OmniDoc

Henrique José Antunes Fernandes

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisor: Prof. Daniel Jorge Viegas Gonçalves

Examination Committee

Chairperson: Prof. Pedro Tiago Gonçalves Monteiro
Supervisor: Prof. Daniel Jorge Viegas Gonçalves
Member of the Committee: Prof. Sandra Pereira Gama

June 2022

Acknowledgments

I would like to thank my parents for their friendship, encouragement and caring over all these years and without whom this project would not be possible.

I would also like to thank my grandparents for their understanding and support throughout all these years.

To my sister, I would like to thank you for your constant support and caring.

To Andreia Jerónimo, for all her love, support, wisdom and patience that always helped me to keep going.

To my friends, thank you for all the help and laughs.

I would also like to thank my dissertation supervisor Prof.Daniel Gonçalves for his support, encouragement and sharing of knowledge that has made this Thesis possible and for trusting me when accepting me for this dissertation.

Lastly, I would like to thank Dr. Clarisse Gonçalves and all the judges from the Portuguese Supreme Court that participated in the usability testing for all the help and availability.

To each and every one of you – Thank you.

Abstract

The Portuguese Supreme Court is the most important part of the Portuguese judicial hierarchy. It is therefore essential that it operates efficiently at the highest level in terms of its tasks and the quality of its decisions. In order to improve the judges' workflow, the IRIS project was launched in cooperation with INESC-ID. The objective of this project is to improve the efficiency of the decision-making process and the publication of decisions to other judges of the Portuguese Supreme Court and the general public. The IRIS project is divided into three different parts: Anonymisation, Summary and Analysis. This dissertation is part of the latter. In order to make it easier for judges to search and navigate through the numerous documents available, it is necessary to create a tool that resembles a digital library that locates relevant documents in an efficient way.

To this end, a web-based system was developed using HTML, CSS, JavaScript and ElasticSearch, which allows users to search for and navigate through different decisions from the various existing courts based on search terms and other metrics such as author or date. The developed solution was tested with Portuguese Supreme Court judges, allowing for feedback and suggestions, which were subsequently implemented. In this usability test, the application had a *SUS* score of 80,75.

Keywords

Analysis; Portuguese Supreme Court; IRIS Project; Web-based system.

Resumo

O Supremo Tribunal de Justiça é a parte mais importante da hierarquia da justiça portuguesa. Isto significa que é imperativo que trabalhe ao mais alto nível de eficiência no desempenho das suas tarefas e na qualidade das suas decisões. Para melhorar o fluxo de trabalho dos seus juízes, foi criado o projeto IRIS em parceria com o INESC-ID. Este projeto consiste em fornecer suporte para análise de processos e publicação de decisões para serem pesquisados por magistrados ou pelo público em geral. O projeto IRIS está dividido em três partes: Anonimização, Sumarização e Análise. Esta dissertação está inserida no último ponto. De maneira a facilitar aos juízes a pesquisa e a navegação pelos numerosos documentos disponíveis, é necessário criar uma ferramenta semelhante a uma biblioteca digital que localize documentos relevantes de forma eficiente.

Foi, então, desenvolvido um website utilizando HTML, CSS, JavaScript e ElasticSearch que permite aos utilizadores pesquisarem por diferentes decisões dos vários tribunais existentes usando termos de pesquisa e outras métricas, como por exemplo o autor ou data e navegar pelos diferentes documentos. A solução desenvolvida foi testada com juízes do Supremo Tribunal de Justiça em que foi possível receber feedback e sugestões que foram aplicadas posteriormente. Nesta avaliação com utilizadores, a aplicação obteve uma pontuação de 80.75 na escala de usabilidade do sistema.

Palavras Chave

Análise; Projeto IRIS; Supremo Tribunal de Justiça; Website.

Contents

1 Introduction **1**

1.1 Objectives 4

1.2 Contributions 4

1.3 Document Organization 5

2 State of the Art **7**

2.1 Legal Text Visualization 9

2.2 Text Visualization 12

2.3 Document Collection Visualization 18

2.4 Digital Libraries for Justice Documents 24

2.4.1 ECLI 24

2.4.2 DGSI 25

2.4.3 Diário da República Eletrónico 26

2.5 Discussion 27

3 Designing the Solution **29**

3.1 Requirements 31

3.2 Pre-Meetings' Approach 32

3.3 Post-Meetings' Approach 35

3.4 Final Approach 37

4 Implementation **39**

4.1 Solution 41

4.1.1 Perform a search 41

4.1.2 Results 42

4.1.2.A Metadata 42

4.1.2.B Visualisation 43

4.1.3 Filters 44

4.1.3.A Drop-down Filters 45

4.1.3.B Date filters 45

4.1.4 Document's page . . . 47
4.2 Architecture . . . 47
4.2.1 Back-End . . . 48
4.2.2 Front-End . . . 48
4.3 Implementation . . . 48
4.3.1 Backend . . . 49
4.3.1.A GET - /api/ . . . 50
4.3.1.B GET - /api/open/:ECLI . . . 50
4.3.2 Database . . . 50
4.3.3 Front-End . . . 54
4.3.3.A Responsive Design . . . 56

5 Evaluation **59**
5.1 User Evaluation . . . 61
5.2 Results . . . 62
5.2.1 Time to Complete Tasks and Judges' Answers . . . 62
5.2.2 User Comments . . . 65
5.3 Changes to the Interface . . . 66
5.3.1 Order from Oldest to Newest . . . 66
5.3.2 Tooltips for Searching and Visualization . . . 67
5.3.3 Pagination . . . 67
5.3.4 Database Index . . . 67
5.4 Discussion . . . 68

6 Conclusion **69**
6.1 System Limitations and Future Work . . . 71

Bibliography **73**

A Single Ease Question (SEQ) **77**

B System Usability Scale (SUS) **79**

List of Figures

2.1 Graph with the articles, entities and concepts found in a specific document. [1] 9

2.2 Graph after clicking in node ”Artigo 17”. The concepts and entities surrounding the center node can also be clicked to update this visualization. [2] 10

2.3 *Parallel Tag Clouds* for court cases. Each column represents a court. [3] 10

2.4 *Lifelines* overview for ”Apple vs Microsoft and Hewlett Packard”. [4] 11

2.5 Factual History for the case mentioned above. [4] 12

2.6 *TileBars* visualization for legal terms. [5] 12

2.7 Scrollbar interface with the words highlighted. [6] 13

2.8 Example of *Word Tree* in *Romeo and Juliet*. [7] 13

2.9 Example of *PhraseNets* in which supernodes were created. [8] 14

2.10 Arc Diagram example for substring ”1234567”. [9] 15

2.11 WordBridge interface with word clouds and their links. [10] 15

2.12 *TextArc* example. 16

2.13 Word highlighting feature of the *AzAA* portal. [11] 16

2.14 *LiteratureVis* being used with different measures. [12] 17

2.15 *Document Cards* displayed on different screens. Image taken from [13] and cropped. . . 18

2.16 *Document Cards* display for a big screen. [13] 18

2.17 Interface showing documents as yellow crosses with keywords above. [14] 19

2.18 *Cartolabe* interface showing document heatmap, labels and search feature. [15] 20

2.19 *Textonic* interface. [16] 20

2.20 *STREAMIT* interface. [17] 21

2.21 *Jigsaw*’s List View and Document Cluster View. [18] 22

2.22 *Contexter* interface. [19] 22

2.23 *Finavistory*’s interface. Image taken from [20] and then cropped. 23

2.24 *EventRiver* interface with dual labels above the bubbles. [21] 24

2.25 ECLI website showing a search for a specific author. 25

2.26 DGSI page for decisions of the Portuguese Supreme Court. 26
2.27 Diário da República Eletrónico page for Portuguese jurisprudence. 27

3.1 First iteration of the interface after adding an annotation. 32
3.2 First iteration of the interface showing where a word appears in the text. 33
3.3 First Iteration's interface. 34
3.4 Tech demo for text manipulation and interaction. 35
3.5 Second iteration search box and result. 36
3.6 Different methods of displaying the vertical bars. 36
3.7 Excerpt of the text with the searched words. 37
3.8 Results page of the final iteration. 37
3.9 Excerpt of the text with the searched words of the final iteration. 38

4.1 Final solution's initial page. 41
4.2 Search for *Cláusula Verbal* using different methods. 42
4.3 Search results for ”*Cláusula Verbal*” 42
4.4 Search relevance. 43
4.5 List of descriptors of a document when searching for ”*Cláusula Verbal*”. 43
4.6 Visualization of a document when searching for ”*Denúncia de Contrato*”. 43
4.7 Excerpt of the text when searching for ”*Denúncia de Contrato*”. The hovered bar is blue, meaning the word that is highlighted is ”*Contrato*”. 44
4.8 Filtered author list with document count. 44
4.9 Court Filter. 45
4.10 Selected documents from January 1 2012 and June 3 2013. 46
4.11 Bar chart with the number of documents for the search term ”Contrato”. 46
4.12 Slider range for years between 2008 and 2018. 47
4.13 Document page for process 132/2002.L1-8. 47
4.14 Digital library's architecture. 48
4.15 Digital library's server components. 50
4.16 Database mappings. 51
4.17 Example of a query. 52
4.18 Example of an aggregation request used in this application. 53
4.19 Initial approach of the document count implementation. 54
4.20 Different layouts according to the screen width. 57

5.1 Time to complete the tasks. 63
5.2 Number of right and wrong answers for each task. 63

5.3 SEQ answers for each task. 65
5.4 SUS answers for each question. 66
5.5 Tooltip explaining how to use the visualization. 67

A.1 SEQ questionnaire 78

List of Tables

5.1 Median, average and standard deviation for the time taken to complete the different tasks (in seconds). 63

1

Introduction

Contents

1.1 Objectives 4

1.2 Contributions 4

1.3 Document Organization 5

The Portuguese Supreme Court is the most important body in the Portuguese judicial hierarchy. In order to fulfil its duties, it must be at the highest level in terms of efficiency of its tasks and the quality of its decisions. As for the analysis of court proceedings, according to the judges of the Portuguese Supreme Court, some of them prefer to do it on paper and although there are tools to perform this task on the computer, they are very outdated making this process very ineffective. In cooperation with INESC-ID, the IRIS project was launched to improve the efficiency of the decision-making process and the publication of decisions to other judges of the Portuguese Supreme Court and the general public by updating digitisation techniques.

When a case is created, it is forwarded to a First Instance court. In this court, after analysing the evidence presented, a decision is made and a judgement is written. The lawyers involved can then appeal this decision if they disagree with it. If they do, then the case goes to a Second Instance court, the court of appeals. There, the appeal is examined and a new judgement is written stating whether the appeal is upheld or not. However, if the lawyers want to appeal against this new decision, the case goes to the Portuguese Supreme Court. However, only a few cases that meet a number of requirements reach this final stage. The Portuguese Supreme Court's decision is final. In order for the Portuguese Supreme Court's judges to write the new and final judgement, they must access the documents of the previous decisions and search and navigate the existing jurisprudence. In addition, the new decision must be made available to other judges so that it can be used in other cases.

Judges also have access to a number of websites with databases populated with various documents from different Portuguese courts. On these websites, they can search for specific terms or filter the results according to a number of criteria such as the date of the document or its author. The results are also ordered according to their relevance to the search or chronologically. The most used websites are DGSI[1] and ECLI[2]. These two websites offer the most complete searches for the magistrates' requirements. However, they have several flaws that judges would like to see addressed. DGSI is said to be very efficient in displaying results, but it is hard to search and filter the results. ECLI is the complete opposite of DGSI. This website has a very appealing interface and searching for documents is easy, but it takes a very long time to display results, so judges prefer to use other websites.

To achieve the above goal of improving the efficiency of the decision-making process, the IRIS project is divided into three different parts: Anonymisation, Summary and Analysis. This dissertation is part of the latter. In order to make it easier for judges to search and navigate through the numerous documents available, it is necessary to create a tool that resembles a digital library that locates relevant documents in an efficient and fast way, so that the process of decision-making is smoother.

[1] http://www.dgsi.pt/

[2] https://jurisprudencia.csm.org.pt/

1.1 Objectives

The aim of this dissertation is to **create an efficient and effective tool to find relevant decisions for judges when deciding on new cases**. To achieve this, a **web technology based tool was designed and implemented**. It provides users with a fast and efficient experience when searching for legal documents by using a powerful search engine tool in ElasticSearch[3] to quickly retrieve results and display a brief summary of each document in an intuitive way in order to help users understand the content and relevance to the search being performed. The user can enter a specific term and the solution will access a database filled with documents from several Portuguese courts and will find the documents in which the term or set of terms occur in several fields. It is also possible to filter the obtained results by specific metrics, e.g. author or descriptor. All versions and updates of the digital library development are stored in a GitHub repository[4].

In order to test the system, judges from the Portuguese Supreme Court performed usability tests. These tests allow us to determine whether the tool meets the intended objectives. Ten judges participated in these tests, in which they were asked to solve a series of seven tasks covering most of the functionalities of the digital library. Based on the results collected and analysed, as well as the feedback from the participants of the user tests, the interface is considered to be easy to learn and user-friendly, and is a helpful tool for the judges' work. The implementation of the digital library as well as the testing phase are described in the next chapters.

1.2 Contributions

The created solution is helpful in searching and navigating the documents of the various Portuguese courts. The system is easy to use and learn and facilitates the judges' workflow. In order to understand how judges perform searches using the existing tools and to draw up a list of requirements for the developed solutions, meetings were held with several judges of the Portuguese Supreme Court.

As it was said before, the IRIS project was created to improve the decision-making process as well as the publication of decisions to other judges of the Portuguese Supreme Court and the general public. The project is divided in three parts: Anonymisation, Summary and Analysis. This dissertation is included in the latter in which its objective is to improve the search and navigation techniques. By developing the digital library, the objective of improving the decision-making process was achieved.

[3] https://www.elastic.co/pt/

[4] https://github.com/HenrySmash/SAMA-IRIS

1.3 Document Organization

This document is divided into six chapters and several appendices. These chapters are Chapter 1 - Introduction, which presents the scope of the work and its involvement in the IRIS project as well as its objectives, Chapter 2 - State of the Art, which presents the research of relevant works for the design and development of the solution, Chapter 3 - Approach, which contains the decision regarding the approach for the development of the solution as well as its architecture, Chapter 4 - Implementation consists of the description of the implementation of the solution, Chapter 5 - Evaluation contains the description of the usability tests and the discussion of the results and finally Chapter 6 - Conclusion contains the final thoughts and discussion. The appendices consist of the forms used in the usability tests. These appendices are appendix A and appendix B.

2

State of the Art

Contents

2.1 Legal Text Visualization 9

2.2 Text Visualization 12

2.3 Document Collection Visualization 18

2.4 Digital Libraries for Justice Documents 24

2.5 Discussion 27

In this chapter, it is presented the research regarding relevant topics to the development of the interface. Since this project will be developed for the Portuguese Supreme Court, it makes sense to research tools and interfaces that allow to display and interact with legal information in an attractive and intuitive manner. The first topic researched was legal text visualization. However, due to the small sample size of interfaces found, it was necessary to widen the scope of the research. For this, text visualizations as well as document collection visualizations were also considered.

Also, since the application resembles a digital library, it is pertinent to research how these are developed and how users perform tasks on them. It is also relevant to understand how the results are displayed to the users and how they can filter such results. This research is focused on digital libraries for justice documents due to the scope of the project.

2.1 Legal Text Visualization

Since this work involves displaying legal documents, it is crucial that we understand what sort of work was done in this area. Below are some examples of tools that use different visualizations to explore and navigate documents to obtain or extract information from them. In the papers presented there are also described various interaction techniques.

The work performed by Carvalho and Barbosa [2] uses multiple visualizations to navigate and analyze documents. The first visualization is a graph as seen on figure 2.1. It is used to display the hierarchy, such as sections, chapters, etc. of the documents.

The navigation and analysis of the document can also be done by semantic browsing. The semantic information is retrieved using natural language processing techniques and is represented by a graph with connected terms. It possesses nodes in the middle that represent articles, on the left represent concepts found in the text and on the right that represent entities.

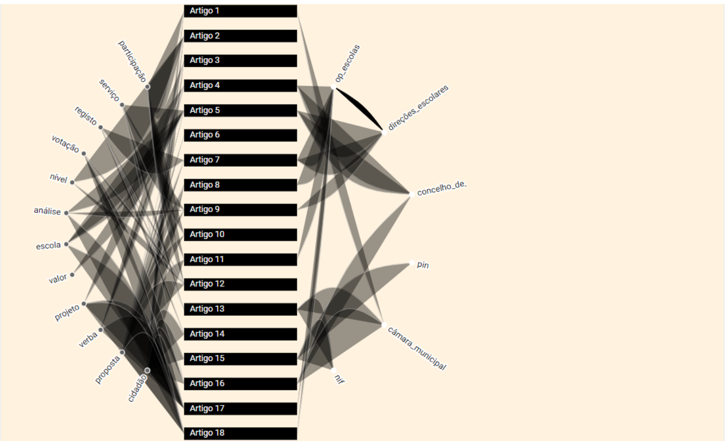


Figure 2.1: Graph with the articles, entities and concepts found in a specific document. [1]

The user can select a specific node and the terms associated with it are displayed with arcs representing connections as seen on figure 2.2. A concept or an entity can also be selected. When an element that is connected with an arc is selected, the visualization is updated, showing it in the middle with its connections.

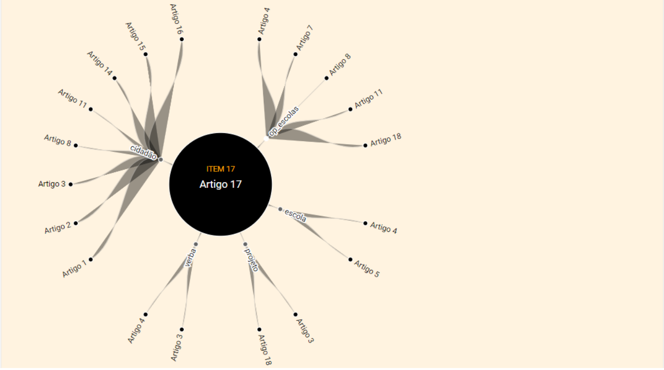
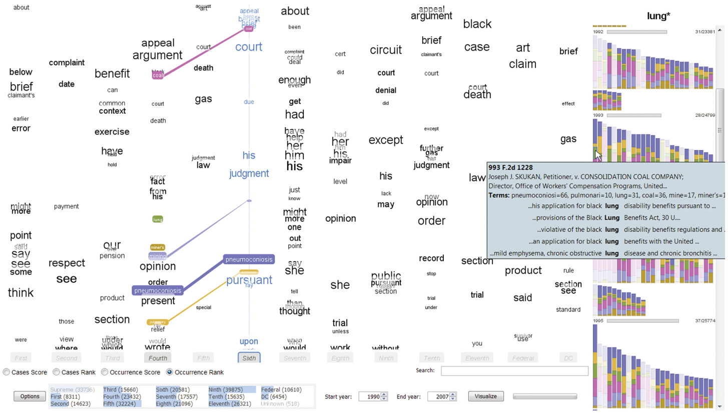


Figure 2.2: Graph after clicking in node "Artigo 17". The concepts and entities surrounding the center node can also be clicked to update this visualization. [2]

A final visualization is a word cloud that is used to understand the most referred topics in the document. The font size is encoded according to the word frequency. This work is interesting in the context of our work since it allows to summarize the content of a document and find connections between entities and terms among the various sections or chapters. A downside is the lost of information that the algorithm may find unimportant.

The *Parallel Tag Clouds* [3] visualization was created to compare large amounts of text. It combines two different approaches: layout techniques from parallel coordinates and word-sizing. Keywords are organized alphabetically into columns, each representing a distinct topic and font size is used to encode frequency. If there are common words among different columns, they are connected using nearest-neighbor edges.

It shows a bar chart (only one word selected) or a stacked bar chart with different colors (multiple words selected) representing the documents where those words appeared. The height of the bars is used to encode the number of occurrences of the term in each document.



993 F.2d 1228
Joseph J. SKUKAN, Petitioner, v. CONSOLIDATION COAL COMPANY;
Director, Office of Workers' Compensation Programs, United...
Terms: pneumoconiosi=66, pulmonari=10, lung=31, coal=36, mine=17, miner's=1
...his application for black **lung** disability benefits pursuant to ...
...provisions of the Black **Lung** Benefits Act, 30 U....
...violative of the black **lung** disability benefits regulations and ...
...an application for black **lung** benefits with the United ...
...mild emphysema, chronic obstructive **lung** disease and chronic bronchitis ...

Figure 2.3: *Parallel Tag Clouds* for court cases. Each column represents a court. [3]

In the figure above, we see an example of this visualization for court cases. Each column represents a court with the keywords associated with them. We can see the stacked bar chart with different colors representing the keywords selected.

Similarly to the previously described work, this is also useful to summarize the content of a document and find connections between terms. It also allows to show the document where the term is located. However, it can become difficult to understand if there are a lot of information.

Lifelines [4] is a visualization used to show relationships of temporal events. It organizes and groups events by topics. When a topic is created, every article or case related to it is aggregated, smoothing the process of accessing documents.

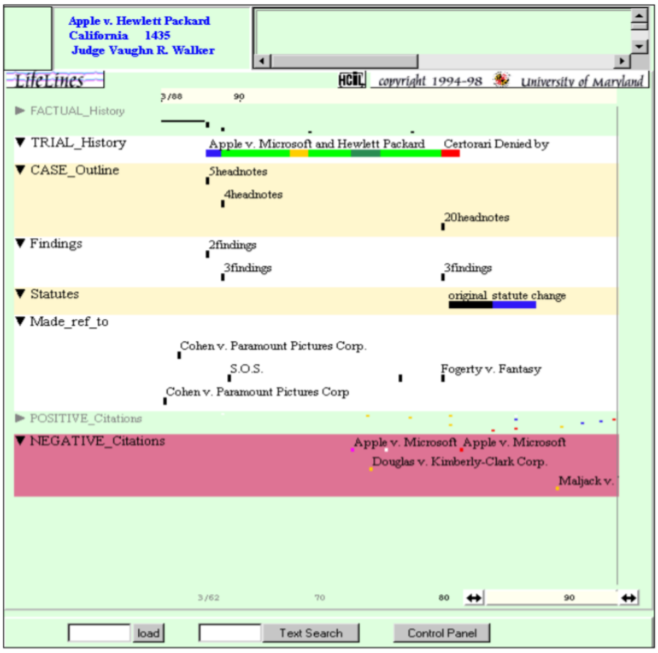
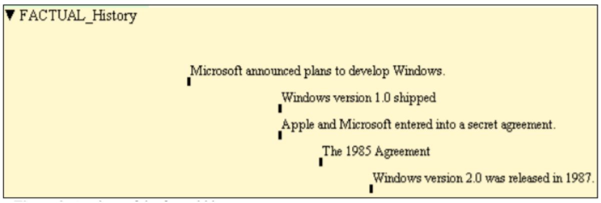


Figure 2.4: *Lifelines* overview for ”Apple vs Microsoft and Hewlett Packard”. [4]

This tool allows interaction by zooming in and out or clicking on the different topics and events. The relationships can be shown in two ways: factual history that includes events leading to the case, showing who, what and when it happened and trial history that includes the events during the course of the case. In the figure below, we can see the factual history for the ”Apple vs Microsoft and Hewlett Packard” case.

This tool is useful to find and access documents in a temporal context by summarizing the events. The downside is that we don’t know the contents of every document.


▼ FACTUAL_History

Microsoft announced plans to develop Windows.

Windows version 1.0 shipped

Apple and Microsoft entered into a secret agreement.

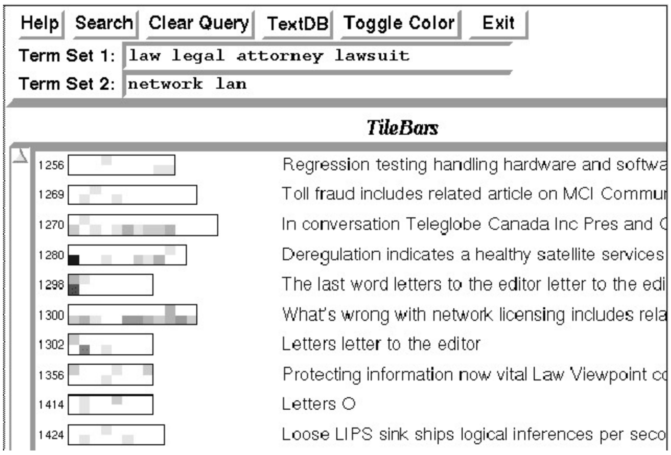
The 1985 Agreement

Windows version 2.0 was released in 1987.

Figure 2.5: Factual History for the case mentioned above. [4]

TileBars [5] is a visualization created to smooth the process of document retrieval based on a query. For this, it uses different ways to encode information. The user inputs a set of desired terms to be found in a collection and in the window is displayed the tiles representing the documents and the title as well as the initial words of said documents (on the right).

Inside the tiles there are squares representing the text segments where those terms were found. The darkness of the square is used to encode the frequency. In the figure below, we see a query of terms for legal documents in which the top row indicates the frequencies from Term Set 1 and the bottom row corresponds to Term Set 2.



Term Set 1: law legal attorney lawsuit

Term Set 2: network lan

1256 Regression testing handling hardware and softwa

1269 Toll fraud includes related article on MCI Commu

1270 In conversation Teleglobe Canada Inc Pres and C

1280 Deregulation indicates a healthy satellite services

1298 The last word letters to the editor letter to the edi

1300 What's wrong with network licensing includes rela

1302 Letters letter to the editor

1356 Protecting information now vital Law Viewpoint co

1414 Letters O

1424 Loose LIPS sink ships logical inferences per seco

Figure 2.6: *TileBars* visualization for legal terms. [5]

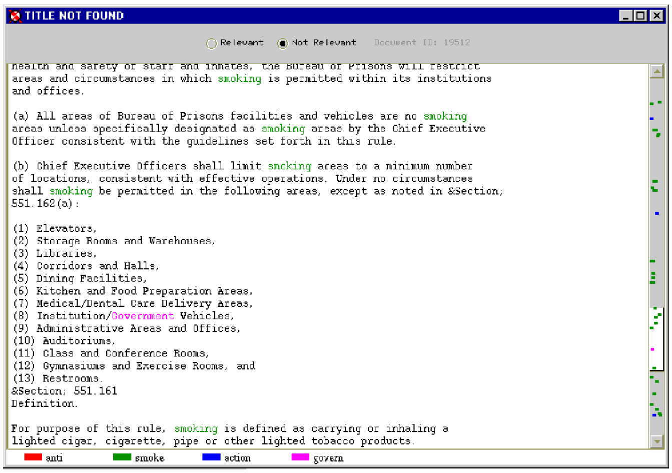
This work is very useful since it collects and displays a part of every document related to the searched terms and provides the frequency of the keywords. The downside is the fact that is impossible to analyze the content of the documents.

2.2 Text Visualization

As it was said before, due to the lack of number of legal text visualizations, it was necessary to expand the research. There are a lot of examples of useful text visualizations to help analyze short texts or long ones, ranging from simple interfaces like the work developed by Byrd [6] or others more complex.

Throughout the Internet, there is a multitude of interfaces and tools that use different visualizations to represent and analyze text [22]. These interfaces and visualizations use various techniques, such as word tags (e.g *Wordle* [23]) that combined with other features help users achieve their goals. These techniques can range from word highlighting to keyword searching, etc. Some visualizations also allow users to interact with the documents by adding annotations or comments to certain parts.

An example of the first feature is a simple tool [6] that was developed in which the user can input words and a window is displayed with the document content with the words highlighted. On the right side of the window, as seen on figure 2.7, little dots represent the highlighted keywords. Different words can be input, having different colors assigned to them. This is a simple example, however is very useful to find keywords in documents. However, the navigation is quite difficult in extensive documents.



```
health and safety of staff and inmates, the Bureau of Prisons will restrict
areas and circumstances in which smoking is permitted within its institutions
and offices.

(a) All areas of Bureau of Prisons facilities and vehicles are no smoking
areas unless specifically designated as smoking areas by the Chief Executive
Officer consistent with the guidelines set forth in this rule.

(b) Chief Executive Officers shall limit smoking areas to a minimum number
of locations, consistent with effective operations. Under no circumstances
shall smoking be permitted in the following areas, except as noted in &Section;
551.162(a):

(1) Elevators,
(2) Storage Rooms and Warehouses,
(3) Libraries,
(4) Corridors and Halls,
(5) Dining Facilities,
(6) Kitchen and Food Preparation Areas,
(7) Medical/Dental Care Delivery Areas,
(8) Institution/Government Vehicles,
(9) Administrative Areas and Offices,
(10) Auditoriums,
(11) Class and Conference Rooms,
(12) Gymnasiums and Exercise Rooms, and
(13) Restrooms.
&Section; 551.161
Definition.

For purpose of this rule, smoking is defined as carrying or inhaling a
lighted cigar, cigarette, pipe or other lighted tobacco products.
```

🟥 anti 🟩 smoke 🟦 action 🟪 govern

Figure 2.7: Scrollbar interface with the words highlighted. [6]

Next, several tools used to navigate and analyze text and their techniques will be presented as well as their advantages and disadvantages.

The *Word Tree* [7] visualization was created to analyze and deal with repetition in text documents. The different words of these documents are arranged in a tree-like structure according to a search term which the user is able to interact with. The font size to represent the number of times a word or phrase appears. The branches of the tree continue until a unique phrase used exactly once is found.

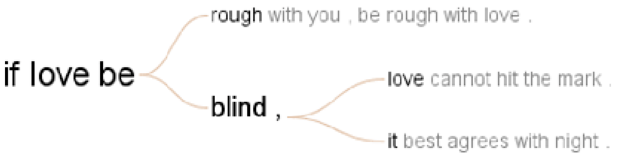

if love be
rough with you , be rough with love .
blind ,
love cannot hit the mark .
it best agrees with night .

Figure 2.8: Example of *Word Tree* in *Romeo and Juliet*. [7]

In the figure above we can see an example of how the *Word Tree* is displayed. "If love be" is considered the root node and it has two children, "rough" and "blind". These two children have other children associated with them and so on.

This work is useful since it shows the most frequent connections between terms, however it can become too difficult to understand for large documents.

Phrase Nets [8] also uses words as nodes, displaying a graph whose edges indicate a relation between two words. The relations may be defined either at the syntactic or lexical level.

To create the network, nodes are defined to be a subset of words occurring in the text, and edges to represent certain relations between these words. There are many ways to choose these node and edge sets. For example, one method is to consider every word as a node and link two terms in immediate succession in the text without punctuation between them.

A filtering process is then performed. The first step is to remove the stopwords, then the remaining words are ranked by relevance. There are several relevant measures. One is the degree in the network and another is the frequency in the text.

An ”edge compression” can be performed to reduce the number of degree-one neighbours that are topologically similar by compressing them into a supernode as shown in figure 2.9.

The advantage of this work resides in the fact that the documents are summarized regarding their keywords. However, for large documents, it can be difficult to read, since the keywords that are not that frequent tend to be hidden.

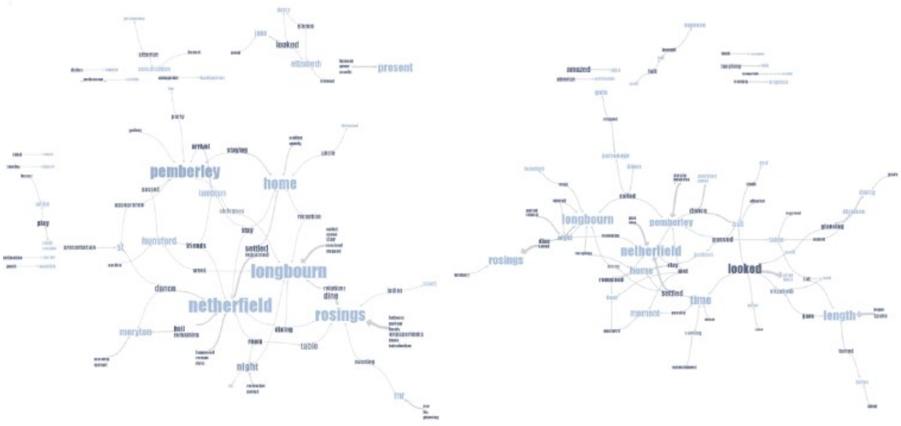


Figure 2.9: Example of *PhraseNets* in which supernodes were created. [8]

Arc diagrams [9] uses a pattern-matching algorithm to find repeated substrings and then represents them as arcs. Two substrings match if they contain the same sequence of symbols, do not intersect and there is no other identical substring that its beginning is the same as the original substrings’ beginning. For example, in the sequence “123a123”, the two “123” substrings form a maximal matching pair, but the two “12” substrings do not. In the figure below we can see a representation of an arc diagram for the substring ”1234567”.

This work is useful to find links between the same entities in different parts of the documents. However, it can be difficult to read if there are a large number of them.

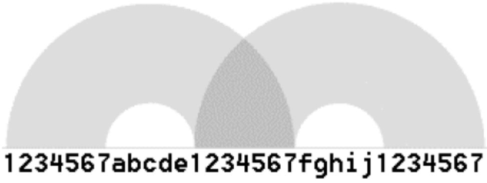


Figure 2.10: Arc Diagram example for substring ”1234567”. [9]

WordBridge [10] was created to analyze relationships between entities in text displaying a graph where both nodes and edges are associated with a set of keywords.

The *WordBridge* technique was implemented in the context of *Apropos*, a web-based text analysis application designed to show relationships between entities extracted from a text corpus. After loading a database, an entity list is created and populated with all of the entities within a dataset. Users can select any of the entities and add them to a canvas. On figure 2.11, we can see various word clouds with different keywords and the links between them.

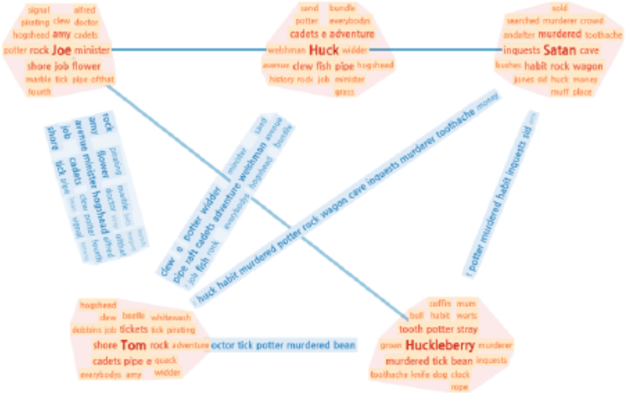


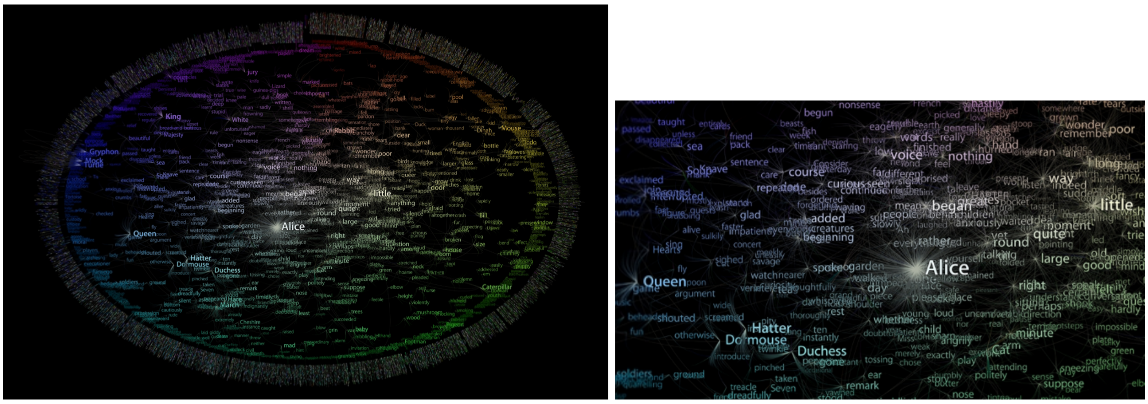
Figure 2.11: WordBridge interface with word clouds and their links. [10]

This technique is useful in the context of this work. For example, a user can search for specific articles of a certain author while finding others that had co-authorship of the same person. However, if there are several connections, it can become confusing.

TextArc [24] is a visualization created to display text and allows users to analyze certain features in an interactive way. This visualization displays keywords in an arc-like shape and uses the word size to demonstrate how frequent a word is. The user can hover over the keywords and the visualization will show the location in the document where that word is referenced.

As we can see in the image below, the words ”Alice”, ”Queen” and ”little” are some of the most common words. These connections show where the word has been mentioned. The user can then click on a specific word from the entire arc and see its location on the text.

This work is useful to summarize documents to find which keywords are most frequent and their location. However, and as seen in the figure above, it can become too difficult to read for extensive documents.



(a) *TextArc* for "Alice in Wonderland" book. [24]

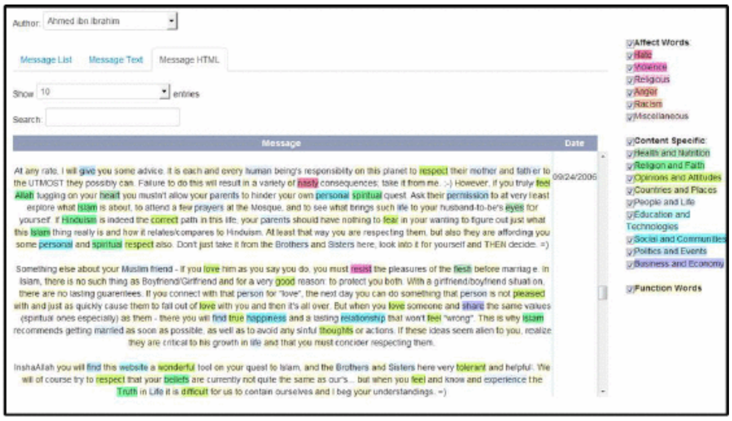
(b) "Alice", "Queen" and "little" are some of the most common words. [24]

Figure 2.12: *TextArc* example.

Another tool was created in the context of cybersecurity for authorship analysis. The *AzAA* [11] portal uses crawlers to extract messages from different Web forums and they can be analyzed in two perspectives: author-level and message-level.

The first perspective is used to identify which authors use specific stylometric features the most. Users can choose a simple summary view or a more detailed heatmap view. From the author-perspective, the user can compare the authorship of two people using a radar chart to summarize authorship differences and similarities.

There is also a feature to analyze the raw message contents in their original form or plain text form. The user can search for specific keywords that will appear highlighted in the message content as seen on figure 2.13.



At any rate, I will give you some advice. It is each and every human being's responsibility on this planet to respect their mother and fath er to the UTMOST they possibly can. Failure to do this will result in a variety of nasty consequences; take it from me. ;-) However, if you truly feel Allah tugging on your heart you mustn't allow your parents to hinder your own personal spiritual quest. Ask their permission to at very least explore what Islam is about, to attend a few prayers at the Mosque, and to see what brings such life to your husband-to-be's eyes for yourself. If Hinduism is indeed the correct path in this life, your parents should have nothing to fear in your wanting to figure out just what this Islam thing really is and how it relates/compares to Hinduism. At least that way you are respecting them, but also they are affording you some personal and spiritual respect also. Don't just take it from the Brothers and Sisters here, look into it for yourself and THEN decide. =)

09/24/2006

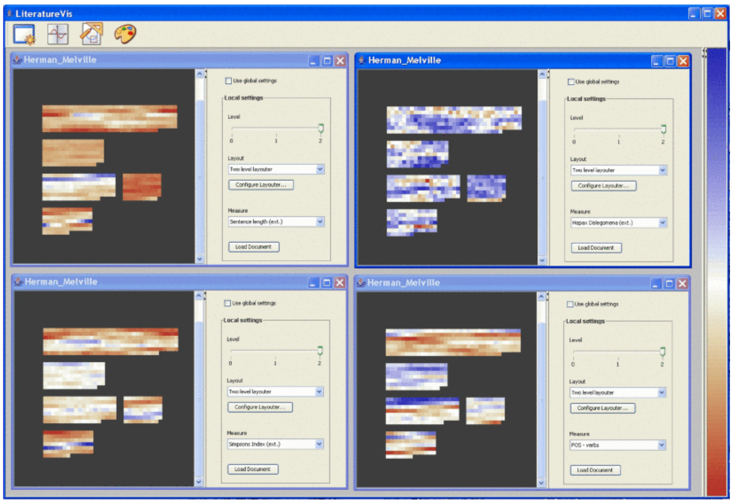
Something else about your Muslim friend - If you love him as you say you do, you must resist the pleasures of the flesh before marriag e. In Islam, there is no such thing as Boyfriend/Girlfriend and for a very good reason: to protect you both. With a girlfriend/boyfriend situat on, there are no lasting guarentees. If you connect with that person for "love", the next day you can do something that person is not pleased with and just as quickly cause them to fall out of love with you and then it's all over. But when you love someone and share the same values (spiritual ones especially) as them - there you will find true happiness and a lasting relationship that won't feel "wrong". This is why Islam recommends getting married as soon as possible, as well as to avoid any sinful thoughts or actions. If these ideas seem alien to you, realize they are critical to his growth in life and that you must concider respecting them.

InshaAllah you will find this website a wonderful tool on your quest to Islam, and the Brothers and Sisters here very tolerant and helpful. We will of course try to respect that your beliefs are currently not quite the same as our's... but when you feel and know and experience the Truth in Life it is difficult for us to contain ourselves and I beg your understandings. =)

Figure 2.13: Word highlighting feature of the *AzAA* portal. [11]

In the context of this work, this tool is useful if we plan to compare two different authors to analyze or find documents related to them. We can also take advantage of the highlighting words feature to find important words within the document content.

A tool was created to analyze text using visual fingerprinting. *LiteratureVis* [12] allows to load multiple texts to compare them. To create the visualizations, the user can select the measures to be used. These measures include sentence length, Simpsons Index, parts of speech, etc. The color of the pixels is used to encode the information received from the algorithms. In the figure below we can see an example of this tool being used.



□ Use global settings
Local settings
Level: 2
Layout: Two level layouter
Measure: Sentence length (ext.)

□ Use global settings
Local settings
Level: 2
Layout: Two level layouter
Measure: Hapax Dislegomena (ext.)

□ Use global settings
Local settings
Level: 2
Layout: Two level layouter
Measure: Simpsons Index (ext.)

□ Use global settings
Local settings
Level: 2
Layout: Two level layouter
Measure: POS - verbs

Figure 2.14: *LiteratureVis* being used with different measures. [12]

This work is useful to analyze several features of the documents, such as complexity of the text as well as syntax. Since the information is encoded using colored pixels, we do not have access to the text.

To provide access to large document collections, *Document Cards* [13] was created. This interface was designed to visualize documents on both small screens such as handheld devices and big screens, for example, a desktop. In the first one, it displays a single card while on a bigger screen it represents a collection of documents.

The card consists of the title of the document, a collection of images, and keywords. To extract the keywords, their frequency and a base form reduction of words algorithm are used. A max of four images are shown and the number of keywords depends on the space that the images occupy. The font size of each word indicates the term weight.

This tool could be of use to this work since it summarizes and displays the content of a document in an attractive way while keeping the important information relevant for the user. The summarization aspect of this tool helps anyone who interacts with it to understand the documents without having to read them completely. On figure 2.15, we can perceive how a document collection is displayed on different screens.

Figure 2.15: *Document Cards* displayed on different screens. Image taken from [13] and cropped.

2.3 Document Collection Visualization

Usually, documents are apart of large document collections in which the user has to navigate and find the desired text. This procedure can become tedious and very hard to manage, so visualizations are created to smooth the process.

Similar to individual document visualization, there are a plethora of visualizations that can be used to represent these collections. The most common visualizations found were the representation of the documents as clusters.

Most of the visualizations found with this research used natural language processing techniques such as bag of words [25] and dimensionality reduction [26] algorithms like Singular Value Decomposition [27] or Latent Semantic Indexing [28] to discover the most used keywords and others similar to them in the collections and project them onto a two-dimensional plane.

The last tool described in the previous section can be included here as well since it deals with multiple documents. *Document Cards* organizes its documents in a grid-like manner in which the user can hover over each one and analyze if they are of interest or not. This tool provides various ways to interact with its elements such as showing the abstract of a document by hovering over the card or highlighting words on the text by clicking a certain term. On figure 2.16, the grid of documents is presented similarly to the way it is displayed on a desktop or laptop screen.

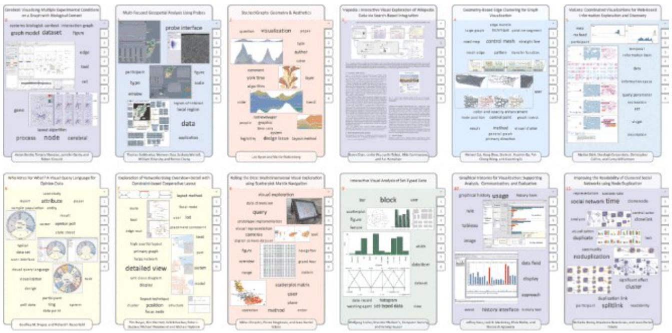


Figure 2.16: *Document Cards* display for a big screen. [13]

In [14] and *Cartolabe* [15], the Latent Semantic Indexing algorithm is used to extract keywords. The documents are then positioned onto a two-dimensional plane.

The documents are represented as yellow crosses on the plane and after they are mapped onto it, two techniques are used to facilitate the visualization: landscape generation and keywords. The first one is achieved by using the density of points (the lighter the color, the higher the density) as seen in figure 2.17. For the second one, each point is assigned a set of most frequent keywords with that document.

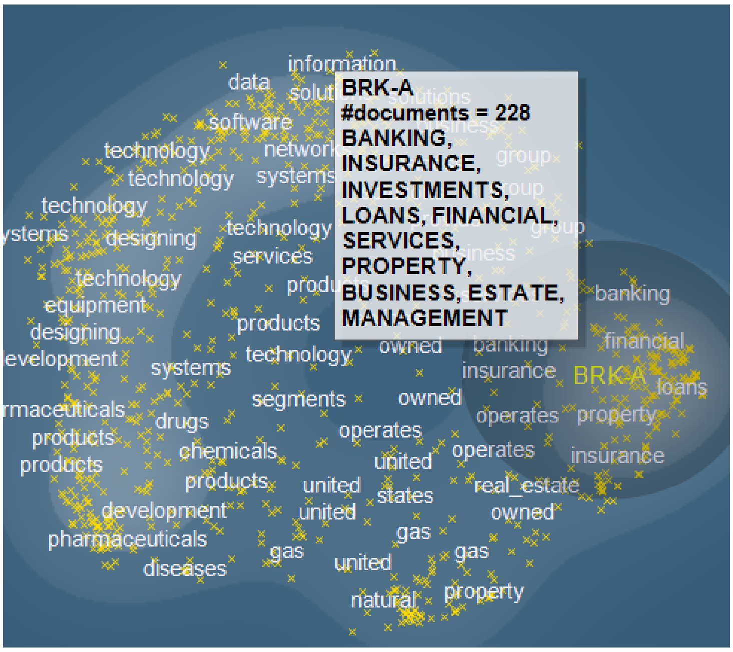

BRK-A
#documents = 228
BANKING, INSURANCE, INVESTMENTS, LOANS, FINANCIAL, SERVICES, PROPERTY, BUSINESS, ESTATE, MANAGEMENT

Figure 2.17: Interface showing documents as yellow crosses with keywords above. [14]

To avoid the clutter of information, the plane only shows the most frequently used words for the area that is being visualized. The user can zoom-in to analyze keywords relative to that area in more detail and when the mouse is moved around, a list of the most common terms is displayed. When a user clicks on a certain cross representing a document, more information is shown on the left side of the interface.

Cartolabe maps the documents according to topic contents and provides a general overview of the document collection. The main entities are authors (red dots) and documents (blue dots) and are both represented as a heatmap with denser areas being brighter as seen in figure 2.18. Labels are displayed on top of these density maps to give contextual information. Yellow labels characterize thematic regions and white labels indicate the most important entities in the map, depending on the zoom level.

Users can select an item on the map and read the associated details on the search box, or search for an item based on its name and see where it is located on the map. Another feature is to retrieve the entities (articles or authors) close to a given entity.

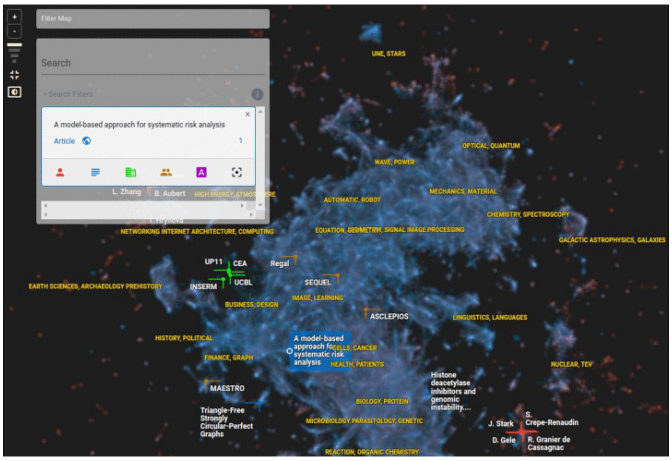

Filter Map
Search
Search Filters
A model-based approach for systematic risk analysis
Article
1
L. Zhang
B. Aubert
LINE, STARS
OPTICAL, QUANTUM
WAVE, POWER
MECHANICS, MATERIAL
AUTOMATIC, ROBOT
CHEMISTRY, SPECTROSCOPY
GALACTIC ASTROPHYSICS, GALAXIES
EQUATION, GEOMETRY, SIGNAL IMAGE PROCESSING
NETWORKING INTERNET ARCHITECTURE, COMPUTING
UP11
CEA
Regal
UCBL
INSERM
EARTH SCIENCES, ARCHAEOLOGY PREHISTORY
SEQUEL
BUSINESS, DESIGN
IMAGE, LEARNING
ASCLEPIOS
LINGUISTICS, LANGUAGES
HISTORY, POLITICAL
FINANCE, GRAPH
A model-based approach for systematic risk analysis
CELLS, CANCER
HEALTH, PATIENTS
NUCLEAR, TEV
MAESTRO
Histone deacetylase inhibitors and genomic instability...
BIOLOGY, PROTEIN
Triangle-Free Strongly Circular-Perfect Graphs
MICROBIOLOGY PARASITOLOGY, GENETIC
J. Stark
S. Crepe-Renaudin
D. Gele
R. Granier de Cassagnac
REACTION, ORGANIC CHEMISTRY

Figure 2.18: *Cartolabe* interface showing document heatmap, labels and search feature. [15]

Both tools are useful to understand and analyze collections of text and both their similarities and topics. However, for large collections they may become overwhelming.

TexTonic [16] was designed to explore very large and unstructured text collections through user-driven analytics. It uses Rapid Automatic Keyword Extraction [29] (RAKE) with a common stop-word list to extract the terms. It combines dimension reduction and a force-directed layout to anchor clusters to a certain position on a 2D plane.

Clusters are identifiable by the shared background, the font color of associated terms, and term position. Term size is determined by term weight and position in the hierarchical cluster. The figure below shows how clusters are displayed.

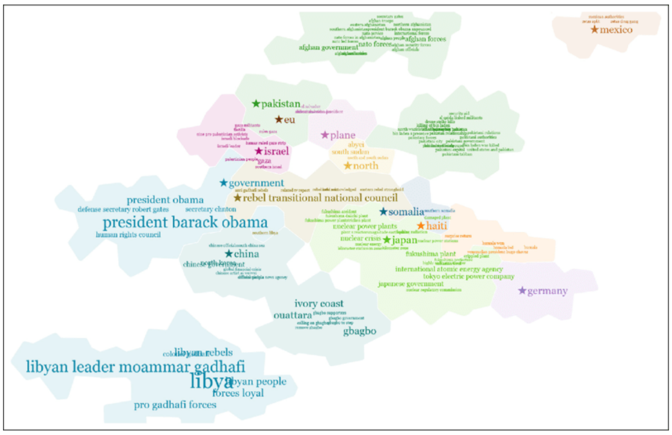
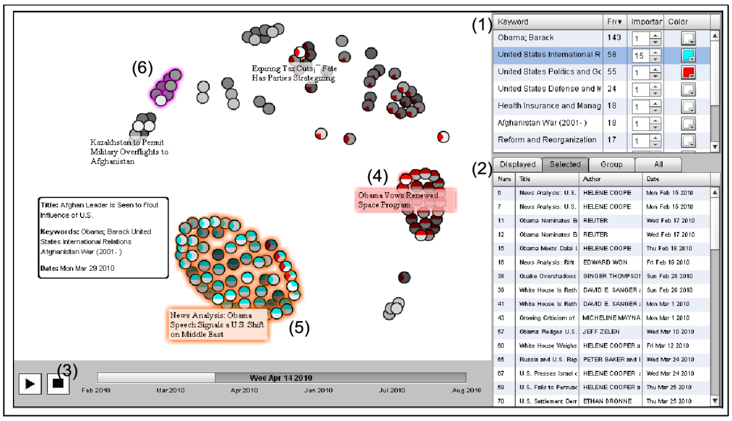


Figure 2.19: *Textonic* interface. [16]

STREAMIT [17] is based on a dynamic force-directed plane into which documents are inserted. Similar documents will have their particles closer (see figure 2.20). Similarity is influenced by keyword importance, meaning the number of times the keyword is referred or the importance it has in the document. It can be modified at any time.

Regarding interaction, users can highlight documents or keywords of interest in the dynamic visualization. They can also retrieve documents through a variety of interactions, such as rubber band selection, search by example, and search by keywords which can be sent to a shoebox so users can investigate them in full detail.


(1)
(2)
(3)
(4)
(5)
(6)
Expiring Tax Cuts¡¯ Fate Has Parties Strategizing
Kazakhstan to Permit Military Overflights to Afghanistan
Obama Vows Renewed Space Program
News Analysis: Obama Speech Signals a U.S. Shift on Middle East
Wed Apr 14 2010
Feb 2010
Mar 2010
Apr 2010
Jun 2010
Jul 2010
Aug 2010
Displayed
Selected
Group
All

Title: Afghan Leader Is Seen to Flout Influence of U.S.

Keywords: Obama; Barack United States International Relations Afghanistan War (2001-)

Date: Mon Mar 29 2010

Keyword	Fre▾	Importar	Color
Obama; Barack	143	1	□
United States International R	58	15	🟦
United States Politics and Gc	55	1	🟥
United States Defense and M	24	1	□
Health Insurance and Manag	18	1	□
Afghanistan War (2001-)	18	1	□
Reform and Reorganization	17	1	□

Num	Title	Author	Date
6	News Analysis: U.S.	HELENE COOPE	Mon Feb 15 2010
7	News Analysis: U.S.	HELENE COOPE	Mon Feb 15 2010
11	Obama Nominates B	REUTER	Wed Feb 17 2010
12	Obama Nominates B	REUTER	Wed Feb 17 2010
15	Obama Meets Dalai L	HELENE COOPE	Thu Feb 18 2010
16	News Analysis: Rift	EDWARD WON	Fri Feb 19 2010
38	Quake Overshadows	GINGER THOMPSON	Sun Feb 28 2010
39	White House Is Reth	DAVID E. SANGER a	Sun Feb 28 2010
41	White House Is Reth	DAVID E. SANGER a	Mon Mar 1 2010
43	Growing Criticism of	MICHELINE MAYNA	Mon Mar 1 2010
57	Obama Pledges U.S.	JEFF ZELEN	Wed Mar 10 2010
60	White House Weighs	HELENE COOPER a	Fri Mar 12 2010
65	Russia and U.S. Rep	PETER BAKER and I	Wed Mar 24 2010
67	U.S. Presses Israel c	HELENE COOPER a	Wed Mar 24 2010
69	U.S. Fails to Persuac	HELENE COOPER a	Thu Mar 25 2010
70	U.S. Settlement Dem	ETHAN BRONNE	Thu Mar 25 2010

Figure 2.20: *STREAMIT* interface. [17]

Both this work and the last are useful since they summarize the content of the documents in a collection and groups them together according to their similarity. The downside lies in the large collections because there are terms that become hidden.

Another interface that clusters documents of a collection is Jigsaw [18]. This tool summarizes a document using most significant sentence, most frequent words across the selected documents, and label sets of grouped documents based on word frequency in each set, word uniqueness across sets, or a combination of both.

This tool has 4 different views. In List View, the user can select the desired entities to be visualized, such as year, author, etc., and order them alphabetically or by frequency of occurrence. The user can then see connections between entities.

The Document Cluster View shows all documents divided into clusters according to their similarity. Each cluster has a different color and is labeled with three keywords that commonly occur. The Cluster View provides a word frequency slider.

The Document View presents a list of documents with the selected document's text and related information. Below the text are the associated entities, and above the text is the one-sentence summary of the document computed by Jigsaw's summary analysis. The word cloud at the top shows the most common words (with highlighted keywords and concepts) in the abstracts of these loaded papers.

The Document Grid View displays all the documents and can sort them by various text metrics, one being similarity to a base document. In the figure below, we can observe *Jigsaw*'s List and Document

Cluster views.

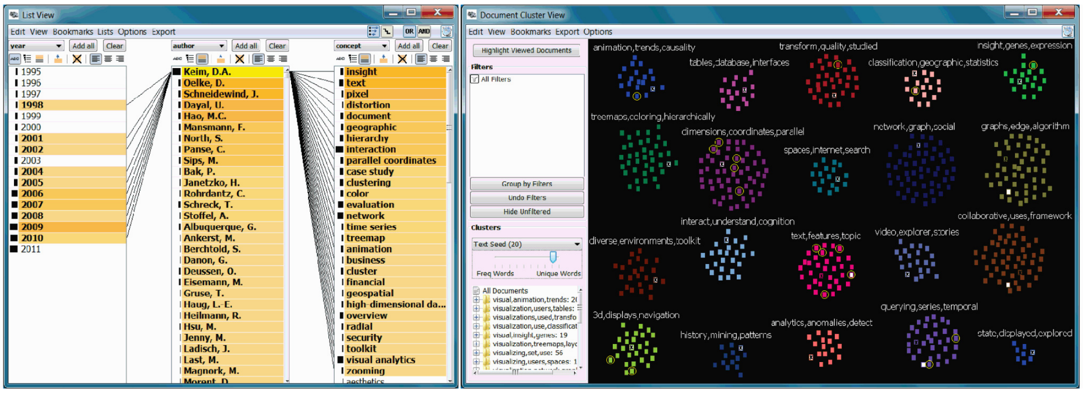

List View
Document Cluster View
Edit View Bookmarks Lists Options Export
Edit View Bookmarks Export Options
year
author
concept
Add all
Clear
Highlight viewed Documents
Filters
All Filters
Group by Filters
Undo Filters
Hide Unfiltered
Clusters
Text Seed (20)
Freq Words
Unique Words
All Documents
animation,trends,causality
tables,database,interfaces
transform,quality,studied
classification,geographic,statistics
insight,genes,expression
treemaps,coloring,hierarchically
dimensions,coordinates,parallel
spaces,internet,search
network,graph,social
graphs,edge,algorithm
interact,understand,cognition
collaborative,uses,framework
diverse,environments,toolkit
text,features,topic
video,explorer,stories
3d,displays,navigation
analytics,anomalies,detect
querying,series,temporal
history,mining,patterns
static,displayed,explored

Figure 2.21: *Jigsaw*'s List View and Document Cluster View. [18]

This tool is useful in the context of this work since it helps to navigate large collections and explore documents as well as find connections between different entities.

During this research, visualizations for news articles were also discovered. For example, *Contexter* [30] processes documents generating name-entity and bag-of-words representations along with the original text representation of the document. As seen in figure 2.22, the entities and their links are displayed as a network.

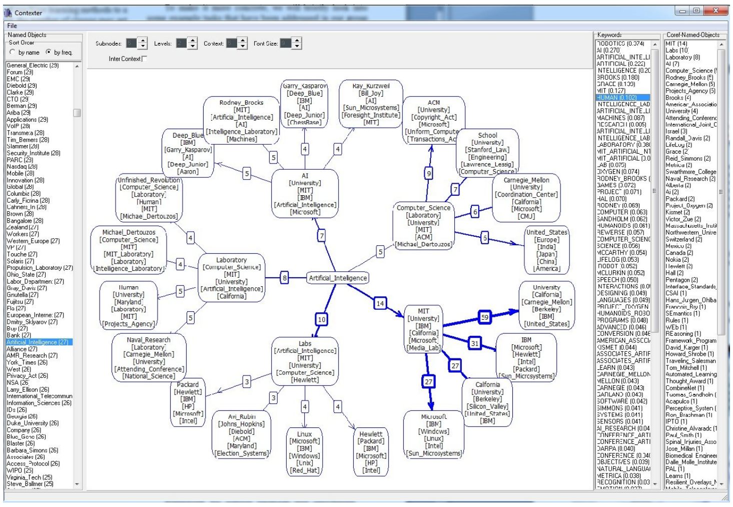


Figure 2.22: *Contexter* interface. [19]

The user can select an entity that will be displayed in the center of the window and will show keywords that are related to said entity. It will also show the number of documents associated with that entity and keywords.

The context of a name-entity is shown in three different ways: a set of keywords usually placed with the selected name-entity, a set of other name-entities, or a set of keywords collocated with the simultaneous appearance of the selected and most frequent other name-entities.

This work is useful since it allows to explore a collection by displaying the different entities of the several documents on the screen as well as their connections.

FinaVistory [20] is an interface to visualize financial news. It is a set of different visualizations that show information regarding different topics.

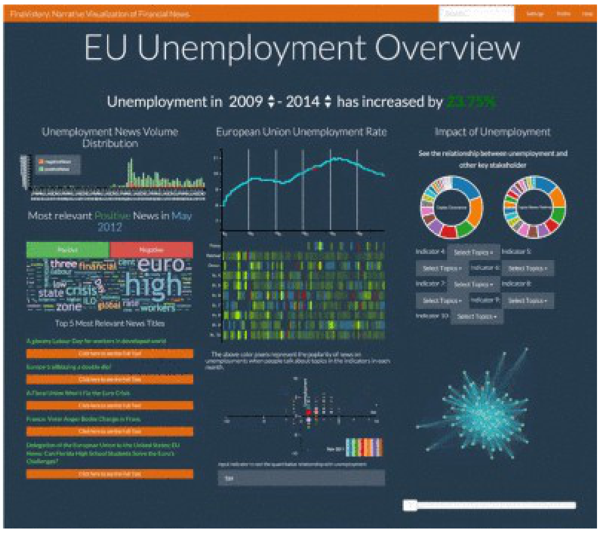

EU Unemployment Overview
Unemployment in 2009 - 2014 has increased by
Unemployment News Volume Distribution
European Union Unemployment Rate
Impact of Unemployment
See the relationship between unemployment and other key stakeholder
Most relevant Positive News in May 2012
Positive
Negative
three
financial
euro
low
state
crisis
high
ILO
zone
global
rate
workers
Top 5 Most Relevant News Titles
Indicator 4:
Select Topics
Indicator 5:
Indicator 6:
Indicator 7:
Indicator 8:
Indicator 9:
Indicator 10:

Figure 2.23: *Finavistory*'s interface. Image taken from [20] and then cropped.

The stacked bar chart is used to show the time distribution of positive and negative news. Positive news means the article conveys a message of increasing index while negative news means the opposite. The word cloud gives more insight to the user by displaying the most frequent keywords presented in the documents. The force-directed layout diagram represents the topics of various articles and the relationships between them. Finally, the bubble chart shows the relationship between user input variables and the price. In the figure below, we can see the interface of this tool.

The multiple visualizations of this tool allow for a complete analysis of a collection. It summarizes a document displaying the most common keyword and shows the documents according to their similarity.

EventRiver [21] is a tool used to detect, track and investigate events of real-life situations. It integrates event-based automated text analysis and visualization to reveal the events motivating the text generation. This tool displays an event as a bubble laid out in a river-like display whose width and height represent the time span that it drew continuous attention and the number of documents reporting or discussing such events. Every bubble has a different color to differentiate stories.

An event is labeled by dual-labels, where core-keywords and context-keywords are displayed in parallel and have distinct background colors. Finally, the users can analyze an event by clicking on it. An Index Panel lists all documents within the cluster, a Detail Panel allows users to read a document and an Evidence Box allows users to save documents of interest. In figure 2.24, the *EventRiver* interface is pictured with labels of some events.

This tool is useful since the river display is an attractive and intuitive way to navigate documents. Documents related to each other would be assigned colors to distinguish them. The downside of this interface is the process of finding the desired document since they are displayed as a bubble.

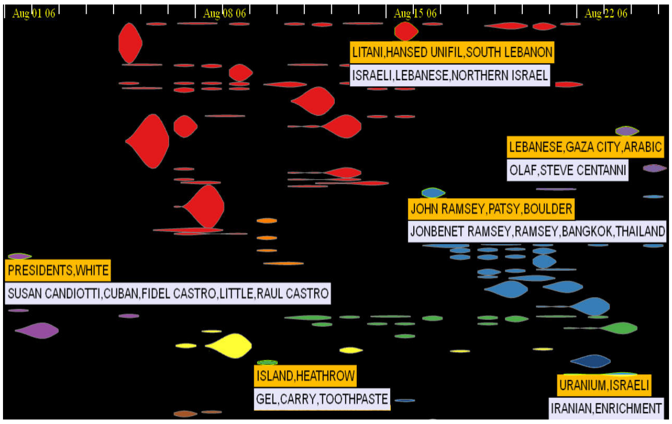

Aug 01 06
Aug 08 06
Aug 15 06
Aug 22 06
LITANI,HANSED UNIFIL,SOUTH LEBANON
ISRAELI,LEBANESE,NORTHERN ISRAEL
LEBANESE,GAZA CITY,ARABIC
OLAF,STEVE CENTANNI
JOHN RAMSEY,PATSY,BOULDER
JONBENET RAMSEY,RAMSEY,BANGKOK,THAILAND
PRESIDENTS,WHITE
SUSAN CANDIOTTI,CUBAN,FIDEL CASTRO,LITTLE,RAUL CASTRO
ISLAND,HEATHROW
GEL,CARRY,TOOTHPASTE
URANIUM,ISRAELI
IRANIAN,ENRICHMENT

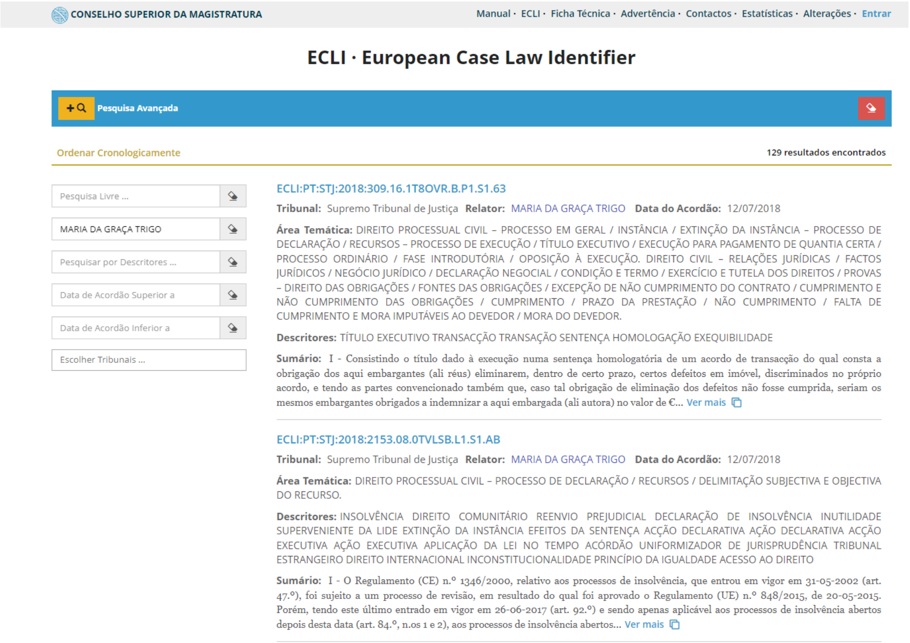
Figure 2.24: *EventRiver* interface with dual labels above the bubbles. [21]

2.4 Digital Libraries for Justice Documents

There are various different digital libraries that are used by magistrates to search for decisions on a daily basis. In order to develop the interface, it is necessary to analyze the advantages and disadvantages of these libraries so that we can adapt and incorporate them in the newly created one. As it was said before, this project will be developed for the Portuguese Supreme Court, so it is required to understand how the magistrates use the existing libraries and what sort of tasks they perform on them. Below are presented examples of the most utilized digital libraries as well as their strong and weak points.

2.4.1 ECLI

ECLI is one of the most used digital libraries to search for decisions on par with DGSI. This website allows the user to search for terms on the Portuguese jurisprudence and sorts the results according to its relevance or chronologically. The user can also search for a specific author or descriptor without entering a search term. ECLI allows the results to be filtered by the different Portuguese courts as well as by date by choosing a range of a time period. The user can combine these different filters to make the search more specific.



CONSELHO SUPERIOR DA MAGISTRATURA
Manual · ECLI · Ficha Técnica · Advertência · Contactos · Estatísticas · Alterações · Entrar
Pesquisa Avançada
Ordenar Cronologicamente
Pesquisa Livre ...
Pesquisar por Descritores ...
Data de Acordão Superior a
Data de Acordão Inferior a
Escolher Tribunais ...
Ver mais

ECLI · European Case Law Identifier

129 resultados encontrados

MARIA DA GRAÇA TRIGO

ECLI:PT:STJ:2018:309.16.1T8OVR.B.P1.S1.63

Tribunal: Supremo Tribunal de Justiça **Relator:** MARIA DA GRAÇA TRIGO **Data do Acordão:** 12/07/2018

Área Temática: DIREITO PROCESSUAL CIVIL – PROCESSO EM GERAL / INSTÂNCIA / EXTINÇÃO DA INSTÂNCIA – PROCESSO DE DECLARAÇÃO / RECURSOS – PROCESSO DE EXECUÇÃO / TÍTULO EXECUTIVO / EXECUÇÃO PARA PAGAMENTO DE QUANTIA CERTA / PROCESSO ORDINÁRIO / FASE INTRODUTÓRIA / OPOSIÇÃO À EXECUÇÃO. DIREITO CIVIL – RELAÇÕES JURÍDICAS / FACTOS JURÍDICOS / NEGÓCIO JURÍDICO / DECLARAÇÃO NEGOCIAL / CONDIÇÃO E TERMO / EXERCÍCIO E TUTELA DOS DIREITOS / PROVAS – DIREITO DAS OBRIGAÇÕES / FONTES DAS OBRIGAÇÕES / EXCEPÇÃO DE NÃO CUMPRIMENTO DO CONTRATO / CUMPRIMENTO E NÃO CUMPRIMENTO DAS OBRIGAÇÕES / CUMPRIMENTO / PRAZO DA PRESTAÇÃO / NÃO CUMPRIMENTO / FALTA DE CUMPRIMENTO E MORA IMPUTÁVEIS AO DEVEDOR / MORA DO DEVEDOR.

Descritores: TÍTULO EXECUTIVO TRANSACÇÃO TRANSAÇÃO SENTENÇA HOMOLOGAÇÃO EXEQUIBILIDADE

Sumário: I - Consistindo o título dado à execução numa sentença homologatória de um acordo de transacção do qual consta a obrigação dos aqui embargantes (ali réus) eliminarem, dentro de certo prazo, certos defeitos em imóvel, discriminados no próprio acordo, e tendo as partes convencionado também que, caso tal obrigação de eliminação dos defeitos não fosse cumprida, seriam os mesmos embargantes obrigados a indemnizar a aqui embargada (ali autora) no valor de €... Ver mais

ECLI:PT:STJ:2018:2153.08.0TVLSB.L1.S1.AB

Tribunal: Supremo Tribunal de Justiça **Relator:** MARIA DA GRAÇA TRIGO **Data do Acordão:** 12/07/2018

Área Temática: DIREITO PROCESSUAL CIVIL – PROCESSO DE DECLARAÇÃO / RECURSOS / DELIMITAÇÃO SUBJECTIVA E OBJECTIVA DO RECURSO.

Descritores: INSOLVÊNCIA DIREITO COMUNITÁRIO REENVIO PREJUDICIAL DECLARAÇÃO DE INSOLVÊNCIA INUTILIDADE SUPERVENIENTE DA LIDE EXTINÇÃO DA INSTÂNCIA EFEITOS DA SENTENÇA ACÇÃO DECLARATIVA AÇÃO DECLARATIVA ACÇÃO EXECUTIVA AÇÃO EXECUTIVA APLICAÇÃO DA LEI NO TEMPO ACÓRDÃO UNIFORMIZADOR DE JURISPRUDÊNCIA TRIBUNAL ESTRANGEIRO DIREITO INTERNACIONAL INCONSTITUCIONALIDADE PRINCÍPIO DA IGUALDADE ACESSO AO DIREITO

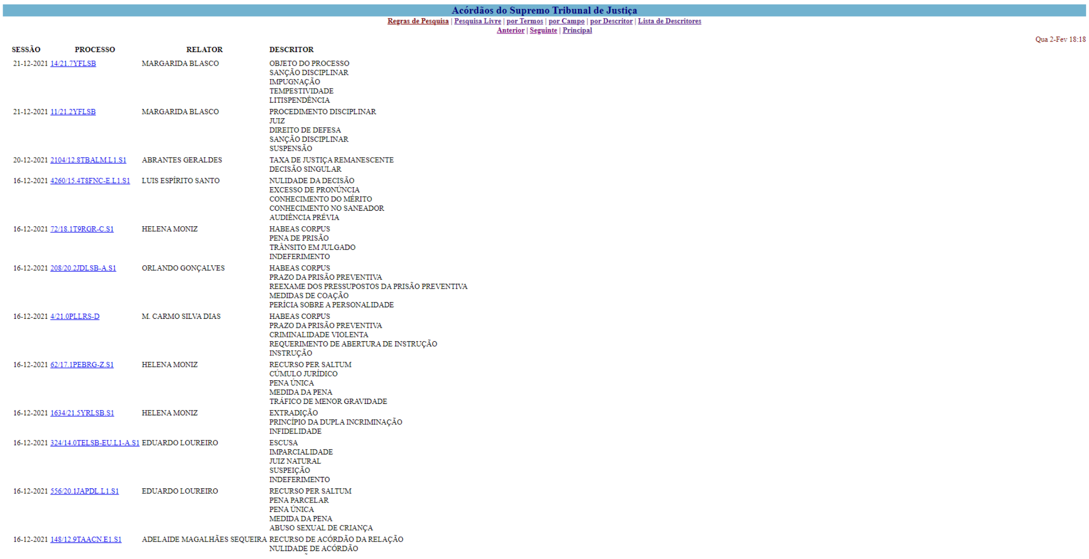
Sumário: I - O Regulamento (CE) n.º 1346/2000, relativo aos processos de insolvência, que entrou em vigor em 31-05-2002 (art. 47.º), foi sujeito a um processo de revisão, em resultado do qual foi aprovado o Regulamento (UE) n.º 848/2015, de 20-05-2015. Porém, tendo este último entrado em vigor em 26-06-2017 (art. 92.º) e sendo apenas aplicável aos processos de insolvência abertos depois desta data (art. 84.º, n.os 1 e 2), aos processos de insolvência abertos... Ver mais

Figure 2.25: ECLI website showing a search for a specific author.

This website has been praised for its intuitive interface since it is easy to search for terms as well as filter the results. It is also an attractive interface since it is not cluttered with information and displays some information regarding the document without showing the full text of the decision. However, the magistrates have demonstrated concerns regarding this website because it is slow retrieving the search results.

2.4.2 DGSI

DGSI is, according to a number of magistrates, the most used digital library for decision searching. Similarly to the ECLI website, it allows the user to search for terms from different courts, however it is different because a specific must be chosen first. The documents regarding that court will be displayed on the front page and then the user has the ability to search for specific terms. This website does not have a filter system implemented, it only allows to search for terms, descriptors or field.



Acórdãos do Supremo Tribunal de Justiça

Regras de Pesquisa | Pesquisa Livre | por Termos | por Campo | por Descritor | Lista de Descritores
Anterior | Seguinte | Principal

Qua 2-Fev 18:18

SESSÃO	PROCESSO	RELATOR	DESCRITOR
21-12-2021	14/21.7YFLSB	MARGARIDA BLASCO	OBJETO DO PROCESSO SANÇÃO DISCIPLINAR IMPUGNAÇÃO TEMPESTIVIDADE LITISPENDÊNCIA
21-12-2021	11/21.2YFLSB	MARGARIDA BLASCO	PROCEDIMENTO DISCIPLINAR JUIZ DIREITO DE DEFESA SANÇÃO DISCIPLINAR SUSPENSÃO
20-12-2021	2104/12.8TBALM.L1.S1	ABRANTES GERALDES	TAXA DE JUSTIÇA REMANESCENTE DECISÃO SINGULAR
16-12-2021	4260/15.4T8FNC-E.L1.S1	LUIS ESPÍRITO SANTO	NULIDADE DA DECISÃO EXCESSO DE PRONÚNCIA CONHECIMENTO DO MÉRITO CONHECIMENTO NO SANEADOR AUDIÊNCIA PRÉVIA
16-12-2021	72/18.1T9RGR-C.S1	HELENA MONIZ	HABEAS CORPUS PENA DE PRISÃO TRÂNSITO EM JULGADO INDEFERIMENTO
16-12-2021	208/20.2JDLSB-A.S1	ORLANDO GONÇALVES	HABEAS CORPUS PRAZO DA PRISÃO PREVENTIVA REEXAME DOS PRESSUPOSTOS DA PRISÃO PREVENTIVA MEDIDAS DE COAÇÃO PERÍCIA SOBRE A PERSONALIDADE
16-12-2021	4/21.0PLLRS-D	M. CARMO SILVA DIAS	HABEAS CORPUS PRAZO DA PRISÃO PREVENTIVA CRIMINALIDADE VIOLENTA REQUERIMENTO DE ABERTURA DE INSTRUÇÃO INSTRUÇÃO
16-12-2021	62/17.1PEBRG-Z.S1	HELENA MONIZ	RECURSO PER SALTUM CÚMULO JURÍDICO PENA ÚNICA MEDIDA DA PENA TRÁFICO DE MENOR GRAVIDADE
16-12-2021	1634/21.5YRLSB.S1	HELENA MONIZ	EXTRADIÇÃO PRINCÍPIO DA DUPLA INCRIMINAÇÃO INFIDELIDADE
16-12-2021	324/14.0TELSB-EU.L1-A.S1	EDUARDO LOUREIRO	ESCUSA IMPARCIALIDADE JUIZ NATURAL SUSPEIÇÃO INDEFERIMENTO
16-12-2021	556/20.1JAPDL.L1.S1	EDUARDO LOUREIRO	RECURSO PER SALTUM PENA PARCELAR PENA ÚNICA MEDIDA DA PENA ABUSO SEXUAL DE CRIANÇA
16-12-2021	148/12.9TAACN.E1.S1	ADELAIDE MAGALHÃES SEQUEIRA	RECURSO DE ACÓRDÃO DA RELAÇÃO NULIDADE DE ACÓRDÃO

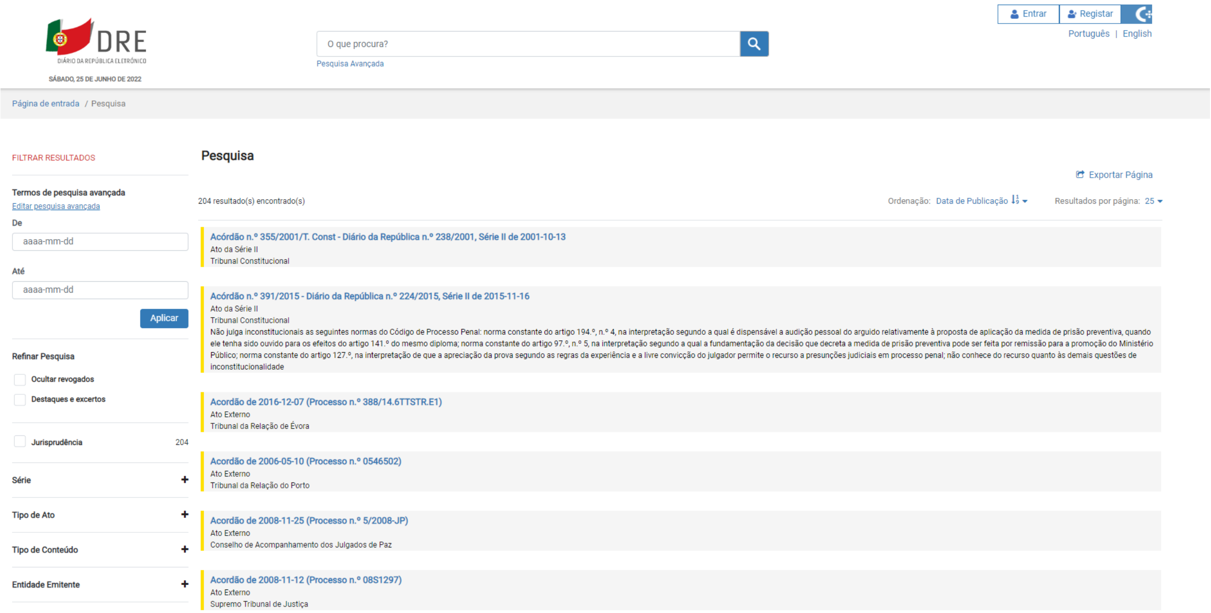
Figure 2.26: DGSI page for decisions of the Portuguese Supreme Court.

DGSI is mainly used because it is faster at displaying results than ECLI is. However, its interface is not as attractive since it is very simplistic and cluttered with information. The front page only displays links to the different courts and does not allow to search for terms unless with choose a specific court. It is not very intuitive as well because it is hard to find the search fields and to figure out how to search.

2.4.3 Diário da República Eletrónico

Another website to find Portuguese jurisprudence that is not as used is the Diário da República Eletrónico. This website allows to quickly search terms in the search box located in the front page or specify certain fields in the advanced search. In here, the user can search for terms in the full text or summary as well as choose the type of document, process or author.

This website is mainly used as as a source of information for Portuguese legislation. For decision searching, magistrates admitted to use the ECLI and DGSI websites.



204 resultado(s) encontrado(s)

Acórdão n.º 355/2001/T. Const - Diário da República n.º 238/2001, Série II de 2001-10-13
Ato da Série II
Tribunal Constitucional

Acórdão n.º 391/2015 - Diário da República n.º 224/2015, Série II de 2015-11-16
Ato da Série II
Tribunal Constitucional
Não julga inconstitucionais as seguintes normas do Código de Processo Penal: norma constante do artigo 194.º, n.º 4, na interpretação segundo a qual é dispensável a audição pessoal do arguido relativamente à proposta de aplicação da medida de prisão preventiva, quando ele tenha sido ouvido para os efeitos do artigo 141.º do mesmo diploma; norma constante do artigo 97.º, n.º 5, na interpretação segundo a qual a fundamentação da decisão que decreta a medida de prisão preventiva pode ser feita por remissão para a promoção do Ministério Público; norma constante do artigo 127.º, na interpretação de que a apreciação da prova segundo as regras da experiência e a livre convicção do julgador permite o recurso a presunções judiciais em processo penal; não conhece do recurso quanto às demais questões de inconstitucionalidade

Acórdão de 2016-12-07 (Processo n.º 388/14.6TTSTR.E1)
Ato Externo
Tribunal da Relação de Évora

Acórdão de 2006-05-10 (Processo n.º 0546502)
Ato Externo
Tribunal da Relação do Porto

Acórdão de 2008-11-25 (Processo n.º 5/2008-JP)
Ato Externo
Conselho de Acompanhamento dos Julgados de Paz

Acórdão de 2008-11-12 (Processo n.º 08S1297)
Ato Externo
Supremo Tribunal de Justiça

Figure 2.27: Diário da República Eletrónico page for Portuguese jurisprudence.

2.5 Discussion

Having researched the state of the art, it is possible to draw conclusions regarding the techniques and features that are used. There are several possible techniques that can be used to display text to the user that range from Word and Tag clouds to clustered networks, tree-maps, arc diagrams or fingerprinting. The most common techniques seen throughout the survey were tag and word clouds. These approaches address the most frequent keywords in the text and help summarize its content preventing the user from reading the entire document to understand it, however we lose the perception of its structure.

Word Trees, clustered networks and arc diagrams are also useful to find links between keywords. Similarly to word clouds, this technique helps to summarize the documents' content while showing relationships between keywords. *WordBridge* and *TextArc* are great examples of this visualization that allows to analyze the relationships between entities as well as their location in the text. The downside is the loss of information for massive texts.

Another visualization that is useful for text analysis is literature fingerprinting. This allows to analyze certain metrics such as sentence length or text syntax using heatmaps. They can also be used to display other text properties such as the occurrence of certain entities or terms.

For documents with multiple chapters or sections, a technique that allows the user to browse them without searching through the whole document is also useful. This can be achieved by transforming each section into drop-down menus with the title of each section or chapter to help the user to navigate. To smooth the process of navigation, a common feature is to highlight keywords in the text by assigning different colors to differentiate them.

Similarly to text visualization, several techniques were found to visualize document collections. These techniques include clustering documents, force-directed layouts and graphs. The most common approach found was to represent the documents as clusters and map them to a 2D plane with labels or associated keywords. This helps the user to have an idea of the size of the collection and to find the desired documents in an easier way. Clusters can be adapted for large collections, since they can display multiple documents as a single cluster. However, we do not have the perception of the content of the documents.

Document Cards is a good approach to text and document collection visualization since it is able to summarize the content of each document including images and display the entire collection to the user. Another effective approach is *EventRiver* since it is able to display the importance of each collection of news using the size of the bubbles.

Regarding digital libraries, *ECLI*, *DGSI* and *Diário da República Eletrónico* are the most used websites when searching for Portuguese jurisprudence. The first and last cases are the most useful when filtering information since they offer a range of options to filter a specific search. They also display the results in an attractive way that shows the relevant information of a document without cluttering the screen. *DGSI* is the most used because it shows results faster compared to the other two websites. However, the interface is not attractive.

In conclusion, there are a lot of elements that can be used to develop the proposed solution. Regarding the visualizations of documents, it is interesting to analyze the occurrences of words in the text as well as their position. For the digital library interface, it is important to look at *ECLI* and the *Diário da República Eletrónico* as inspiration for searching as well as filtering results.

3

Designing the Solution

Contents

3.1 Requirements 31

3.2 Pre-Meetings' Approach 32

3.3 Post-Meetings' Approach 35

3.4 Final Approach 37

After analysing the state of the art in different areas, a number of requirements can be identified that the interface should fulfil. It is also important to consider the scope of the IRIS project and its objectives. Several meetings were held with different judges of the Portuguese Supreme Court to understand what was the goal of the final developed solution and to draw up a list of requirements that the developed solution should meet. In these meetings, the judges described the different steps of a case until it reaches the Portuguese Supreme Court. A case contains all the decisions taken by the courts of first and second instance, so a single case can include documents adding up to thousands of pages. To write a new decision, judges also use documents from other cases that have reached the Portuguese Supreme Court. The judges showed the current applications used to search and analyse the different documents of each case. The judges explained that their work is currently very difficult because the applications used are very slow and outdated. The INESC-ID group was shown step by step how the current applications work and they were able to confirm that they need to be updated. After these meetings, the list of objectives changed considerably and a prototype of an application was designed. This prototype is still subject to change as the INESC-ID is still holding meetings with judges. Below is a list of requirements that was created after all the meetings were completed.

3.1 Requirements

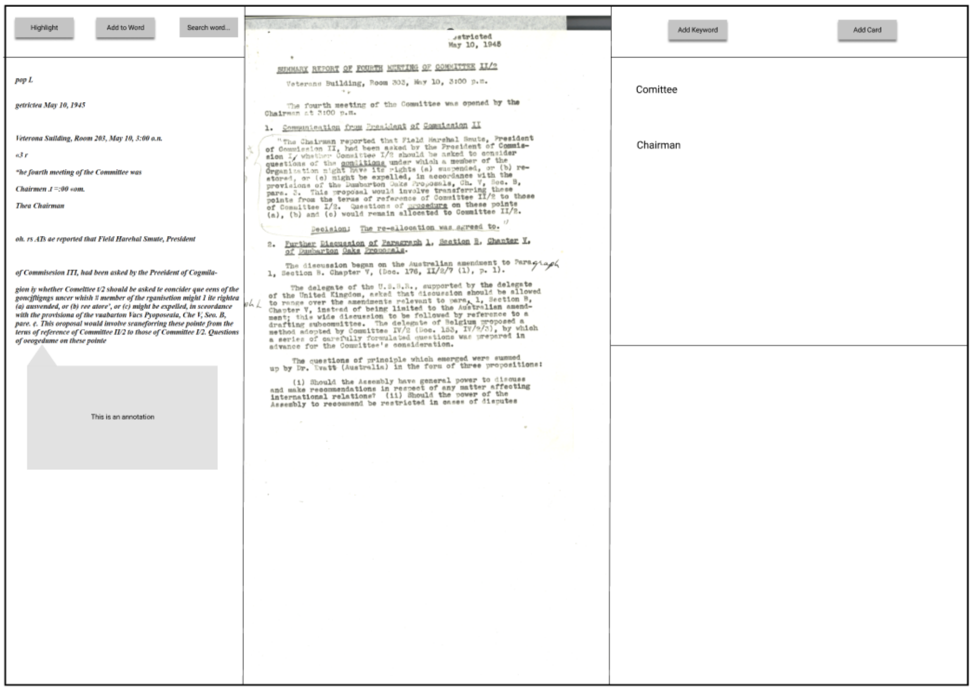
As mentioned above, the objective of the project is to improve the efficiency of judges' workflow. When judges draft a new decision, they need to access and navigate documents from previous decisions as well as documents from other courts. For this, a system similar to a digital library must be developed. This system must allow judges to search for different documents in a way that:

- **allows access to documents from other Portuguese courts** so that judges have access to decisions from similar cases;
- **displays results efficiently and quickly**, which is important to achieve the goal of this project, namely to improve the efficiency of judges' workflow;
- **it is easy to understand and use**;
- **it allows filtering the results by a number of metrics such as court or author** to get more precise results and find the right document faster;
- **allows quick access to the content of a document** to reduce the time spent searching for a specific term;
- **does not fill the screen with unnecessary information**, as the user interface must be readable and easy to use;

- **allows search results to be shared with different people**, which helps judges to share documents with each other;

3.2 Pre-Meetings' Approach

Before the goals of the IRIS project were set in the meetings mentioned above, the original ones looked quite different. Based on the initial information we had received from the judges, it was deemed necessary to develop a solution to navigate and analyse the content of a PDF document so that the potentially thousands of pages could be summarised and analysed by the system rather than by the judges, making this process much faster and more efficient. It would also be used to help navigating the document and writing decisions. The design process began with the creation of a low-fidelity prototype. This phase is the first step in the development of a solution and helps the developers to turn the initial concept ideas into a testable product and get feedback to improve it in the next phases. The low-fidelity prototype can be created using a variety of tools, from a simple pen and paper to powerful web tools. In this case, the website Figma[1] was used. This website allows users to design interfaces and test the different features that could be implemented in the future.



pop L

getrictea May 10, 1945

Veterona Sulldng, Room 203, May 10, 3:00 o.n.

«3 r

"he fourth meeting of the Committee was

Chairmen .t =:00 «om.

Thea Chairman

oh. rs ATs ae reported that Field Harehal Smute, President

of Commisesion ITI, had been asked by the Preeident of Cogmila-

gion iy whether Comelttee t/2 should be asked te eoncider que eens of the goncjftigngs uncer whish ‡ member of the rganisetion might 1 ite rightea (a) ausvended, or (b) ree atore', or (c) might be expelled, in sceordance with the provisiona of the vuabarton Vacs Pyoposeala, Che V, Seo. B, pare. ¢. This oroposal would involve sraneforring these pointe from the terus of reference of Committee II/2 to those of Committee I/2. Questions of oeogedume on these pointe

estricted
May 10, 1945

SUMMARY REPORT OF FOURTH MEETING OF COMMITTEE II/2

Veterans Building, Room 303, May 10, 3:00 p.m.

The fourth meeting of the Committee was opened by the Chairman at 3:00 p.m.

1. Communication from President of Commission II

"The Chairman reported that Field Marshal Smuts, President of Commission II, had been asked by the President of Commission I, whether Committee I/2 should be asked to consider questions of the conditions under which a member of the Organization might have its rights (a) suspended, or (b) restored, or (c) might be expelled, in accordance with the provisions of the Dumbarton Oaks Proposals, Ch. V, Sec. B, para. 3. This proposal would involve transferring these points from the terms of reference of Committee II/2 to those of Committee I/2. Questions of procedure on these points (a), (b) and (c) would remain allocated to Committee II/2.

Decision: The re-allocation was agreed to."

2. Further Discussion of Paragraph 1, Section B, Chapter V, of Dumbarton Oaks Proposals.

The discussion began on the Australian amendment to Paragraph 1, Section B. Chapter V, (Doc. 176, II/2/7 (1), p. 1).

The delegate of the U.S.S.R., supported by the delegate of the United Kingdom, asked that discussion should be allowed to range over the amendments relevant to para. 1, Section B, Chapter V, instead of being limited to the Australian amendment; this wide discussion to be followed by reference to a drafting subcommittee. The delegate of Belgium proposed a method adopted by Committee IV/2 (Doc. 153, IV/2/3), by which a series of carefully formulated questions was prepared in advance for the Committee's consideration.

The questions of principle which emerged were summed up by Dr. Evatt (Australia) in the form of three propositions:

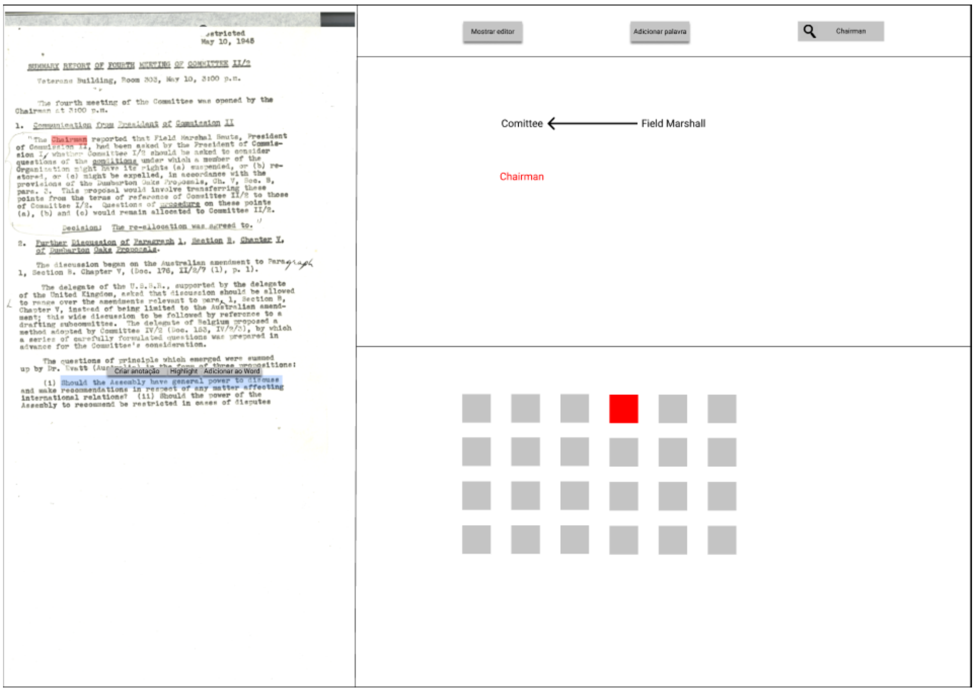
(i) Should the Assembly have general power to discuss and make recommendations in respect of any matter affecting international relations? (ii) Should the power of the Assembly to recommend be restricted in cases of disputes

Figure 3.1: First iteration of the interface after adding an annotation.

As can be seen in Figure 3.1, the original interface used features of visualisation tools to summarise

[1] https://www.figma.com/

the content of a document and annotations to facilitate analysis. The user could add various annotations to the text to make notes on topics they found relevant. The user could also enter words to search for in the document. These were highlighted in the corpus to identify the number of occurrences and their position. The tool allowed keywords to be added to a particular section and links to be made between them, creating a mind map. This feature was also used to make connections between sentences. Another feature of this interface was the heat map, which showed in which paragraph or section of the text a particular word was located. Finally, the user could paste certain parts of the text into a separate Microsoft Word document. This would be useful for a judge when drafting a decision.



estricted
May 10, 1945

SUMMARY REPORT OF FOURTH MEETING OF COMMITTEE II/2

Veterans Building, Room 303, May 10, 3:00 p.m.

The fourth meeting of the Committee was opened by the Chairman at 3:00 p.m.

1. Communication from President of Commission II

"The Chairman reported that Field Marshal Smuts, President of Commission II, had been asked by the President of Commission I, whether Committee I/2 should be asked to consider questions of the conditions under which a member of the Organization might have its rights (a) suspended, or (b) restored, or (c) might be expelled, in accordance with the provisions of the Dumbarton Oaks Proposals, Ch. V, Sec. B, para. 3. This proposal would involve transferring these points from the terms of reference of Committee II/2 to those of Committee I/2. Questions of procedure on these points (a), (b) and (c) would remain allocated to Committee II/2.

Decision: The re-allocation was agreed to."

2. Further Discussion of Paragraph 1, Section B, Chapter V, of Dumbarton Oaks Proposals.

The discussion began on the Australian amendment to Paragraph 1, Section B, Chapter V, (Doc. 176, II/2/7 (1), p. 1).

The delegate of the U.S.S.R., supported by the delegate of the United Kingdom, asked that discussion should be allowed to range over the amendments relevant to para. 1, Section B, Chapter V, instead of being limited to the Australian amendment; this wide discussion to be followed by reference to a drafting subcommittee. The delegate of Belgium proposed a method adopted by Committee IV/2 (Doc. 183, IV/2/3), by which a series of carefully formulated questions was prepared in advance for the Committee's consideration.

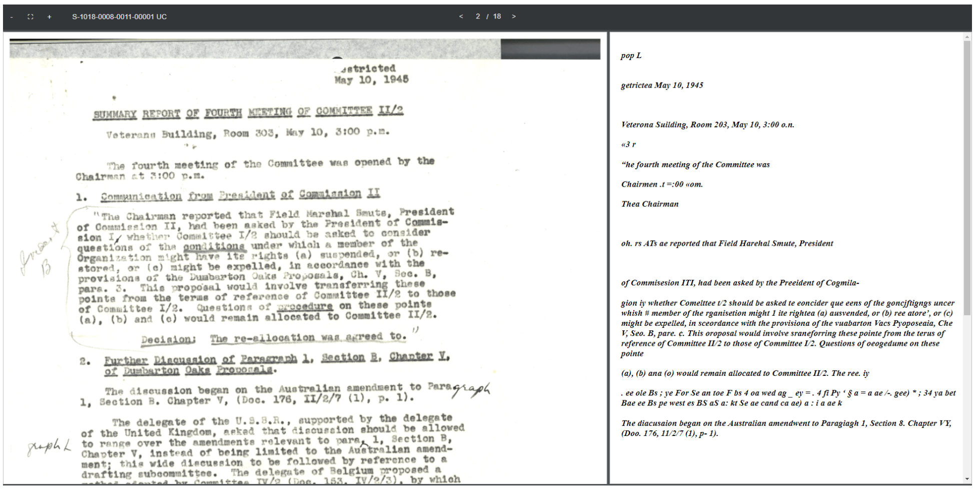
The questions of principle which emerged were summed up by Dr. Evatt (Australia) in the form of three propositions:

(i) Should the Assembly have general power to discuss and make recommendations in respect of any matter affecting international relations? (ii) Should the power of the Assembly to recommend be restricted in cases of disputes

Figure 3.2: First iteration of the interface showing where a word appears in the text.

Originally, the objective was to navigate and interact with the content of a PDF document. A tool was needed that would open the document in the browser and allow certain tasks to be performed, such as highlighting words or searching for terms. Since there are not many tools that have these features built in for free, the solution was developed using a library[2] that performs OCR on the images of the individual pages of the document. OCR stands for Optical Character Recognition and is the process of converting images into text. This process was needed since the documents sent by the judges to the members of the INESC-ID group were all *PNG* files instead of a normal PDF. The solution developed consisted of one page divided into two sections. The first section is the original image of the PDF document and the second contains the text extracted by the OCR process.

[2] https://github.com/not-implemented/hocr-proofreader



Restricted
May 10, 1945

SUMMARY REPORT OF FOURTH MEETING OF COMMITTEE II/2

Veterans Building, Room 303, May 10, 3:00 p.m.

The fourth meeting of the Committee was opened by the Chairman at 3:00 p.m.

1. Communication from President of Commission II

"The Chairman reported that Field Marshal Smuts, President of Commission II, had been asked by the President of Commission I, whether Committee I/2 should be asked to consider questions of the conditions under which a member of the Organization might have its rights (a) suspended, or (b) restored, or (c) might be expelled, in accordance with the provisions of the Dumbarton Oaks Proposals, Ch. V, Sec. B, para. 3. This proposal would involve transferring these points from the terms of reference of Committee II/2 to those of Committee I/2. Questions of procedure on these points (a), (b) and (c) would remain allocated to Committee II/2.

Decision: The re-allocation was agreed to."

Present B

2. Further Discussion of Paragraph 1, Section B, Chapter V, of Dumbarton Oaks Proposals.

The discussion began on the Australian amendment to Paragraph 1, Section B. Chapter V, (Doc. 176, II/2/7 (1), p. 1).

The delegate of the U.S.S.R., supported by the delegate of the United Kingdom, asked that discussion should be allowed to range over the amendments relevant to para. 1, Section B, Chapter V, instead of being limited to the Australian amendment; this wide discussion to be followed by reference to a drafting subcommittee. The delegate of Belgium proposed a method adopted by Committee IV/2 (Doc. 153, IV/2/3), by which

graph L

pop L

getrictea May 10, 1945

Veterona Suilding, Room 203, May 10, 3:00 o.n.

«3 r

“he fourth meeting of the Committee was

Chairmen .t =:00 «om.

Thea Chairman

oh. rs ATs ae reported that Field Harehal Smute, President

of Commisesion ITI, had been asked by the Preeident of Cogmila-

gion iy whether Comelttee t/2 should be asked te eoncider que eens of the goncjftigngs uncer whish # member of the rganisetion might 1 ite rightea (a) ausvended, or (b) ree atore’, or (c) might be expelled, in sceordance with the provisiona of the vuabarton Vacs Pyoposeaia, Che V, Seo. B, pare. ¢. This oroposal would involve sraneforring these pointe from the terus of reference of Committee II/2 to those of Committee I/2. Questions of oeogedume on these pointe

(a), (b) ana (o) would remain allocated to Committee II/2. The ree. iy

*. ee ole Bs ; ye For Se an toe F bs 4 oa wed ag _ ey = . 4 fi Py ‘ § a = a ae /-. gee) * ; 34 ya bet Bae ee Bs pe west es BS aS a: kt Se ae cand ca ae) a : i a ae k*

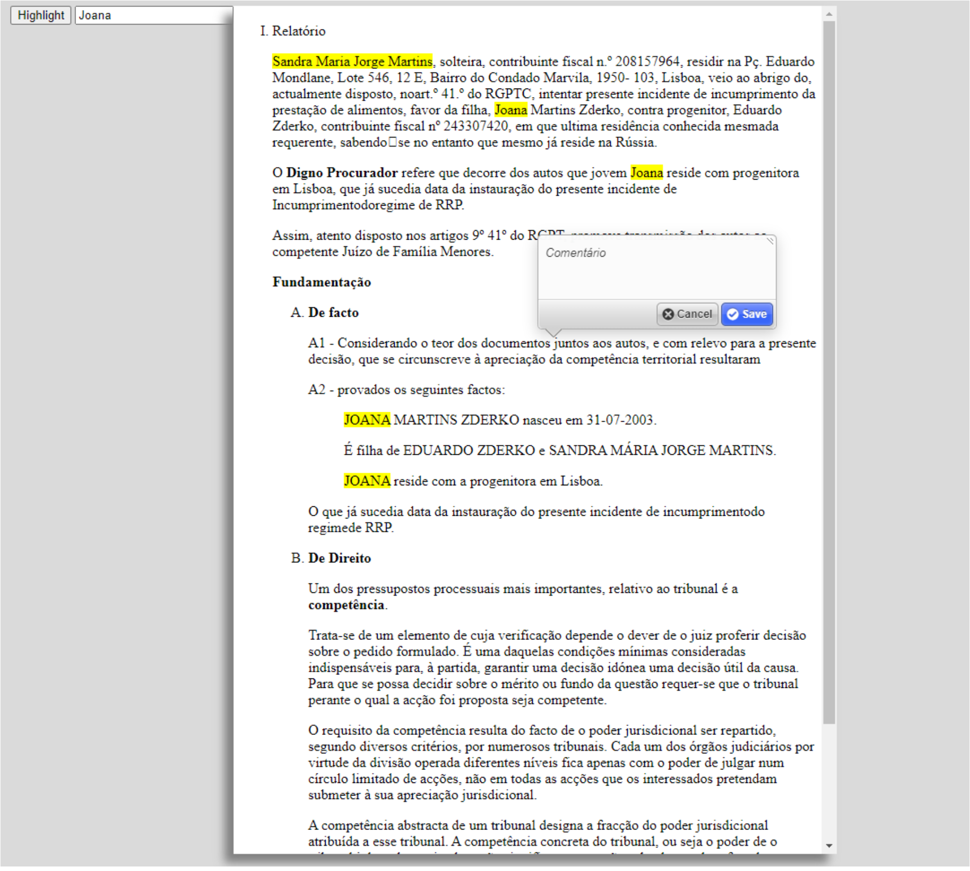
The diacusaion began on the Australian amendwent to Paragiagh 1, Section 8. Chapter VY, (Doo. 176, 11/2/7 (1), p- 1).

Figure 3.3: First Iteration’s interface.

After developing this solution, it was necessary to create features for text interaction. These functionalities were to be merged with the interface mentioned above, but due to changes in the project goals, the whole solution was discarded. In this tool, the user had a central area where the text of a document was located and where the user could perform some tasks. These tasks included highlighting words, word searching and adding annotations to the text. These functions were implemented using three different JavaScript libraries, namely mark.js[3], HR.js[4] and Annotator[5].

As it can be seen in figure 3.4, the user could highlight words by selecting them with the mouse. When the button with the text ”Highlight” in the upper left corner was then clicked, the background of the word turned yellow. Next to the button was a search box where the user could enter words to search for in the text. When a match was found, the background of the words also turned yellow. Finally, the user could also click anywhere on the page, whereupon a field appeared in which it could be written annotations. These annotations remained visible on the screen until the user closed the page.

[3] https://markjs.io/
[4] https://mburakerman.github.io/hrjs/
[5] http://annotatorjs.org/



I. Relatório

Sandra Maria Jorge Martins, solteira, contribuinte fiscal n.º 208157964, residir na Pç. Eduardo Mondlane, Lote 546, 12 E, Bairro do Condado Marvila, 1950- 103, Lisboa, veio ao abrigo do, actualmente disposto, noart.º 41.º do RGPTC, intentar presente incidente de incumprimento da prestação de alimentos, favor da filha, Joana Martins Zderko, contra progenitor, Eduardo Zderko, contribuinte fiscal nº 243307420, em que ultima residência conhecida mesmada requerente, sabendo□se no entanto que mesmo já reside na Rússia.

O **Digno Procurador** refere que decorre dos autos que jovem Joana reside com progenitora em Lisboa, que já sucedia data da instauração do presente incidente de Incumprimentodoregime de RRP.

Assim, atento disposto nos artigos 9º 41º do R
competente Juízo de Família Menores.

Fundamentação

A. **De facto**

A1 - Considerando o teor dos documentos juntos aos autos, e com relevo para a presente decisão, que se circunscreve à apreciação da competência territorial resultaram

A2 - provados os seguintes factos:

JOANA MARTINS ZDERKO nasceu em 31-07-2003.

É filha de EDUARDO ZDERKO e SANDRA MÁRIA JORGE MARTINS.

JOANA reside com a progenitora em Lisboa.

O que já sucedia data da instauração do presente incidente de incumprimentodo regimede RRP.

B. **De Direito**

Um dos pressupostos processuais mais importantes, relativo ao tribunal é a **competência**.

Trata-se de um elemento de cuja verificação depende o dever de o juiz proferir decisão sobre o pedido formulado. É uma daquelas condições mínimas consideradas indispensáveis para, à partida, garantir uma decisão idónea uma decisão útil da causa. Para que se possa decidir sobre o mérito ou fundo da questão requer-se que o tribunal perante o qual a acção foi proposta seja competente.

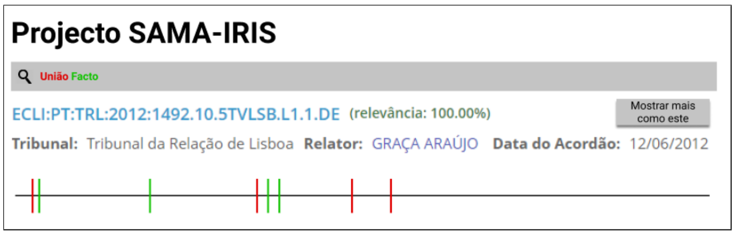
O requisito da competência resulta do facto de o poder jurisdicional ser repartido, segundo diversos critérios, por numerosos tribunais. Cada um dos órgãos judiciários por virtude da divisão operada diferentes níveis fica apenas com o poder de julgar num círculo limitado de acções, não em todas as acções que os interessados pretendam submeter à sua apreciação jurisdicional.

A competência abstracta de um tribunal designa a fracção do poder jurisdicional atribuída a esse tribunal. A competência concreta do tribunal, ou seja o poder de o

Figure 3.4: Tech demo for text manipulation and interaction.

3.3 Post-Meetings' Approach

Once the list of objectives had been updated, the approach to the interface also had to be changed to meet the new objectives described in section 3.1. The Figma website was used again to design a low-fidelity prototype that incorporated some key features that the judges considered important. Taking inspiration from the ECLI website and the various existing search engines, an interface was designed to allow users to search for one or more specific terms, with results displayed in order of relevance to the search or by other metrics such as the date of the documents. As we can see on Figure 3.5, users are provided with a search box in which they can enter words, much like a search engine. The application then returns several documents divided by pages. Each page contains a maximum number of results.



Projecto SAMA-IRIS

União Facto

ECLI:PT:TRL:2012:1492.10.5TVLSB.L1.1.DE (relevância: 100.00%)

Tribunal: Tribunal da Relação de Lisboa **Relator:** GRAÇA ARAÚJO **Data do Acordão:** 12/06/2012

Figure 3.5: Second iteration search box and result.

Each result displays information about each document, such as the author, date and court. This gives the user a brief description of the document before reading its contents. It also displays the title of the document and its relevance to the search. The title name is the ECLI link and opens a separate page to the document itself so that the user can read the full text. In order to analyse the text corpus without reading it in its entirety, it was decided to create a visualisation based on the related work, with a horizontal bar representing the content of the document and several vertical bars representing the number of occurrences of the searched words (Figure 3.6).

As can be seen in the figure below, three methods were tested to see which was the best methodology to distinguish the vertical bars and identify which word the bars represented. To choose the best method for this feature, an informal test was held with a group of five users in which they stated their preferred method. The first method was to assign a colour to a word and give the bars the same colour. This proved to be a good method as long as the colours did not destroy the overall aesthetics of the user interface and blended well together. The second and third methods are similar. Different coloured symbols were used to represent the words, with different symbols for different words. The second method placed the symbols on top of the horizontal bars and the third method was a combination of the first two, where there were vertical bars with the symbols on top. The last two methods were not chosen because the second was difficult to understand when several symbols were close together, and in the third one the symbols were too small to see.

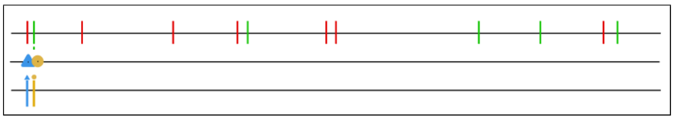
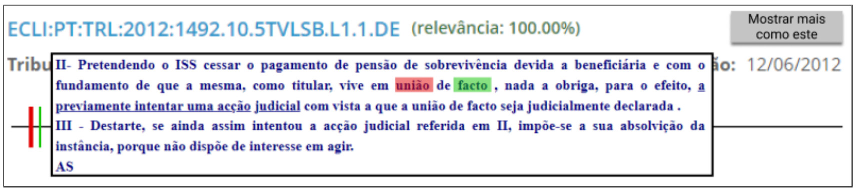


Figure 3.6: Different methods of displaying the vertical bars.

Another feature incorporated into the prototype was the ability to hover over the vertical bars and see an excerpt of the text in which the word is located, as seen in Figure 3.7. This gives the user an idea of the content of the document and the context of the word in the text without the user having to read the whole corpus. This is useful to give a brief summary of the content of the document and let the user decide whether the document is actually relevant to what they are looking for. The words searched for are highlighted in the part of the text with the same colour that was assigned at the beginning. This way, the user can quickly find the words they were looking for and read the relevant paragraph. It is also possible to press a button in the right-hand corner of each document to search for similar documents. This provides a new set of results with documents that mention the searched words in several fields and are thematically related to the selected document.



ECLI:PT:TRL:2012:1492.10.5TVLSB.L1.1.DE (relevância: 100.00%)

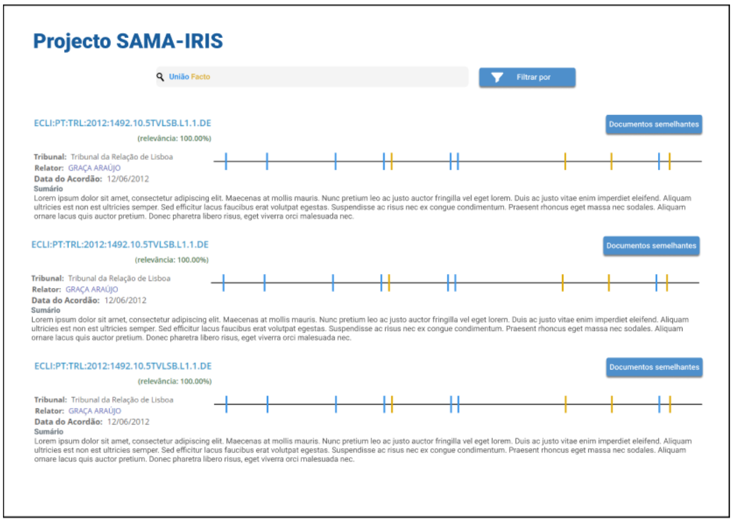
Tribu ... ão: 12/06/2012

II- Pretendendo o ISS cessar o pagamento de pensão de sobrevivência devida a beneficiária e com o fundamento de que a mesma, como titular, vive em união de facto, nada a obriga, para o efeito, a previamente intentar uma acção judicial com vista a que a união de facto seja judicialmente declarada .
III - Destarte, se ainda assim intentou a acção judicial referida em II, impõe-se a sua absolvição da instância, porque não dispõe de interesse em agir.
AS

Figure 3.7: Excerpt of the text with the searched words.

3.4 Final Approach

Several changes were made to improve the visual aspect of the user interface of the solution. The layout of each result was changed to make it clearer, and a document summary field was added. This helps the user to further analyse the content of the document without having to read the whole corpus. The title of the user interface and the button to find similar documents were changed to a blue colour to be more aesthetically pleasing. The colour of the vertical bars in the visualisation was also changed to a more subtle colour scheme so that they do not distract the user from the rest of the information on the screen.



Projecto SAMA-IRIS

União Facto

ECLI:PT:TRL:2012:1492.10.5TVLSB.L1.1.DE
(relevância: 100.00%)

Tribunal: Tribunal da Relação de Lisboa
Relator: GRAÇA ARAÚJO
Data do Acordão: 12/06/2012
Sumário
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas at mollis mauris. Nunc pretium leo ac justo auctor fringilla vel eget lorem. Duis ac justo vitae enim imperdiet eleifend. Aliquam ultricies est non est ultricies semper. Sed efficitur lacus faucibus erat volutpat egestas. Suspendisse ac risus nec ex congue condimentum. Praesent rhoncus eget massa nec sodales. Aliquam ornare lacus quis auctor pretium. Donec pharetra libero risus, eget viverra orci malesuada nec.

ECLI:PT:TRL:2012:1492.10.5TVLSB.L1.1.DE
(relevância: 100.00%)

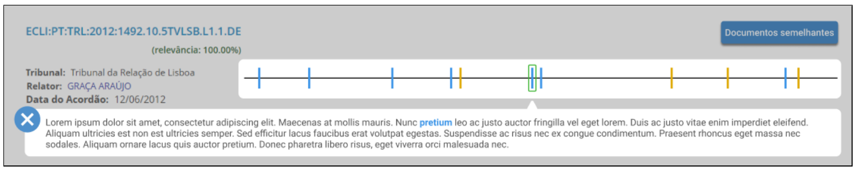
Tribunal: Tribunal da Relação de Lisboa
Relator: GRAÇA ARAÚJO
Data do Acordão: 12/06/2012
Sumário
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas at mollis mauris. Nunc pretium leo ac justo auctor fringilla vel eget lorem. Duis ac justo vitae enim imperdiet eleifend. Aliquam ultricies est non est ultricies semper. Sed efficitur lacus faucibus erat volutpat egestas. Suspendisse ac risus nec ex congue condimentum. Praesent rhoncus eget massa nec sodales. Aliquam ornare lacus quis auctor pretium. Donec pharetra libero risus, eget viverra orci malesuada nec.

ECLI:PT:TRL:2012:1492.10.5TVLSB.L1.1.DE
(relevância: 100.00%)

Tribunal: Tribunal da Relação de Lisboa
Relator: GRAÇA ARAÚJO
Data do Acordão: 12/06/2012
Sumário
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas at mollis mauris. Nunc pretium leo ac justo auctor fringilla vel eget lorem. Duis ac justo vitae enim imperdiet eleifend. Aliquam ultricies est non est ultricies semper. Sed efficitur lacus faucibus erat volutpat egestas. Suspendisse ac risus nec ex congue condimentum. Praesent rhoncus eget massa nec sodales. Aliquam ornare lacus quis auctor pretium. Donec pharetra libero risus, eget viverra orci malesuada nec.

Figure 3.8: Results page of the final iteration.

As we can see in Figure 3.8, the search box has been centred and a filter button has also been added so that the user can perform a more specific search. The filters include the different courts, authors and time periods. The user can click on the button and a drop-down list of each filter will appear. The options in the list are check boxes and when the user selects them, the results are updated accordingly. The latest change to the prototype can be seen in Figure 3.9 and relates to the excerpt from the text. The excerpt appears above the summary so that it is easier to hover over a new bar, and it minimises the extent to which the user loses context of the state of the user interface. The words being searched for are highlighted with the appropriate colour instead of being surrounded by a box, and the bar is also highlighted so that the user always knows which bar is being analysed.



ECLI:PT:TRL:2012:1492.10.5TVLSB.L1.1.DE

(relevância: 100.00%)

Tribunal: Tribunal da Relação de Lisboa
Relator: GRAÇA ARAÚJO
Data do Acordão: 12/06/2012

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas at mollis mauris. Nunc pretium leo ac justo auctor fringilla vel eget lorem. Duis ac justo vitae enim imperdiet eleifend. Aliquam ultricies est non est ultricies semper. Sed efficitur lacus faucibus erat volutpat egestas. Suspendisse ac risus nec ex congue condimentum. Praesent rhoncus eget massa nec sodales. Aliquam ornare lacus quis auctor pretium. Donec pharetra libero risus, eget viverra orci malesuada nec.

Figure 3.9: Excerpt of the text with the searched words of the final iteration.

4

Implementation

Contents

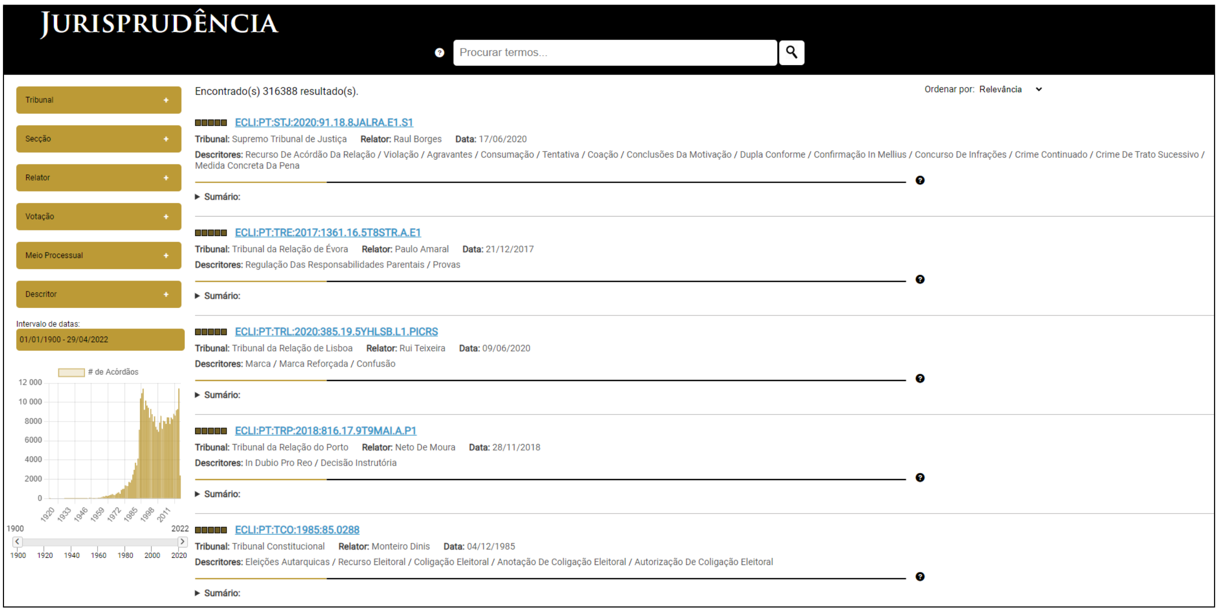
4.1 Solution 41

4.2 Architecture 47

4.3 Implementation 48

4.1 Solution

Following the design of the prototype on the Figma website, a solution was developed to meet the updated objectives of the project. In this section we describe how the interface works and how to use it. The colour scheme of the interface was originally the same as that of the Figma website. However, to maintain consistency with the Portuguese Supreme Court website, the colour scheme was changed to match it. Below you can see the front page of the developed solution.



Encontrado(s) 316388 resultado(s).

ECLI:PT:STJ:2020:91.18.8JALRA.E1.S1

Tribunal: Supremo Tribunal de Justiça **Relator:** Raul Borges **Data:** 17/06/2020

Descritores: Recurso De Acórdão Da Relação / Violação / Agravantes / Consumação / Tentativa / Coação / Conclusões Da Motivação / Dupla Conforme / Confirmação In Mellius / Concurso De Infrações / Crime Continuado / Crime De Trato Sucessivo / Medida Concreta Da Pena

▶ **Sumário:**

ECLI:PT:TRE:2017:1361.16.5T8STR.A.E1

Tribunal: Tribunal da Relação de Évora **Relator:** Paulo Amaral **Data:** 21/12/2017

Descritores: Regulação Das Responsabilidades Parentais / Provas

▶ **Sumário:**

ECLI:PT:TRL:2020:385.19.5YHLSB.L1.PICRS

Tribunal: Tribunal da Relação de Lisboa **Relator:** Rui Teixeira **Data:** 09/06/2020

Descritores: Marca / Marca Reforçada / Confusão

▶ **Sumário:**

ECLI:PT:TRP:2018:816.17.9T9MAI.A.P1

Tribunal: Tribunal da Relação do Porto **Relator:** Neto De Moura **Data:** 28/11/2018

Descritores: In Dubio Pro Reo / Decisão Instrutória

▶ **Sumário:**

ECLI:PT:TCO:1985:85.0288

Tribunal: Tribunal Constitucional **Relator:** Monteiro Dinis **Data:** 04/12/1985

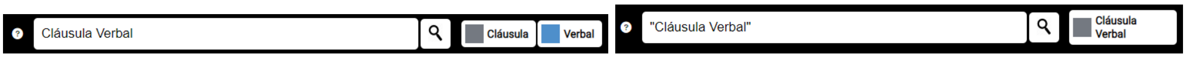
Descritores: Eleições Autarquicas / Recurso Eleitoral / Coligação Eleitoral / Anotação De Coligação Eleitoral / Autorização De Coligação Eleitoral

▶ **Sumário:**

Figure 4.1: Final solution's initial page.

4.1.1 Perform a search

After opening the page, the user can search for one or more terms by entering them in the search box at the top of the page. These terms are matched against the various fields indexed in the database and the results are sent to the client. There are several ways in which the user can perform the search. The first is to simply enter the search terms. They are then matched individually against the text of the summary and the decision itself. They are also matched against other document-related information such as the author, the list of descriptors and the process number. The second option is to search for words between inverted commas and these are matched together. For example, if a user searches for *Cláusula Verbal*, as seen in Figure 4.2 the words are matched individually. However, when the inverted commas are used, they are matched as an expression. The third method of searching is to specify the field in which the user wants to search. For example, if the user wants to find documents by a particular author, they must begin the search with "*Relator:*" and include the author's name in inverted commas after the colon. Several fields are available, e.g. descriptor, court, date, vote, etc. To search in these fields, users must enter the name of the field in the search box, followed by a colon. The different search methods can also be used together. The user only has to insert an operator e.g. *AND* or *OR* between the different methods. This way users can do a more specific search.



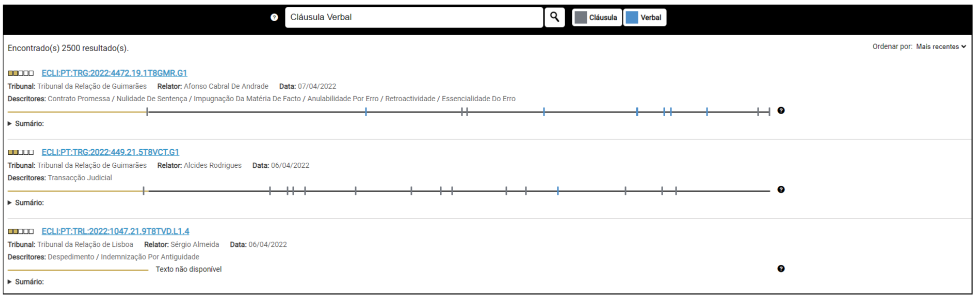
(a) Search without quotation marks. The words are matched individually.

(b) Search with quotation marks. The words are matched together.

Figure 4.2: Search for *Cláusula Verbal* using different methods.

4.1.2 Results

After the search is performed, the interface is updated with the results retrieved from the server. The total number of results is also displayed. However, the maximum number of results displayed is 500, divided into 25 pages with 20 documents each. When the user changes pages, a new request is made to the database with the number of the currently displayed page, so that the next 20 documents are retrieved. The maximum number of documents was chosen because the relevance of the documents after this number was very low and therefore it did not make sense to display them. The user can sort the documents by relevance and by date, from most recent to least recent. To do this, click on the drop-down list on the right-hand side of the interface and select the desired option. In Figure 4.3 you can see the results for the search *"Cláusula Verbal"* sorted according to the most recent date.



Encontrado(s) 2500 resultado(s).

ECLI:PT:TRG:2022:4472.19.1T8GMR.G1
Tribunal: Tribunal da Relação de Guimarães **Relator:** Afonso Cabral De Andrade **Data:** 07/04/2022
Descritores: Contrato Promessa / Nulidade De Sentença / Impugnação Da Matéria De Facto / Anulabilidade Por Erro / Retroactividade / Essencialidade Do Erro
▸ Sumário:

ECLI:PT:TRG:2022:449.21.5T8VCT.G1
Tribunal: Tribunal da Relação de Guimarães **Relator:** Alcides Rodrigues **Data:** 06/04/2022
Descritores: Transacção Judicial
▸ Sumário:

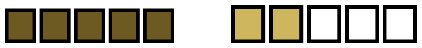
ECLI:PT:TRL:2022:1047.21.9T8TVD.L1.4
Tribunal: Tribunal da Relação de Lisboa **Relator:** Sérgio Almeida **Data:** 06/04/2022
Descritores: Despedimento / Indemnização Por Antiguidade
Texto não disponível
▸ Sumário:

Figure 4.3: Search results for "*Cláusula Verbal*"

4.1.2.A Metadata

Each result is a different document relating to the words entered. It contains information that allows you to determine which court the document belongs to, who its author is and what descriptors it contains. It also contains a summary of the document so that the user can read the topics to which the document refers.

The first line of a result contains an indicator that shows the relevance of the document to the search. Relevance is measured by five bars that are filled with the same colour, but vary in lightness depending on the relevance value. Each bar represents 20%, i.e. if the relevance of a document is 90%, five bars are painted, at 30% only two bars are painted.



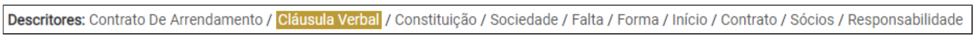
(a) Relevance between 80% and 100%

(b) Relevance between 20% and 40%

Figure 4.4: Search relevance.

Next to the relevance indicator is the title of the document. It is a link that opens a separate tab in the browser with the full document information. See section 4.2.4 - Document's page, for more details on this topic.

Below this line are the fields that allow the user to identify the document, e.g. the court, the author, the date of the document and the list of descriptors. If the latter contains words that have been searched for, the background colour is changed to gold. In this way, you can quickly find documents that cover specific topics. In the figure below you can see an example of a search for "*Cláusula Verbal*", where the words of a document have a different background colour.



Descritores: Contrato De Arrendamento / Cláusula Verbal / Constituição / Sociedade / Falta / Forma / Início / Contrato / Sócios / Responsabilidade

Figure 4.5: List of descriptors of a document when searching for "*Cláusula Verbal*".

The last section for a result is the summary of the document. This section works like a drop-down list. When the user clicks on it, the box expands to show the summary of the document. If it contains the words searched for, the background colour of these words changes to gold so that they can be easily identified, similar to the display of the descriptors.

4.1.2.B Visualisation

For each result, there is also a visualisation that allows the user to analyse the number of occurrences of the searched words and their distribution in the documents. This visualisation consists of two horizontal bars in different colours. The first gold-coloured bar represents the summary part of the document and the second black line represents the full text of the document.

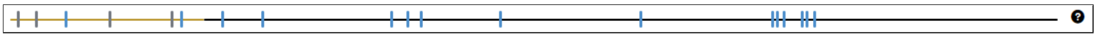
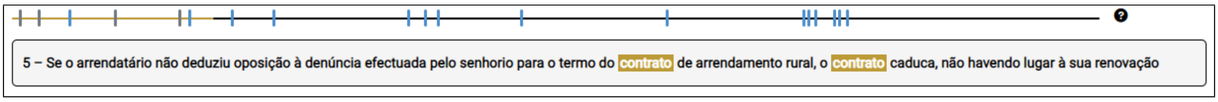


Figure 4.6: Visualization of a document when searching for "*Denúncia de Contrato*".

After the client side has received the documents returned by the server, the search is filtered so that stop words are not shown in the visualisation or in the colour legend. Then, depending on the relative position of the words in the document, vertical bars are superimposed on the horizontal ones. The vertical bars are also coloured differently so that they are easy to recognise. There are five different colours, but it is possible to have even more vertical bars in the visualisation. However, they will then be coloured black. The reason for this is that users do not usually look for long sentences in the text, so it did not make sense to use more than five colours.

In Figure 4.6 it is possible to analyse a visualisation of a document when searching for "*Denúncia de Contrato*". As you can see, the search has been filtered and the word "*de*" does not appear and the remaining words have been assigned different colours (grey for "*Denúncia*" and blue for "*Contrato*").

Another feature of this visualisation is the possibility to display an excerpt from the text in which the word occurs. The user only needs to move the mouse pointer over a vertical bar and a box with the extract is displayed. Similarly to the list of descriptors and the document summary, the words searched for in the extract also have the golden background colour to help locate them in the text, as seen in Figure 4.7. This excerpt feature allows the user to analyse the context in which the searched words are inserted and to understand the relevance of the discussed topics in the document.

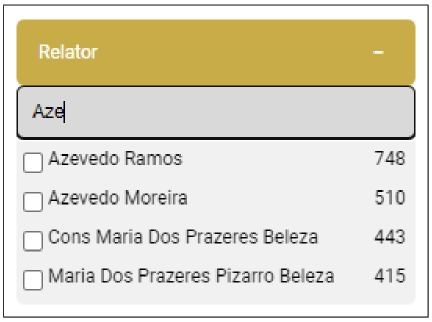


5 – Se o arrendatário não deduziu oposição à denúncia efectuada pelo senhorio para o termo do contrato de arrendamento rural, o contrato caduca, não havendo lugar à sua renovação

Figure 4.7: Excerpt of the text when searching for "*Denúncia de Contrato*". The hovered bar is blue, meaning the word that is highlighted is "*Contrato*".

4.1.3 Filters

To enable a more specific search, the user can apply filters before or after searching for documents. These filters consist of drop-down lists that correspond to the various existing fields in the database. The filter for the date is divided in three parts: a chart showing the number of documents for a given year, a slider that allows you to select an interval of years, and a date picker that allows you to select a specific interval of days, months or years.



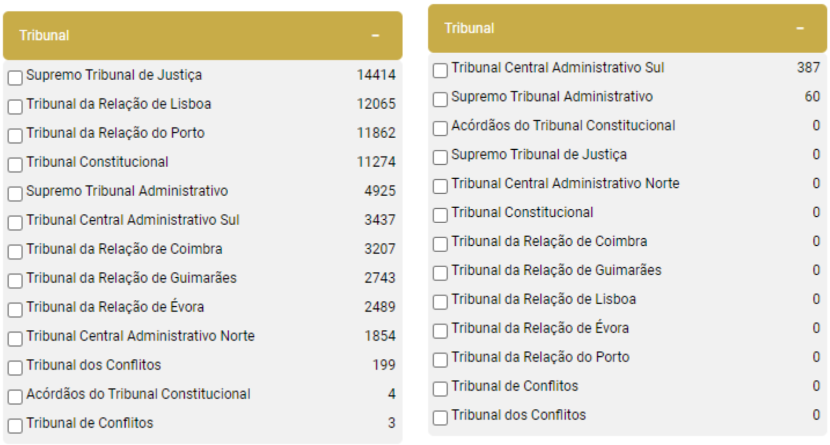
Aze

□ Azevedo Ramos	748
□ Azevedo Moreira	510
□ Cons Maria Dos Prazeres Beleza	443
□ Maria Dos Prazeres Pizarro Beleza	415

Figure 4.8: Filtered author list with document count.

4.1.3.A Drop-down Filters

The drop-down filters allow you to select multiple values from different fields, e.g. court, section of the court, author, voting decision, procedural means and descriptor. When the user selects one of these dropdown lists, it expands to show the different options for filtering the search. Each list contains twenty options corresponding to the number of results per page. At the end of these options is a field with the number of options that were not displayed. Each option contains the number of documents for each value of the filters. When the user selects a specific value, the number of documents for each option is updated with the number of documents available for the new search (Figure 4.9). This gives the user an idea of the dimension of the search. All drop-down lists except the court filter also have a search field where the user can enter terms. The list is then updated with the options that match the input value entered (Figure 4.8).



Tribunal	–
□ Supremo Tribunal de Justiça	14414
□ Tribunal da Relação de Lisboa	12065
□ Tribunal da Relação do Porto	11862
□ Tribunal Constitucional	11274
□ Supremo Tribunal Administrativo	4925
□ Tribunal Central Administrativo Sul	3437
□ Tribunal da Relação de Coimbra	3207
□ Tribunal da Relação de Guimarães	2743
□ Tribunal da Relação de Évora	2489
□ Tribunal Central Administrativo Norte	1854
□ Tribunal dos Conflitos	199
□ Acórdãos do Tribunal Constitucional	4
□ Tribunal de Conflitos	3

(a) Court filter before selecting an option.

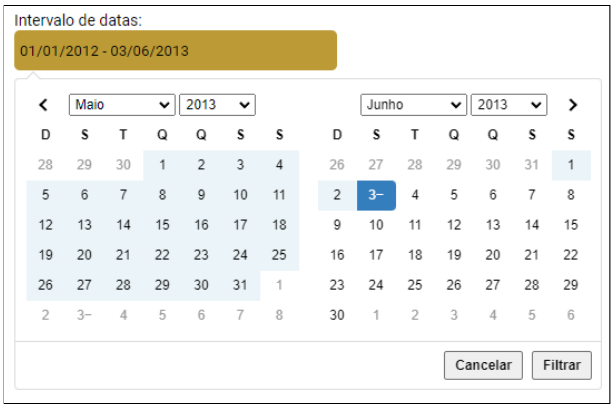
Tribunal	–
□ Tribunal Central Administrativo Sul	387
□ Supremo Tribunal Administrativo	60
□ Acórdãos do Tribunal Constitucional	0
□ Supremo Tribunal de Justiça	0
□ Tribunal Central Administrativo Norte	0
□ Tribunal Constitucional	0
□ Tribunal da Relação de Coimbra	0
□ Tribunal da Relação de Guimarães	0
□ Tribunal da Relação de Lisboa	0
□ Tribunal da Relação de Évora	0
□ Tribunal da Relação do Porto	0
□ Tribunal de Conflitos	0
□ Tribunal dos Conflitos	0

(b) Court filter after selecting an author. The document count is updated.

Figure 4.9: Court Filter.

4.1.3.B Date filters

The date filters' interface is different from the others. The date data type requires a different type of interface since it does not work as the others. Therefore, the filter uses three different methods, but they update each other and the other filters according to the new results received. The first method is a date picker. The user can select a specific day, month and year for a start and end date. Then the server returns the results that correspond to that time interval. The user can easily jump between months and years by selecting the desired value in the drop-down lists at the top of the date picker (Figure 4.10).



Intervalo de datas:

01/01/2012 - 03/06/2013

Maio 2013

D	S	T	Q	Q	S	S
28	29	30	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	1
2	3–	4	5	6	7	8

Junho 2013

D	S	T	Q	Q	S	S
26	27	28	29	30	31	1
2	3–	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	1	2	3	4	5	6

Figure 4.10: Selected documents from January 1 2012 and June 3 2013.

The next method has two features. It displays the number of documents for a given interval of years as a bar chart and allows the user to request documents for a given year from the server by clicking on a single bar, updating the information on the other filters of the page. In the image below we see the number of documents for the search term ”Contrato”.

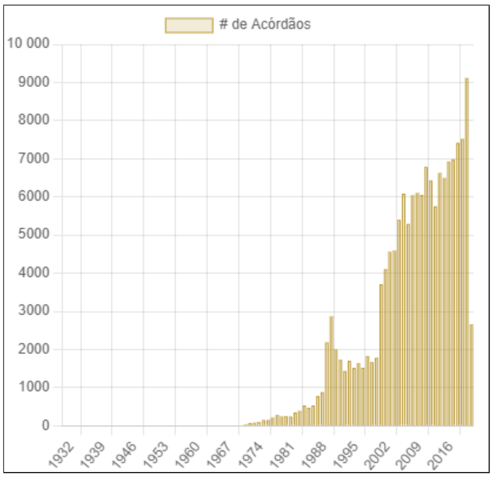


Figure 4.11: Bar chart with the number of documents for the search term ”Contrato”.

The final method of filtering by date is a range slider. The user can move two handles on opposite sides of the slider and the results will update according to the years selected. The smallest year is 1900 and the largest year is the current year. Above each handle is a tooltip that shows which year is selected. They are updated as the handles move. In the image below, the user has requested documents from 2008 to 2018, as the tooltips show.

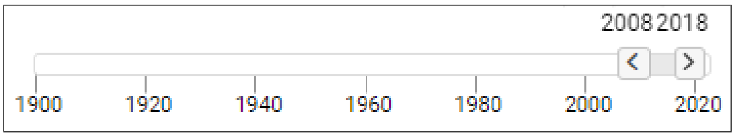
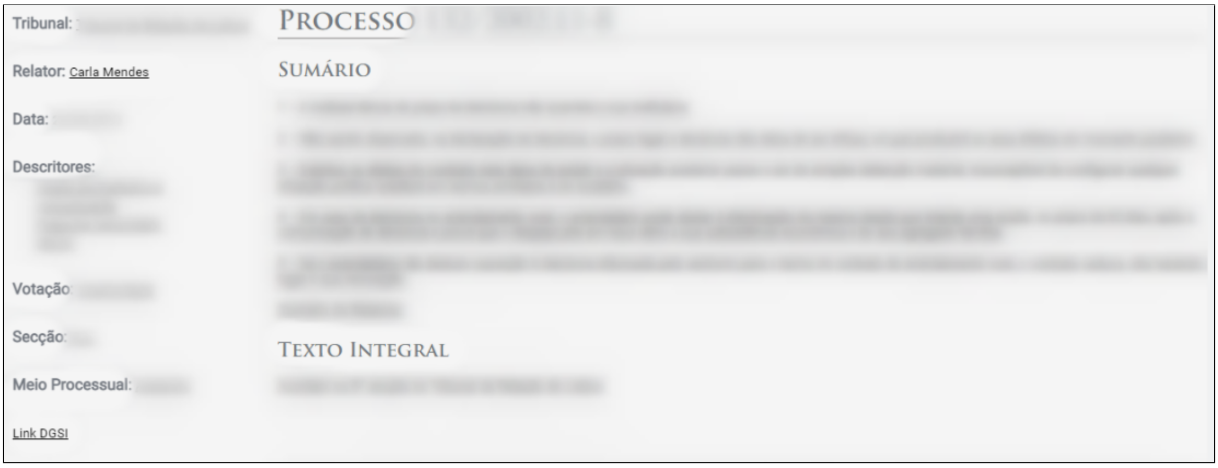


Figure 4.12: Slider range for years between 2008 and 2018.

4.1.4 Document's page

As described in section 4.2.2.A - Metadata, each result title is a link that opens a new tab in the browser with the full information of the document. On this page, the user can see all the metadata associated with the particular document, as well as the summary and full text. In this way, the user can fully explore the content of a document and use it at their own discretion. On the left side of the summary and full text is a column with the rest of the metadata, i.e. court, author, etc. Each value of these fields is a link and acts as a filter, as the user can click on it and the page returns to the search interface, where a new search is performed with the specific filter. For example, in Figure 4.13 the user has a document open whose author is "*Carla Mendes*". When the author's link is clicked, the client requests more documents by the same author. There is also a link below these fields that takes the user to the document's page on the DGSI website.



Tribunal:

Relator: Carla Mendes

Data:

Descritores:

Votação:

Secção:

Meio Processual:

Link DGSI

PROCESSO

SUMÁRIO

TEXTO INTEGRAL

Figure 4.13: Document page for process 132/2002.L1-8.

4.2 Architecture

Using the list of requirements described in section 3.1, it was possible to design the architecture of the application. As we can see in Figure 4.14, the solution is divided into two main components: Front-End and Back-End.

4.2.1 Back-End

As you can see in the figure below, the back-end is divided into two sections: Server and Database. The database contains all the information that needs to be displayed to the user, while the server is responsible for handling requests from the client and communicating with the database to get the required information and return it to the user. The server uses ElasticSearch's API to execute the queries to the database and receives a JavaScript Object Notation (JSON) object from it, which is sent to the client. ElasticSearch[1] is a powerful search engine that allows users to perform several tasks such as search for documents, aggregations and document count or even log analysis. It works with several programming languages such as Python, JavaScript, PHP, C#, etc. It has an API that allows users to accomplish the tasks mentioned before. To access said API, we need to create an instance of ElasticSearch in the server using a node with the URL where the database is located.

4.2.2 Front-End

The front-end processes the events performed by the user (e.g. clicking on a filter or searching for a term) and displays the information received from the server. The communication between the client and the server is done through REST API requests and the interface is built based on the JSON object mentioned above. REST API is used because it is easy to implement and works well with the JSON format. The next section explains the implementation of this architecture in more detail.

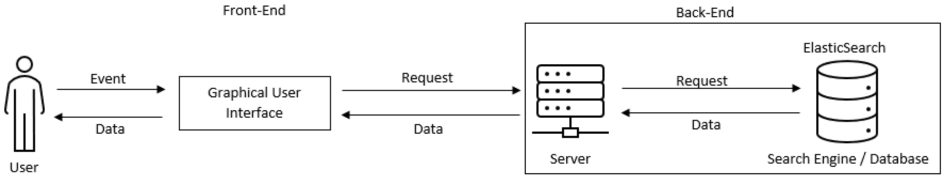


Figure 4.14: Digital library's architecture.

4.3 Implementation

Once the meetings with the judges of the Portuguese Supreme Court were completed and the final design was decided, a prototype was developed. The front-end component of the digital library was implemented using HTML, CSS and JavaScript. In this way, a simple but effective interface could be created. Although a JavaScript framework such as React.js or Vue.js could have been used, I had trouble implementing the filters and document count for each option, because when an option was selected, the remaining ones with zero documents were hidden. We did not want this to happen, since the goal of the

[1] https://www.elastic.co/pt/

filters was to always display their options even if their document count was zero to prevent the user from losing context of the existing filters. The backend component was developed using Node.js[2] and the Express.js[3] Framework. This framework is suitable for simple solutions and is very powerful. It provides the necessary tools to efficiently process requests and return the required information, facilitating the communication between the server and the database. EJS[4] was also used to display the page of the documents. This tool allows us to have a predefined structure of an HTML page (called a template) and fill it with the required information when a document is found. This was helpful to create the page template and serve it directly from the server without overloading the front-end. We chose EJS because it is a simple tool that is easy to learn and use. There was another option, *Pug*[5], but it has a steeper learning curve. ElasticSearch was also used as a database to index all documents retrieved from the DGSI website. It provides powerful and efficient search tools that can be useful in the context of this work, such as its API that was used to make calls to the database and get the desired results returned to the frontend.

4.3.1 Backend

During the development and testing phase, the server was hosted at INESC-ID. This is a temporary solution to share the progress of the development between the group members while the digital library is not yet fully developed. The goal is to host the application on the Portuguese Supreme Court servers so that it is only available to the judges.

To better organise the project, the back-end was split into two components: `server.js` and `search.js`. The `server.js` component receives the requests from the client and forwards them to the `search.js` component. In order to make the queries to the ElasticSearch database, several parameters were required, such as the search term or the filters to apply. So there were a lot of variables to analyse. Originally, several nested if statements were used to test the different possible combinations of filters. However, this was extremely inefficient. Therefore, the server analyses the parameters sent in the client's request by checking whether they are present or not, and then passes them on to the API as a JSON object. This process is used as well to check the current page of results.

The only type of request used is the GET request, so it is only possible to retrieve information and not add it to the database. When you made a request to the server, sometimes the results came back empty. This was because the server sent the results before ElasticSearch had finished retrieving the documents. To solve this problem, the queries were made asynchronously. This means that the system waits until the data has been retrieved from the database before sending it. This improves efficiency and

[2] https://nodejs.org/en/
[3] https://expressjs.com/
[4] https://ejs.co/
[5] https://pugjs.org/api/getting-started.html

ensures that the results are always ready to be sent to the client.

The two main difficulties on the backend of the project were setting up an Express.js server and connecting to the ElasticSearch API, since I had to familiarize myself with these two tools.

4.3.1.A GET - /api/

This is the path that is called when the user wants to retrieve documents from the database. The endpoint analyses the parameters of the request to check whether they exist or not. These parameters are then added to a created object that is sent as an argument to the *ElasticSearch*'s search[6] API. A JSON object is then returned with the documents (called ”hits”) as well as the aggregations containing the number of documents for each query.

4.3.1.B GET - /api/open/:ECLI

This path is used to render the page of the document that corresponds to the *ECLI* title given as the query parameter. It uses the same API as the one mentioned above to search for the document. Using the EJS library, it then fills the `document.ejs` file containing the template of the page with the information of the found document. A new tab is opened in which the template is already filled in and stylised.

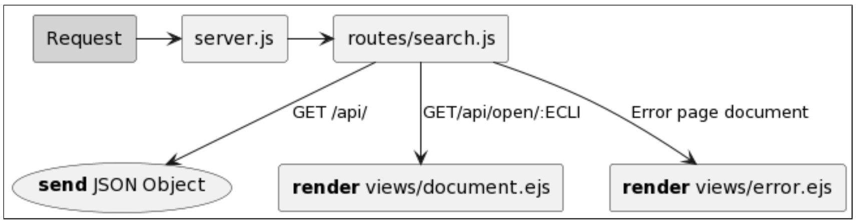


Figure 4.15: Digital library's server components.

4.3.2 Database

As mentioned earlier, ElasticSearch was used to create the database. ElasticSearch is a distributed search and analytics engine built on Apache Lucene[7]. It accepts JSON queries and returns the results in JSON format as well. We chose this tool because it is fast, scalable and provides the necessary tools to efficiently search for documents and analyse their information. This system contains all the documents retrieved from the DGSI website and indexes multiple fields for each document, as can be seen in Figure 4.16.

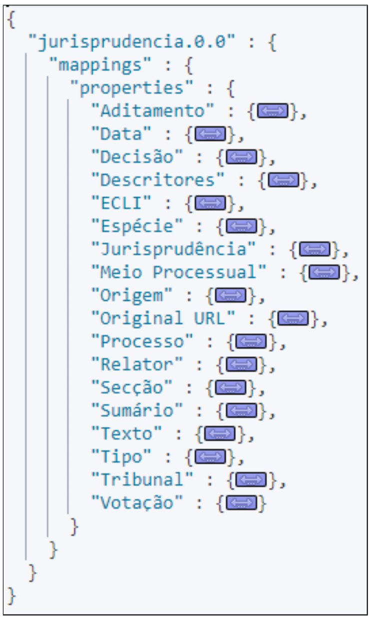
At the beginning of the development phase, the documents were divided into several indexes representing the different Portuguese courts. However, this was difficult to handle when several courts were

[6] https://www.elastic.co/guide/en/elasticsearch/reference/current/search-search.html

[7] https://lucene.apache.org/

selected. Therefore, in order to facilitate the search for these documents, they were combined into a single index. At the beginning, a single node is created with a specific URL. At the moment, the URL given is *localhost:9200*, but it can be changed to a better option in the future. This node could also be protected with ElasticSearch's authentication features. However, it has not been implemented for this development as it is easier to test the different versions of the program. This feature can be added to the database at a later date for security reasons.

After we had set up the database, we had to figure out how queries were made to it. As mentioned earlier, ElasticSearch receives queries in the form of a JSON object. Within this object there are several fields that indicate where the search needs to take place and what to search for. We needed to ensure that the search would be performed, but at the same time guarantee that the filters would be applied when selected. The only solution we found was the boolean query. In this query, operators (”*AND*”, ”*OR*”, ”*NOR*” and ”*NAND*”) can be mixed to get better results. In this case, the field `must` is used to ensure that the searched terms occur in the document. It also ensures that when searching for two or more terms, they all occur in the same document. This field is the equivalent of the operator ”*AND*”.



```
{
  "jurisprudencia.0.0" : {
    "mappings" : {
      "properties" : {
        "Aditamento" : {↔},
        "Data" : {↔},
        "Decisão" : {↔},
        "Descritores" : {↔},
        "ECLI" : {↔},
        "Espécie" : {↔},
        "Jurisprudência" : {↔},
        "Meio Processual" : {↔},
        "Origem" : {↔},
        "Original URL" : {↔},
        "Processo" : {↔},
        "Relator" : {↔},
        "Secção" : {↔},
        "Sumário" : {↔},
        "Texto" : {↔},
        "Tipo" : {↔},
        "Tribunal" : {↔},
        "Votação" : {↔}
      }
    }
  }
}
```

Figure 4.16: Database mappings.

In order to search for the terms entered, we had to specify the fields in which to search for them and ensure that they all occur in the document if possible. Initially, the ”`multi_match`”[8] query was used. This analysed the text to find and match the searched terms in multiple fields. However, since no search operators could be specified in this query, the search was performed with the ”`query_string`”[9]. This query allows documents to be returned based on a given string and uses a parser to parse it using operators such as ”*AND*”. It then analyses the words independently. This query provides a field called ”`default_operator`” where you can specify the default search operator that the query uses to find the words. In this case, the operator ”*AND*” has been defined.

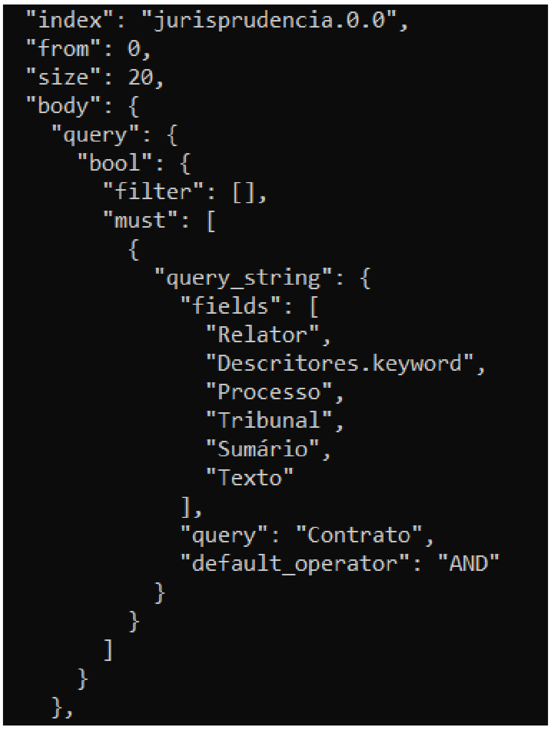
After searching for the terms, we need a way to refine the search. In earlier stages of development, the field ”*filter*” was not used. Instead, we tried to do multiple ”`match`” queries within the main query. This attempted to match the different filters with the specified fields. However, since this method did not give good results from the database and returned errors most of the time, the field ”*filter*” was used within the bool query. This field accepts two types of filters: the ”*range*” and the ”*terms*”. The first is used to

[8] https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-multi-match-query.html

[9] https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-query-string-query.html

filter the results by a specific date interval and the second is used for the other type of filters. This field allows you to search for multiple terms within a specific database field.

Ordering the results, either by relevance or by date, is an important issue in the drafting of decisions according to Portuguese Supreme Court judges. It is important to check the date of the document to understand if it is really relevant to the case being analysed. ElasticSearch offers only one option to sort the results, namely the `sort` field. In the first phase of development, there was no option to sort the documents, as the use of this ElasticSearch feature prevented the score of the documents from being sent to the client. However, after researching the ElasticSearch documentation, the flag `track_scores` was discovered which, when set to true, allows the score to be sent. In Figure 4.17 an example of a query to the database can be seen.



```
"index": "jurisprudencia.0.0",
"from": 0,
"size": 20,
"body": {
  "query": {
    "bool": {
      "filter": [],
      "must": [
        {
          "query_string": {
            "fields": [
              "Relator",
              "Descritores.keyword",
              "Processo",
              "Tribunal",
              "Sumário",
              "Texto"
            ],
            "query": "Contrato",
            "default_operator": "AND"
          }
        }
      ]
    }
  },
```

Figure 4.17: Example of a query.

In terms of document scoring, when a search is performed, ElasticSearch automatically calculates the score for each document, which represents its relevance to the search. This score is calculated using Lucene[10]'s Practical Scoring Function[11]. This function is based on the Term Frequency/Inverse Document Frequency (TF/IDF) model, i.e. it takes into account the number of times the terms appear in the document and the total number of documents in the index to determine the documents' score.

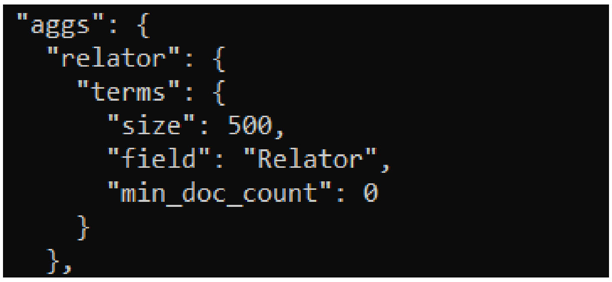
Since the database retrieves hundreds or even thousands of documents from a single search, it does not make sense to display them all on a single page as it would be too long and take a long time to load. There was the possibility of setting up an infinite scrolling feature, similar to the social media applications. However, this would not work in this case as it would be extremely slow to make multiple queries, so it was important to create the pagination feature. Pagination means dividing the data into pages. In this case, we divide the 500 results into 25 pages. To do the pagination, the server requests 20 documents per request because the performance is limited and it is easier to display on the screen. This is done using the *from* and *size* fields. The first allows the selection of the index number from which the database should start searching and the second specifies the number of documents to be sent to

[10] https://lucene.apache.org/

[11] https://www.elastic.co/guide/en/elasticsearch/guide/current/practical-scoring-function.html

the client. When the user changes the page by clicking on the buttons at the bottom of the page, a new request is made to the database with the new number of the current page.

The number of documents is also an important topic to display when performing a search and using the filters, as it provides context to the user. Every option of the filters has a document count indicating the number of documents associated with that value. To do this, ElasticSearch's aggregations are used. They consist of several groups of values based on the different fields and related to the search performed, with a document count associated with each value. Originally, the number of values for each aggregated group was 1000, but due to performance limitations when entering terms in the filters' search field, the number had to be reduced to 500. When a new query is made to the database, the aggregations are calculated according to the terms and filters entered.



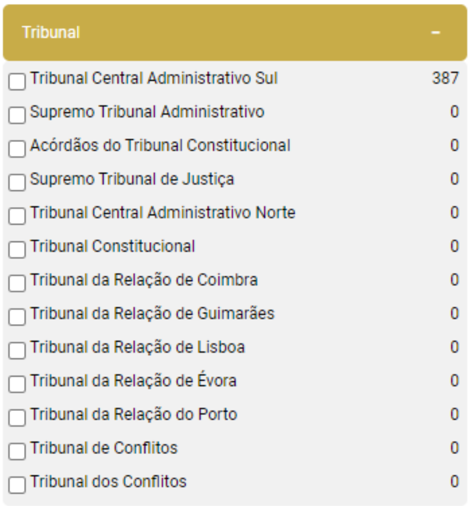
```
"aggs": {
  "relator": {
    "terms": {
      "size": 500,
      "field": "Relator",
      "min_doc_count": 0
    }
  },
```

Figure 4.18: Example of an aggregation request used in this application.

When multiple options of a filter of the same category are selected, we wanted to use the *OR* operator in the search so that the database can retrieve documents that match the options selected. For example, for the court filter, when a user selects the option ”Supremo Tribunal de Justiça” and ”Supremo Tribunal Administrativo”, the database must be able to find documents that have the first or the second options. When filters of different categories are selected, we wanted to use the *AND* operator, in order to make sure that the retrieved documents have the filters associated to them. For example, for the court and author filters, when a user selects the court option ”Supremo Tribunal de Justiça” and the author option ”Francisco Rothes”, the database must be able to find all the documents of that court that were written by that author.

With this concept in mind, the original approach had to be altered because for every filter (except the descriptor), the document count of each unselected option was set to zero when a new option was chosen. For example, for the court filter, if a user selected the option ”Supremo Tribunal Administrativo Sul”, the document count of the remaining options of the filter was set to zero (Figure 4.19). This is due to the fact that each document only has one court associated to it, so when an option is selected, the aggregation ignores the remaining options. When a second court is selected, the aggregations count the number of documents that have the first court associated to it or the second.

For this reason, the document count for filters of the same category was incorrect, as there were documents available, despite showing the number zero. To fix this, the ElasticSearch's `post_filter` feature was used. Instead of creating the aggregations with the filtered search, this feature creates the aggregations first and then performs the filtering, ensuring that the options for filters of the same category remain with the correct document count and the remaining filters are updated.



Tribunal	–
□ Tribunal Central Administrativo Sul	387
□ Supremo Tribunal Administrativo	0
□ Acórdãos do Tribunal Constitucional	0
□ Supremo Tribunal de Justiça	0
□ Tribunal Central Administrativo Norte	0
□ Tribunal Constitucional	0
□ Tribunal da Relação de Coimbra	0
□ Tribunal da Relação de Guimarães	0
□ Tribunal da Relação de Lisboa	0
□ Tribunal da Relação de Évora	0
□ Tribunal da Relação do Porto	0
□ Tribunal de Conflitos	0
□ Tribunal dos Conflitos	0

Figure 4.19: Initial approach of the document count implementation.

4.3.3 Front-End

As mentioned above, the digital library was implemented using HTML, CSS and JavaScript. After a request is made to the server, the front-end creates all the elements displayed on the page. All elements are created using native JavaScript, except for the time filters. Since there was no easy way to create these filters natively, the jQuery UI[12] library was used. This library provides the necessary tools to create the filters and process the various events as their state changes. The bar chart was created using the *Chart.js*[13] library. This allowed to quickly create a chart with the information available. The requests to the server are made through the axios[14] API. This API was implemented because it has proven to be easy to use and provides backward compatibility for browsers. Requests to the server are always asynchronous to ensure that results are retrieved before they are displayed to the user on the screen. To achieve this, JavaScript's `async` and `await` functions were used.

The interface needed to be visually appealing similar to the ECLI website. For this purpose, the color scheme and fonts of this website were initially used. However, the group members of the IRIS project felt that since the interface will be used by judges of the Portuguese Supreme Court, it should be similar

[12] https://jqueryui.com/
[13] https://www.chartjs.org/
[14] https://axios-http.com/docs/intro

in color and font to the Portuguese Supreme Court website. Therefore, the fonts TrajanPro[15] is used for titles and Lato[16] is used for the rest of the page's body. For the colors, gold and black are used. This ensures that consistency is maintained between the websites.

Displaying the relevance of a document to the search is considered important because it can be a deciding factor in navigating the results list. Originally, relevance was displayed as a percentage. However, after a discussion with the group members of INESC-ID, it was agreed that the different percentages do not make a difference when selecting a document, so a new approach was introduced. The relevance of each document consists of five `<div>` tags, which are painted according to the value of the score returned by the database. In the first iteration of this approach, the colour scale used was red (least relevant), orange, yellow, dark green and green (most relevant). However, as there were problems with distinguishing the colours and several people from the INESC-ID group, who are colour blind, stated that it was difficult to distinguish the colours, the implementation of the relevance had to be changed. So the colour was changed to the same golden colour used throughout the user interface and the lightness of the colour represents the relevance of the document. The higher the lightness, the higher the relevance.

The choice of colour scheme for the visualisation was also difficult, as there were not many colours that together made the visualisation easy to interpret. Several colours were tested until they were chosen. Originally, there was only one horizontal bar in black, representing the summary and full text of the document. This was changed after group members from INESC-ID explained that it was better to split the bars in order to distinguish where the words were. So the horizontal bar was divided into two and the bar representing the summary was given a golden colour.

When performing a search, the visualisation may contain a large number of vertical bars. To calculate their position, the relative position of the word in the text is used. The position is then converted into a percentage value and is used to position the vertical bar in the visualisation. This is not the most efficient way to create a visualisation. However, this was the best option found.

Each time a word occurs, the excerpt of the text in which it occurs must also be displayed. Originally, the excerpt was only displayed when a vertical bar was clicked, but this method was not very intuitive as there was no indication on the user interface that the bars could be clicked. Therefore, the approach was changed so that the excerpt is displayed when the mouse is hovered over a bar. This is a more intuitive way of displaying the text because the first thing users do is hover the mouse over elements, as found in the usability tests.

The filters are one of the most important parts of the user interface and it is important that they are visible and easily accessible to users. Originally, the filters were arranged in a drop-down list next to the search box, which displayed their options when the user hovered over it. However, this made users lose track of what was already selected and forced them to hover over the list again to review it. With the

[15] https://fontsgeek.com/fonts/Trajan-Pro-Regular

[16] https://fonts.google.com/specimen/Lato

current approach, each filter is an expandable <div> that displays the options when the user clicks on it. This way, the filter can always be open so that the user can see which options have been selected. Each filter consists of a list of twenty options to keep the consistency between the number of results per page. This number of options also helps to keep the user interface clear and easy to understand. Below these options is a field that shows the number of options not displayed, so that the user knows how many have been retrieved from the server.

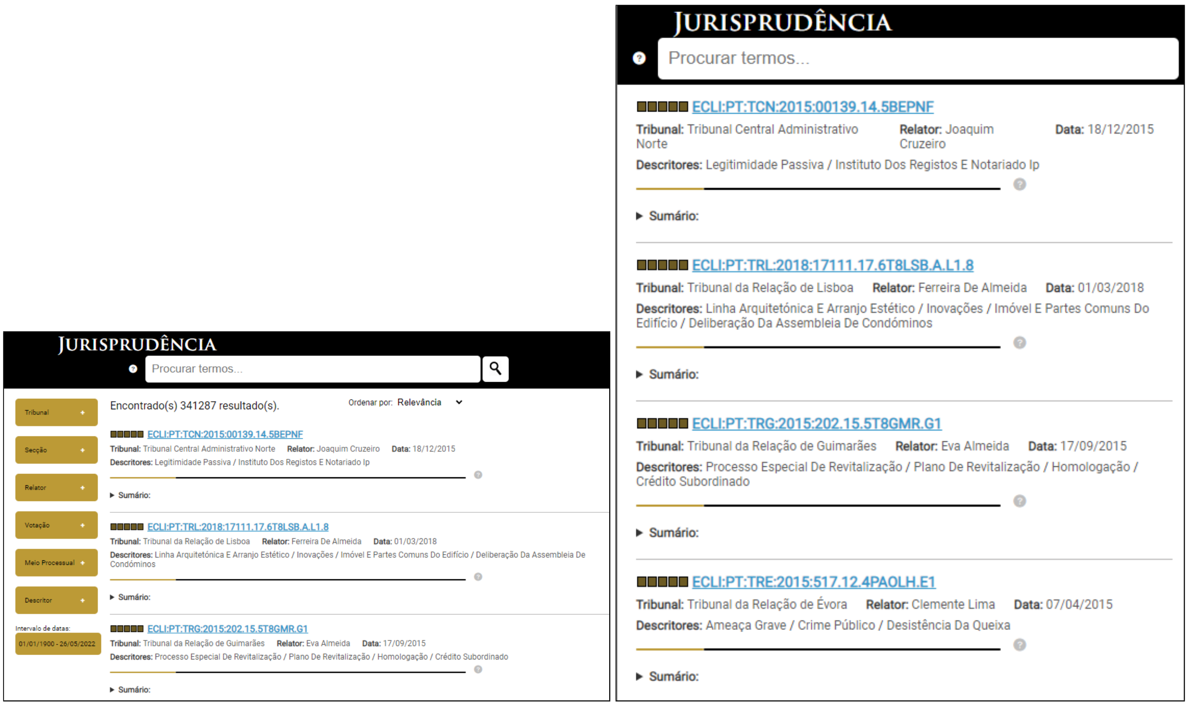
In addition to these twenty options, each filter also has an input field, as mentioned earlier. This was done to avoid having too much information on the screen and to make it easier to access the different values of the filters. When a user starts typing in the field, the values of the filters that match the input appear. This is done by creating a regular expression with the typed input and matching it with the different options of the filters. The development of this feature went through several phases. The first was to perform queries to the server using the ElasticSearch search API with a `wildcard` for each character typed by the user, but this was quickly discarded due to performance issues. The second phase was to create an HTML <datalist>. After the first twenty values were added to the filters, the rest were added to this list, which was displayed with the options that matched the input when the user started typing. However, when the user started entering terms, the context of the filters was lost and they took up a lot of space on the screen. Therefore, the last stage was implemented.

The document page contains all the information about a single document, therefore a lot of information is displayed on the screen. The layout of this page was also based on the ECLI website. The data is presented in an attractive way and allows the user to quickly find information about the document. In the area next to the summary and the full text, the user can perform another search by clicking on the link of the respective filter. This makes a new call to the server with the value of the filter.

Judges also found it useful to have an option to find similar documents to those found with a search, so that they have access to documents that are related to each other or to a particular topic. Originally this was done using the `more_like_this` function of ElasticSearch. Similar to the document score, this function is also based on the TF/IDF methodology to find the most similar documents. If a term occurs frequently in a large number of documents, it is considered similar by the function. However, this function did not give good results and no solution to this problem was found, so the feature had to be removed from the digital library.

4.3.3.A Responsive Design

To ensure that the user interface is suitable for different devices such as computer, tablet, mobile phone, etc., the <meta> tag was added along with the `viewport` method. This helps in adapting the information to the size of the screen. Also, the `@media` rule for multiple screen widths has been added in the CSS file. This allows you to show or hide multiple elements on the user interface depending on the size of the screen. In Figure 4.20 we can see how the interface adapts to the different widths of the screen. For the width of 950px, the bar chart and the slider are hidden because they become hard to interpret and use. For a width of 600px, all filters are hidden.



Encontrado(s) 341287 resultado(s).

■■■■■ ECLI:PT:TCN:2015:00139.14.5BEPNF
Tribunal: Tribunal Central Administrativo Norte **Relator:** Joaquim Cruzeiro **Data:** 18/12/2015
Descritores: Legitimidade Passiva / Instituto Dos Registos E Notariado Ip
▶ **Sumário:**

■■■■■ ECLI:PT:TRL:2018:17111.17.6T8LSB.A.L1.8
Tribunal: Tribunal da Relação de Lisboa **Relator:** Ferreira De Almeida **Data:** 01/03/2018
Descritores: Linha Arquitetónica E Arranjo Estético / Inovações / Imóvel E Partes Comuns Do Edifício / Deliberação Da Assembleia De Condóminos
▶ **Sumário:**

■■■■■ ECLI:PT:TRG:2015:202.15.5T8GMR.G1
Tribunal: Tribunal da Relação de Guimarães **Relator:** Eva Almeida **Data:** 17/09/2015
Descritores: Processo Especial De Revitalização / Plano De Revitalização / Homologação / Crédito Subordinado
▶ **Sumário:**

■■■■■ ECLI:PT:TCN:2015:00139.14.5BEPNF
Tribunal: Tribunal Central Administrativo Norte **Relator:** Joaquim Cruzeiro **Data:** 18/12/2015
Descritores: Legitimidade Passiva / Instituto Dos Registos E Notariado Ip
▶ **Sumário:**

■■■■■ ECLI:PT:TRL:2018:17111.17.6T8LSB.A.L1.8
Tribunal: Tribunal da Relação de Lisboa **Relator:** Ferreira De Almeida **Data:** 01/03/2018
Descritores: Linha Arquitetónica E Arranjo Estético / Inovações / Imóvel E Partes Comuns Do Edifício / Deliberação Da Assembleia De Condóminos
▶ **Sumário:**

■■■■■ ECLI:PT:TRG:2015:202.15.5T8GMR.G1
Tribunal: Tribunal da Relação de Guimarães **Relator:** Eva Almeida **Data:** 17/09/2015
Descritores: Processo Especial De Revitalização / Plano De Revitalização / Homologação / Crédito Subordinado
▶ **Sumário:**

■■■■■ ECLI:PT:TRE:2015:517.12.4PAOLH.E1
Tribunal: Tribunal da Relação de Évora **Relator:** Clemente Lima **Data:** 07/04/2015
Descritores: Ameaça Grave / Crime Público / Desistência Da Queixa
▶ **Sumário:**

(a) Interface layout for a width of 950px. **(b)** Interface layout for a width of 600px.

Figure 4.20: Different layouts according to the screen width.

5

Evaluation

Contents

5.1 User Evaluation 61

5.2 Results 62

5.3 Changes to the Interface 66

5.4 Discussion 68

In this chapter, the user tests and their results are described and discussed in detail. To evaluate the application, meetings with several judges were held in the Portuguese Supreme Court building. The aim of these tests was to find out whether the digital library meets the requirements stated in section 3.1 and to collect the judges' opinions about the user interface and the general experience with it. These tests were also useful for finding bugs to fix at a later date, and for getting feedback and suggestions on specific features that the judges might like to see implemented. The judges were asked to complete a number of tasks, which are mentioned in the next sections. The tests were conducted using the version of the interface found at this URL: `https://pe.inesc-id.pt/`. This was the most recent version of the project. After the tests were completed, some changes were made according to user feedback. These changes are also explained in more detail in the following sections.

5.1 User Evaluation

As mentioned earlier, meetings were held to conduct the user tests. These tests were conducted with ten different judges of the Portuguese Supreme Court. Nine of the ten tests were conducted individually on each judge's computer in their respective office, as there was no way for computers from outside the Portuguese Supreme Court to access the Wi-Fi network and no internet hotspot was available and the last test was performed on my laptop. The judges were asked to perform a series of seven tasks and the time it took them to complete the task and whether the answer was correct or not were collected. These tasks were to cover all the functionalities of the digital library user interface.

At the beginning of each test session, there was a brief explanation of the digital library and how it was integrated into the scope of the IRIS project. After that, the judges had five minutes to use the functions of the digital library and familiarise themselves with how it worked. Then the test began. The list of tasks is given below.

- **Task 1** - Indicate the date of the oldest document that refers to the word ”Contract”.
- **Task 2** - For the first document in which the word ”Debt” appears, indicate in which section of the court the decision was made.
- **Task 3** - Identify the author who has the most documents in which the word ”Forfeiture” occurs.
- **Task 4** - For the word ”Contract” and documents of the Supreme Administrative Court between January and February 2015, indicate the last descriptor of the second document.
- **Task 5** - For the word ”Debt” in the second occurrence of the word of the first document, indicate the value of the debt in euros.
- **Task 6** - How many documents refer to the word ”Guilt” for the year 2021?

- **Task 7** - Indicate to which court the most recent document referring to the word ”Pawn” belongs.

Each task is independent, which means that the result of one task does not depend on the result of the previous task. For each individual test session, the order of the tasks was randomised and each task started on the home page of the user interface because we did not want all users to perform the tasks in the same order since this could alter the results (learning effect[1]). Some tasks are expected to be easy, such as task number 2, number 3 or number 7, due to the fact that judges may already be familiar with the user interface before starting the tests and the path to the correct answer is very simple as they only require a few clicks. The other tasks are considered more difficult because the users have to use several functions to achieve the expected result.

As mentioned earlier, for each task, the time taken to complete it and whether the answer was correct or not was recorded. The number of clicks was not counted, as some tasks could be performed using multiple methods, such as tasks 4 and 6. Each task had a time limit of five minutes. At the end of each task, users were asked to fill in a Single Ease Question form (SEQ). This questionnaire asked how difficult the task was on a scale of 1 (Very Difficult) to 5 (Very Easy). They were also asked to give feedback/suggestions on the particular task. After all tasks were completed, the judges were asked to fill out a System Usability Scale form (SUS[2]). These two forms can be found in the appendix A and B.

5.2 Results

This section reviews and discusses the various findings from the test sessions with the judges. The test sessions were conducted with ten assistant judges of the Portuguese Supreme Court. Of the ten participants, one was a man and nine were women. All the judges use their computers on a daily basis and are moderately comfortable with technology. From the results collected, only two tasks were answered correctly in each test session.

5.2.1 Time to Complete Tasks and Judges’ Answers

Although some tasks are easy to perform, some judges may find it difficult to perform others. As you can see from the boxplot and table below, tasks number one, four and six are the most time-consuming to complete. The first two are considered the most challenging tasks, so it is normal that they take the longest to complete. There is also the number of correct and incorrect answers given for each task. As you can see, task four has a high number of wrong answers, while easier tasks like two and seven were answered correctly by all judges in a short amount of time. Task number five also has a high number of wrong answers. However, judges were quicker to indicate the desired answer in this task than in others.

[1] https://sportscienceinsider.com/what-is-the-learning-effect/#Why_is_the_learning_effect_a_problem

[2] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Median	64,81	10,65	22,52	36,66	23,01	34,78	11,85
Average	80,45	12,45	24,47	43,59	46,55	38,29	13,79
Standard Deviation	57,55	5,03	12,70	17,63	48,54	20,59	7,10

Table 5.1: Median, average and standard deviation for the time taken to complete the different tasks (in seconds).

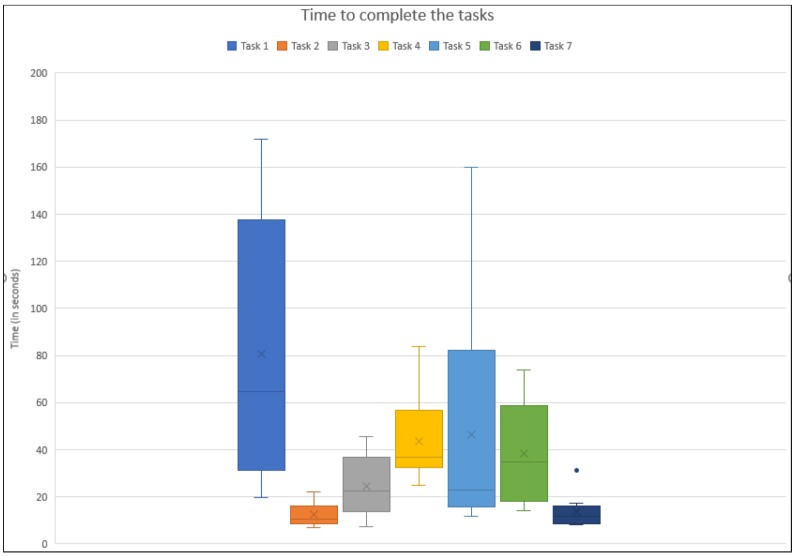


Figure 5.1: Time to complete the tasks.

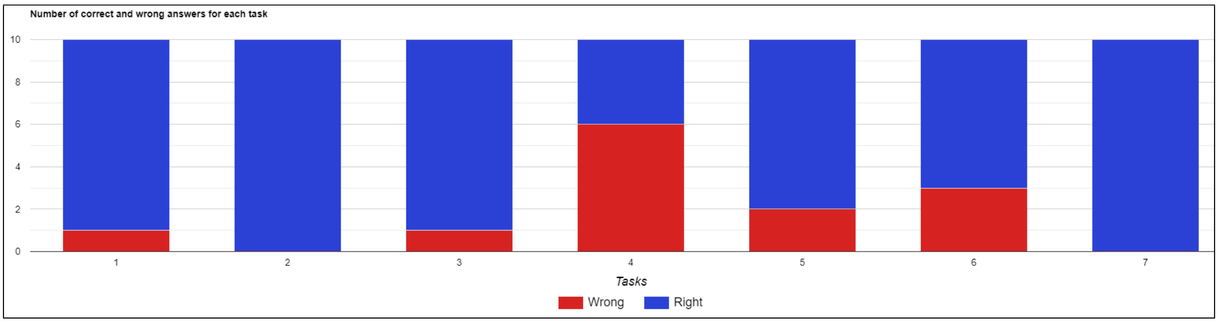


Figure 5.2: Number of right and wrong answers for each task.

Task one (Indicate the date of the oldest document that refers to the word ”Contract”.) was considered an easy task, as users only had to use the bar chart on the left of the screen to click on the first bar that appeared or use the slider to create an interval that included the desired document. However, when the tests began, the judges had difficulty finding the bar in the chart due to the size of the screen. The bar was not visible enough and users did not expect to be able to click on the bars to filter the documents by year. As a result, they lost a lot of time trying to find a way to find older documents. The judges tried all the ways to navigate the results: from the first to the last page of documents, using the other time filters or using the drop-down list to order the results. The latter did not work because there was no way to filter from the oldest to the newest document. After the tests, the judges said that this option should definitely be added and that it would have helped them to complete the task much faster and easier.

Task four (For the word ”Contract” and the Supreme Administrative Court documents between January and February 2015, indicate the last descriptor of the second document) was considered a more difficult task as it required the use of multiple filters to get the correct result. The aim of this task was to find out if the judges were able to use the filters on the left side of the screen. To get the correct result, the judges have to search for the word ”contract” and then use the filters for court and time. There were a high number of incorrect answers in this task because the judges did not understand how the filter for selecting an interval of days worked. When they clicked on it, they simply selected the months and years they wanted, but did not know that they had to select the first day of January and the last day of February. However, after the test was completed, most judges admitted that the function was not as complicated as they thought and that they would have done better on the task without the pressure of the test. There were also a small number of judges who stated that they would have preferred to do the filter differently.

Task five (For the word ’Debt’, give the value of the debt in euros on the second occurrence of the word in the first document.) was considered one of the most difficult tasks, as members of the INESC-ID group were aware that the judges were not familiar with the visualisation function. Although the time needed for the task and the number of wrong answers were not as high as expected, the judges found it difficult to find the answer. During the tests, it was confirmed that the judges did not pay attention to the visualisation, even if the functionality was explained to them before the test started. They did not understand that the vertical bars represented the occurrences of the terms they were looking for and that they could hover over them with the mouse pointer to display the text extract. However, knowing that they could open the page of the document, they used the browser’s `find` function to search for the occurrence of the term required in the task.

Task six (How many documents refer to the word ”Guilt” for the year 2021?) also required the judges to use the time filters. Similar to task four, the easiest way to get the correct answer was to click on the bar representing the year 2021 in the bar chart. Another option was to use the slider, but this would take longer. At the beginning of the tests, the judges found it difficult to use the time filters to select a single year. They did not understand that they could interact with the different bars on the bar chart. After the tests were completed, every effort was made to minimise the problem with the size of the bars in the bar chart. However, as there are a large number of documents for several years, some bars are smaller to best accommodate all the information.

The remaining tasks were solved with a high number of correct answers and in an acceptable time. The judges found the tasks easy enough to solve and were able to figure out how to use the available features of the user interface to achieve the best possible result.

5.2.2 User Comments

As mentioned earlier, judges were asked to fill out a *SEQ* form indicating how difficult a task was to complete on a scale of one (Very Hard) to five (Very Easy). This helped to understand the judges' thoughts about the interface. The boxplot showing the responses to each task is shown below.

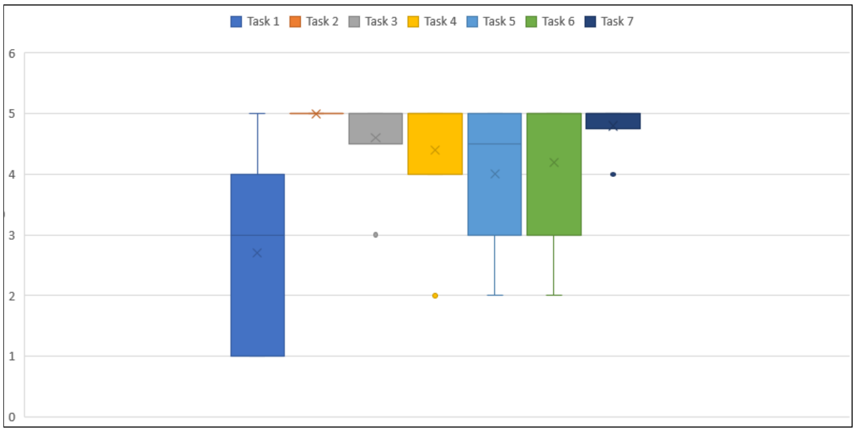


Figure 5.3: SEQ answers for each task.

From this figure it can be seen that the judges found the first task the most difficult to solve. Tasks five and six were also considered the most difficult, but not as much as the first. For the first task (Indicate the date of the oldest document in which that refers to the word ”Contract”), the judges suggested that the tooltips for the years should not overlap, as this makes it difficult to read the selected numbers. For tasks four and five, the judges suggested implementing a tooltip explaining the different ways of searching and how the visualisation works. This would help in understanding how the digital library works and help new users get used to the interface quickly. One judge also suggested that excerpts should be clickable to jump to where the excerpt is on the document page. The terms searched for should also be highlighted on this page. For task six (How many documents refer to the word ”Guilt” for the year 2021?), some judges suggested adding a filter to sort the documents from older to newer. In the tasks where the time filters were used, judges indicated that the filters were not very noticeable, but they acknowledged that they would get used to them over time and liked the idea and implementation.

Some tasks were quick to complete and scored highly in the *SEQ* questionnaire as the judges had time to use the interface before the tests began. For example, tasks two and seven had a low completion time due to this. After completing the tests, judges were asked to fill in a *SUS* form in which they rated how they felt about using the interface and the overall experience. The *SUS* score of the interface was 80.75, so the application is considered *excellent*, as any system with a rating above 80.3 is considered excellent[3].

In the *SUS* questionnaire, low ratings are expected for the odd questions, while high ratings are

[3] https://uxplanet.org/how-to-measure-product-usability-with-the-system-usability-scale-sus-score-69f3875b858f

expected for the even questions. This is the case with this system, except for the last question ("I wanted to do something, but I did not find a way to do it"), where there were a number of high-scoring answers. This means that the judges felt that they needed a function that was not available in the digital library to do a particular task, in this case was the ability to order documents from oldest to newest. The scores in the remaining answers are expected for each question (odd numbered questions have high scores and even numbered questions have low scores).

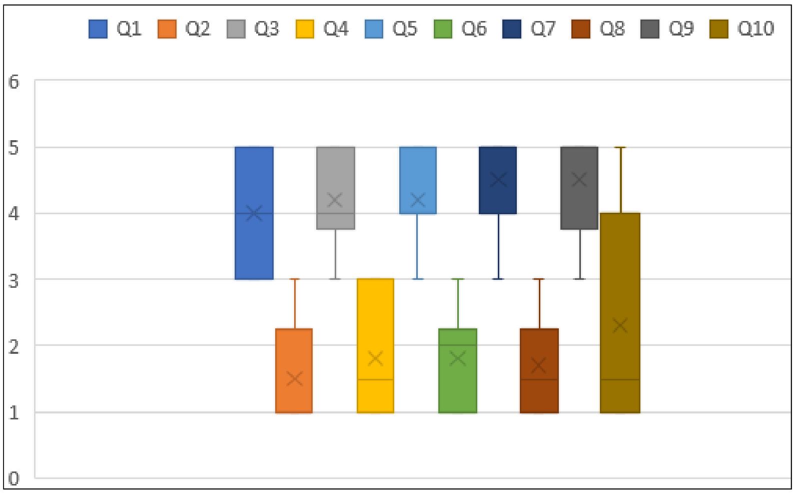


Figure 5.4: SUS answers for each question.

5.3 Changes to the Interface

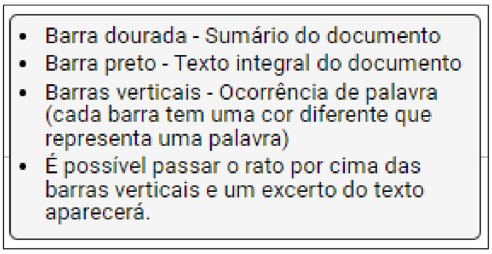
After the usability tests were completed, the judges gave their feedback on how the digital library's user interface and overall experience could be improved. These suggestions were written down and then analysed. Some features were implemented after the tests, but there are more that can be developed in future versions of the system. There was also a need to change the index of the database. This change was made after the usability tests. However, this was not suggested by the judges. The implemented changes are described in the following subsections.

5.3.1 Order from Oldest to Newest

Taking into account the judges' feedback, some changes were made to the digital library interface. The biggest complaint from the judges was the lack of an option to sort documents from oldest to newest. Since there was an option to sort from newest to oldest, it made sense to have the other option as well. Therefore, an option was added to the drop-down list on the right side of the user interface. When this option is selected, a flag is sent to the server and the correct order is chosen.

5.3.2 Tooltips for Searching and Visualization

Another feature strongly requested by almost all judges were two tooltips explaining how to perform a search and how to use visualisation. Two icons, taken from the *Font Awesome* website[4], were placed next to the search box and each horizontal bar of the visualisation. Hovering the mouse pointer over these icons displays the tooltips.



- Barra dourada - Sumário do documento
- Barra preto - Texto integral do documento
- Barras verticais - Ocorrência de palavra (cada barra tem uma cor diferente que representa uma palavra)
- É possível passar o rato por cima das barras verticais e um excerto do texto aparecerá.

Figure 5.5: Tooltip explaining how to use the visualization.

The tooltip explaining how to perform the search consists of four bullet points describing the different searches mentioned in section 4.1.1 (normal search, expression search and super search). The tooltip also contains a bullet point explaining that the operator *AND* or *OR* can be used. The second tooltip (Figure 5.5) also consists of an enumeration of bullet points describing how to interpret the different bars present in the visualisation. It describes what the horizontal bars represent (the golden bar is the summary of the document and the black bar is the full text) as well as the vertical bars (occurrences of the searched term or terms). It is also explained to the user that it is possible to move the mouse pointer over the vertical bars to display an excerpt of the text.

5.3.3 Pagination

Although it was not necessary, the judges used pagination during the usability tests. They stated that they wished there was a way to jump to the beginning and end of the results. Therefore, these two new options were added. When a user clicks the button to jump to the beginning, a new request is made to the server with the number one representing the first page of results. This number is used by ElasticSearch in the field "*from*". When a user clicks the button to go to the end of the results, the number twenty-five is sent, representing the last page of results.

5.3.4 Database Index

A new index has been created in the database with new fields. These new fields provide more information to the user and are therefore more helpful to judges when searching for a document. All these new fields are displayed on the documents page if they are available for the document in question.

[4] https://fontawesome.com/icons

5.4 Discussion

By analysing the time taken by all the judges to complete each task as well as the answers given, it is possible to draw some conclusions about the results. As the internet speed in the Portuguese Supreme Court was slow, this affected some test results. As seen in section 5.2.1, tasks one, four, five and six took the longest, while at the same time the most incorrect answers were given.

After analysing the information gathered during the usability tests, it can be confirmed that the judges found the experience of the digital library user interface pleasant. The system is well integrated, does not have many inconsistencies and for the most part correctly fulfilled the objectives of the tasks. Although the judges found some tasks more difficult than originally thought, they will get used to the features over time as they use the system more frequently.

When asked if the functions are well integrated (question 5), the judges responded with an average score of 4,2 in the *SUS* questionnaire. In addition, the judges felt that the system is not unnecessarily complex (average score 1,5), that it is easy to use (average score 4,2) and that they would quickly learn to use it (average score 4,5). However, some judges felt that they would need the help of experts (average score of 3,8) and that they would need to learn a lot before they could use the system (average score of 3,7). The rating of the overall *SUS* questionnaire (80,75) also shows that the interface is user-friendly and understandable. Therefore, it can be said that the digital library is able to fulfil the objective of facilitating the process of searching and analysing documents.

6

Conclusion

Contents

6.1 System Limitations and Future Work 71

It is important that the Portuguese Supreme Court works as efficiently as possible. To do so, its judges must have access to tools that facilitate their work. In this dissertation, as part of the IRIS project, a digital library was designed and developed that allows judges to search and navigate legal documents from all Portuguese courts.

A series of interviews were conducted with several Portuguese Supreme Court judges to find out what tools and systems are currently available to them and how they use them in their daily work, and to obtain a list of requirements for the new solution to be developed.

Once the list of requirements was finalised, designing the interface was the next step in the development process. The architecture of the system was also determined. The application was designed to be fast and efficient and to match the judges' existing tools. To this end, the solution was based on the interface of existing search tools such as the ECLI and DGSI websites, as these were considered by judges to be the best systems available and were the most widely used.

After the implementation phase, the system was tested with ten different judges of the Portuguese Supreme Court. These usability tests made it possible to understand how the judges perceived the user interface and to identify errors and areas for improvement. From the results of these tests, it could be concluded that the interface was easy to use and user-friendly. The judges liked their experience with it and the overall aesthetics of the system. They also gave their feedback and made suggestions on certain features they would like to see implemented. Some changes were made, but due to lack of time, others may be implemented in future versions of the system.

6.1 System Limitations and Future Work

As the IRIS project is still ongoing at the moment, this interface is still being subject to changes. There are several features that can be added to further improve the system. One feature that should be added is the ability to find similar documents, as the originally implemented approach did not work as expected. One issue that was identified during user testing and has not been resolved is the size of the bars of the bar chart. On smaller screens, the bars corresponding to a low number of documents are very small, making it difficult for users to see and click on them.

Based on the feedback and suggestions from the judges after the tests, there are several features that can be implemented. The first feature mentioned by a judge would be the ability to click on the vertical bars and open the page of the document at the location of the excerpt. This would make it easier to find a particular extract, so you do not waste time searching for it. Also on the page of the document, the searched terms should be highlighted. A final suggestion was that the search bar should always be visible even when scrolling, so that users do not lose the context of what was being searched.

There are also features that can be implemented that were not suggested by the judges. These features were considered important by the members of the INESC-ID group to make the judges' workflow even easier. At the moment, the tool is available to any user, but as this project is intended for the Portuguese Supreme Court, it is important to ensure the security of the users. Therefore, in the future, it would be important to implement an authentication system to access the tool. Another feature is the possibility to create bookmarks. This allows judges to save several documents that could be related to each other and quickly access them. There is also the option to implement a new filter that allows documents to be selected based on their relevance. If you want to ignore less relevant documents, this would be a good option. To increase efficiency in creating the visualisation, you could also use the ElasticSearch function for term vectors. This allows you to store the offset and position of each word and use it in a query in the database.

Bibliography

[1] "Nsm viz website," http://nsm-viz.nrc.pt/concept-map/cmg-op-2015.

[2] N. R. Carvalho and L. S. Barbosa, "Transforming legal documents for visualization and analysis," in *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 23–26. [Online]. Available: https://doi.org/10.1145/3209415.3209424

[3] C. Collins, F. Viegas, and M. Wattenberg, "Parallel tag clouds to explore and analyze faceted text corpora," 11 2009, pp. 91 – 98.

[4] A. Harris, R. Allen, C. Plaisant, and B. Shneiderman, "Temporal visualization for legal case histories," 10 1999.

[5] M. A. Hearst, "Tilebars: Visualization of term distribution information in full text information access," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '95. USA: ACM Press/Addison-Wesley Publishing Co., 1995, p. 59–66. [Online]. Available: https://doi.org/10.1145/223904.223912

[6] D. Byrd, "A scrollbar-based visualization for document navigation," in *Proceedings of the Fourth ACM Conference on Digital Libraries*, ser. DL '99. New York, NY, USA: Association for Computing Machinery, 1999, p. 122–129. [Online]. Available: https://doi.org/10.1145/313238.313283

[7] M. Wattenberg and F. B. Viegas, "The word tree, an interactive visual concordance," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1221–1228, 2008.

[8] F. van Ham, M. Wattenberg, and F. B. Viegas, "Mapping text with phrase nets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1169–1176, 2009.

[9] M. Wattenberg, "Arc diagrams: visualizing structure in strings," in *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, 2002, pp. 110–116.

[10] K. Kim, S. Ko, N. Elmqvist, and D. S. Ebert, “Wordbridge: Using composite tag clouds in node-link diagrams for visualizing content and relations in text corpora,” in *2011 44th Hawaii International Conference on System Sciences*, 2011, pp. 1–8.

[11] V. Benjamin, W. Chung, A. Abbasi, J. Chuang, C. A. Larson, and H. Chen, “Evaluating text visualization for authorship analysis,” *Security Informatics*, vol. 3, no. 1, p. 10, Sep 2014. [Online]. Available: https://doi.org/10.1186/s13388-014-0010-8

[12] D. A. Keim and D. Oelke, “Literature fingerprinting: A new method for visual literary analysis,” in *2007 IEEE Symposium on Visual Analytics Science and Technology*, 2007, pp. 115–122.

[13] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, and O. Deussen, “Document cards: A top trumps visualization for documents,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1145–1152, 2009.

[14] B. Fortuna, M. Grobelnik, and D. Mladenić, “Visualization of text document corpus.” *Informatica (Slovenia)*, vol. 29, pp. 497–504, 11 2005.

[15] P. Caillou, J. Renault, J. D. Fekete, A. C. Letournel, and M. Sebag, “Cartolabe: A web-based scalable visualization of large document collections,” *IEEE Computer Graphics and Applications*, pp. 1–1, 2020.

[16] C. L. Paul, J. Chang, A. Endert, N. Cramer, D. Gillen, S. Hampton, R. Burtner, R. Perko, and K. A. Cook, “Textonic: Interactive visualization for exploration and discovery of very large text collections,” *Information Visualization*, vol. 18, no. 3, pp. 339–356, 2019. [Online]. Available: https://doi.org/10.1177/1473871618785390

[17] J. Alsakran, Y. Chen, Y. Zhao, J. Yang, and D. Luo, “Streamit: Dynamic visualization and interactive exploration of text streams,” in *2011 IEEE Pacific Visualization Symposium*, 2011, pp. 131–138.

[18] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko, “Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 10, pp. 1646–1663, 2013.

[19] D. Mladenic and M. Grobelnik, “Automatic text analysis by artificial intelligence,” *Informatica (Slovenia)*, vol. 37, pp. 27–33, 2013.

[20] Y. Chan and H. Qu, “Finavistory: Using narrative visualization to explain social and economic relationships in financial news,” in *2016 International Conference on Big Data and Smart Computing (BigComp)*, 2016, pp. 32–39.

[21] D. Luo, J. Yang, M. Krstajic, W. Ribarsky, and D. Keim, “Eventriver: Visually exploring text collections with temporal references,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 1, pp. 93–105, 2012.

[22] K. Kucher and A. Kerren, “Text visualization techniques: Taxonomy, visual survey, and community insights,” in *2015 IEEE Pacific Visualization Symposium (PacificVis)*, 2015, pp. 117–121.

[23] F. B. Viegas, M. Wattenberg, and J. Feinberg, “Participatory visualization with wordle,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1137–1144, 2009.

[24] “Textarc,” http://wbradfordpaley.com/live, accessed: 2021-03-29.

[25] Y. Zhang, R. Jin, and Z.-H. Zhou, “Understanding bag-of-words model: A statistical framework,” *International Journal of Machine Learning and Cybernetics*, vol. 1, pp. 43–52, 12 2010.

[26] A. Sarveniazi, “An actual survey of dimensionality reduction,” *American Journal of Computational Mathematics*, vol. 04, pp. 55–72, 01 2014.

[27] V. Klema and A. Laub, “The singular value decomposition: Its computation and some applications,” *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980.

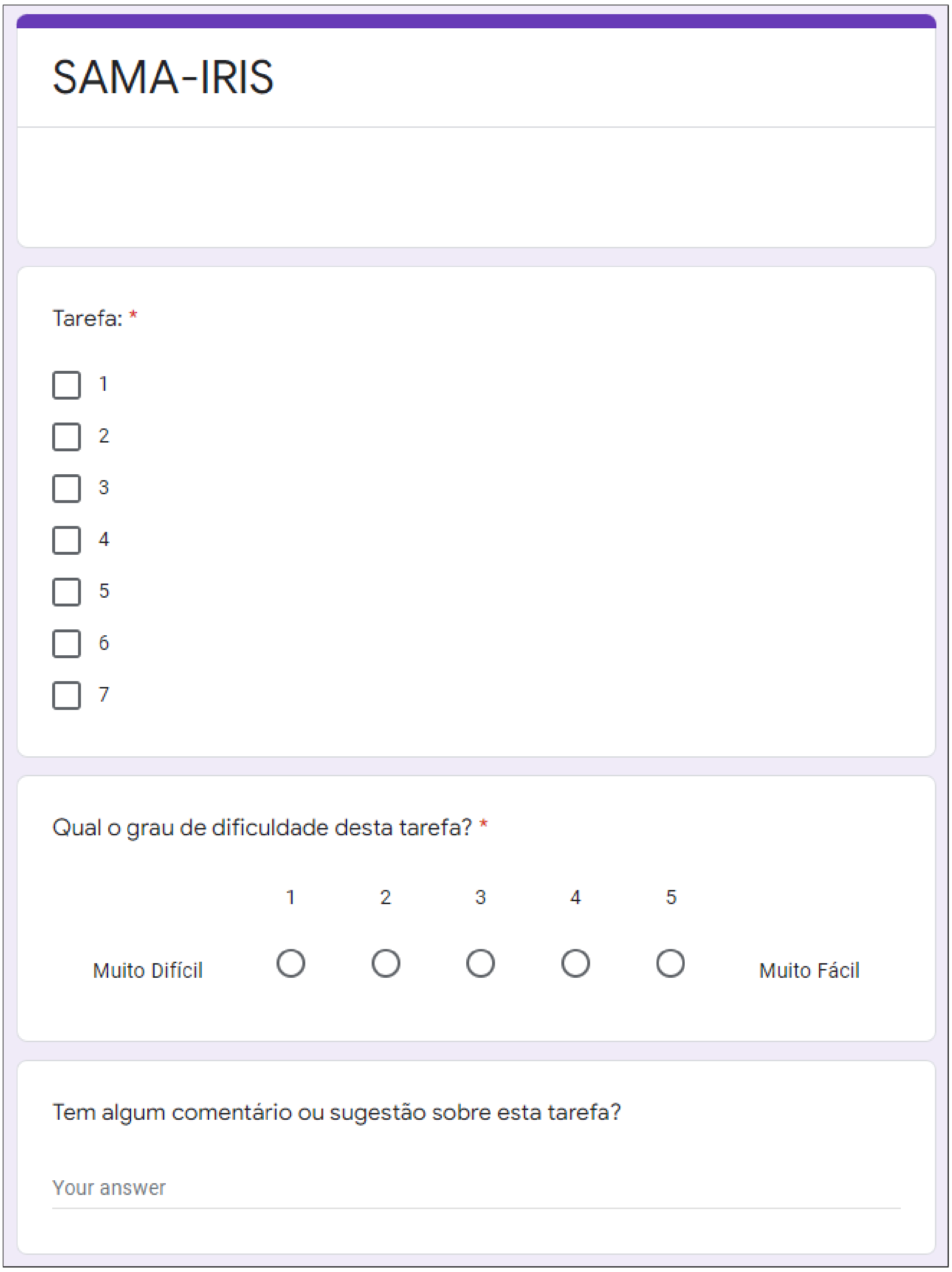
[28] A. Moldovan, R. Boţ, and G. Wanka, “Latent semantic indexing for patent documents,” *International Journal of Applied Mathematics and Computer Science*, vol. 15, pp. 551–560, 01 2005.

[29] S. Rose, Cowley, W. E, V. Crow, and N. Cramer, “Rapid automatic keyword extraction for information retrieval and analysis,” 01 2009.

[30] M. Grobelnik and D. Mladenic, “Visualization of news articles.” *Informatica (Slovenia)*, vol. 28, no. 4, pp. 375–380, 2004.

A

Single Ease Question (SEQ)



SAMA-IRIS

Tarefa: *

□ 1

□ 2

□ 3

□ 4

□ 5

□ 6

□ 7

Qual o grau de dificuldade desta tarefa? *

	1	2	3	4	5	
Muito Difícil	○	○	○	○	○	Muito Fácil

Tem algum comentário ou sugestão sobre esta tarefa?

Your answer

Figure A.1: SEQ questionnaire

B

System Usability Scale (SUS)

SUS questionnaire

*mandatory

The scale goes from one (Strongly disagree) to five (Strongly Agree).

1. Gostaria de usar esta aplicação frequentemente.*
2. A aplicação era desnecessariamente complexa.*
3. A aplicação era fácil de usar.*
4. Precisaria de apoio de outra pessoa com conhecimentos técnicos para usar a aplicação.*
5. As funcionalidades da aplicação estavam bem integradas.*
6. Existiam muitas inconsistências na aplicação.*
7. A maioria dos utilizadores aprenderiam rapidamente a usar a aplicação.*
8. Precisei de ter vários conhecimentos técnicos antes de usar a aplicação.*
9. Senti-me confiante a usar a aplicação.*
10. Senti que precisava de fazer uma certa ação mas não tinha maneira de a executar.*