPCIONAL] Para questões/comentários relativos a qualquer um dos elementos anteriores, use a seguinte caixa de texto:
Short answer text

No âmbito de uma notícia centrada na morte de ex-Presidente da República Mário Soares, indique se achou importante, ou não, a visualização na storyline do momento cronológico destacado de 1973.

1973
Mário Soares

☐ Sim, é importante
☐ Sim, é importante mas alguns componentes (título, data ou imagem) não estão corretos
☐ Não, não é importante

[OPCIONAL] Para questões/comentários relativos a qualquer um dos elementos anteriores, use a seguinte caixa de texto:
Short answer text

(AQ) Separator for questions on Mário Soares

(AQ) Separator for questions on Mário Soares

(AR) Question 35

Figure B.1: Complete survey on Text2storyline’s evaluation
B. SURVEY: Text2Storyline Evaluation

FIGURE B.1: Complete survey on Text2Storyline’s evaluation
TEXT2STORYLINE: GENERATING ENRICHED STORYLINES FROM TEXT

Figure B.1: Complete survey on Text2Storyline’s evaluation
B. SURVEY: TEXT2STORYLINE EVALUATION

FIGURE B.1: Complete survey on Text2Storyline’s evaluation
FIGURE B.1: Complete survey on Text2Storyline’s evaluation
FIGURE B.1: Complete survey on Text2Storyline’s evaluation
Figure B.1: Complete survey on Text2Storyline’s evaluation
Figure B.1: Complete survey on Text2Storyline’s evaluation
Bibliography


American Chapter of the Association for Computational Linguistics (NAACL), 2021. [Cited on page 22.]


