UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DOUGLAS ÁLISSON MARQUES DE SÁ VITÓRIO

Ulysses-RFSQ: improving Information Retrieval through Relevance Feedback for similar queries

Recife

2025

DOUGLAS ÁLISSON MARQUES DE SÁ VITÓRIO

Ulysses-RFSQ: improving Information Retrieval through Relevance Feedback for similar queries

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador: Adriano Lorena Inácio de Oliveira

Coorientadora: Ellen Polliana Ramos Souza Pereira

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Vitório, Douglas Álisson Marques de Sá.
Ulysses-RFSQ: improving information retrieval through relevance feedback for similar queries / Douglas Alisson Marques de Sá Vitorio. - Recife, 2025.
240f.: il.

Tese (Doutorado)- Universidade Federal de Pernambuco, Centro de Informática, Programa de Pós-Graduação em Ciência da Computação, 2025.
Orientação: Adriano Lorena Inácio de Oliveira.
Coorientação: Ellen Polliana Ramos Souza Pereira.

1. Recuperação de informação; 2. Feedback de relevância; 3. Consultas similares; 4. Re-ranqueamento; 5. Domínio legislativo. I. Oliveira, Adriano Lorena Inácio de. II. Pereira, Ellen Polliana Ramos Souza. III. Título.

UFPE-Biblioteca Central

Douglas Álisson Marques de Sá Vitório

“Ulysses-RFSQ: improving Information Retrieval through Relevance Feedback for similar queries”

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Inteligência Computacional.

Aprovada em: 27/11/2025.

Orientador: Prof. Dr. Adriano Lorena Inacio de Oliveira

BANCA EXAMINADORA

Prof. Dr. Cleber Zanchettin
Centro de Informática / UFPE

Profa. Dra. Renata Maria Cardoso Rodrigues de Souza
Centro de Informática / UFPE

Profa. Dra. Nádia Félix Felipe da Silva
Instituto de Informática / UFG

Prof. Dr. Márcio de Souza Dias
Instituto de Biotecnologia / UFCAT

Prof. Dr. George Gomes Cabral
Departamento de Estatística e Informática / UFRPE

ACKNOWLEDGEMENTS

First, I would like to thank my parents, Iraneide and Fernandes, and my cousins, Isabela and Ingrid, for all of the support and for being present in my life.

To all my friends and lovers, without whom I would have never made it this far. Even from a distance, they are always here with me.

To my cat, Phoebe (a.k.a., Bibo), for keeping me company and bringing a smile to my face even in the darkest days.

To my advisors, Adriano and Ellen, the members of the Ulysses project, and the Brazilian Chamber of Deputies, for making this research possible.

Finally, I would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for funding this research.

RESUMO

O uso do Feedback de Relevância é capaz de aperfeiçoar o desempenho da Recuperação de Informação (RI), mas esse método é comumente utilizado apenas para melhorar o processo de recuperação para a consulta que está correntemente sendo processada. Quando a informação de relevância de buscas passadas está disponível, essa informação pode ser utilizada para auxiliar buscas futuras. Se duas consultas são suficientemente similares, os documentos julgados como relevantes para uma podem também ser relevantes para a outra. Entretanto, poucos estudos foram encontrados na literatura lidando com esse uso da informação de relevância de consultas passadas, pois há uma falta de bases de dados de benchmark contendo essa informação para consultas similares. Dessa forma, este estudo apresenta Ulysses-RFSQ, um novo método de RI que visa aprimorar os resultados para consultas futuras a partir do uso da informação do Feedback de Relevância de buscas passadas similares. Seu funcionamento se dá pelo re-ranqueamento da lista de documentos recuperada por um algoritmo de RI base através da adição de um bônus ou uma penalidade ao escore dos documentos. Assim, esse método pode ser utilizado com qualquer algoritmo que calcule um escore para os documentos, tais como o algoritmo BM25 ou modelos Sentence-BERT. Para avaliar o método Ulysses-RFSQ, uma base de dados de Feedback de Relevância, chamada Ulysses-RFCorpus, foi construída junto com a Câmara dos Deputados brasileira e disponibilizada para a comunidade. Além do Ulysses-RFCorpus, o método proposto também foi avaliado em uma base de dados maior, também fornecida pela Câmara (o corpus da Pesquisa Prévia), a qual não pôde ser disponibilizada publicamente. A avaliação desse método no cenário legislativo é justificada pelo fato de que a maioria das consultas utilizadas no processo legislativo brasileiro é redundante. Como resultados, os achados apontaram que o Ulysses-RFSQ é capaz de usar a informação de feedback de consultas passadas similares para aprimorar o desempenho do algoritmo base para consultas futuras. Melhorias nas métricas de MAP, MRP, MRR e nDCG mostraram que o método proposto pôde re-ranquear os documentos relevantes nas primeiras posições enquanto recuperava documentos relevantes que não foram recuperados pelo algoritmo de RI base. As melhorias puderam ser melhor observadas em cenários nos quais o algoritmo base não obteve resultados muito bons e utilizando um maior conjunto de consultas passadas armazenadas. Por exemplo, as melhorias observadas nos resultados de MAP variaram de 0,0384 a 0,0773 para o corpus da Pesquisa Prévia — em alguns casos, mais do que dobrando o desempenho do algoritmo utilizado como *baseline*.

Palavras-chave: Recuperação de Informação. Feedback de Relevância. Consultas similares. Re-ranqueamento. Domínio legislativo.

ABSTRACT

The use of Relevance Feedback can enhance the Information Retrieval (IR) performance, but this method is often used only to improve the retrieval for a specific query: the one currently being processed. When there is available relevance information from past searches, this information may be useful to help future searches. If two queries are sufficiently similar, the relevant documents for one may also be relevant for the other. However, only a few studies were found in the literature dealing with this use of relevance information from past queries, as there is a lack of benchmark datasets containing this information for similar queries. In this sense, this study presents Ulysses-RFSQ, a novel IR method that aims to improve the results for future queries by using the Relevance Feedback information from past similar ones. It works by re-ranking the list of documents retrieved by a base IR algorithm through the addition of a bonus or a penalty to the documents' score. Therefore, it can be used with any algorithm that computes a score for the documents, such as BM25 or Sentence-BERT models. To evaluate the Ulysses-RFSQ method, a Relevance Feedback dataset, called Ulysses-RFCorpus, was built together with the Brazilian Chamber of Deputies and made available to the community. Besides Ulysses-RFCorpus, the proposed method was also evaluated in larger dataset (the Preliminary Search corpus) provided by the Chamber, which could not be made available. The method's evaluation in the legislative scenario is justified by the fact that most of the queries used in the Brazilian legislative process are redundant. As results, the findings pointed out that Ulysses-RFSQ can use the past feedback information from similar queries to improve the base algorithm's performance for future queries. Improvements in MAP, MRP, MRR, and nDCG showed that the proposed method could re-rank the retrieved documents list in a way that can rearrange the relevant documents in the first positions while fetching relevant documents not retrieved by the base IR algorithm. The improvements could be better seen in scenarios in which the base IR algorithm did not achieve great results and while using a larger set of stored queries. For instance, the observed improvements in the MAP results ranged from 0.0384 to 0.0773 for the Preliminary Search corpus — in some cases, more than doubling the baseline's performance.

Keywords: Information Retrieval. Relevance Feedback. Similar queries. Re-ranking. Legislative domain.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AP Average Precision

BERT Bidirectional Encoder Representations for Transformers

CD Critical Difference

Conle Legislative Consulting

DCG Discounted Cumulative Gain

IDCG Ideal Discounted Cumulative Gain

IR Information Retrieval

LaBSE Language-agnostic BERT Sentence Encoder

LIR Legal Information Retrieval

LLMs Large Language Models

LMs Language Models

LSA Latent Semantic Analysis

MAP Mean Average Precision

MRP Mean R-Precision

MRR Mean Reciprocal Rank

nDCG Normalized Discounted Cumulative Gain

NER Named Entity Recognition

NLP Natural Language Processing

PRF Pseudo-Relevance Feedback

QE Query Expansion

RF Relevance Feedback

RQ Research Question

RQs Research Questions

RR Reciprocal Rank

SBERT Sentence-BERT

STJ Brazilian Superior Court of Justice

STS Semantic Textual Similarity

SVM Support Vector Machine

VSM Vector Space Model

CONTENTS

1 INTRODUCTION

1.1 CONTEXT AND MOTIVATION

Information Retrieval (IR) systems have been used to fetch and present relevant documents to a user based on their query. IR techniques can be used wherever there is a need for information to be retrieved, being applied to different areas and domains. Nevertheless, one of the first areas to apply IR algorithms was the legal one, with the first legal retrieval system appearing in the 60s (MAXWELL; SCHAFER, 2008).

The legal area's interest in IR systems can be explained due to the fact that all the work in this area is based on textual documents, whereas the access and possession of relevant knowledge is crucial. There is also an increase in the number of legal documents being produced (SOUZA et al., 2021b), which makes it complex for professionals to work with such a large amount of data. Thus, a specific IR subarea, called Legal Information Retrieval (LIR), has been developed to assist with tasks within the legal domain (MAXWELL; SCHAFER, 2008).

LIR includes tasks such as jurisprudence analysis, as well as the support to the law-making process. Within the legislative scenario, which focuses on the process of making new legislation, automated IR techniques are necessary to keep up with the increasing growth in the number of documents created by parliamentarians. Organizing, accessing, and retrieving this kind of data pose significant challenges due to the unstructured nature of these documents (CANTADOR; SÁNCHEZ, 2020).

As an example, the Legislative Consulting (Conle)[1] department of the Brazilian Chamber of Deputies plays a crucial role in the law-making process in Brazil. Before a parliamentarian makes a legislative proposal which may become a bill to be voted in the Congress, they must consult Conle for previously submitted proposals and existing bills.

The process of retrieving old documents is highly time-consuming, given the substantial volume of legislative proposals that Conle must handle annually. Since the decade of 1930, the Chamber has processed over 144,000 bills (BRANDT, 2020), with most of them being redundant. Nowadays, Conle uses an IR system based on Natural Language Processing (NLP) techniques to retrieve documents according to a parliamentarian's query (SOUZA et al., 2021b). This automation in the search process enables Conle to deal with its great demand.

[1] https://www2.camara.leg.br/a-camara/estruturaadm/diretorias/diretoria-legislativa/estrutura-1/conle

The performance of documents retrieval may be improved by using the relevance information of the documents through a process known as Relevance Feedback (RF). The RF method uses the judgment made by users, commonly experts (GUTIÉRREZ-SOTO, 2016), iteratively to enhance the IR system's performance generally in two ways: by expanding the query with terms from the relevant documents, or by training a Supervised Machine Learning algorithm with the relevant and non-relevant documents information. Both usages can be found in the literature for LIR (RISSLAND; DANIELS, 1996; ZHANG et al., 2020).

However, Pseudo-Relevance Feedback (PRF), which involves automatically labeling the top-ranked retrieved documents as *relevant* to mimic real feedback, is more commonly found in the IR literature. This approach is simple and computationally efficient, but it has clear disadvantages compared to feedback provided by experts (CARPINETO; ROMANO, 2012). Its usage is due to the lack of sufficient datasets containing relevance judgments, as the users are usually reluctant to give this feedback information (MANNING; RAGHAVAN; SCHÜTZE, 2008).

Although operational environments have stored query logs, these logs only contain what can be called *implicit feedback* (MANNING; RAGHAVAN; SCHÜTZE, 2008). This kind of feedback information is given by the users in an indirect way — as they click in a link, for example — and should not be used directly as judgments of relevance. As stated by Joachims et al. (2007), clicks are informative but biased, and it is difficult to interpret them as an absolute relevance judgment.

Therefore, as IR approaches should not be based on log files, they need collections composed by a set of documents, a set of queries, and a set of relevance judgments. The creation of appropriate IR collections, though, is very costly in time and efforts (GUTIÉRREZ-SOTO, 2016).

Both methods, either using real feedback or pseudo-feedback information, commonly improve the retrieval for a specific query. Only the feedback information given for the query currently being processed is used. In addition, this feedback is used only in that specific session (YIN et al., 2002). In this sense, in cases in which feedback information can be stored for queries processed in the past, this information could be used to enhance the IR system for the processing of new queries, but it is generally not used at all. Usually, all of the information about a retrieval is lost after the presentation of the results list to the user (HUBERT; MOTHE, 2007).

It is appealing to think that the results of past searches may be useful to help future searches (FITZPATRICK; DENT, 1997). For instance, an alternative for RF to be used in order to make the retrieval model better in a way that impacts other searches is the storage and utilization of this feedback information to improve the IR process for similar queries. If there are past queries sufficiently similar to the one being processed, the documents judged as *relevant* for those queries might also be relevant to the current one.

Few studies, though, have been performed aiming to use past queries information for this purpose — namely improving IR for future queries. This is also due to the aforementioned lack of available benchmark datasets containing relevance information (GUTIÉRREZ-SOTO, 2016). If it is difficult to find RF datasets suited for IR, it is even harder to find collections containing sets of similar queries. Most of the existing IR collections are composed of independent queries, as they lack of relevance judgments, not being appropriate to evaluate approaches based on the feedback information given for past queries.

A feasible alternative is to simulate collections containing similar queries (GUTIÉRREZ-SOTO, 2016). For instance, new simulated queries can be generated from past real ones by removing or changing terms from them or by extracting terms from their top-ranked documents (CETINTAS; SI; YUAN, 2011). However, it is difficult to simulate the relevance judgment for these queries.

Another alternative is to use data from domains in which there are redundancy in the queries, such as the legislative scenario (BRANDT, 2020). Given that, and due to the aforementioned dependency on the retrieval of useful information in this specific domain, it was used as the main focus for the method presented in this study.

This research was made possible through an agreement between Universidade de São Paulo and the Brazilian Chamber of Deputies (agreement no. 20.1.405.55.0). It was conducted in the context of the Ulysses project, an institutional set of Artificial Intelligence initiatives aiming to increase transparency, improving the Chamber's relationship with citizens and supporting the legislative activity (ALMEIDA, 2021).

1.2 OBJECTIVES

1.2.1 General objective

In this sense, the main goal of this study is to present a model-independent Information Retrieval method that uses the Relevance Feedback information from past similar queries aiming to improve the IR process for future queries.

1.2.2 Specific objectives

In order to achieve the general goal of this study, some specific objectives were listed:

- To build a corpus for the legislative scenario containing relevance information given by experts, in order to make the evaluation of the proposed method possible;
- To propose a method, called Ulysses-RFSQ, which uses the RF information from past similar queries to re-rank the documents retrieved by an IR algorithm, improving the retrieval results;
- To evaluate the proposed method, comparing its results to a baseline's which do not use past relevance information.

1.3 RESEARCH QUESTIONS AND HYPOTHESES

In order to achieve the goals presented in Section 1.2, this study is guided by a set of Research Questions (RQs):

1. **RQ1) Is it feasible to use Relevance Feedback information from stored past queries to improve the documents retrieval process for new queries?** This more generic RQ aims to confirm the feasibility of this study. We hypothesize that the use of RF information from past queries can be used to improve the performance of the IR process if there are stored queries sufficiently similar to the one currently being processed, as can be seen in the literature for specific scenarios.
2. **RQ2) Can a method that utilizes the RF information from similar past queries to re-rank the retrieved documents improve the IR results within the Brazilian legislative domain?** As aforementioned, there is a redundancy in the documents and queries generated within the Brazilian legislative process, thus we hypothesize that the use of the RF information from similar stored queries may improve the IR algorithms results for this specific domain.
3. **RQ3) What is the trade-off between the use of the RF information from a greater number of past queries and the use of this information from a smaller set of highly similar ones?** The set of similar past queries must be selected from a

stored database in view of their similarity to the query currently being processed. Thus, we hypothesize that the selection of a smaller set containing queries with a higher level of similarity may be better than selecting a larger set of queries that are less similar to the current one.

4. **RQ4) What is the best method to find, within a database of stored queries, the queries that are similar to the one currently being processed?** As we are using the RF information from similar past queries, the first task is to find this set of queries. We hypothesize that the use of Language Models to capture context and semantic characteristics of the texts may be the best way to select the similar queries, rather than comparing the queries only by the presence or absence of terms.

5. **RQ5) Is the irrelevant documents information from past queries useful for re-ranking the retrieved documents for a new query?** As some IR datasets present different levels of relevancy for the documents — e.g., *irrelevant*, *somewhat relevant*, and *very relevant* —, we hypothesize that this information can also be used to improve the IR process.

1.4 SCIENTIFIC CONTRIBUTIONS

In this section, the scientific contributions regarding this research or related to Legal Information Retrieval are presented.

First, we created a corpus containing a set of queries and their respective RF information judged by experts. It was named Ulysses-RFCorpus, was built together with the Brazilian Chamber of Deputies, and made publicly available[2],

1.4.1 As first author regarding this research

Some papers containing the main contributions of this research were published: Vitório et al. (2022) presented a preliminary version of Ulysses-RFSQ and its evaluation, whereas the paper of Vitório et al. (2025a) presents the construction process of Ulysses-RFCorpus. In addition, a third study has assessed and compared the use of BM25 algorithms and SBERT models for documents retrieval within the Brazilian legislative scenario (VITÓRIO et al., 2025b).

2 https://github.com/ulysses-camara/Ulysses-RFCorpus

1. **Title:** Ulysses-RFSQ: A Novel Method to Improve Legal Information Retrieval Based on Relevance Feedback (VITÓRIO et al., 2022)
   Authors: Douglas Vitório, Ellen Souza, Lucas Martins, Nádia F. F. da Silva, André Carlos Ponce de Leon de Carvalho, and Adriano L. I. Oliveira
   Venue: 11th Brazilian Conference on Intelligent Systems (BRACIS) - Campinas-BRA, 2022
   DOI: https://doi.org/10.1007/978-3-031-21686-2_6

2. **Title:** Building a Relevance Feedback Corpus for Legal Information Retrieval in the Real-Case Scenario of the Brazilian Chamber of Deputies (VITÓRIO et al., 2025a)
   Authors: Douglas Vitório, Ellen Souza, Lucas Martins, Nádia F. F. da Silva, André Carlos Ponce de Leon de Carvalho, Adriano L. I. Oliveira, and Francisco Edmundo de Andrade
   Venue: Language Resources and Evaluation, Volume 59, pages 1257–1277, 2025 (first published 18 August 2024)
   DOI: https://doi.org/10.1007/s10579-024-09767-3

3. **Title:** BM25 x Vila Sésamo: avaliando modelos Sentence-BERT para Recuperação de Informação no cenário legislativo brasileiro (VITÓRIO et al., 2025b)
   Authors: Douglas Vitório, Ellen Souza, José Antônio dos Santos, André Carlos Ponce de Leon Ferreira de Carvalho, Adriano L. I. Oliveira, and Nádia F. F. da Silva
   Venue: Linguamática, Volume 17 (1), pages 17-33, 2025
   DOI: https://doi.org/10.21814/lm.17.1.474

1.4.2 Related to Legal Information Retrieval

Besides the publication of the main contributions of this study, 10 other papers related to LIR were published within the scope of the Ulysses Project (ALMEIDA, 2021). The researches presented in those papers were conducted alongside the research presented in this study.

The paper of Souza et al. (2021b) presented the pipeline of the IR system used by Conle, which also worked as one of the baselines for this study. Meanwhile, Souza et al. (2021a) expanded the first work, evaluating Stemming techniques for the same scenario. Rocha et al. (2023) evaluated IR frameworks in the scenario of the Brazilian Chamber of Deputies and Santos et al. (2024) created a hybrid IR system combining BM25 with BERT-based models.

Albuquerque et al. (2022) and Costa et al. (2022) built a Named Entity Recognition (NER) corpus, called UlyssesNER-Br, for the Brazilian legal scenario, with the goal to assist the LIR process. Albuquerque et al. (2023) used UlyssesNER-Br to evaluate Deep Learning models for NER, Albuquerque et al. (2024) performed Query Expansion using the UlyssesNER-Br entities, and Gouveia et al. (2025) evaluated Active Learning techniques for NER corpora expansion..

Finally, Siqueira et al. (2025) built and presented Ulysses Tesemõ, a large corpus for the Brazilian legal and governmental domains.

1. **Title:** An Information Retrieval Pipeline for Legislative Documents from the Brazilian Chamber of Deputies (SOUZA et al., 2021b)
   Authors: Ellen Souza, **Douglas Vitório**, Gyovana Moriyama, Luiz Santos, Lucas Martins, Mariana Souza, Márcio Fonseca, Nádia Félix, André C. P. L. F. Carvalho, Hidelberg O. Albuquerque, and Adriano L. I. Oliveira
   Venue: 34th International Conference on Legal Knowledge and Information Systems (JURIX) - Online, 2021
   DOI: https://doi.org/10.3233/FAIA210326

2. **Title:** Assessing the Impact of Stemming Algorithms Applied to Brazilian Legislative Documents Retrieval (SOUZA et al., 2021a)
   Authors: Ellen Souza, Gyovana Moriyama, **Douglas Vitório**, André C. P. L. F. de Carvalho, Nádia Félix, Hidelberg O. Albuquerque, and Adriano L. I. Oliveira
   Venue: 8th Symposium in Information and Human Language Technology (STIL) - Online, 2021
   DOI: https://doi.org/10.5753/stil.2021.17802

3. **Title:** UlyssesNER-Br: A Corpus of Brazilian Legislative Documents for Named Entity Recognition (ALBUQUERQUE et al., 2022)
   Authors: Hidelberg O. Albuquerque, Rosimeire Costa, Gabriel Silvestre, Ellen Souza, Nádia F. F. da Silva, **Douglas Vitório**, Gyovana Moriyama, Lucas Martins, Luiza Soezima, Augusto Nunes, Felipe Siqueira, João P. Tarrega, Joao V. Beinotti, Marcio Dias, Matheus Silva, Miguel Gardini, Vinicius Silva, André C. P. L. F. de Carvalho, and Adriano L. I. Oliveira
   Venue: 15th International Conference on Computational Processing of Portuguese

(PROPOR) - Fortaleza-BRA, 2022
DOI: https://doi.org/10.1007/978-3-030-98305-5_1

4. **Title:** Expanding UlyssesNER-Br Named Entity Recognition Corpus with Informal User-Generated Text (COSTA et al., 2022)
   Authors: Rosimeire Costa, Hidelberg Oliveira Albuquerque, Gabriel Silvestre, Nádia Félix F. Silva, Ellen Souza, **Douglas Vitório**, Augusto Nunes, Felipe Siqueira, João Pedro Tarrega, João Vitor Beinotti, Márcio de Souza Dias, Fabíola S. F. Pereira, Matheus Silva, Miguel Gardini, Vinicius Silva, André C. P. L. F. de Carvalho, and Adriano L. I. Oliveira
   Venue: 21st EPIA Conference on Artificial Intelligence (EPIA) - Lisbon-POR, 2022
   DOI: https://doi.org/10.1007/978-3-031-16474-3_62

5. **Title:** On the Assessment of Deep Learning Models for Named Entity Recognition of Brazilian Legal Documents (ALBUQUERQUE et al., 2023)
   Authors: Hidelberg O. Albuquerque, Ellen Souza, Adriano L. I. Oliveira, David Macêdo, Cleber Zanchettin, **Douglas Vitório**, Nádia F. F. da Silva, and André C. P. L. F. de Carvalho
   Venue: 22nd EPIA Conference on Artificial Intelligence (EPIA) - Faial Island-POR, 2023
   DOI: https://doi.org/10.1007/978-3-031-49011-8_8

6. **Title:** Avaliação de frameworks para Recuperação de Documentos Legislativos: um Estudo de Caso na Câmara dos Deputados Brasileira (ROCHA et al., 2023)
   Authors: Flávio C. Rocha, Ellen Souza, **Douglas Vitório**, Nádia F. F. da Silva, André C. P. L. F. de Carvalho, and Adriano L. I. Oliveira
   Venue: XI Workshop de Computação Aplicada em Governo Eletrônico (WCGE) - João Pessoa-PB, 2023
   DOI: https://doi.org/10.5753/wcge.2023.229925

7. **Title:** HIRS: A Hybrid Information Retrieval System for Legislative Documents (SANTOS et al., 2024)
   Authors: José Antônio dos Santos, Ellen Souza, Carmelo Bastos- Filho, Hidelberg O. Albuquerque, **Douglas Vitório**, Danilo Carlos Gouveia de Lucena, Nádia Silva, and André de Carvalho
   Venue: 23nd EPIA Conference on Artificial Intelligence (EPIA) - Viana do Castelo-POR,

2024

DOI: https://doi.org/10.1007/978-3-031-73497-7_26

8. **Title:** UlyssesNERQ: Expanding Queries from Brazilian Portuguese Legislative Documents through Named Entity Recognition (ALBUQUERQUE et al., 2024)

   Authors: Hidelberg O. Albuquerque, Ellen Souza, Tainan Silva, Rafael P. Gouveia, Flavio Junior, **Douglas Vitório**, Nádia F. F. da Silva, André C. P. L. F. de Carvalho, Adriano L. I. Oliveira, and Francisco Edmundo Andrade

   Venue: 16th International Conference on Computational Processing of Portuguese (PROPOR) - Santiago de Compostela-ESP, 2024

9. **Title:** Ulysses Tesemõ: a new large corpus for Brazilian legal and governmental domain (SIQUEIRA et al., 2025)

   Authors: Felipe A. Siqueira, **Douglas Vitório**, Ellen Souza, José A. P. Santos, Hidelberg O. Albuquerque, Márcio S. Dias, Nádia F. F. Silva, André C. P. L. F. de Carvalho, Adriano L. I. Oliveira, and Carmelo Bastos-Filho

   Venue: Language Resources and Evaluation, Volume 59, pages 1685–1704, 2025 (first published 18 July 2024)

   DOI: https://doi.org/10.1007/s10579-024-09762-8

10. **Title:** Applying Active Learning in Named Entity Recognition Corpora Expansion in Legal Domain (GOUVEIA et al., 2025)

    Authors: Rafael P. Gouveia, André C. P. L. F. de Carvalho, Ellen Souza, Hidelberg O. Albuquerque, **Douglas Vitório**, Nádia F. F. Silva

    Venue: 26th Annual International Conference on Digital Government Research (dg.o) - Porto Alegre-RS, 2025

    DOI: https://doi.org/10.59490/dgo.2025.945

1.5 STRUCTURE

The remaining of this study is organized as follows:

- Chapter 2 brings an overview of the key elements related to this study, such as IR, Legal Information Retrieval, and Relevance Feedback, as well as the related work;
- Chapter 3 details the Brazilian legislative scenario in which this research was conducted;

- Chapter 4 introduces Ulysses-RFSQ, the proposed method;
- Chapter 5 describes the construction process of Ulysses-RFCorpus;
- Chapter 6 presents the experimental setup used to evaluate Ulysses-RFSQ;
- Chapter 7 reports and discusses the results obtained in this study, in order to answer the Research Questions;
- Finally, Chapter 8 draws the conclusions and presents the future work.

2 BACKGROUND

In this chapter, the key elements that are pertinent to this study are explained. Section 2.1 presents an overview of Information Retrieval and Legal Information Retrieval. Section 2.2 explains Relevance Feedback and its use for similar queries, including the found related work (Section 2.2.2).

2.1 INFORMATION RETRIEVAL

Information Retrieval (IR) involves finding unstructured material, from a large collection, that satisfies an information need (MANNING; RAGHAVAN; SCHÜTZE, 2008). This material usually is composed by text documents — in this sense, the process can also be called *documents retrieval* —, whereas a user expresses their information need through a query.

The main goal is to present to the user a set of documents that are relevant for them, in the view of their information need. Nevertheless, the concept of relevance is subjective, as the same document may be relevant for a user and irrelevant for another (CASELI; NUNES, 2024). Thereby, the user always judges the document relevance according to how it satisfies their specific query.

As consequence, in order to present pertinent information to the user, IR algorithms can only estimate the relevance of the documents. Most of these algorithms perform this estimation by computing the similarity between the documents and the query. The similarity is often computed based on the occurrence of query terms within the document. Thus, the IR algorithm scores and ranks the documents in response to the query, presenting the top-ranked ones to the user (MANNING; RAGHAVAN; SCHÜTZE, 2008).

For instance, the scoring function of the Okapi BM25 (ROBERTSON et al., 1994) algorithm — the most well-known scoring function for documents retrieval — estimates the relevance of a document d to a query q based on the query terms appearing in d, regardless of their proximity within d. Its formula is presented in Equation 2.1 (KAMPHUIS et al., 2020):

$$score(d,q) = \sum_{t \in q} \log \left(\frac{N - DF(t) + 0.5}{DF(t) + 0.5} \right) \cdot \frac{TF(t,d)}{k_1 \left(1 - b + b \cdot \frac{|d|}{L}\right) + TF(t,d)}, \tag{2.1}$$

in which N is the number of documents in the dataset, $DF(t)$ (document frequency) is the number of documents containing the query term t, $TF(t,d)$ is the frequency of t in document

d, $\mid d \mid$ is the number of terms in document d, L is the average number of terms per document, and b and k_1 are parameters that can be adjusted for each dataset. k_1 helps to control the term frequency (TF) scale, while b works on the normalization as a function of the document's size (CASELI; NUNES, 2024).

The matching between the query's and document's terms, however, causes a problem known as *vocabulary mismatch*, as the terms used in the query may not be present in the document (CASELI; NUNES, 2024). In addition, the same term may have different meanings and its presence within the document may not be related to the user's information need expressed through the query.

An alternative to mitigate the vocabulary mismatch problem is to take the context and semantic of the terms into account. This can be performed using Language Models (LMs) to extract contextual embeddings from the texts, based on, for instance, the words preceding and following a particular term within the sentence (WANG et al., 2024).

The application of neural networks to generate contextual embeddings and their subsequent use for NLP has been substantially growing in the last years, becoming the state-of-the-art for many tasks (WOLF et al., 2020). These architectures, known as *Transformers* (VASWANI et al., 2017), have in Bidirectional Encoder Representations for Transformers (BERT) (DEVLIN et al., 2019) its most well-known and utilized model, including for IR and other text-ranking tasks (LIN; NOGUEIRA; YATES, 2022).

The BERT model uses the Transformers architecture to capture bidirectional contexts in texts (DEVLIN et al., 2019). It was initially pre-trained using two tasks: Masked Language Modeling, in which random terms in a sentence are masked and the model is trained to predict these terms, and Next Sentence Prediction, in which the model understands the relation between two sentences by predicting the next sentence. This training step made BERT very suitable for generation tasks, such as Question Answering. Nevertheless, its great capacity for language comprehension has also made researchers to apply BERT-based approaches to IR (WANG et al., 2024).

Due to the LMs' high computational cost to find similar sentences, Reimers and Gurevych (2019) proposed a modification of BERT known as Sentence-BERT (SBERT). SBERT uses triplet and siamese networks to derive semantically significant sentence embeddings, which can be compared using distance measures, such as cosine similarity or the Manhattan/Euclidean distance. This method has allowed BERT-based models to be easily used for tasks such as Semantic Textual Similarity (STS) and IR with semantic search, reducing the computational complexity of the systems.

Using a distance measure to compute the similarity between the embeddings generated for two pieces of text allows the SBERT models to compute a score for each document according to the query. Thus, the set of documents can be ranked in a similar way as algorithms such as BM25 do, in which the top-ranked ones may be more relevant to the user, as they are more semantically similar to the query.

In the past few years, the use of Large Language Models (LLMs) have gained attention for NLP tasks, due to their ability to generate human-like text. However, Wang et al. (2024) pointed out many remaining challenges in using LLMs in real-world problems, such as for IR. The LLMs require significant computational resources for both training and inference, they can also generate responses based only on their pre-training knowledge, and there are privacy-related concerns on using the LLMs' APIs. On the other hand, BERT-based models do not pose privacy-related risks and require significantly less computational resources, being more well-suited for real-world tasks such as documents retrieval. In addition, BERT can be adapted for specific tasks through pre-training and fine-tuning. Vitório et al. (2025b) pointed out the importance of fine-tuning for IR within the Brazilian legislative domain, as BERT models fine-tuned with Brazilian legislative data achieved better results than zero-shot ones.

An issue while using BERT models to perform IR, though, lies on their use for the retrieval of large documents, such as the legal ones. BERT has an input limitation of 512 tokens, while Brazilian legislative documents, for instance, comprise about 700 tokens, on average (VITÓRIO et al., 2025a). This limitation makes that documents bigger than 512 tokens must be truncated, losing information that may be important for the retrieval process.

2.1.1 Legal Information Retrieval

Legal Information Retrieval (LIR) has become a prominent issue in the application of Artificial Intelligence techniques in the area of law. The proper functioning of legal institutions requires the retrieval of relevant documents from extensive datasets. The information revolution and the Open Data movement have further emphasized this requirement, as there has been a significant increase in the availability of legal data, especially on the Internet. Data accessibility, however, did not keep up with this growth (OPIJNEN; SANTOS, 2017).

In this work, we classify legal documents in two main categories: judicial and legislative. Judicial documents comprise court decisions — which are also referred to as *jurisprudence*

—, lawsuits, and other documents created and used by courts within the judicial process. Meanwhile, legislative data is composed of the various documents created during the law-making process, including legislation and bills.

Those categories of documents are written using different structures, jargon, and domain-specific languages, which makes necessary the development of specific resources for each document type. Resources built from judicial documents may not be useful for processing legislative ones, and vice versa.

In the judicial realm, it is crucial for judges and lawyers to retrieve and provide access to similar cases, as court decisions for cases similar to the current one should be taken into account. Nevertheless, the concept of similarity between the cases is not well-defined, requiring input from specialists (BHATTACHARYA et al., 2020). The retrieval is commonly performed using the courts' computational systems, which are, however, usually inefficient legacy systems based on Boolean logic (GOMES; LADEIRA, 2020). Using keywords and operators to construct the query, these systems are complex and rely on the user's knowledge to choose the appropriate keywords (RUSSELL-ROSE; CHAMBERLAIN; AZZOPARDI, 2018).

Gomes and Ladeira (2020) evaluated the Boolean-based legacy system of the Brazilian Superior Court of Justice (STJ). The authors compared that approach with IR methods based on document similarity, such as TF-IDF, BM25, and word embedding LMs. They used jurisprudential data from STJ and found out that the IR techniques surpassed the legacy system in terms of both performance and usability.

For the legislative scenario, the situation is even more complex. The law-making process produces crucial information that can also have a significant impact on the lives of citizens, leading to changes in society. This information, nevertheless, must be properly organized, stored, and made available for both citizens and parliamentarians, as immediate access is necessary for it to be well-used (BRANDT, 2018). Thus, more efficient IR methods must be developed in order to keep up with the growing demand for information and to efficiently obtain legislative data.

Cantador and Sánchez (2020) proposed a novel method for the retrieval of legislative texts. They assessed documents from the Spanish Congress of Deputies, which are part of the *Parlamento2030*[1] dataset. This dataset consists of debate transcripts and legislative proposals. The authors improved the retrieval process by incorporating a semantic relation measure into

[1] https://www.parlamento2030.es/about-en

the Vector Space Model (VSM) (SALTON; WONG; YANG, 1975) and combining it with an ontology-based document representation model.

Both studies conducted by Gomes and Ladeira (2020) — using Brazilian Portuguese judicial documents — and Cantador and Sánchez (2020) — using Spanish legislative data — aimed to evaluate the performance of their models by employing experts to assess the relevance of the retrieved documents. However, none of them made this feedback information available. This shows that, even when researchers use experts to provide RF information for their experiments, they usually only use the feedback to evaluate their techniques, without making these corpora available to the community. Thus, publicly available corpora, like the one also presented in this study, are highly important as they allow for the evaluation of IR techniques in various scenarios.

Finally, Souza et al. (2021a, 2021b) investigated IR algorithms and presented a pipeline for the retrieval of legislative documents within the context of the Brazilian Chamber of Deputies. Evaluating the use of three variants of the BM25 algorithm, along with different preprocessing techniques, they built the IR system currently employed by the Chamber to retrieve bills and other queries similar to a parliamentarian's request. Santos et al. (2024), on their turn, proposed a hybrid system combining a BM25 variant and a BERT-based model fine-tuned with legislative data to also deal with the Brazilian Chamber of Deputies scenario. They evaluated five SBERT models and the BM25L variant (LV; ZHAI, 2011) with and without pre-processing techniques in order to select the best combination for their hybrid system.

2.2 RELEVANCE FEEDBACK AND ITS USE FOR SIMILAR QUERIES

As aforementioned, the real relevance of a document must be judged by the user, according to their information need. Thus, the user can point out whether a document is relevant to them or not, or even the decree of relevance of that document, e.g., whether the document is completely relevant or only partially relevant to them.

The Relevance Feedback (RF) method consists of utilizing a user's annotation on the relevance of a document to enhance IR for a specific query. Usually, this information is used to select specific terms and expressions from the relevant documents in order to add them to the query or to adjust the weights of terms in the original query (CARPINETO; ROMANO, 2012), in a process called Query Expansion (QE). As most users find it difficult to formulate a good query to express their actual information need, the RF method started suggesting that the query formulation process should be iterative, expanding the query after each search (SALTON; BUCKLEY, 1990).

The first search should be performed with an initial query treated as a tentative, a trial run only designed to retrieve a few useful documents. Then, these few documents could have their relevance judged and the documents considered relevant should be used to improve the query formulation, hoping that, in the subsequent searches, more and more useful documents will be retrieved by the system (SALTON; BUCKLEY, 1990). Thus, this process should be repeated a few times in order to achieve better results for a query.

Another way to use the RF information is through Supervised Machine Learning, where IR is treated as a two-class classification problem: *relevant* and *irrelevant* (OKABE; YAMADA, 2005). A classifier is trained using the user's judgments as a training set, then it is used to label new documents as either relevant or irrelevant for that query. Onoda, Murata and Yamada (2007) used RF to interactively train a Support Vector Machine (SVM) classifier aiming to improve the documents retrieval performance. The authors utilized a straightforward IR technique based on VSM to perform the first search and retrieve the initial list of documents, which were manually judged based on their relevance. Subsequently, the SVM classifier was trained using this data and used to generate the final list of documents.

As the users are typically reluctant to provide the feedback information in real search contexts and as the good performance of RF techniques depends on the existence of sufficient relevance judgments (ALY, 2008), the Pseudo-Relevance Feedback (PRF) method is often used. The PRF method uses pseudo-feedback — also called *retrieval feedback* or *blind feedback* — to simulate the feedback given by users, when this is not available. This local feedback information mimics relevance feedback by assuming that the top-ranked documents retrieved by the IR system are relevant (MANNING; RAGHAVAN; SCHÜTZE, 2008).

The PRF approach is simple and computationally efficient, but it has clear disadvantages compared to feedback provided by experts (CARPINETO; ROMANO, 2012). If many of the top documents are actually relevant, the process achieves results similar to the use of real relevance judgments. However, if none of the top documents are relevant and they are used for QE, for instance, the expansion will have negative effects as the new query will emphasize the same mistakes that caused the poor initial retrieval (BUCKLEY et al., 1995). Thus, the success of PRF depends on how good is the IR system being used.

Finally, a third kind of relevance feedback is obtained from indirect sources, such as user clicks or measuring the time a user spent on a webpage or reading a document. This is called *implicit feedback* and can be obtained by monitoring and interpreting the user's actions and behavior (JANNACH; LERCHE; ZANKER, 2018). It can be used either for IR systems and for recommendation systems, and is more available for real-world applications than explicit feedback. However, we cannot be always sure that this kind of feedback is correctly interpreted, as the user is not explicitly stating their preferences or explicitly measuring the relevance of a document based on their information need. Therefore, it is less reliable than feedback from explicit sources, although it may be more useful than pseudo-feedback, as it contains some kind of judgment by the user (MANNING; RAGHAVAN; SCHÜTZE, 2008).

2.2.1 Looking for similar past queries

The use of either real relevance feedback or pseudo-feedback generally focuses on the improvement for a specific query. The feedback information is used within a specific session and/or to iteratively enhance the search results for the query currently being processed (YIN et al., 2002). Nevertheless, stored past relevance data could also be used to improve the retrieval performance for future queries.

The IR process for each query is unique, since the documents that are relevant to a query may not be relevant to any other, as the actual relevance of the documents must be judged by the user, according to their information need. There is an alternative, however, to consider past queries that are similar to the current one and use the RF information given for them. If there are very similar queries in the system's usage history, the documents judged as relevant to those queries may also be relevant to the query currently being processed (GUTIÉRREZ-SOTO; HUBERT, 2014)

Although it is possible to observe the use of historical feedback for image retrieval for at least two decades (YIN et al., 2002), there are only a few studies in the literature that deal with this kind of use of RF information for textual document retrieval. According to Gutiérrez-Soto (2016), this is due to the fact that there are no available benchmark datasets with relevance information for similar queries. Popular IR evaluation collections — such as the ones from TREC[2] and CLEF[3] — only provide sets of dissimilar queries and topics.

Therefore, most of the studies that perform IR regarding historical data forcus in the Personalized Information Retrieval area, in which past information about a user's preferences

[2] https://trec.nist.gov
[3] http://www.clef-initiative.eu

is used to improve the search for that user, within a session. This process is commonly used in search engines on the Internet (GUTIÉRREZ-SOTO, 2016). In this study, however, the goal is to use the historical data outside a user's session, considering feedback given by different users.

For Moshfeghi, Velinov and Triantafillou (2016), the main challenge dealing with past queries is to find the similar ones. While Cetintas, Si and Yuan (2011) stated that there are many approaches to measure the similarity between queries, from which two are more used: term-based and retrieval-based. The former computes the similarity between the terms presents in the query, such as using cosine similarity or edit distance, whereas the latter compares the retrieved documents list for each query. Gutiérrez-Soto et al. (2021) evaluated the semantic match, using a set of semantically similar terms, between the user's query and past stored ones, pointing out the advantages of using the queries' semantics for web search engines caching. More recently, BERT and SBERT allowed to find paraphrases, i.e, texts with similar or identical meaning, through the generation and comparison of embeddings containing contextual and semantic meanings (REIMERS; GUREVYCH, 2019). This use of LMs may help the search for similar queries.

2.2.2 Related work

In this section, the related work using past queries information to improve the IR process outside user sessions is presented. First, we present an overview of studies that used the RF information to improve the retrieval process regardless of the method used, from which we could conclude that most studies use the past information to reformulate the queries. Later, we discuss two works that used this information to respond directly to a new query, thus are more related to our work.

Fitzpatrick and Dent (1997) analyzed the effect of using past similar queries to perform automatic QE. They compared the retrieval process using no feedback with two methods to expand the query: using top-document feedback and past-query feedback. The former was performed in the same way as the standard Query Expansion using RF method, in which the query is submitted to the system, the system retrieves a list of documents, and terms selected from the top-ranked documents are added to the query in order to perform another search. Meanwhile, the latter was performed by computing the similarity between the current query and past stored ones, aiming to create an affinity pool of queries, from which the top-ranked documents were selected to perform the expansion. The use of past-query feedback improved the results, specially when analyzing only the queries which have affinity pools.

The authors (FITZPATRICK; DENT, 1997), however, used automatic feedback, i.e. pseudo-feedback, to select the relevant documents for each query, thus they did not use human-generated relevance judgments. Also, they created a specific measure to compute the inter-query similarity, which compares the list of retrieved documents for each query using the probability of relevance for each position in the TREC datasets used. Therefore, it is a measure that depends on an extensively analysis of the dataset and cannot be easily used, also requiring an initial retrieval for a new query to find similar past ones, which may bias the expansion.

Instead of only extracting terms from the documents considered relevant for past queries, Billerbeck et al. (2003) looked for queries associated to the relevant documents in order to extract the terms. They performed an initial search for the current query to get the top-ranked documents for it, then looked for a set of queries associated to those documents and built surrogates using these sets. Thus, the expansion terms used for QE were selected from the surrogates. The association process also uses pseudo-feedback, as it considers the top-ranked documents retrieved for a query as the associated documents to that query. As the IR algorithm, they used Okapi BM25 (ROBERTSON et al., 1994).

Dealing with Collaborative Information Retrieval, Hust (2004) evaluated several techniques for QE using the relevance information from past queries submitted by different users. The cosine similarity was used to measure the similarity between the queries and a threshold was used to decide which ones are sufficiently similar. The author found out that benchmark IR datasets, such as SMART[4] and the ones from TREC, do not have many queries with highly correlated similarities. As the results could not allow a conclusion about the effectiveness of the methods proposed, his justification was the lack of similar queries in the datasets used. Okapi BM25 was used for the IR process and the study was also based on Pseudo-Relevance Feedback.

Cetintas, Si and Yuan (2011) evaluated cosine similarity and a retrieval-based measure to find similar queries to be used for resource selection in the domain of Distributed Information Retrieval. They also highlighted the lack of available collections containing similar queries, which they mitigated by generating simulated queries from TREC datasets.

El-Ghali and El-Qadi (2017) also performed QE using the relevant list of documents from past similar queries. First, the authors used a language model to find the most related past queries to the current one, in a phase they called Query Recommendation. The language model

[4] ftp://ftp.cs. cornell.edu/pub/smart

computed the similarity between two queries by the capacity of a past query to generate the new one, considering the terms present in both queries and the list of documents clicked for the past query. After that, they used the Latent Semantic Analysis (LSA) (DEERWESTER et al., 1990) model to select candidate terms to perform the expansion. The LSA method was applied using: 1) the query to be expanded; 2) the top-ranked documents for it; 3) the recommended queries; and 4) the top-clicked documents for every recommended query. Thus, they used both pseudo-feedback and implicit feedback.

Within the scope of Legal Information Retrieval, Schweighofer and Geist (2007) commented on the use of stored search context information, such as users interactions with the system, to perform QE. They stated that this information is stored in LIR systems for billing purposes and that this approach would be tested in a dataset of Austrian law. However, no other paper from the authors was found considering the use of feedback information for QE and this was the only reference to the use of relevance information from past queries found in the legal domain.

As can be seen, most of the studies working with RF information from past queries use it to perform QE, reformulating the queries. Nevertheless, two other studies found in the literature are closer to this work, as they reuse past relevant documents to respond directly to a new query, without modifying it.

Song and Myaeng (2012) proposed a novel term weighting method which considers the relevant and non-relevant documents from past retrieval results. The authors assumed that the role of a term in past queries could predict its value in future queries, through a measure called *discrimination power*. The discrimination power value was computed using the ranks or the similarity values of the retrieved documents for a past query — i.e., the relevance judgment was not given by users —, then it was added to the TF-IDF weighting function (ROBERTSON, 2004). This information, however, was obtained from all the term's history in the dataset and its capacity to separate relevant from non-relevant documents. All the queries that contained a particular term were considered, disregarding whether they are similar or not to the current one.

As they (SONG; MYAENG, 2012) looked for the set of terms that appear in past queries to compute the discrimination power value and this set was not large, Query Expansion with PRF was also used to increase the number of terms in the queries. Without it, most of the queries would not be influenced by the novel weighting method, as none of their terms appeared in past queries. The results showed that this method improved TF-IDF-based algorithms, such as BM25.

To deal with the lack of datasets containing relevance judgments for similar queries, Gutiérrez-Soto (2016) had to simulate IR collections, including not only documents and queries, but also the judgments of relevance, to evaluate his methods. He presented four Monte Carlo algorithms to assign a probability of relevance in the ranking of documents based on the position of relevant documents obtained from the most similar past query, using the assumption that the relevant documents tend to appear at the top of the list. Using the cosine similarity to search for the similar queries, he also used this measure as baseline to perform the retrieval of documents, reaching better results with the Monte Carlo algorithms.

As conclusion, Table 1 presents a summary of the related work and their comparison with this study. As can been seen, none of the studies found in the literature using the RF information from past queries used it to re-rank the retrieved documents. Also, they did not used relevance judgments from experts, as this information is not available for similar queries in the benchmark datasets that can be found.

Table 1 – Summary of related work and comparison with this study.

Work	Use of past queries	Similarity measure	RF source	IR algorithms
Fitzpatrick and Dent (1997)	for query expansion	retrieval-based	pseudo	OpenText search engine
Billerbeck et al. (2003)	for query expansion	retrieval-based	pseudo	Okapi BM25
Hust (2004)	for query expansion	cosine	pseudo	Okapi BM25
Cetintas, Si and Yuan (2011)	for resource selection	cosine, retrieval-based	pseudo	ReDDE
Song and Myaeng (2012)	in the retrieval (term weighting)	-	pseudo	TFIDF, DFR_BM25, Hiemstra model
Gutiérrez-Soto (2016)	in the retrieval (new algorithms)	cosine	simulated	own algorithms (Monte Carlo)
El-Ghali and El-Qadi (2017)	for query expansion	term-based	pseudo, implicit	own algorithm (LSARQ)
This study	in the retrieval (re-ranking)	cosine, SBERT embeddings	experts	Okapi BM25, BM25L, BERT-based

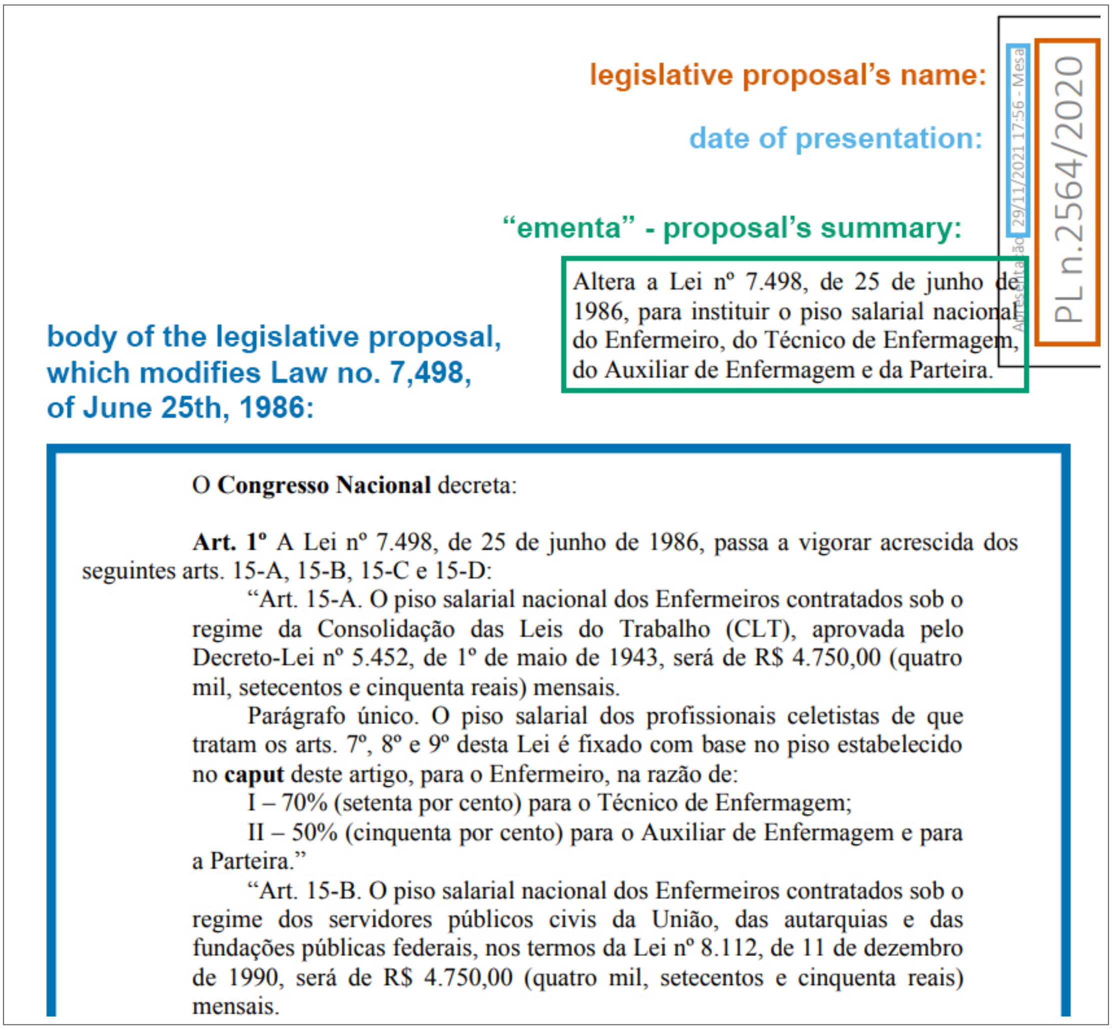
Source: Created by the author (2025)

The studies presented in this subsection helps to answer **RQ1** as they show that the use of the RF information from stored past queries can improve the results of the IR process for new queries. However, to achieve this improvement, it is necessary a set of queries sufficiently similar to the one currently being processed, which is difficult to find in the literature, as pointed out by works such as the ones from Hust (2004), Cetintas, Si and Yuan (2011), and Gutiérrez-Soto (2016).

3 THE SCENARIO OF THE BRAZILIAN CHAMBER OF DEPUTIES

The Brazilian legislative process comprises the drafting, analysis, and voting of various types of legislative proposals, such as bills of law (in Portuguese, *Projeto de Lei* or, to simplify, *PL*), provisional measures, constitutional amendments (in Portuguese, *Proposta de Emenda à Constituição* or *PEC*), legislative decrees, among others. These proposals are pointed out by Brandt (2020) as the key element of the legislative process and each one of their types follows a different procedure and can produce different effects, such as to create a new law, to modify an existing law, or to promote changes in the Constitution. However, all of them can be referred to as *bills*. Figure 1 presents an illustrative example of a Bill of Law (*Projeto de Lei*) formulated and ratified within the Brazilian legislative process, highlighting its main parts. This legislative proposal aims to modify a previous law.

Figure 1 – Example of a legislative proposal formulated and ratified within the Brazilian legislative process.



Source: Created by the author (2025)

The Chamber of Deputies is one of the two chambers that compose the Brazilian National Congress, alongside the Federal Senate. Both chambers are responsible for the legislative process: creating, voting, and revising the legislative proposals. A bill laid before the Chamber of Deputies must be revised by the Senate, and vice versa.

The Chamber also has another crucial role in the legislative scenario. Its Legislative Consulting (Conle) department works on retrieving similar, previously submitted proposals in order to assist parliamentarians in the law-making process. Before making a legislative proposal, a parliamentarian submits a query to request a list of similar documents, including active or inactive bills and similar requests made by other parliamentarians. This process is called *preliminary search* (in Portuguese, *pesquisa prévia*), and the parliamentarians' requests are known as *legislative consultations* or *job requests* (*solicitações de trabalho)*. The list of retrieved documents, as a result of the preliminary search, helps to verify if there are similar bills already being discussed in the Chamber. The preliminary search also offers support to the parliamentarian in making a new legislative proposal.

It is also worth to mention that this process of consultancy and other advisory services are confidential, according to Article 13 of Resolution of the Chamber of Deputies No. 48, of 1993 (Brazilian Chamber of Deputies, 1993). For instance, this restriction made it impossible for the dataset used by Souza et al. (2021a, 2021b) and Santos et al. (2024) to be made available. Table 2 gives examples of real legislative consultations.

Until 2021, the preliminary search was performed in a manual and very time-consuming way. The consultants from Conle had to identify keywords within the legislative consultation and use a Boolean system to retrieve a set of documents based on those keywords. They, then, had to read all of the retrieved documents and select those that fit the request, providing a list of documents to the parliamentarian.

Nowadays, however, Conle uses an IR system, whose pipeline was proposed by Souza et al. (2021b), to automatically retrieve pertinent documents. The system is based on BM25L (LV; ZHAI, 2011) and the use of a set of preprocessing techniques: the removal of punctuation, accentuation, and stopwords, stemming — with the Savoy algorithm (SAVOY, 2006) —, and a combination of unigram and bigram. This preprocessing configuration was the best for this scenario, as reported by Souza et al. (2021b, 2021a), which evaluated several combinations in order to find the best pipeline for the system.

Table 2 – Examples of legislative consultations and their English translation.

Legislative consultation	**English translation**
obrigar a União e os estados a implementar compensações financeiras aos municípios que abrigam unidades prisionais Federais.	oblige the Union and the states to implement financial compensation to the municipalities that in which Federal prison units are located.
AGRADEÇO O ESTUDO REALIZADO PELO NOBRE CONSULTOR [OMITIDO] TODAVIA O DEPUTADO REQUER A FEITURA DE PROJETO DE LEI QUE OBRIGUE QUE AS DOAÇÕES DE ARMAMENTOS POSSAM SER FEITAS SOMENTE PARA AS FORÇAS POLICIAIS.	I WOULD LIKE TO THANK THE STUDY CARRIED OUT BY THE NOBLE CONSULTANT [OMITTED] HOWEVER, THE PARLIAMENTARIAN REQUESTS THE MAKING OF A BILL THAT REQUIRES THAT WEAPONS DONATIONS CAN ONLY BE MADE TO THE POLICE FORCES
Solicito um projeto no sentido de tornar hediondo o crime de corrupção de menor previsto no ECA.	I request a project to make heinous the crime of corruption of minors provided for in the ECA.
Prezados colegas, a pedido da deputada [OMITIDO] solicitamos a confecção de um PL que garanta aos usuários do passe livre o direito de um percentual de pelo menos 10% (ou percentual razoável) dos assentos em aeronaves e transporte terrestre, ferroviário e marítimo. Assegurando que o usuário poderá agendar a ida e volta no mesmo ato. E que o beneficio terá que ser fornecido em todas as modalidades de transportes terrestre (convencionais/executivos) e transporte aéreo (domésticos). Estamos à disposição para demais esclarecimentos. Encaminho em anexo documentos de uma ação judicial relacionadas ao tema, para devidas inclusões no texto. Att. [OMITIDO]	Dear colleagues, at the request of Congresswoman [OMITTED], we request the drafting of a bill that guarantees free pass users the right to at least 10% (or a reasonable percentage) of seats on aircraft and on land, rail, and sea transportation. This bill ensures that users can schedule their round trip journeys simultaneously. The benefit must be provided on all modes of land transportation (conventional/executive) and air transportation (domestic). We are available for any further clarification. I am attaching documents from a related lawsuit for inclusion in the text. Att. [OMITTED]

Source: Created by the author (2025)

BM25L is a BM25 variant that fixes the Okapi's (ROBERTSON et al., 1994) preference for shorter documents by changing its scoring function (KAMPHUIS et al., 2020):

$$score(d,q) = \sum_{t \in q} \log\left(\frac{N+1}{DF(t)+0.5}\right) \cdot \frac{(k_1+1)\cdot(C(t,d)+\delta)}{k_1 + C(t,d) + \delta}, \tag{3.1}$$

in which δ is a parameter, and $C(t,d)$ is computed by:

$$C(t,d) = \frac{TF(t,d)}{1 - b + b \cdot \frac{|d|}{L}}. \tag{3.2}$$

Thus, the aforementioned IR system automatically retrieves the documents from the Chamber's database, which contains more than 144,000 bills (BRANDT, 2020), estimating their relevance to the user's query, i.e., a parliamentarian's consultation. It also retrieves past legislative consultations similar to the current one.

At the end of the retrieval process using the IR system, a Conle consultant should select the documents which actually respond to the parliamentarian's request. In this way, feedback information given by experts is automatically stored in the system and can be used to improve the model for future queries.

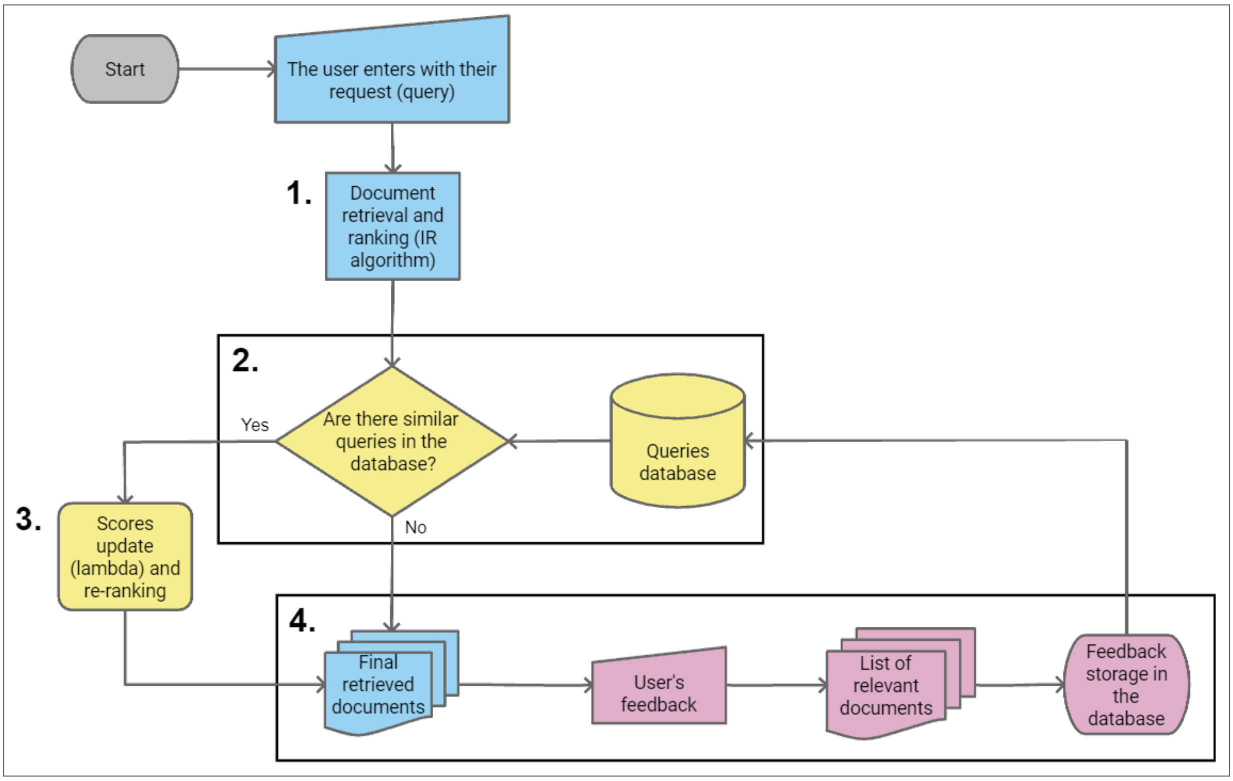
4 ULYSSES-RFSQ

In this chapter, the proposed method, called *Ulysses-RFSQ* (**RFSQ**: **R**elevance **F**eedback for **S**imilar **Q**ueries), is described.

It consists of re-ranking the documents that were judged for old queries similar to the current one, by the addition of a bonus or a penalty, according to their relevance judgment. It is composed by four steps, which are detailed in the following sections: 1) the preliminary ranking of documents by an IR algorithm; 2) the similar queries selection; 3) the ranking update; and 4) the Relevance Feedback information acquisition.

Figure 2 presents the method's pipeline, in which the re-ranking parts added by Ulysses-RFSQ are in yellow, whereas the blue elements represent the standard IR process and the RF stages are represented by the pink ones. The numbers point out the four mentioned steps.

Figure 2 – Ulysses-RFSQ's pipeline, pointing out the re-ranking stages added by it (in yellow) to the standard IR process (in blue), as well as the RF stages (in pink).



Source: Created by the author (2025)

4.1 STEP 1: RANKING THE DOCUMENTS

The first step consists in the scoring and ranking of the documents by a base IR algorithm. Any algorithm that results in a score for the documents can be used. As explained in Chapter 2, in the standard IR process, the algorithm computes a score for each document and, according to it, the documents are ranked from the highest to the lowest. Then, a list containing the top n documents usually is presented to the user.

The scores computed by some IR algorithms, such as BM25, however, don't have an upper limit value, thus they need to be normalized in the range of [0, 1], in order to be used for Ulysses-RFSQ. Without this normalization, the value added to the documents' scores might not be sufficient to have an effect in the posterior documents re-ranking. For this, the Min-Max Normalization (Equation 4.1) can be used, as it preserves the relationships among the original values.

$$normalized\_score(d, q) = \frac{score(d, q) - min(all\_scores(q))}{max(all\_scores(q)) - min(all\_scores(q))}, \tag{4.1}$$

in which d is a document, q is a query, and all_scores is the set of scores for all documents.

4.2 STEP 2: SELECTING SIMILAR QUERIES

As Ulysses-RFSQ focuses on using the feedback information given for past queries to improve the documents retrieval, it is necessary to maintain and store the old queries in a database. This database must contain the query text and its RF information, with data about the documents judged for that query.

In the second step, the similarity between the current query and each query stored in the database is computed. After that, those queries that have a similarity greater than a cut-off threshold are selected. This threshold, which was called cut, is a parameter of the method that needs to be set and can vary according to the chosen similarity measure. If there are no queries that have a similarity greater than the threshold, the third step is skipped and the standard IR algorithm list of ranked documents is presented to the user, without any modification.

It is important to assess the value of the cut-off threshold to ensure that a sufficient number of queries will be used, while guaranteeing that they are similar to the current one.

4.3 STEP 3: UPDATING THE RANKING

After selecting the similar queries, the documents judged for them have their scores updated by the addition of a value called $lambda$ (λ).

The idea behind Ulysses-RFSQ is to give a bonus for those documents judged as *relevant* for past similar queries, while giving a penalty for those judged as *irrelevant*. Based on using or not the *irrelevant* information, as well as different levels of relevance, four versions of this method were created. These versions are detailed in the following subsections, alongside the preliminary version of Ulysses-RFSQ proposed in (VITÓRIO et al., 2022). All of the versions are summarized in Table 3.

Regardless of the version used, the final score for each document is computed by Equation 4.2:

$$final\_score(d, q) = normalized\_score(d, q) + \lambda(d, q). \tag{4.2}$$

With the final scores, the documents are re-ranked and the new ranked list is the result of the IR process.

4.3.1 Ulysses-RFSQ-v1: the preliminary version

The first version of Ulysses-RFSQ was proposed and evaluated in (VITÓRIO et al., 2022) and it used only the *relevant* documents information. Thus, the λ's value was always positive, adding a bonus to the score of those documents.

In this version, $lambda$ was computed by Equation 4.3:

$$\lambda(d, q) = \ln\left(\sum_{q_j \in Q} (sim(q, q_j) \cdot normalized\_score(d, q_j)) + 1\right), \tag{4.3}$$

in which q is the current query, Q is the set of similar past queries, and $sim(q, q_j)$ is the similarity between q and q_j. The natural logarithm was used to keep the λ's value in a small range, and the addition of 1 to the sum prevented the value from being negative.

As can be noticed by Equation 4.3, λ is computed based on two factors: the similarity between the past query and the current one, and the IR algorithm's (e.g. BM25) normalized score computed for that document according to the past query. This score also have to be stored in the database. Thus, the value of λ is directly proportional to the similarity between the queries and the estimation of relevance for the document with regard to that past query.

Another characteristic of this equation is that if a document is present in more than one similar query, its bonus increases, as $lambda$ considers the sum of all occurrences of that document in the similar queries set. On the other hand, if the document is not present in the judged list of any similar query, its bonus is 0, i.e., its score will remain the same.

4.3.2 Ulysses-RFSQ-OR: using only the *relevant* information

The second version of Ulysses-RFSQ, called **Ulysses-RFSQ-OR**, also uses only the *relevant* documents to compute λ. However, its formula changed from the preliminary one:

$$\lambda(d,q) = \tanh\left(\sum_{q_j \in Q}(sim(q,q_j) \cdot normalized\_score(d,q_j) \cdot rel(d,q_j))\right) \times \delta. \tag{4.4}$$

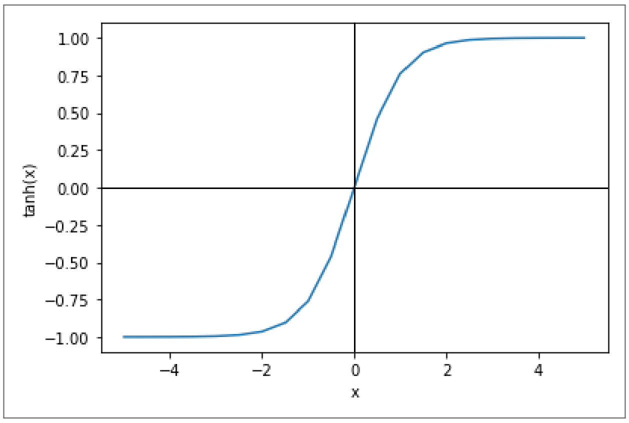
Three main changes can be observed between Equations 4.3 and 4.4: 1) the $rel(d,q_j)$ factor was added to present the relevance judgment of document d for query q_j; 2) the natural logarithm function was replaced by the hyperbolic tangent function; and 3) a δ parameter was added to control the importance of the past relevance information.

For the Ulysses-RFSQ-OR version, the $rel(d,q_j)$ factor is not important, as all documents used were judged as *relevant*, thus having a rel value of 1. It is essential, nevertheless, for the following versions, which use different categories of relevance.

The hyperbolic tangent function was also chosen for the use of different relevance categories, as it maps positive inputs into positive outputs and negative inputs into negative outputs, producing results in the range of [-1, 1]. On the other hand, the natural logarithm function used in the preliminary version cannot deal with negative inputs. Figure 3 presents a graph of the hyperbolic tangent function.

Finally, the δ parameter is used to control the importance of λ for the documents ranking update: if it is set with a large value, the RF information will have a greater importance for the IR process, giving a large bonus for the documents. Using this parameter, the user can have more control on the use of past information.

Figure 3 – Hyperbolic tangent function.



Source: Created by the author (2025)

4.3.3 Ulysses-RFSQ-RI: using both *relevant* and *irrelevant* information

The third version of Ulysses-RFSQ, **Ulysses-RFSQ-RI**, uses the same Equation 4.4, but considering both *relevant* and *irrelevant* documents. Thus, the λ's value may be either a bonus or a penalty for the documents. If a document was judged as *irrelevant* for most of the similar past queries, it will receive a penalty in its score. In this version, $rel(d, q_j)$ assumes the value of 1 if the document d was judged as *relevant* for query q_j and -1 otherwise.

4.3.4 Ulysses-RFSQ-DRL: using different relevance levels

The fourth version, **Ulysses-RFSQ-DRL**, was created for scenarios in which the documents were judged using different levels of relevance, e.g., *very relevant* and *somewhat relevant*. However, it ignores the *irrelevant* documents. Documents from different levels have different values for the $rel(d, q_j)$ factor, ranging from 0 to 1 according to how high in the relevance hierarchy they were judge.

4.3.5 Ulysses-RFSQ-ALL: using all relevance information available

Ulysses-RFSQ-ALL is the combination of the other versions, using the *irrelevant* information in addition to the different levels of relevance. For instance, in a scenario with two levels of relevance (*very relevant* and *somewhat relevant*), Ulysses-RFSQ-ALL uses three categories of information: *very relevant*, *somewhat relevant*, and *irrelevant*; thus the $rel(d, q_j)$ factor has three different values. This version can only be used in scenarios in which all this information is available.

Table 3 – Summary of Ulysses-RFSQ versions.

Version	λ formula	RF information used	Study that proposed
Ulysses-RFSQ-v1	Equation 4.3	only *relevant*	Vitório et al. (2022)
Ulysses-RFSQ-OR	Equation 4.4	only *relevant*	this study
Ulysses-RFSQ-RI	Equation 4.4	*relevant* and *irrelevant*	this study
Ulysses-RFSQ-DRL	Equation 4.4	different relevance levels (e.g., *very relevant* and *somewhat relevant*)	this study
Ulysses-RFSQ-ALL	Equation 4.4	*irrelevant* and different relevance levels (e.g., *very relevant* and *somewhat relevant*)	this study

Source: Created by the author (2025)

4.4 STEP 4: ACQUIRING THE RELEVANCE FEEDBACK INFORMATION

Finally, the n documents with the highest final scores are presented to the user. The user, then, provides feedback information, judging the retrieved documents based on their relevance to their request. This list of judged documents is stored with their respective scores in the database, as well as the query.

The document's score must be stored because these data will be used for future requests. Thus, the IR system is always being improved by the relevance information provided by users, feeding itself back. It is worth mentioning that this method can be used in two ways: 1) without any previous stored queries, thus, for the first use, the queries database is empty and Step 3 is skipped until the IR system is sufficiently used; or 2) using a previous feedback database, with which the $lambda$'s value might impact the performance from the start.

5 ULYSSES-RFCORPUS

In order to create a legislative IR dataset containing RF information that could be use to evaluate Ulysses-RSFQ and also be made available to the community, a corpus named **Ulysses-RFCorpus** was built together with Conle. It is composed by a set of queries simulating real legislative consultations created by the Conle consultants and their lists of legislative proposals judged for them. Thus, as the queries were not created by parliamentarians and do not contain private information, this corpus is publicly available[1]. A paper detailing the construction process of this corpus was also published (VITÓRIO et al., 2025a)

A total of 703 queries were created and provided to us by the Conle team. A group of 54 consultants worked on those queries, with an average of 13 queries per consultant and a standard deviation of 5.17. Four of them built more than 20 queries, while 11 consultants built less than 10. The minimum amount of queries built per a consultant was two and the maximum was 28.

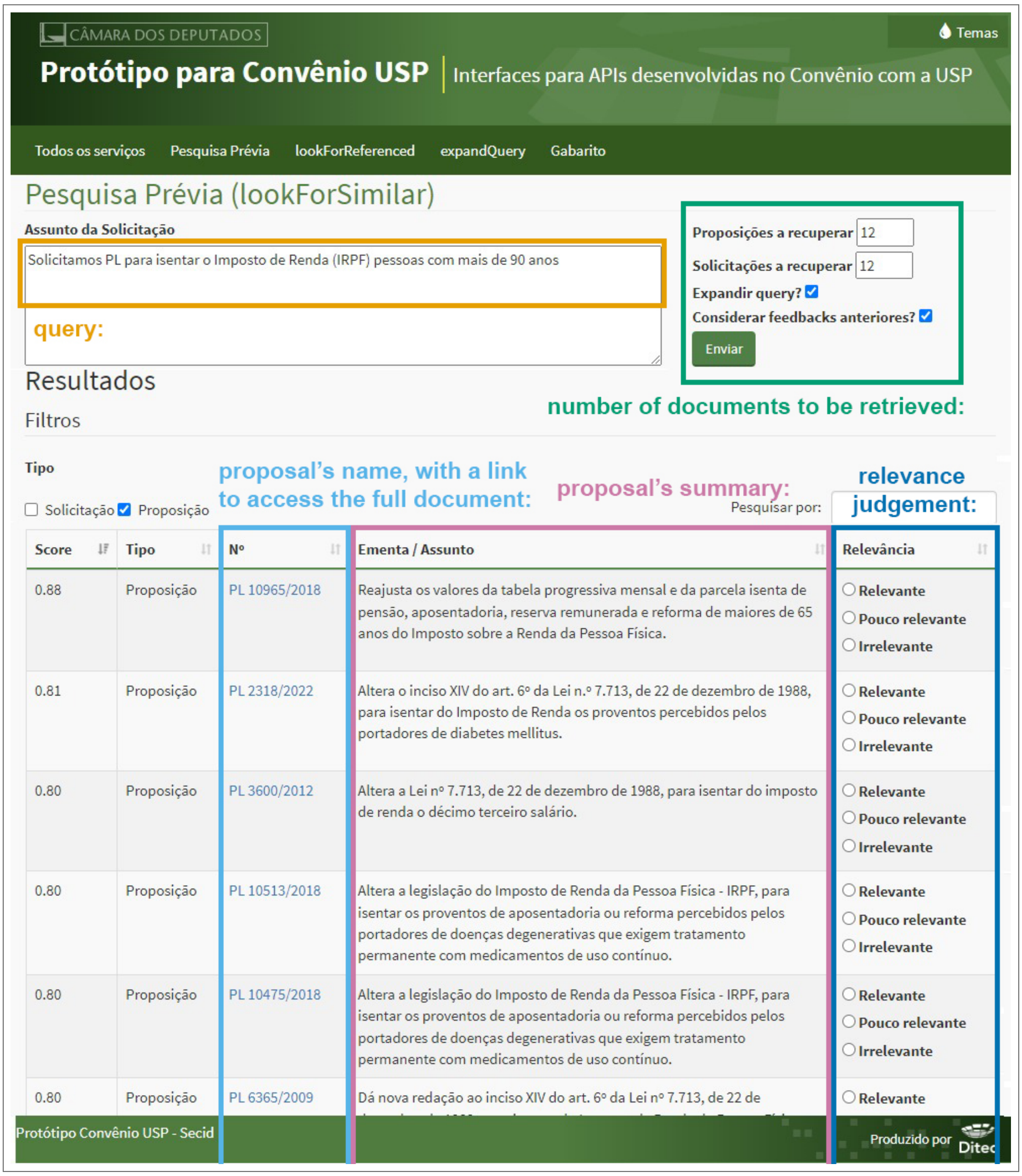
The consultants used the Conle's IR system (SOUZA et al., 2021b) to retrieve 12 similar bills and 12 other legislative consultations for each query. They, then, provided feedback on the relevance of the retrieved documents, categorizing them as either *very relevant* (*relevante*), *somewhat relevant* (*pouco relevante*), or *irrelevant* (*irrelevante*). As BM25L retrieves the top-ranked n documents, the Conle team decided to set the value of n at 12. Figure 4 presents the interface used to retrieve and judge the documents, in which the main parts are highlighted and explained. Each consultant provided feedback only for their own queries.

First, since we are unable to make the actual legislative consultations dataset available, due to the consultancy confidentiality explained in Chapter 3, the feedback information received for these documents were not used. We also opted to focus on the retrieval of large documents, such as bills, whereas legislative consultations are short. Therefore, this portion of the feedback information — regarding legislative consultations — was excluded from Ulysses-RFCorpus.

In addition, there were instances in which the consultant did not provide feedback for all 12 retrieved bills. As a result, queries for which they did not evaluate at least 10 documents were also excluded. As shown in Table 4, the consultants did not give sufficient feedback for 10 of the 703 queries. Finally, two of the queries made available by the Conle team were identical, thus one of them was excluded, resulting in a total of 692 queries in the corpus.

[1] https://github.com/ulysses-camara/Ulysses-RFCorpus

Figure 4 – Interface used by the Conle team to judge the retrieved documents.



Source: Created by the author (2025)

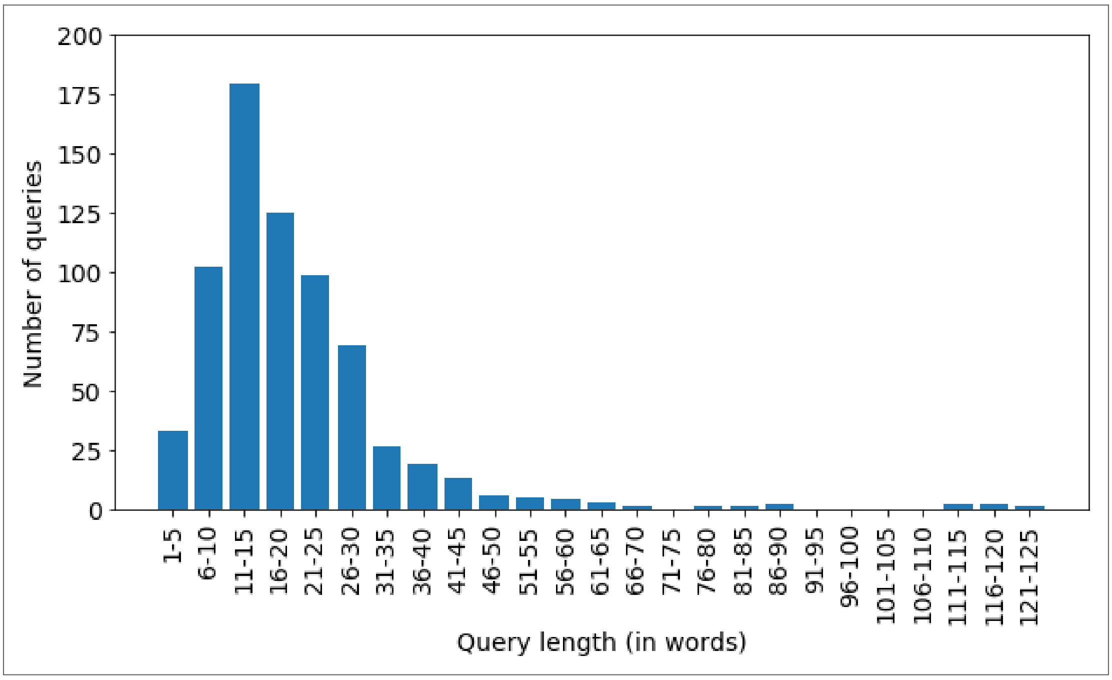
Table 4 – Number of queries by the quantity of bills judged for them.

# of bills judged	# of queries
12	609
11	69
10	15
9	1
5	1
2	1
0	7
Total	**703**

Source: Created by the author (2025)

Analyzing the queries text, it was found out that they have an average length of 19.93 words and a standard deviation of 14.30, with the shortest query consisting of only two words, whereas the largest query contains 121 words. Figure 5 presents the histogram of query length, indicating that the majority of queries consist of between six and 30 words. This shows that the simulated queries are similar to actual queries, as their length is in the range pointed out by Souza et al. (2021b) for actual legislative consultations: between 10 and 40 words.

Figure 5 – Histogram illustrating the distribution of query length in Ulysses-RFCorpus.



Source: Created by the author (2025)

Table 5 presents examples of the queries formulated by the Conle team, accompanied by their English translation. It was observed that legislative consultations, although frequently employing similar terminology, can exhibit distinct structural variations. The content of these texts can vary, ranging from meticulously crafted and formal requests to basic lists of keywords. Additionally, it is not uncommon for these texts to contain typographical and grammatical errors, as can be seen in the first two rows of Table 5.

Finally, the cosine similarity between the queries was computed aiming to verify their redundancy. Table 6 presents the number and percentage of queries that are similar to any other and the number of pairs of similar queries, considering different levels of similarity. Almost every query is similar to another one considering a cosine similarity of 0.1. Meanwhile, more than 11% of them may benefit from the use of Ulysses-RFSQ considering a cosine similarity greater than 0.4. One of the only two identical queries was removed from the dataset, as previously explained, thus there was no pair of queries with a similarity greater than 0.9.

Table 5 – Examples of queries built by the Conle team for Ulysses-RFCorpus and their English translation.

Query	English translation
Solicitamos PL para isentar o Imposto de Renda (IRPF) pessoas com mais de 90 anos [sic]	We request PL to exempt the Income Tax (IRPF) people over 90 years old [sic]
Solicitao [sic] PEC para alterar o art. 12 da CF, evitando a perda da nacionalidade brasilira [sic] do cidadão que adquiriu outra nacionalidade.	Requestthe [sic] PEC to amend art. 12 of the CF, avoiding the loss of Brazilan [sic] nationality of the citizen who acquired another nationality.
De ordem do sr. Deputado Fulano de tal , solicito projeto de lei proibindo o uso de fogos de artifício com barulho em todo País.	By order of mr. Deputy So-and-so , I request a bill banning the use of loud fireworks throughout the Country.
Alterar o art. 206 do código civil para aumentar para 2 anos o prazo da prescrição do segurado contra o segurador	Change art. 206 of the civil code to increase the statute of limitations of the insured against the insurer to 2 years
Alterar a lei 11.952/19 regularizando a posse de terra de fazendeiros anteriores a Constituição.	Amend law 11,952/19 regulating land ownership by farmers prior to Constitution.
prezado consultor solicito o estudo da possibilidade de elaboração de instrumento jurídico para permitir o cultivo de transgênicos em terras indigenas	dear consultant I request the study of the possibility of making a legal instrument to allow transgenic cultivation in indigenous lands
Estabelece o Dia Nacional do Contador.	Establishes National Accountant Day.
tributar altos lucros	tax high profits
proibir refrigerante escolas	ban soda schools
PENSÃO ESPECIAL ÓRFÃOS COVID	SPECIAL PENSION ORPHANS COVID

Source: Created by the author (2025)

Table 6 – Number of queries from Ulysses-RFCorpus that are similar to others and number of pairs of similar queries, by level of similarity.

Cosine similarity	# queries	% of queries	# pairs of queries
≥ 0.1	691	99.86	10,714
≥ 0.2	503	72.69	765
≥ 0.3	175	25.29	141
≥ 0.4	79	11.42	47
≥ 0.5	25	3.61	14
≥ 0.6	6	0.87	3
≥ 0.7	4	0.58	2
≥ 0.8	2	0.29	1
≥ 0.9	0	0.00	0

Source: Created by the author (2025)

As will be pointed out in Section 6.1.2, actual legislative consultations — when disregarding cases in which the same query was processed by different consultants — have a degree of redundancy similar to the simulated ones built for this corpus.

6 EXPERIMENTAL SETUP

In this chapter, the experimental setup used to evaluate the different versions of Ulysses-RFSQ is detailed. The goal is to answer the Research Questions presented in Chapter 1.

6.1 DATASETS

Three datasets containing legislative data were used to evaluate the proposed method: one containing legislative documents (bills) to be retrieved and other two containing queries and their respective lists of judged documents, including Ulysses-RFCorpus — presented and detailed in Chapter 5. All corpora were made available by the Brazilian Chamber of Deputies and built together with the Conle team.

The use of legislative data is justified due to the lack of publicly available IR datasets containing RF information for similar queries, as pointed out by Hust (2004), Cetintas, Si and Yuan (2011), Gutiérrez-Soto (2016), and Gutiérrez-Soto et al. (2021). Thus, the use of parliamentarian's queries, which are often redundant, makes the evaluation of Ulysses-RFSQ possible.

6.1.1 Bills corpus

For the IR process, a dataset containing 105,669 bills was used. Therefore, the goal of the IR algorithms was to retrieve documents from this dataset. These bills are publicly available on the Internet and can be found in the Chamber of Deputies' website[1].

It contains eight different types of bills: Recommendation (*Indicação* - INC); Bill of Legislative Decree (*Projeto de Decreto Legislativo* - PDL or PDC); Bill of Law (*Projeto de Lei* - PL or PLC); Proposal of Oversight and Control (*Proposta de Fiscalização e Controle* - PFC); Bill of Supplementary Law (*Projeto de Lei Complementar* - PLP); Bill of Resolution (*Projeto de Resolução* - PRC or PRN); Bill of Conversion (*Projeto de Lei de Conversão* - PLV); and Constitutional Amendment Bill (*Proposta de Emenda à Constituição* - PEC).

[1] https://www.camara.leg.br/busca-portal/proposicoes/pesquisa-simplificada

6.1.2 Preliminary Search corpus

In addition to Ulysses-RFCorpus, the other dataset provided by Conle and used in this study contains a set of legislative consultations (queries) and their list of relevant documents selected by a Conle consultant, as a result of the manual preliminary search — explained in Chapter 3. The legislative consultations present in this corpus are real ones created by parliamentarians, thereat this dataset could not be made available due to the confidentiality rules also explained in Chapter 3.

The queries and their lists of relevant documents were extracted from .DOC, .PDF, and .HTML files, which were sent to the parliamentarians as the result of the preliminary search. A total of 2,420 queries could be extracted from these files.

It is worth to mention that there were cases during the preliminary search in which the same query was processed by different consultants, presenting different lists of relevant documents. This fact denotes a problematic characteristic of the manual preliminary search: without an automatic and reliable method to retrieve the documents, the parliamentarians may obtain different lists of relevant documents for the same query — depending on which consultant have processed their request — and none of them may be complete. Thus, several important documents may be left out in this process.

Therefore, for this study, these identical queries were removed, but their lists of relevant documents were merged. In other words, in the cases in which the same query was processed by different consultants and resulted in different sets of relevant documents, these different sets were combined and every document judged as *relevant* for that query now composes its list of relevant documents, whereas one of the the duplicated queries was removed. This step, together with removing other problematic queries — such as queries for which the extraction was not successful —, resulted in a corpus with 1,990 queries able to be used.

The number of relevant documents for each query varies from one to 66, but a small percentage of them (only 56 queries) have more than 20 relevant documents in their lists, which were also removed from this evaluation. As the IR algorithms used in this study require to set the number of documents to be retrieved, it was set for 20 documents and queries with more than 20 relevant documents would harm the performance metrics. Thus, the final Preliminary Search corpus contained 1,934 queries.

In order to confirm the redundancy of the legislative consultations, the cosine similarity between the queries was also computed for the Preliminary Search corpus. Table 7 presents the number and the percentage of queries that are similar to any other query and the number of pairs of similar queries, considering different levels of similarity. It shows that only one query do not have another similar to it considering a cosine similarity greater than 0.1, which also occurred in Ulysses-RFCorpus.

Table 7 – Number of queries from the Preliminary Search corpus that are similar to others and number of pairs of similar queries, by level of similarity.

Cosine similarity	# of queries	% of queries	# of pairs of queries
≥ 0.1	1,933	99.95	129,351
≥ 0.2	1,421	73.47	4,194
≥ 0.3	597	30.87	607
≥ 0.4	208	10.75	151
≥ 0.5	78	4.03	50
≥ 0.6	43	2.22	27
≥ 0.7	16	0.83	11
≥ 0.8	10	0.52	7
≥ 0.9	4	0.21	2

Source: Created by the author (2025)

Examining this scenario, it can also be concluded that, even disregarding all of the identical queries, there still are two nearly-identical pairs. In addition, more than 30% of the corpus may benefit from the use of Ulysses-RFSQ considering a cut parameter of 0.3. It is worth remembering that the assessment of the cut parameter is crucial to determine which similarity threshold is the most suitable for each scenario.

6.1.3 Ulysses-RFCorpus

Besides the building and cleaning processes detailed in Chapter 5, some other queries were disregarded from Ulysses-RFCorpus for the experiments conducted in this study. From the 692 queries contained in the final version of Ulysses-RFCorpus, 46 do not present any relevant document in their lists — i.e., from the retrieved documents for those queries, the consultants judged none as *very relevant* or *somewhat relevant*. Thus, these queries were removed and, for the experiments conducted in this study, 646 queries were used.

Therefore, Table 8 summarizes the two corpora used to evaluate the proposed method, pointing out their sizes.

Table 8 – Summary of the queries datasets used in this study.

Dataset	# of queries
Ulysses-RFCorpus	646
Preliminary Search corpus	1,934

Source: Created by the author (2025)

6.2 EXPERIMENTS

In order to evaluate the proposed method and answer the Research Questions presented in Chapter 1, its performance was compared with a baseline without using the past queries relevance information. Thus, the Ulysses-RFSQ results were compared with the base IR algorithms results without re-ranking, using only the standard IR process — as described in the blue parts of Figure 2.

All the experiments were built and performed using the Python language and the Euler cluster[2] from the Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI) of the Universidade de São Paulo (USP). This cluster allows code execution using GPU processing with the following specifications: two Intel Xeon E5-2650v4 processors at 2.2 GHz with 12 cores each, 128 GB DDR3 1866MHz memory, and 1 Nvidia Tesla P100 GPU - 3584 CUDA cores - 16 GB.

6.2.1 Base IR algorithms

Two approaches were used as the base IR algorithm for which Ulysses-RFSQ was applied. Between the two approaches, 14 different techniques were evaluated in this study, which are summarized in Table 9 and are explained in the following subsections.

6.2.1.1 BM25 approach

First, two variants of BM25 — Okapi BM25 (ROBERTSON et al., 1994) and BM25L (LV; ZHAI, 2011) — were used in order to check if Ulysses-RFSQ could improve the results for different versions of this algorithm. Both variants were also evaluated with and without pre-processing techniques — identified by, respectively, “PRE” and “NP” in Table 9.

[2] https://euler.cemeai.icmc.usp.br

Table 9 – Summary of the IR algorithms used in this study.

#	Algorithm	Approach	Domain	Language
1.	BM25L_NP	BM25	-	-
2.	BM25L_PRE	BM25	-	-
3.	Okapi BM25_NP	BM25	-	-
4.	Okapi BM25_PRE	BM25	-	-
5.	BERTimbau	SBERT	various	Brazilian Portuguese
6.	Legal-BERTimbau	SBERT	legal	European Portuguese
7.	JurisBERT	SBERT	legal	Brazilian Portuguese
8.	BERTimbauLaw	SBERT	legal	Brazilian Portuguese
9.	LegalBert-pt	SBERT	legal	Brazilian Portuguese
10.	LaBSE	SBERT	various	multilingual
11.	Paraphrase Multilingual MPNet	SBERT	various	multilingual
12.	Paraphrase Multilingual MiniLM	SBERT	various	multilingual
13.	FT BERTimbau	SBERT	legislative	Brazilian Portuguese
14.	FT LegalBert-pt	SBERT	legislative	Brazilian Portuguese

Source: Created by the author (2025)

The choice for BM25 lies in the simplicity of this algorithm, in its usage for retrieving legal documents (OLIVEIRA; JUNIOR, 2018; GOMES; LADEIRA, 2020; CHALKIDIS et al., 2021), and in its good performance in this specific scenario (SOUZA et al., 2021b; SANTOS et al., 2024; VITÓRIO et al., 2025b). As aforementioned, the IR model used by Conle was built from BM25L.

For the preprocessed version of each variant, both documents and queries were processed using the same techniques as presented by Souza et al. (2021b):

- punctuation, accentuation, and stopwords removal;
- stemming with the Savoy algorithm (SAVOY, 2006);
- a combination of unigram and bigram.

The Python libraries NLTK[3] and scikit-learn[4] were used for this preprocessing step. As for the BM25 parameters' values, the recommendations of the original papers' authors were followed: $k_1 = 1.5$, $b = 0.75$, and $\delta = 0.5$.

6.2.1.2 SBERT approach

The second approach uses the SBERT architecture to generate contextual embeddings from the query and the documents and, then, computing their similarity. The cosine measure

[3] https://www.nltk.org
[4] https://scikit-learn.org/stable/

was chosen to compute the similarity between the embeddings, as it was used by the SBERT authors (REIMERS; GUREVYCH, 2019) and results in a score in the range of [0, 1].

A total of 10 different publicly available BERT-based models were selected to perform the IR process, following the models evaluated by Vitório et al. (2025b) shich achieved the best results. Four of them were trained using Portuguese data from the legal domain: Legal-BERTimbau, JurisBERT, BERTimbauLaw, and LegalBert-pt; while two were fine-tuned using Brazilian legislative data: FT BERTimbau and FT LegalBert-pt. BERTimbau was also selected, as it was trained for Brazilian Portuguese, in addition to three multilingual models: LaBSE, Paraphrase Multilingual MPNet, and Paraphrase Multilingual MiniLM. All 10 models can be found in the HuggingFace[5] platform and can be used with the SBERT architecture.

BERTimbau: BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020) is a version of BERT trained for the Portuguese language. Using brWac (FILHO et al., 2018) — a big and diverse corpus of web pages —, BERT was pre-trained for three NLP tasks: Semantic Textual Similarity, Recognizing Textual Entailment, and NER. Two different-sized models were made available: Base[6] (110M parameters) e Large[7] (330M parameters). The Large version was used in this study.

Legal-BERTimbau: Legal-BERTimbau[8] (MELO; SANTOS; DIAS, 2023) is a fine-tuned version of BERTimbau for the legal Portuguese domain. In order to perform the domain adaptation, pairs of legal sentences from the Supremo Tribunal de Justiça of Portugal were used.

JurisBERT: JurisBERT[9] (VIEGAS; COSTA; ISHII, 2023) is a BERT-based model trained from scratch for the Brazilian judicial domain. First, it was pre-trained using publicly available legal Brazilian documents, such as laws, decrees, and *acórdãos*. Later, pairs of *acórdãos* extracted from the Brazilian Supremo Tribunal Federal and other regional courts were used for fine-tuning.

BERTimbauLaw: Viegas, Costa and Ishii (2023) also made available[10] a fine-tuned version of BERTimbau. For this model, called BERTimbauLaw, they performed fine-tuning in the same way as JurisBERT.

LegalBert-pt: LegalBert-pt (SILVEIRA et al., 2023) was pre-trained using 1.5 million documents from 10 Brazilian legal courts. It was created to deal with NER and classification

[5] https://huggingface.co
[6] https://huggingface.co/neuralmind/bert-base-portuguese-cased
[7] https://huggingface.co/neuralmind/bert-large-portuguese-cased
[8] https://huggingface.co/rufimelo/Legal-BERTimbau-large
[9] https://huggingface.co/alfaneo/jurisbert-base-portuguese-sts
[10] https://huggingface.co/alfaneo/bertimbaulaw-base-portuguese-sts

tasks within the legal domain. Two versions of this model were built: LegalBert-pt SC[11], which was trained from scratch, and LegalBert-pt FP[12], which is an adaptation of BERTimbau. As Silveira et al. (2023) reported better results using LegalBert-pt FP, it was the version chosen for this study.

LaBSE: the Language-agnostic BERT Sentence Encoder (LaBSE)[13] (FENG et al., 2022) supports 109 languages, including Portuguese. It was trained using monolingual data from CommonCrawl and Wikipedia, in addition to bilingual pairs translated from web pages. Although it was primarily evaluated for Bitext Retrieval, it is also used for STS.

Paraphrase Multilingual MPNet: the Paraphrase Multilingual MPNet[14] model was built through the MPNet (SONG et al., 2020) Knowledge Distillation process (REIMERS; GUREVYCH, 2020), using the XLM-RoBERTa (CONNEAU et al., 2020) — which was pre-trained for 100 different languages — as the student model. Thus, this model is capable of generating embeddings to be used for a variety of NLP tasks, such as clusterization and semantic search, as well as for IR.

Paraphrase Multilingual MiniLM: also originated from the Knowledge Distillation process, the Paraphrase Multilingual MiniLM[15] model was built through training a multilingual version of MiniLM (WANG et al., 2021) — which is based on XLM-RoBERTa —, while using the monolingual version of MiniLM (WANG et al., 2020) as the teacher model. It can also be used for different NLP tasks and for IR.

FT BERTimbau: in the study of Santos et al. (2024), the authors made available a version of BERTimbau fine-tuned with Brazilian legislative data[16]. The fine-tuning was performed using pairs of related legislative proposals, from a tree of proposals containing the relationship between them. In this work, this model is referenced to as “FT BERTimbau”.

FT LegalBert-pt: Santos et al. (2024) also made available[17] a version of LegalBert-pt fine-tuned with the same technique as the FT BERTimbau model. In this work, this model is referenced to as “FT LegalBert-pt”.

[11] https://huggingface.co/raquelsilveira/legalbertpt_sc
[12] https://huggingface.co/raquelsilveira/legalbertpt_fp
[13] https://huggingface.co/sentence-transformers/LaBSE
[14] https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2
[15] https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2
[16] https://huggingface.co/josedossantos/bertimbau-tuned
[17] https://huggingface.co/josedossantos/legalbertpt-tuned

6.2.2 Similar queries search

For the second step of Ulysses-RFSQ — the search for the set of similar past queries —, 12 strategies were compared. In order to answer **RQ4**, the same 10 BERT-based models summarized in Table 9 were evaluated to compute the similarity between the queries, in addition to the cosine measure with and without preprossessing. This evaluation aimed to compare the use of contextual semantic embeddings to compute the similarity with the use of just the presence and absence of terms.

The preprocessing techniques used to compute the cosine similairty were the same used to preprocess the documents and queries for the IR process: those presented by Souza et al. (2021b).

6.2.3 Parameters assessment

As explained in Chapter 4, the use of Ulysses-RFSQ depends on two main parameters: cut and δ. The cut parameter defines the selection of the set of similar queries based on their similarity — either using contextual embeddings or just the cosine similarity measure —, thus this parameter was assessed considering the values from 0.1 to 0.9 for both corpora. This assessment helps to answer **RQ3** and to compare the selection of a greater number of past queries with the selection of a smaller number of more similar ones.

On the other hand, for the δ parameter, which helps to control the importance of the RF information for the re-ranking step, the values of 0.1, 0.5, 1.0, and 2.0 were assessed. The preliminary study presented in (VITÓRIO et al., 2022) lacks of these assessments. In it, the cut parameter was set as 0.3 and the δ parameter was not present in the λ's formula.

Finally, the $rel(d, q_j)$ value for the *somewhat relevant* documents present in Ulysses-RFCorpus was set to 0.5 for the Ulysses-RFSQ versions that use the different levels of relevance information (Ulysses-RFSQ-DRL and Ulysses-RFSQ-ALL). For these versions, a value of 1 was used for the *very relevant* documents and -1 for the *irrelevant* ones.

For the parameters assessment experiments, we opted for using a smaller set of queries from each corpora in order to choose the best pair of parameters' values for each scenario. For the Preliminary Search corpus, we used 20% of it, i.e., 386 randomly chosen queries. This choice was inspired by evaluation techniques that split the dataset and uses 20% of the data for validation. Meanwhile, as Ulysses-RFCorpus is much smaller, we used 50% of this corpus, i.e., 323 queries. We noticed that a smaller set would not comprise a sufficient number of similar queries to the Ulysses-RFSQ re-ranking step be effective.

6.2.4 Evaluation

The IR process using either BM25 of SBERT embeddings is deterministic, i.e., the same list of documents will always be retrieved for a specific query. The same goes for Ulysses-RFSQ if the database of past queries did not have changed. In this sense, to perform the experiments using the datasets described in Chapter 5 and Section 6.1.2, a leave-one-out strategy was applied: for each evaluated query, all the other queries were used as the past queries database (Figure 2).

This strategy is better than the one used in the preliminary evaluation (VITÓRIO et al., 2022), which was performed using the 10-fold cross-validation technique. Using the leave-one-out strategy, a real-world utilization of Ulysses-RFSQ can be simulated, in which each query is processed at a time and all the past ones may be used to compute the λ's value.

As the metrics to evaluate the performance of the IR process, Mean Average Precision (MAP), Mean R-Precision (MRP), Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (nDCG) were used. In this study, both *very relevant* and *somewhat relevant* documents from Ulysses-RFCorpus were understood as *relevant* documents to compute the metrics, except for nDCG — for which they were understood separately. As aforementioned, the retrieval of 12 documents for Ulysses-RFCorpus and 20 documents for the Preliminary Search corpus were considered for the evaluation.

6.2.4.1 Mean Average Precision (MAP)

Combining Recall and Precision, Average Precision (AP) (ZHANG; ZHANG, 2009) is a suitable metric to evaluate IR systems, as it computes the mean of the precision scores after each relevant document is retrieved:

$$AP = \frac{\sum_r Precision@r}{\#(\text{relevant documents})}, \tag{6.1}$$

in which r is the rank of each relevant document.

As AP is calculated for each query, the average AP considering all queries in a dataset is called Mean Average Precision (MAP) (BEITZEL; JENSEN; FRIEDER, 2009b):

$$MAP = \frac{\sum_1^n AP_n}{n}, \tag{6.2}$$

in which n is the number of queries.

6.2.4.2 Mean R-Precision (MRP)

R-Precision computes Precision considering the quantity of documents that were judged as *relevant* (R) for a query (BEITZEL; JENSEN; FRIEDER, 2009a). Equations 6.3 and 6.4 present the formulas for R-Precision and MRP, which is the arithmetic mean of R-Precision for a set of n queries.

$$R - Precision = \frac{Precision@R}{R}, \tag{6.3}$$

$$MRP = \frac{\sum_1^n RP_n}{n}. \tag{6.4}$$

6.2.4.3 Mean Reciprocal Rank (MRR)

The Reciprocal Rank (RR) metric computes the reciprocal of the rank at which the first relevant document was retrieved (CRASWELL, 2009). In other words, it computes how soon the first relevant document appeared in the retrieved list.

As this metric only looks at the first relevant document, it is often used in systems in which the user needs only one document. However, it can also be used to measure if the IR algorithm retrieves the relevant documents in low ranks.

Equation 6.5 computes RR for a query, while Equation 6.6 computes the average RR for a set of n queries, called Mean Reciprocal Rank (MRR):

$$RR = \frac{1}{\text{rank of the first relevant document}}, \tag{6.5}$$

$$MRR = \frac{\sum_1^n RR_n}{n}. \tag{6.6}$$

6.2.4.4 *Normalized Discounted Cumulative Gain (nDCG)*

As the feedback information given in Ulysses-RFCorpus uses non-binary notions of relevance — i.e., the documents were judged as either *irrelevant*, *very relevant*, or *somewhat relevant* —, the Normalized Discounted Cumulative Gain (nDCG) metric could be applied for this corpus.

nDCG is based on the assumption that highly relevant documents are more valuable for the user than marginally relevant documents. Thus, since all documents are not equally relevant, the most relevant ones should be identified and ranked first for presentation to the user (JÄRVELIN; KEKÄLÄINEN, 2009). In the case of Ulysses-RFCorpus, for instance, documents judged as *very relevant* should be retrieved in smaller ranks than the *somewhat relevant* documents.

This metric is computed based on the Discounted Cumulative Gain (DCG) and the Ideal Discounted Cumulative Gain (IDCG) for a query:

$$nDCG = \frac{DCG}{IDCG}, \tag{6.7}$$

in which DCG can be computed by Equation 6.8 and IDCG is the best possible rank for the query, i.e., the maximum DCG value that can be obtained if the results were ideally ranked.

$$DCG = \sum_{1}^{n} \frac{2^{rel_i} - 1}{log_2(i+1)}, \tag{6.8}$$

in which n is the number of retrieved documents, i is the position of the document in the retrieved documents list, and rel_i is the degree of relevance of the document in position i.

6.2.4.5 *Statistical significance evaluation*

Finally, in order to evaluate the method with statistical significance, the Student's t-test (STUDENT, 1908) was applied for the comparison between Ulysses-RFSQ and the baselines. As pointed out by Urbano, Lima and Hanjalic (2019), Student's is the most robust significance test for IR. In their work, the authors used the MAP measure to perform the evaluation, thus we also utilized this metric.

In addition, we also used the Nemenyi post-hoc test (NEMENYI, 1963) to compare the methods used to search for the past similar queries.

7 RESULTS

In this chapter, we present and discuss the results obtained from the experiments. The goal is to answer the Research Questions introduced in Chapter 1, thus the results are organized by RQ. To increase readability, Appendix A summarizes every different configuration and how they are referenced to throughout this study.

7.1 PARAMETERS ASSESSMENT (RQ3)

First, the parameters cut and δ had to be assessed in order to choose the best values for the experiments. In addition, we aimed to answer **RQ3) What is the trade-off between the use of the RF information from a greater number of past queries and the use of this information from a smaller set of highly similar ones?**

The Ulysses-RFSQ-OR version was used to execute the experiments, while the MAP metric was chosen to perform this assessment and to select the best pair of values for each scenario. Appendices B and C summarize the selected values for each configuration for Ulysses-RFCorpus and the Preliminary Search corpus, respectively.

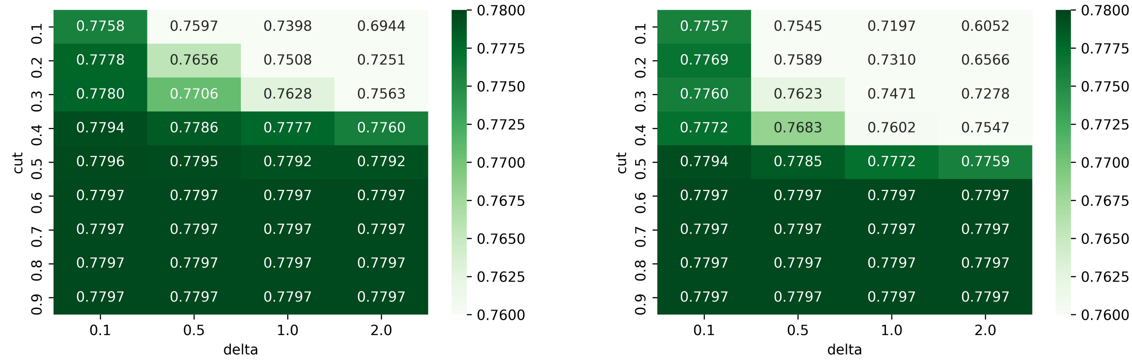
7.1.1 Ulysses-RFCorpus

In this section, the parameters assessment is presented and discussed for Ulysses-RFCorpus, first for the BM25 approach used to retrieve the documents and later for the SBERT approach.

7.1.1.1 BM25 approach

Figures 6 and 7 present the MAP result achieved by each combination of parameters's values with the BM25L algorithm and using only the cosine similarity to select the similar queries. Analyzing Figure 6, which brings the heatmaps for BM25L_PRE_PRE and BM25L_PRE_NP, we can see that, for values of cut greater than 0.5, there were an insufficient number of similar queries to impact the MAP results. Therefore, the best pair of values for cut and δ was $\{cut = 0.5, \delta = 0.1\}$ for both scenarios — using the cosine similarity with and without preprocessing —, although the use of Ulysses-RFSQ-OR has worsened the results for this algorithm.

Figure 6 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_PRE (a) and BM25L_PRE_NP (b) using the cosine similarity to select the similar queries from Ulysses-RFCorpus.

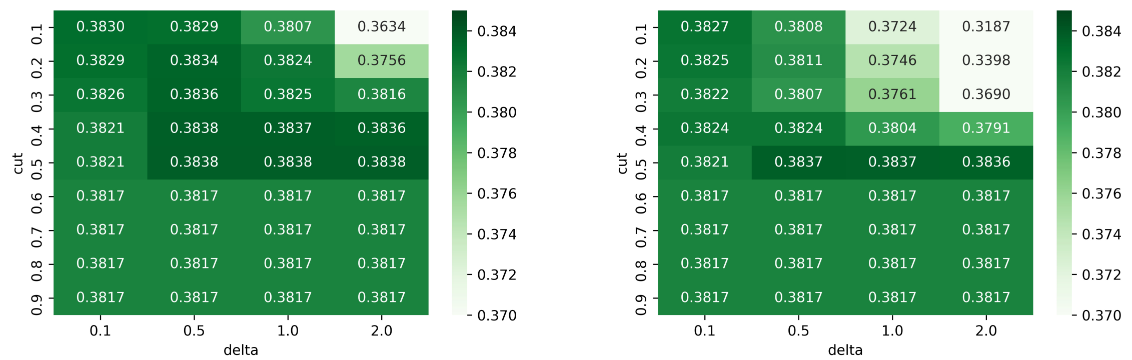


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 7 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_NP_PRE (a) and BM25L_NP_NP (b) using the cosine similarity to select the similar queries from Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

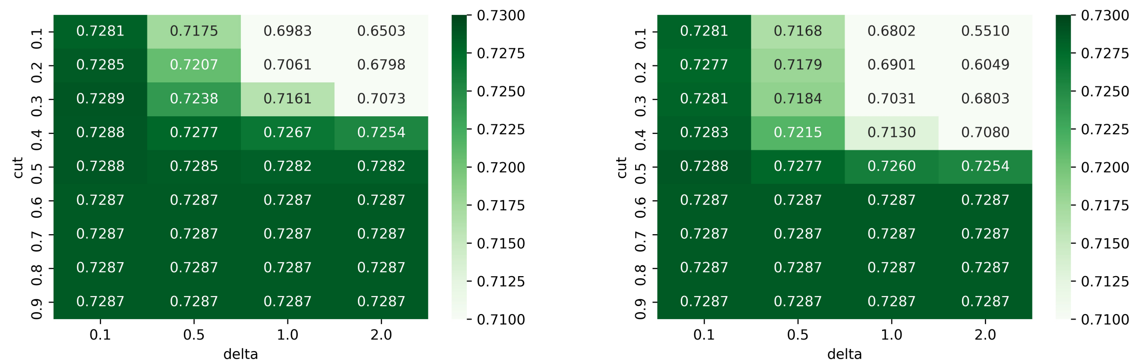
The heatmaps for BM25L_NP_PRE and BM25L_NP_NP (Figure 7) confirm this finding, as the best performances were achieved using $cut = 0.5$, while greater values of cut did not change the MAP result of the base IR algorithm. Therefore, for both these scenarios, the selected pair of values was $\{cut = 0.5,\ \delta = 0.5\}$. When there was a tie between the results achieved by different pairs of parameters' values, we opted to select the greater value of cut and the smallest value of δ.

The fact that the BM25L version without preprocessing did not reach the same performance as the preprocessed version could explain the use of a greater value of δ (0.5) for this case.

In scenarios for which the base IR algorithm doesn't achieve state-of-the-art results, there is more room for improvement, thus giving a higher importance to the RF information can lead to a greater impact on the results.

The same findings could be observed for both versions of Okapi BM25 using the cosine similarity to search for the similar past queries (Figures 8 and 9). The only difference was observed for OkapiBM25_PRE_PRE. As can be seen in Figure 8a, the best performance for this algorithm was achieved by the pair $\{cut = 0.3,\ \delta = 0.1\}$. For the other configurations that used Okapi BM25, the best results were achieved by the same combination of parameters' values selected for their BM25L counterparts: $\{cut = 0.5,\ \delta = 0.1\}$ for OkapiBM25_PRE_NP and $\{cut = 0.5,\ \delta = 0.5\}$ for both OkapiBM25_NP_PRE and OkapiBM25_NP_NP.

Figure 8 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_PRE (a) and OkapiBM25_PRE_NP (b) using the cosine similarity to select the similar queries from Ulysses-RFCorpus.



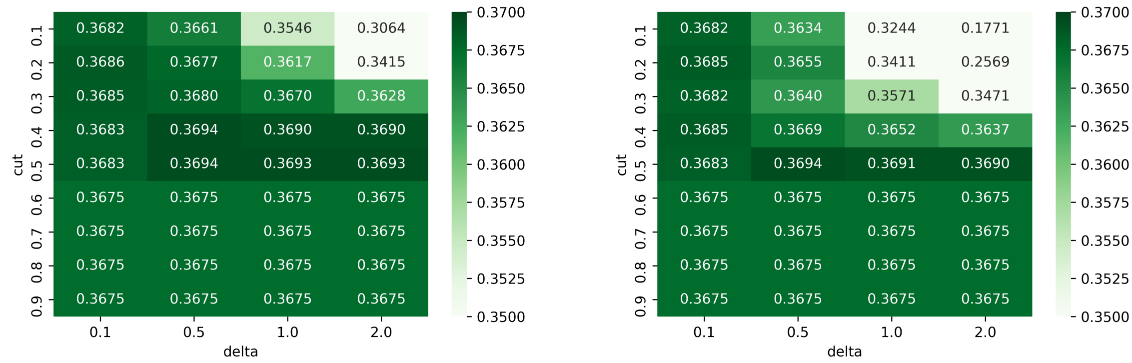
(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Appendices D and E present the heatmaps for the parameters evaluation using the BERT-based models to select the similar queries with, respectively, BM25L and Okapi BM25 as the base IR algorithm. As can be seen in the results for both scenarios, the use of $\delta = 0.1$ and a high value of cut (ranging from 0.7 to 0.9, depending on which of the LMs was used) achieved the best performance for most of the experimental configurations. This corroborates the finding using only the cosine similarity: that the good performance of the BM25 variants for Ulysses-RFCorpus, mainly when the documents were preprocessed, implies the use of a smaller number of similar queries. A larger set may deteriorate the results, while using a small set of highly similar queries can have a positive impact on them.

Figure 9 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_NP_PRE (a) and OkapiBM25_NP_NP (b) using the cosine similarity to select the similar queries from Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Nevertheless, some SBERT models that achieved the best results for the versions of BM25 algorithms without preprocessing — which did not reach the same performance as the preprocessed versions — used lower values of cut, such as Legal-BERTimbau and LaBSE, or greater values of δ, such as Multilingual MPNet and Multilingual MiniLM. This might imply that the use of contextual embeddings can select a larger set of similar queries and give more valuable RF information — which can be used with greater values of δ.

7.1.1.2 SBERT approach

For the use of the SBERT architecture as the base IR algorithm, the heatmaps obtained from the parameters evaluation using only cosine to compute the similarity between the queries can be found in Appendix F. The results achieved with the BERT-based models are different than those using the BM25 variants with respect to which combinations of cut and δ's values reached the best performances. They showed that lower values of cut with greater values of δ are preferable for this approach. This may be explained by the fact that the SBERT models reached a very poor performance when compared to the BM25 variants, thus the use of a larger set of similar queries, while giving a higher importance to the RF information, can have a greater positive impact on the IR process while using the cosine similarity to select the similar queries.

Evaluating the BERT-based models being used to search for the similar queries, though, each SBERT model used as the base IR algorithm must be analyzed depending on its performance. Appendix G brings the heatmaps for all combinations of BERT-based models: used as the IR algorithm and used for the similar queries search. The results show that, for the models that achieved the worst MAP results — BERTimbau and LegalBert-pt —, the best value of δ ranged from 0.1 to 2.0, depending on which model was used for the queries selection step. Meanwhile, the selected δ's value for the best LMs used as the IR algorithm — LegalBERTimbau, BERTimbauLaw, Multilingual MPNet, Multilingual MiniLM, FTBERTimbau, and FTLegalBert-pt — was 0.1 for the vast majority of configurations.

This confirms the findings from the BM25 approach: that, for IR algorithms with poor performance, a greater importance can be given for the RF information to largely improve the results. On the other hand, for algorithms that already achieve substantial results, Ulysses-RFSQ uses the RF information from past queries to improve the performance in a modest way.

As for the cut's value, the results showed that its choice depends on the strategy used to select the queries, at least for Ulysses-RFCorpus. Strategies based solely on the cosine similarity require small values of cut in order to use a larger amount of queries. Meanwhile, strategies based on contextual embeddings can find a sufficient number of queries using greater values for the cut-off threshold. For instance, the use of Ulysses-RFSQ-OR with BERTimbau as the queries selector required $cut = 0.9$ for every experimental configuration in order to have a positive impact on the results.

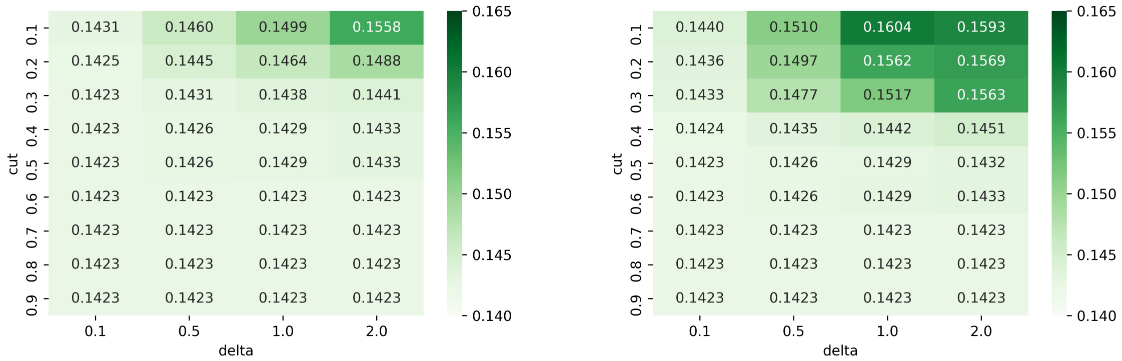
7.1.2 Preliminary Search corpus

In this section, we present and discuss the parameters assessment for the Preliminary Search corpus, first using the BM25 approach to perform the documents retrieval and later using the SBERT approach.

7.1.2.1 BM25 approach

The MAP results of the parameters evaluation for the versions of BM25L with and without preprocessing and using only the cosine similarity to retrieve the similar queries can be found in Figures 10 and 11, respectively. For this dataset, the results show a preference to use larger sets of past similar queries and to give a greater importance to the RF information. As can be seen in the heatmaps, the best results for this scenario were achieved with the smallest values of cut (0.1 or 0.2) and the greatest values of δ (1.0 or 2.0).

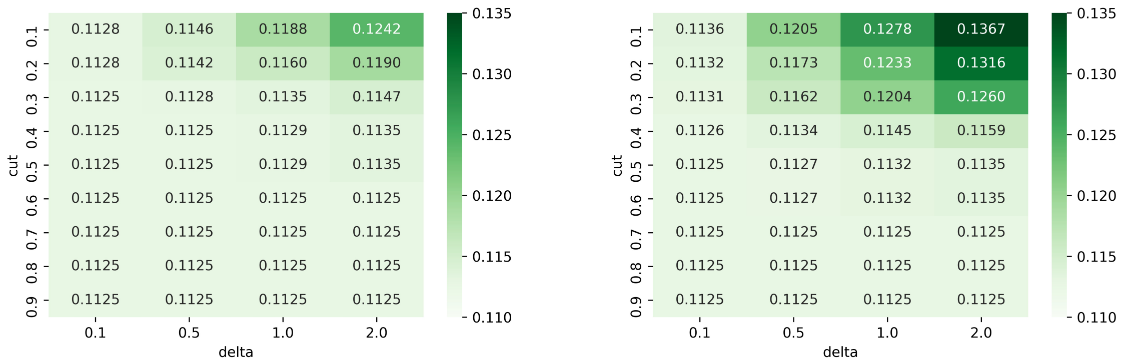
Figure 10 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_PRE (a) and BM25L_PRE_NP (b) using the cosine similarity to select the similar queries from the Preliminary Search corpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 11 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_NP_PRE (a) and BM25L_NP_NP (b) using the cosine similarity to select the similar queries from the Preliminary Search corpus.



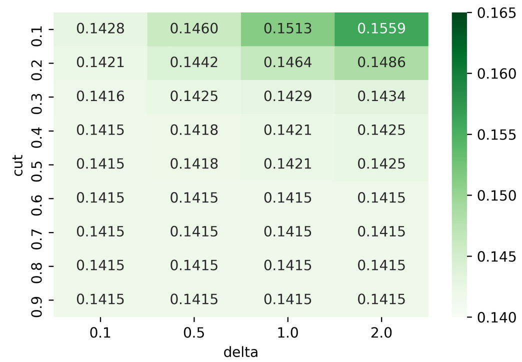
(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

The same could be observed for the two versions of Okapi BM25 (Figures 12 and 13), for which the highest value of cut that achieved a best result was 0.3, for OkapiBM25_NP_NP. This may be explained by the fact that the evaluated configurations did not achieve, for the Preliminary Search corpus, results as great as for Ulysses-RFCorpus, thus the use of the RF information from a large number of past queries could be more useful for this scenario.

It can also be noticed that the number of selected past similar queries was insufficient to have any impact on the MAP results using the cosine similarity with values of the cut-off threshold greater than 0.5 — the same occurred for Ulysses-RFCorpus. In addition, with cut's values greater than 0.3, a substantial impact only could be seen when using $\delta = 2.0$. Nevertheless, it is worth remembering that, for the parameters evaluation, only 20% of the Preliminary Search dataset was used.

Figure 12 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_PRE (a) and OkapiBM25_PRE_NP (b) using the cosine similarity to select the similar queries from the Preliminary Search corpus.

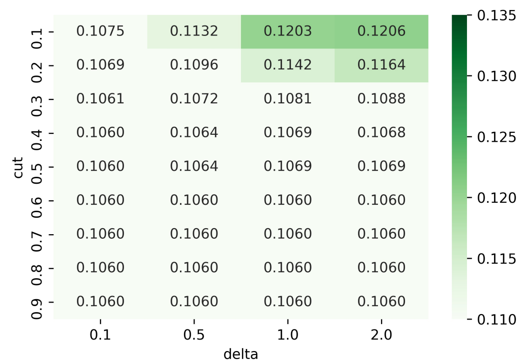


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 13 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_NP_PRE (a) and OkapiBM25_NP_NP (b) using the cosine similarity to select the similar queries from the Preliminary Search corpus.



(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Analyzing the results of the use of SBERT models to select the past similar queries (Appendices H and I), we can conclude that a larger number of queries is selected with greater values of the cut-off threshold when the similarity between the queries is computed from contextual embeddings. They show that, when comparing the similarity between the embeddings, the best results are usually obtained with values of cut from 0.7 to 0.9, differing from the results obtained when the comparison was made using only the cosine similarity.

As for the best values of δ for this corpus, a pattern could not be observed from the experiments. The results showed that the choice for this parameter's value may depend on both the base IR algorithm and the BERT-based model chosen to retrieve the similar queries, i.e., it may vary for each individual scenario.

7.1.2.2 SBERT approach

The findings for using the SBERT approach as the base IR algorithms are similar to the ones from the use of the BM25 approach. As can be seen in Appendices J and K — which bring the heatmaps for the use of BERT-based models to retrieve the documents with the use of, respectively, cosine similarity and LMs to select the similar queries —, the use of cosine required lower values of the cut-off threshold to have a positive impact on the results. On the other hand, the use of contextual embeddings achieved better results with greater cut's values.

Meanwhile, the best values of δ could range from 0.1 to 2.0 depending on the experimental configuration, thus it doesn't seem to follow a pattern for this corpus. The only exception is for the fine-tuned BERT-based models — FT BERTimbau and FT LegalBert-pt —, for which the best results were achieved using $\delta = 0.1$ for almost every configuration. This may be explained by the fact that these fine-tuned models reached better performances when compared to the other BERT-based models.

7.1.3 Discussion

In this sense, based on what was observed for both the BM25 and SBERT approaches, we can conclude that the assessment of the parameters cut and δ should be performed for each individual scenario, in order to select the best pair of values for each case.

Some findings, though, are useful to consider when performing this selection, such as the fact that, for every evaluated scenario — either using BM25 or BERT-based models and with both corpora —, the strategy chosen to retrieve the similar queries set impacted the cut's value choice. When using just the cosine measure to compute the similarity between the queries, it was necessary to set the cut parameter with a small value to retrieve a number of queries that was sufficient to impact the results.

This can be explained by Tables 10 and 11, which present the percentage of queries that are considered similar to any other by each algorithm according to each value of cut for Ulysses-RFCorpus and the Preliminary Search corpus, respectively. In other words, they show the percentage of queries that have at least one other similar query, thus being affected by Ulysses-RFSQ. For values of the cut-off threshold from 0.6 and greater, no query from Ulysses-RFCorpus and less than 2% of the queries from the Preliminary Search corpus have another query considered similar to them while using only the cosine similarity measure.

Table 10 – Percentage of queries from Ulysses-RFCorpus used for the parameters assessment that are similar to any other, by algorithm used to select the similar queries and by level of similarity. The colors represent the cut's value usage by the best configurations.

Algorithm	$\geq$ **0.1**	$\geq$ **0.2**	$\geq$ **0.3**	$\geq$ **0.4**	$\geq$ **0.5**	$\geq$ **0.6**	$\geq$ **0.7**	$\geq$ **0.8**	$\geq$ **0.9**
Cosine (preprocessed)	71.8%	20.7%	5.3%	1.2%	0.6%	0%	0%	0%	0%
Cosine	99.7%	65.0%	14.6%	5.3%	1.2%	0%	0%	0%	0%
BERTimbau	100%	100%	100%	100%	100%	100%	100%	96.9%	45.2%
Legal-BERTimbau	100%	100%	100%	96.0%	67.2%	25.1%	5.3%	0%	0%
JurisBERT	100%	100%	100%	100%	100%	99.4%	86.7%	11.8%	0%
BERTimbauLaw	100%	100%	100%	100%	100%	96.9%	76.5%	21.1%	0%
LegalBert-pt	100%	100%	100%	100%	100%	99.4%	92.0%	60.4%	0%
LaBSE	100%	100%	100%	99.7%	94.7%	65.6%	14.2%	1.2%	0%
Multilingual MPNet	100%	100%	100%	99.1%	97.2%	74.0%	30.0%	4.6%	0%
Multilingual MiniLM	100%	100%	99.4%	98.5%	90.4%	67.5%	22.6%	4.3%	0%
FT BERTimbau	100%	100%	100%	100%	100%	100%	96.6%	79.3%	19.2%
FT LegalBert-pt	100%	100%	100%	100%	100%	100%	96.6%	50.2%	5.0%

Source: Created by the author (2025)

Table 11 – Percentage of queries from the Preliminary Search corpus used for the parameters assessment that are similar to any other, by algorithm used to select the similar queries and by level of similarity. The colors represent the cut's value usage by the best configurations.

Algorithm	$\geq$ **0.1**	$\geq$ **0.2**	$\geq$ **0.3**	$\geq$ **0.4**	$\geq$ **0.5**	$\geq$ **0.6**	$\geq$ **0.7**	$\geq$ **0.8**	$\geq$ **0.9**
Cosine (preprocessed)	52.3%	17.4%	5.7%	2.1%	1.6%	0.5%	0.5%	0%	0%
Cosine	99.7%	64.2%	22.8%	6.2%	2.6%	1.6%	0.5%	0.5%	0%
BERTimbau	100%	100%	100%	100%	100%	100%	100%	100%	90.7%
Legal-BERTimbau	100%	100%	100%	99.7%	87.3%	42.0%	11.1%	3.6%	0%
JurisBERT	100%	100%	100%	100%	100%	100%	88.3%	16.3%	0.5%
BERTimbauLaw	100%	100%	100%	100%	100%	99.5%	87.3%	37.6%	1.0%
LegalBert-pt	100%	100%	100%	100%	100%	100%	100%	94.8%	33.4%
LaBSE	100%	100%	100%	100%	96.1%	76.7%	21.2%	3.4%	0%
Multilingual MPNet	100%	100%	100%	100%	99.7%	95.9%	67.4%	24.6%	2.1%
Multilingual MiniLM	100%	100%	100%	100%	99.2%	92.0%	62.2%	17.4%	0.5%
FT BERTimbau	100%	100%	100%	100%	100%	100%	99.5%	86.3%	40.9%
FT LegalBert-pt	100%	100%	100%	100%	100%	100%	99.2%	83.2%	18.9%

Source: Created by the author (2025)

On the other hand, the use of contextual embeddings made almost every query to be affect by Ulysses-RFSQ while using values of cut up to 0.4. For instance, the BERTimbau model generated highly similar embeddings for the queries. This algorithm assumed that almost every query had at least one similar other while considering a similarity lower than 0.9 — for the Preliminary Search corpus, all of the queries had a similar one with this level of similarity. Meanwhile, Legal-BERTimbau was the most selective model for considering a query similar to some other. From the Preliminary Search corpus, only 42% of the queries were considered similar to any other query by this algorithm with a similarity greater than 0.6, while only 25% of the queries from Ulysses-RFCorpus had another similar query while using $cut = 0.6$.

Tables 10 and 11 also inform the usage of each value of cut by the best configurations for each algorithm: darker shades of red indicate that this value of the cut-off threshold was used more times to achieve the best results. For Ulysses-RFCorpus, the best performance tended to be achieved by updating the ranking for the least amount of queries that could impact the results. As this amount depends on the algorithm used, for this scenario it varied between 4% and 60%. It is worth pointing out, though, that the use of LegalBert-pt, which achieved most best results re-ranking the retrieved documents for 60.4% of the queries, also deteriorated the base IR algorithm's performance in some cases. This shows that the use of a smaller set of similar queries would be more suitable for this specific scenario.

Nevertheless, the same findings could not be observed for the Preliminary Search corpus. For this dataset, larger amounts of queries were often used to reach the best performance for each experimental configuration. In some cases, the best results were achieved while updating the ranking for 100% of the queries, which did not occur for Ulysses-RFCorpus. Some SBERT models — such as LaBSE, Multilingual MPNet, and Multilingual MiniLM — also performed better in cases in which they have selected a larger set of queries than using a smaller though sufficient number. This may be explained by the fact that the Preliminary Search corpus is a more difficult one, for which the evaluated IR algorithms did not achieved the same good performance as for Ulysses-RFCorpus, thus there are more queries that could benefit from the Ulysses-RFSQ re-ranking.

Therefore, answering **RQ3**, we can point out that the trade-off between the use of a greater or a smaller number of past similar queries within the Ulysses-RFSQ process depends on the scenario in which it is applied. Using this method in a dataset for which the IR algorithms already achieve great results, it is better to use a small set of highly similar queries. For scenarios for which there is more room for improvement, however, the documents ranking can be updated for a larger amount of queries. This choice also depends on the algorithm used to retrieve the similar queries, as some BERT-based models generated embeddings for this task better than others.

7.2 ULYSSES-RFSQ AND BASELINES COMPARISON (RQ2)

In order to evaluate the performance of the proposed method, we compared the results achieved by Ulysses-RFSQ with those from the base IR algorithms without the use of the past RF information. In this section, we assessed the Ulysses-RFSQ-OR version, which uses only the *relevant* information, aiming to answer **RQ2) Can a method that utilizes the RF information from similar past queries to re-rank the retrieved documents improve the IR results within the Brazilian legislative domain?**

Appendix L presents the results for every configuration and for both Ulysses-RFCorpus and the Preliminary Search corpus. In it, the green color indicates a better result than the baseline's, while the red color indicates the cases in which the Ulysses-RFSQ-OR configuration performed worse than the base IR algorithm.

7.2.1 Ulysses-RFCorpus

The experimental results showed that the use of Ulysses-RFSQ-OR improved the performance of the IR process in most scenarios. For each algorithm applied to retrieve the documents, the best results were achieved by a configuration that used the past RF information, overcoming the baseline.

However, we could notice that, for specific algorithms such as the BM25 variants, a considerable number of configurations deteriorated the baselines' results for Ulysses-RFCorpus. This can be explained by the great performance achieved in this corpus by the BM25 algorithms. It is worth to remember that the BM25L_PRE version is the state-of-the-art for this dataset (VITÓRIO et al., 2025a; VITÓRIO et al., 2025b). Nonetheless, the largest portion of the different configurations that used Ulysses-RFSQ-OR still were capable of improving the results, mainly for the MRP and MRR metrics. Five of them also performed equal to or better than the BM25L_PRE baseline for all of the metrics, including MAP.

On the other hand, when using SBERT models to perform the documents retrieval, almost every Ulysses-RFSQ-OR configuration could improve the retrieval results. As the SBERT approach did not achieve the same performance as the use of BM25, there was more room for improvement, in which the use of RF information from similar past queries could be more beneficial. The only exception was observed using FT BERTimbau as the IR algorithm — the best BERT-based model to retrieve documents for this legislative scenario (VITÓRIO et al., 2025b) —, for which seven configurations improved the results, while five did not. For all of the other algorithms, the vast majority of configurations performed better than the baseline for most of the evaluated performance metrics. Moreover, the improvements observed for this scenario were also greater than the ones for the use of the BM25 approach, ranging from 0.0122 to 0.0391 for the MAP metric, while the improvements achieved for the use of the BM25 variants ranged from 0.0015 to 0.0055.

We could also realize that, for this corpus, the most difficult metric to improve was MRR. Some configurations from different scenarios deteriorated the MRR results while improving the other three metrics. This may indicate that Ulysses-RFSQ-OR performs the re-ranking in a way that more relevant documents are selected and presented to the user, but also resulting in some irrelevant ones being placed in the first positions of the retrieved list.

In order to evaluate the observed improvements in a statistically significant way, we compared the AP results of Ulysses-RFSQ-OR with the baselines' using the Student's t-test (STUDENT, 1908) with a confidence level of 95%. Tables 12 and 13 present the p-value obtained for each comparison between a Ulysses-RFSQ-OR configuration and the IR algorithm used as baseline. The underlined p-values indicate the cases in which there was a statistically significant difference between the performances.

The Student's t-test showed that the improvements achieved by Ulysses-RFSQ-OR on the preprocessed versions of BM25L and Okapi BM25 were not statistically significant. This was expected, as these versions of the BM25 variants already reached a great performance for Ulysses-RFCorpus, thus the improvements observed for these algorithms were only marginal. Moreover, the use of LegalBert-pt as the algorithm to search for the similar queries harmed the BM25_PRE's performance in a statistically significant way.

Nevertheless, some of the configurations that improved the performance of the BM25 algorithms without preprocessing proved to be statistically better than the baselines. Eight different configurations could improve the BM25L_NP's results with statistical significance, while three did the same for the Okapi BM25_NP's. As these versions of BM25 did not achieve the same performance as the preprocessed versions, Ulysses-RFSQ-OR could have a greater and more significant impact on their MAP results.

Table 12 – P-values obtained with the Student's t-test for the comparison between the Ulysses-RFSQ-OR configurations and the baselines for Ulysses-RFCorpus. The columns indicate the algorithms used to search for the similar queries. The green color represents the configurations for which Ulysses-RFSQ-OR improved the MAP results, while the red color represents those which harmed the base IR algorithm performance. The p-values that show statistical significant differences between the results are underlined.

IR algorithm	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw
BM25L_PRE	0.072	0.923	0.060	0.634	0.802	0.083
BM25L_NP	<u>0.038</u>	0.274	<u>0.008</u>	0.929	<u>0.001</u>	<u>0.026</u>
Okapi BM25_PRE	0.109	0.127	0.952	0.488	0.243	0.087
Okapi BM25_NP	0.860	0.121	0.730	<u>0.006</u>	<u>0.016</u>	0.459
BERTimbau	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.001</u>	<u>0.000</u>
Legal-BERTimbau	<u>0.000</u>	<u>0.000</u>	<u>0.001</u>	0.276	0.082	<u>0.001</u>
JurisBERT	<u>0.003</u>	<u>0.003</u>	<u>0.005</u>	<u>0.000</u>	0.702	0.076
BERTimbauLaw	<u>0.000</u>	<u>0.000</u>	0.060	<u>0.000</u>	<u>0.048</u>	<u>0.035</u>
LegalBert-pt	<u>0.000</u>	<u>0.000</u>	0.184	<u>0.000</u>	0.069	<u>0.015</u>
LaBSE	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.031</u>	<u>0.000</u>	<u>0.000</u>
Multilingual MPNet	<u>0.000</u>	<u>0.017</u>	0.405	<u>0.000</u>	<u>0.044</u>	<u>0.000</u>
Multilingual MiniLM	0.193	<u>0.006</u>	<u>0.011</u>	0.303	<u>0.004</u>	<u>0.003</u>
FT BERTimbau	<u>0.000</u>	<u>0.000</u>	0.122	0.898	0.304	0.936
FT LegalBert-pt	<u>0.003</u>	0.168	0.078	<u>0.000</u>	0.129	<u>0.018</u>

Source: Created by the author (2025)

Table 13 – P-values obtained with the Student's t-test for the comparison between the Ulysses-RFSQ-OR configurations and the baselines for Ulysses-RFCorpus. The columns indicate the algorithms used to search for the similar queries. The green color represents the configurations for which Ulysses-RFSQ-OR improved the MAP results, while the red color represents those which harmed the base IR algorithm performance. The p-values that show statistical significant differences between the results are underlined.

IR algorithm	LegalBert-pt	LaBSE	MPNet	MiniLM	FT BERTimbau	FT LegalBert-pt
BM25L_PRE	<u>0.016</u>	0.939	0.323	0.952	0.071	0.791
BM25L_NP	<u>0.025</u>	<u>0.002</u>	0.059	0.259	<u>0.001</u>	<u>0.002</u>
Okapi BM25_PRE	0.392	0.254	0.068	0.211	0.728	0.529
Okapi BM25_NP	0.897	0.231	0.988	<u>0.014</u>	0.682	0.352
BERTimbau	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>
Legal-BERTimbau	0.528	<u>0.069</u>	<u>0.003</u>	0.163	<u>0.043</u>	<u>0.000</u>
JurisBERT	0.282	0.170	0.091	0.262	<u>0.016</u>	<u>0.008</u>
BERTimbauLaw	0.101	<u>0.001</u>	<u>0.002</u>	<u>0.002</u>	0.062	0.098
LegalBert-pt	0.832	0.173	<u>0.001</u>	<u>0.000</u>	<u>0.002</u>	<u>0.005</u>
LaBSE	<u>0.005</u>	<u>0.002</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>
Multilingual MPNet	0.533	<u>0.031</u>	<u>0.025</u>	<u>0.005</u>	0.088	0.164
Multilingual MiniLM	0.289	0.743	<u>0.001</u>	0.889	<u>0.003</u>	<u>0.007</u>
FT BERTimbau	0.245	0.117	<u>0.018</u>	0.215	<u>0.034</u>	0.218
FT LegalBert-pt	0.111	<u>0.000</u>	0.980	0.393	0.077	<u>0.008</u>

Source: Created by the author (2025)

Analyzing the results for the use of SBERT models as the IR algorithm, we could notice an even better performance of Ulysses-RFSQ-OR. The p-values indicate that most of the configurations achieved statistically better results than the baselines that used LMs to retrieve the documents. In addition, for the use of BERTimbau and LaBSE, all of the configurations improved the MAP results in a statistical significant way. This can be explained by the aforementioned poor performance reached by these models while performing IR in this scenario.

7.2.2 Preliminary Search corpus

For the Preliminary Search corpus, the experimental results showed that the use of Ulysse-RFSQ-OR improved the results for all of the different configurations and scenarios: for every base IR algorithm and using any strategy to search for the similar queries. This version of Ulysses-RFSQ achieved improvements on the MAP results ranging from 0.0384 to 0.0474 for the BM25 approach and ranging from 0.0546 to 0.0773 for the SBERT approach, more than doubling the baseline's performance in some cases.

This shows the great impact of the use of RF information from similar past queries on scenarios for which the baseline algorithm is not able to achieve a great performance, such as for the Preliminary Search dataset. The best baseline algorithm for this corpus (OkapiBM25_PRE) achieved a MAP of 0.1338 and the use of Ulysses-RFSQ-OR could improve its result, reaching a MAP of 0.1722.

We could also see that the BM25 algorithms and the best BERT-based model (FT BERTimbau) performed similarly for this dataset, which did not occur for Ulysses-RFCorpus. In addition, the use of Ulysses-RFSQ-OR made FT BERTimbau achieve the best results of all, overcoming the BM25 approach: using the cosine similarity without preprocessing to select the similar queries, it could reach a MAP of 0.1839.

Tables 14 and 15 bring the p-values obtained with the Student's t-test for the comparison between the AP results of the Ulysses-RFSQ-OR configurations and the baselines' in the Preliminary Search corpus. Using a significance level of 95%, they show that the differences in the results were not statistically significant only in two scenarios, both using BM25L_NP to retrieve the documents. Therefore, we can conclude that, for the Preliminary Search corpus, Ulysses-RFSQ-OR could improve the baselines' results with statistical significance for the vast majority of cases.

Table 14 – P-values obtained with the Student's t-test for the comparison between the Ulysses-RFSQ-OR configurations and the baselines for the Preliminary Search corpus. The columns indicate the algorithms used to search for the similar queries. The green color represents the configurations for which Ulysses-RFSQ-OR improved the MAP results. The p-values that show statistical significant differences between the results are underlined.

IR algorithm	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw
BM25L_PRE	0.000	0.000	0.000	0.000	0.000	0.000
BM25L_NP	0.000	0.000	0.000	0.000	0.000	0.000
Okapi BM25_PRE	0.000	0.000	0.000	0.000	0.000	0.000
Okapi BM25_NP	0.000	0.000	0.000	0.000	0.012	0.000
BERTimbau	0.000	0.000	0.000	0.000	0.000	0.000
Legal-BERTimbau	0.000	0.000	0.000	0.000	0.000	0.000
JurisBERT	0.000	0.000	0.000	0.000	0.000	0.000
BERTimbauLaw	0.000	0.000	0.000	0.000	0.000	0.000
LegalBert-pt	0.000	0.000	0.004	0.000	0.000	0.000
LaBSE	0.000	0.000	0.000	0.000	0.000	0.000
Multilingual MPNet	0.000	0.000	0.000	0.000	0.000	0.000
Multilingual MiniLM	0.000	0.000	0.000	0.000	0.000	0.000
FT BERTimbau	0.000	0.000	0.000	0.000	0.000	0.000
FT LegalBert-pt	0.000	0.000	0.000	0.000	0.000	0.000

Source: Created by the author (2025)

Table 15 – P-values obtained with the Student's t-test for the comparison between the Ulysses-RFSQ-OR configurations and the baselines for the Preliminary Search corpus. The columns indicate the algorithms used to search for the similar queries. The green color represents the configurations for which Ulysses-RFSQ-OR improved the MAP results. The p-values that show statistical significant differences between the results are underlined.

IR algorithm	LegalBert-pt	LaBSE	MPNet	MiniLM	FT BERTimbau	FT LegalBert-pt
BM25L_PRE	0.000	0.000	0.000	0.000	0.000	0.000
BM25L_NP	0.089	0.000	0.000	0.089	0.000	0.000
Okapi BM25_PRE	0.000	0.000	0.000	0.000	0.000	0.000
Okapi BM25_NP	0.000	0.000	0.000	0.000	0.000	0.000
BERTimbau	0.000	0.000	0.000	0.000	0.000	0.000
Legal-BERTimbau	0.000	0.000	0.000	0.000	0.000	0.000
JurisBERT	0.000	0.000	0.000	0.000	0.000	0.000
BERTimbauLaw	0.000	0.000	0.000	0.000	0.000	0.000
LegalBert-pt	0.001	0.000	0.000	0.000	0.000	0.000
LaBSE	0.000	0.000	0.000	0.000	0.000	0.000
Multilingual MPNet	0.000	0.000	0.000	0.000	0.000	0.000
Multilingual MiniLM	0.000	0.000	0.000	0.000	0.000	0.000
FT BERTimbau	0.000	0.000	0.000	0.000	0.000	0.000
FT LegalBert-pt	0.000	0.000	0.000	0.000	0.000	0.000

Source: Created by the author (2025)

Two factors may have influenced the better performance of Ulysses-RFSQ for this dataset compared with Ulysses-RFCorpus: the difficulty and the size of the Preliminary Search corpus.

As observed in the experiments for Ulysses-RFCorpus, the use of the past RF information achieves greater improvements for scenarios in which the base IR algorithm did not perform very well, thus it could have a greater impact on the results for more difficult datasets. The corpus' size could also have influenced this impact, as the Preliminary Search corpus contains three times more queries than Ulysses-RFCorpus.

7.2.3 Discussion

Analyzing the results obtained for both corpora aiming to answer **RQ2**, we could say that the proposed method is capable of improving the results of documents retrieval for the Brazilian legislative domain. By using RF information from past similar queries to re-rank the documents, Ulysses-RFSQ could improve the IR process in many scenarios. The improvements were better observed in scenarios for which the base IR algorithm did not achieve a great performance, although marginal improvements could also be observed while using the best base IR algorithms, such as the preprocessed versions of BM25L and Okapi BM25.

Moreover, we could see that the larger the amount of past queries within the database, the better the performance of Ulysses-RFSQ-OR. Table 16 brings the improvements in the baselines' MAP results achieved by the best configuration of Ulysses-RFSQ-OR both for the parameter assessment and using the whole corpora. The results show that the improvements were always greater using the whole set of queries than for the smaller subset used for the parameters evaluation — which corresponded of 20% of the Preliminary Search corpus and 50% of Ulysses-RFCorpus.

For the Preliminary Search corpus, the use of the entire set of queries nearly doubled the improvements observed using only 1/5 of the data for all of the IR algorithms. In some cases, the improvements were 2.2 or 2.3 times bigger using 100% of the stored data. The same could be observed for some scenarios using Ulysses-RFCorpus: for Legal-BERTimbau, BERTimbauLaw, and Multilingual MPNet, the improvements achieved using all of the queries were two times greater than using only half of them.

Table 16 – Improvements in the baselines' MAP results by the best configuration for each IR algorithm, using the different subsets of the corpora.

	Ulysses-RFCorpus		**Preliminary Search**	
IR algorithm	**50%**	**100%**	**20%**	**100%**
BM25L_PRE	0.0012	0.0015	0.0203	0.0406
BM25L_NP	0.0038	0.0055	0.0259	0.0415
Okapi BM25_PRE	0.0018	0.0029	0.0200	0.0384
Okapi BM25_NP	0.0041	0.0049	0.0250	0.0474
BERTimbau	0.0214	0.0391	0.0450	0.0773
Legal-BERTimbau	0.0091	0.0210	0.0289	0.0586
JurisBERT	0.0084	0.0161	0.0371	0.0708
BERTimbauLaw	0.0099	0.0223	0.0300	0.0566
LegalBert-pt	0.0142	0.0224	0.0315	0.0692
LaBSE	0.0238	0.0276	0.0269	0.0636
Multilingual MPNet	0.0073	0.0160	0.0271	0.0563
Multilingual MiniLM	0.0071	0.0130	0.0291	0.0546
FT BERTimbau	0.0076	0.0122	0.0293	0.0625
FT LegalBert-pt	0.0123	0.0147	0.0367	0.0747

Source: Created by the author (2025)

7.3 COSINE SIMILARITY AND CONTEXTUAL EMBEDDINGS COMPARISON (RQ4)

In order to answer **RQ4) What is the best method to find, within a database of stored queries, the queries that are similar to the one currently being processed?**, we compared the results achieved by the use of only the cosine similarity to select the similar queries with the use of contextual embeddings generated by BERT-based models. For this evaluation, we used the Ulysses-RFSQ-OR version. The results for each configuration for both corpora can be found in Appendix L.

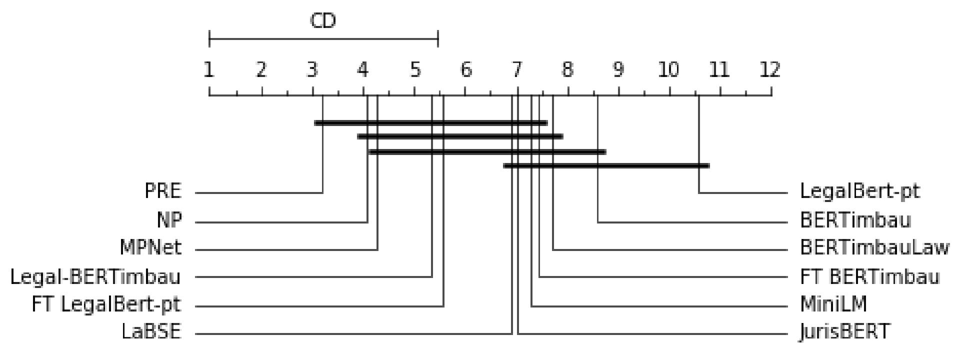
7.3.1 Ulysses-RFCorpus

The results for Ulysses-RFCorpus show that the best technique to select the similar queries varied depending on the algorithm used to perform the documents retrieval. For eight of the IR algorithm, the best MAP result was obtained by the use of only the cosine similarity, while the use of the semantic embeddings was superior for six of them. Analyzing the two different approaches to retrieve the legislative documents separately, we can see that an SBERT model achieved the best result for all of the four BM25 versions. Meanwhile, for eight of the 10 LMs, the use of the cosine similarity overcame the use of contextual embeddings.

In addition, we used the Student's t-test to compare the AP results achieved by each queries selector algorithm in a statistically significant way. Appendix M presents the p-values obtained for every comparison using each of the IR algorithms. As we used a confidence level of 95%, it can be noticed that there was no statistical difference between the strategies used to retrieve the similar queries for the vast majority of the scenarios. Some cases presented statistical significance only while using BERTimbau and LegalBert-pt to retrieve the documents.

This may be explained by the number of queries contained in Ulysses-RFCorpus. Thus, there might be insufficient data to evaluate and point out which is the best strategy to select the past similar queries in this dataset. This can also be seen by the application of the Nemenyi post-hoc test (NEMENYI, 1963) on the MAP results for Ulysses-RFCorpus. The Critical Difference (CD) diagram found in Figure 14 shows that, although the cosine similarity versions achieved the more consistent results for this corpus — reaching the lower mean ranks —, there was no significant difference between them and most of the BERT-based models.

Figure 14 – CD diagram comparing the MAP results of the algorithms used to search for the similar queries in Ulysses-RFCorpus.



Source: Created by the author (2025)

It is also worth noticing that FT BERTimbau, the best BERT-based model for the retrieval of the legislative documents, achieved a poor performance when used to select the similar queries from Ulysses-RFCorpus. Figure 14 shows that this model achieved a high mean rank, surpassing only three other models.

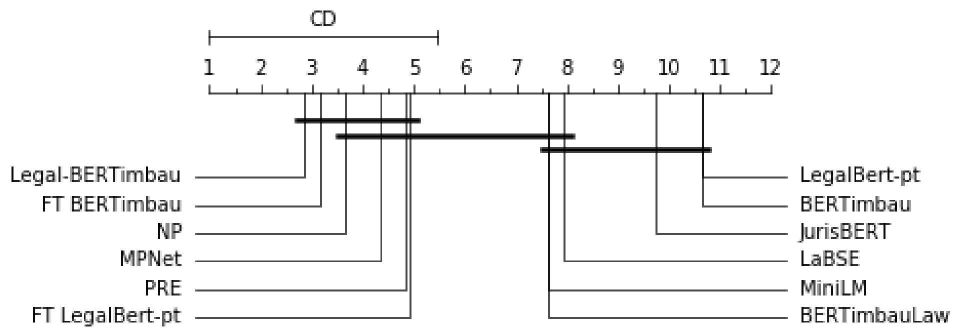
7.3.2 Preliminary Search corpus

For the Preliminary Search corpus, the use of BERT-based models to retrieve the past queries achieved the best results for 11 of the 14 base IR algorithms. Meanwhile, the cosine similarity without preprocessing was the best for the other three algorithms. Thus, in the majority of the evaluated scenarios, the use of contextual embeddings to search for the similar queries performed better for this corpus.

Analyzing the Student's t-test p-values obtained from the comparison between the AP results of the different algorithms used to retrieve the similar queries (Appendix N), we can see that some techniques achieved statistically better results than others in specific scenarios. This shows that, depending on the IR algorithm used, there may be a considerable difference between the choice for using only the cosine similarity and the use of contextual embeddings. For each scenario, the best algorithm used to retrieve the similar queries set performed statistically better than some others.

In general, considering the MAP results of all of the different configurations using the Preliminary Search corpus, there are no significant difference between the use of the two best SBERT models — FT BERTimbau and Legal-BERTimbau — and the use of just the cosine similarity. This can be seen in the CD diagram presented in Figure 15. However, analyzing each individual scenario (Appendix N), we can notice that the use of contextual embeddings performed statistically better than cosine for six of the IR algorithms.

Figure 15 – CD diagram comparing the MAP results of the algorithms used to search for the similar queries in the Preliminary Search corpus.



Source: Created by the author (2025)

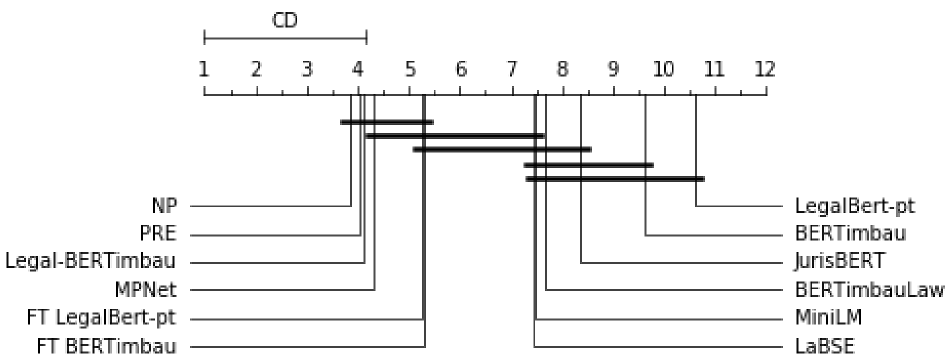
Differently from the results observed using Ulysses-RFCorpus, FT BERTimbau achieved the best performance for half of the IR algorithms for the Preliminary Search dataset. In addition, as can be seen in Figure 15, it performed statistically better than six of the nine other BERT-based models.

Finally, evaluating the use of preprocessing techniques to compute the cosine similarity, there were only three cases in which a significant difference could be noticed between the two versions. The preprocessed version of the cosine measure performed statistically better while using LaBSE and Multilingual MiniLM as the base IR algorithm, while the cosine version without preprocessing techniques was statistically superior for the use of FT LegalBert-pt. For all of the other IR algorithms, there was no statistical significance between the results of the two cosine versions.

7.3.3 Discussion

Analyzing the results for both Ulysses-RFCorpus and the Preliminary Search corpus, we cannot point out the best strategy to select the similar queries. The CD diagram from Figure 16 shows that, considering the MAP results for all configurations and for both corpora, there are no statistical difference between the use of only the cosine similarity and the use of contextual embeddings to retrieve the past queries.

Figure 16 – CD diagram comparing the MAP results of the algorithms used to search for the similar queries considering both corpora.



Source: Created by the author (2025)

By analyzing the individual scenarios, we could see that one strategy performed better than the other for specific cases. For the Ulysses-RFCorpus, the difference between the strategies was very small. For the Preliminary Search corpus, on the other hand, there were scenarios in which the use of SBERT embeddings achieved significant better results.

In addition, some of the BERT-based models were superior than the other models in the majority of the cases. Figure 16 indicates that, considering all evaluated scenarios for Ulysses-RFSQ-OR with both corpora, the two fine-tuned models — FT BERTimbau and FT LegalBert-pt — achieved statistically better results than their zero-shot versions — BERTimbau and LegalBert-pt. This points out the importance of fine-tuning for domain-specific tasks.

Based on the distinct results obtained for each corpora, we could also see that there was no statistical difference between the performance achieved by the preprocessed version of the cosine similarity and the version without the use of preprocessing techniques. For Ulysses-RFCorpus, the PRE version achieved a smaller mean rank than the NP one. The NP version, on its turn, was slightly better than the preprocessed version for the Preliminary Search corpus and when analyzing the results for both corpora together. In this sense, we can point out that may be better to use the cosine similarity without preprocessing, as both strategies achieved similar results and the preprocessing step is expensive.

Therefore, answering **RQ4**, we can conclude that the use of contextual embeddings may improve the IR results in a larger way than the use of the cosine similarity for specific cases. Nonetheless, the use of just the cosine measure without preprocessing achieved the most consistent results, obtaining the smaller mean rank in Figure 16 and being statistically better than more than a half of the SBERT models. This performance, together with the computational cost of using LMs, show that the simplest technique can still be the more suitable for many scenarios.

7.4 EVALUATING THE FOUR ULYSSES-RFSQ VERSIONS (RQ5)

As explained in Chapter 4, four versions of Ulysses-RFSQ — besides the preliminary one proposed in (VITÓRIO et al., 2022) — were developed. They differ by the use or not of the *irrelevant* documents to compute λ, as well as the use of more than one relevance level, such as *very relevant* and *somewhat relevant*. Thus, in this section, we compared the four versions aiming to answer **RQ5) Is the irrelevant documents information from past queries useful for re-ranking the retrieved documents for a new query?**

The results for Ulysses-RFSQ-OR (which uses only the *relevant* documents), Ulysses-RFSQ-RI (which uses both *relevant* and *irrelevant* information), Ulysses-RFSQ-DRL (which uses the different levels of relevance), and Ulysses-RFSQ-ALL (which uses all of the available information) can be found in Appendices L, O, P, and Q, respectively. As the Preliminary Search corpus contains only the list of *relevant* documents for each query, this assessment was performed solely with Ulysses-RFCorpus.

In order to evaluate and compare the four versions, Appendix R summarizes the MAP and nDCG results for each Ulysses-RFSQ version. In it, the green color represents the cases for which that version performed better than Ulysses-RFSQ-OR, which we used as baseline to assess the use of only the *relevant* documents. Meanwhile, the red color indicates the cases in which the use of additional information worsened the retrieval results compared with Ulysses-RFSQ-OR.

For the BM25 approach, we can see that the use of the *irrelevant* information improved the results for the preprocessed variants. In both cases, Ulysses-RFSQ-RI and Ulysses-RFSQ-ALL performed better than Ulysses-RFSQ-OR while using most of the algorithms to select the similar queries. It is worth remembering that BM25L_PRE and Okapi BM25_PRE are the best algorithms for IR in this specific legislative scenario (VITÓRIO et al., 2025b). Nonetheless, Ulysses-RFSQ-RI and Ulysses-RFSQ-ALL could improve their results even further. We noticed an improvement on the nDCG performance for these two algorithms of 0.0022 and 0.0033 using Ulysses-RFSQ-ALL, in comparison with the improvements of 0.0010 and 0.0025 achieved by Ulysses-RFSQ-OR.

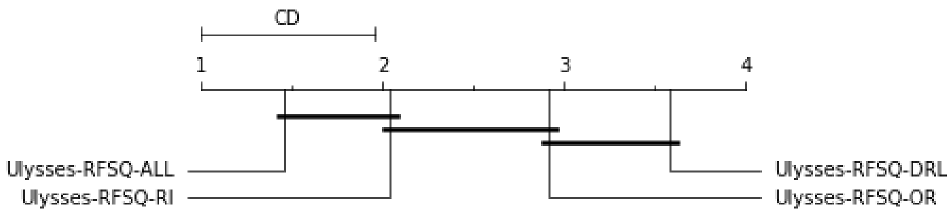
On the other hand, the use of the *irrelevant* information to give a penalty for the documents' scores did not reach the same results for the variants of the BM25 algorithm without preprocessing. For both BM25L_NP and Okapi BM25_NP, the best results were achieved by Ulysses-RFSQ-DRL, which only differentiates the two levels of relevance. Meanwhile, Ulysses-RFSQ-RI and Ulysses-RFSQ-ALL performed worse than Ulysses-RFSQ-OR for most of these configurations.

The same could be observed for the use of BERT-based models to retrieve the documents. For this scenario, Ulysses-RFSQ-DRL surpassed the Ulysses-RFSQ-RI and Ulysses-RFSQ-ALL versions. However, for half of the experimental configurations, the version of Ulysses-RFSQ which don't differentiate the levels of relevance achieved the best results, which did not occur for the BM25 algorithm. Moreover, for the vast majority of the cases using LMs to perform the documents retrieval, Ulysses-RFSQ-RI and Ulysses-RFSQ-ALL harmed the results in comparison with Ulysses-RFSQ-OR, while Ulysses-RFSQ-DRL reached a similar performance to the version which used only the information from one category of relevance.

This may be explained by the performance of each group of base IR algorithms for Ulysses-RFCorpus. While using the preprocessed BM25 variants, it was better to re-rank both *relevant* and *irrelevant* documents and use the different levels of relevance in order to slightly improve the already great results. Figures 17 and 18 present the CD diagrams obtained by performing the Nemenyi post-hoc test on, respectively, the MAP and nDCG results achieved by each Ulysses-RFSQ version for the BM25L_PRE and Okapi BM25_PRE configurations. The diagrams show that, for this scenario, Ulysses-RFSQ-ALL performed statistically better than

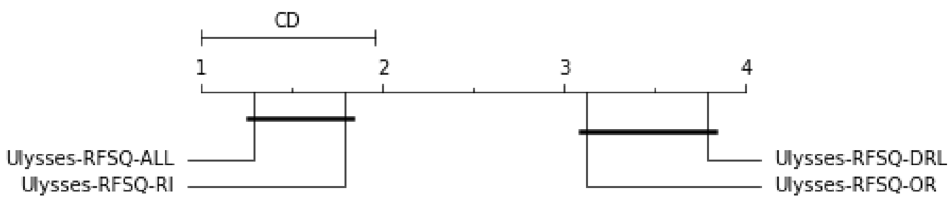
Ulysses-RFSQ-OR for both metrics, while Ulysses-RFSQ-RI overcame Ulysses-RFSQ-OR in a statistically significant way for the nDCG evaluation.

Figure 17 – CD diagram comparing the MAP results of the four Ulysses-RFSQ versions for the preprocessed BM25 variants.



Source: Created by the author (2025)

Figure 18 – CD diagram comparing the nDCG results of the four Ulysses-RFSQ versions for the preprocessed BM25 variants.



Source: Created by the author (2025)

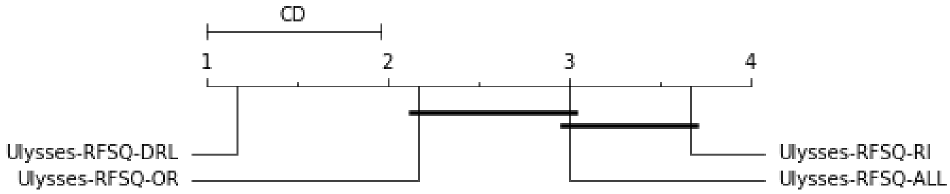
It is worth noticing that, for this scenario, the use of different weights for the two different relevance levels — while disregarding the *irrelevant* information — harmed the nDCG results more than the MAP ones. The nDCG metric which takes into account these different levels to evaluate the retrieved documents.

Unlike the previous scenario, for the use of the BM25 algorithms without preprocessing techniques — which achieved an intermediate performance —, giving different weights to documents with different relevance levels could achieve greater improvements on the baselines' results than the use of only one category of relevance. Observing the CD diagrams (Figures 19 and 20) for the configurations using BM25L_NP and Okapi BM25_NP, we can see that the Ulysses-RFSQ-DRL version was statistically superior than all of the other three versions, both for MAP and nDCG.

In addition, analyzing the nDCG's CD diagram (Figure 20), we can see that, although there was no statistical difference between the results, the Ulysses-RFSQ version using all the information, including the two categories of relevance, achieved a smaller mean rank than

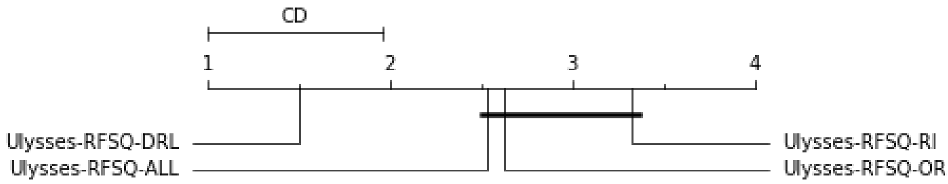
Ulysses-RFSQ-OR. The same did not occur for the MAP metric, showing that, for this scenario, the acknowledgment of the different relevance levels had a bigger importance, at least for the nDCG metric.

Figure 19 – CD diagram comparing the MAP results of the four Ulysses-RFSQ versions for the BM25 variants without preprocessing.



Source: Created by the author (2025)

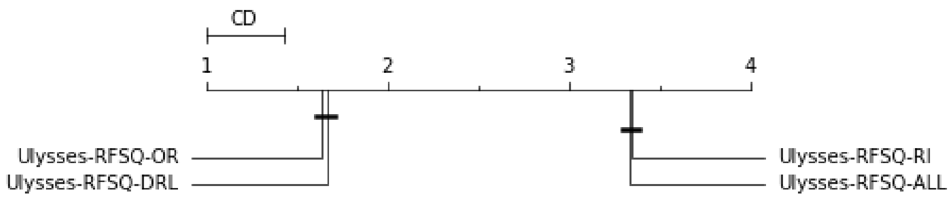
Figure 20 – CD diagram comparing the nDCG results of the four Ulysses-RFSQ versions for the BM25 variants without preprocessing.



Source: Created by the author (2025)

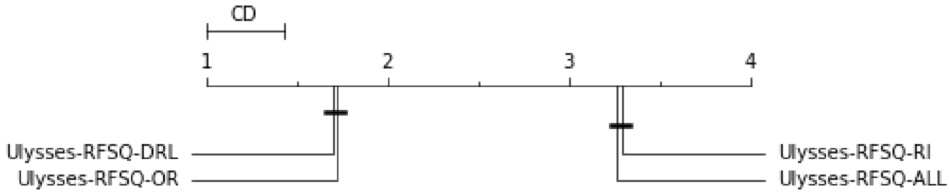
For the configurations using BERT-based models to perform the documents retrieval — which achieved a very poor performance for Ulysses-RFCorpus —, there was no significant difference between the results of Ulysses-RFSQ-OR and Ulysses-RFSQ-DRL, as can be seen in Figures 21 and 22. Moreover, both versions that use the *irrelevant* documents information — Ulysses-RFSQ-RI and Ulysses-RFSQ-ALL — performed statistically worse than those which did not use this information.

Figure 21 – CD diagram comparing the MAP results of the four Ulysses-RFSQ versions for the BERT-based models.



Source: Created by the author (2025)

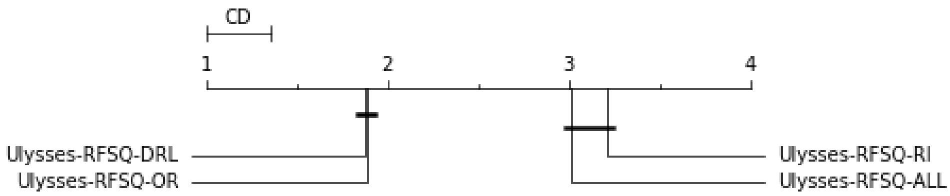
Figure 22 – CD diagram comparing the nDCG results of the four Ulysses-RFSQ versions for the BERT-based models.



Source: Created by the author (2025)

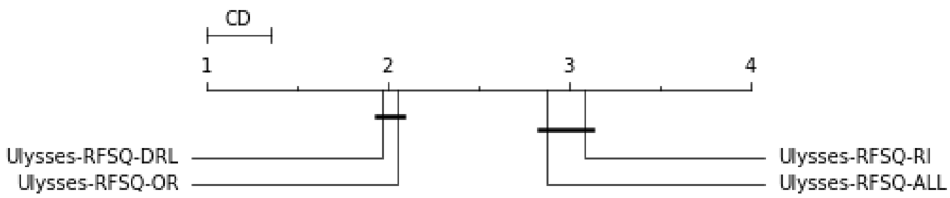
Finally, analyzing all of the different experimental configurations together, the findings are similar to those using the SBERT models. The CD diagrams from Figures 23 and 24 show that Ulysses-RFSQ-OR and Ulysses-RFSQ-DRL achieved better results than the other two versions with statistical significance. On the other hand, there was no significant difference between the results of these best versions.

Figure 23 – CD diagram comparing the MAP results of the four Ulysses-RFSQ versions for all configurations.



Source: Created by the author (2025)

Figure 24 – CD diagram comparing the nDCG results of the four Ulysses-RFSQ versions for all configurations.



Source: Created by the author (2025)

Comparing the performance of the four Ulysses-RFSQ versions while using either only the cosine similarity or contextual embeddings to select the similar queries, we can see that the Ulysses-RFSQ-OR version was the better for using only the cosine measure. For the majority of the cases, the other three versions of the proposed method did not surpassed the version that only re-ranks the *relevant* documents. This may indicate that the cosine similarity alone is not well-suited to retrieve a set of similar queries from which is possible to extract valuable RF information from the *irrelevant* documents and the different relevance levels. As both MAP and nDCG results were harmed by the use of more than just the *relevant* documents, these versions of Ulysses-RFSQ may need different parameters' values to have a positive impact on the retrieval performance.

Therefore, answering **RQ5**, we can conclude that the use of the *irrelevant* information to give a penalty for the retrieved documents' score was useful for only one specific scenario. The versions that use this information were only better than the others while using the Ulysses-RFCorpus' state-of-the-art algorithms to perform the legislative documents retrieval. For this case, the version of Ulysses-RFSQ that uses all of the available information could improve the results more than the Ulysses-RFSQ-OR version. Although the improvements on the performance of the preprocessed BM25 variants were small, it is worth remembering that these algorithms already achieve great results for Ulysses-RFCorpus.

For the other scenarios, it was better to use either only the *relevant* information or to give different weights to the different levels of relevance. In general, there was no significant difference between the use of Ulysses-RFSQ-OR or Ulysses-RFSQ-DRL, except for some specific cases, in which Ulysses-RFSQ-DRL performed better.

8 CONCLUSION

In this study, a novel IR method was proposed and described: *Ulysses-RFSQ*. Ulysses-RFSQ uses the RF information from past queries similar to the one currently being processed, aiming to improve the retrieval performance by re-ranking the documents retrieved by a base IR algorithm. Its idea lies on the assumption that if a document is relevant to a query, it might also be relevant to another query sufficiently similar to the first one. In addition, one of the Ulysses-RFSQ's characteristics is that it can be used together with any IR algorithm that computes a score for the documents, such as BM25 or SBERT models.

The legislative domain was chosen to perform the experiments, as parliamentarians' consultations — which work as queries — are often redundant, making the evaluation of Ulysses-RFSQ possible. There is a lack of benchmark datasets containing RF information for similar queries, which was mitigated by the creation of a corpus in the legislative domain: Ulysses-RFCorpus — another contribution of this work. It was built together with the Conle department of the Brazilian Chamber of Deputies and it is publicly available[1].

Besides Ulysses-RFCorpus, another dataset was made available by the Chamber for this study and both were used to perform the evaluation of the proposed method, which was compared with baselines of IR algorithms without the use of past relevance information. Two variants of BM25 and 10 BERT-based models were used as the base IR algorithms, in order to evaluate the impact of Ulysses-RFSQ in different scenarios.

The experiments were conducted aiming to answer several Research Questions. First, the two parameters of Ulysses-RFSQ (cut and δ) were evaluated in order to select the best values for the other experiments and to assess the trade-off between the use of RF information from a greater number of past queries and the use of a smaller set of highly similar ones. Based on the results, we could conclude that the assessment of the parameters' values should be performed for each individual scenario. Nevertheless, when using a base IR algorithm that already achieves a great performance, it is more advisable to perform the re-ranking in a smaller set of queries in order to improve even further the results.

The use of contextual semantic embeddings — generated by BERT-based models — and the use of just the cosine similarity to search for the similar past queries were also compared. For this assessment, we could not reach a conclusion on which approach was more useful, as

[1] https://github.com/ulysses-camara/Ulysses-RFCorpus

the differences between the results were not statistically significant. Therefore, we can say that the use of just the cosine similarity may be more suitable, as the use of BERT-based models has a higher computational cost.

Using Ulysses-RFCorpus, we could also assess the use of different categories of document relevance, such as *very relevant* and *somewhat relevant*, as well as the use of documents judged as *irrelevant* for the past queries to re-rank the documents for the current query. For these experimental configurations, the findings show that the use of the *irrelevant* information was useful in only one specific case. Thus, re-ranking just the *relevant* documents or giving different weights for the different levels of relevance achieved the best performance in the vast majority of the scenarios.

Finally, through all of the experiments, we could find that the proposed method can improve the performance of different base IR algorithms for this Brazilian legislative scenario. These improvements, however, can be seen more clearly and with statistical significance for cases in which the base algorithm could not reach great results, such as using the BERT-based models and for the Preliminary Search corpus. The observed improvements in the MAP results ranged from 0.0122 to 0.0391 for Ulysses-RFCorpus and from 0.0384 to 0.0773 for the Preliminary Search corpus — in some cases, more than doubling the baseline's performance. It is worth to mention that the Preliminary Search corpus contains three times more queries than Ulysses-RFCorpus, pointing out that the more the number of stored past queries, the greater the improvement on the results.

On the other hand, while using preprocessed BM25 variants to perform the documents retrieval in Ulysses-RFCorpus, the improvements were only marginal, when they could be observed. The reason for this is that these algorithms are the state-of-the-art for this corpus (VITÓRIO et al., 2025a; VITÓRIO et al., 2025b). Nonetheless, even for these algorithms with great performance, the results could be improved in many cases, although without statistical significance.

8.1 LIMITATIONS

The first limitation of this study is the lack of Relevance Feedback benchmark datasets containing similar queries. This issue made it impossible to evaluate Ulysses-RFSQ in a greater number of datasets, harming the scope of this study. The corpora from the Brazilian legislative domain were used to mitigate this problem, however each corpus has some drawbacks.

Ulysses-RFCorpus was built using an IR model almost identical to the one used as one of the baselines for this study. Therefore, the baseline achieved state-of-the-art results — as it retrieved almost every relevant document —, leaving no room for major improvements. Meanwhile, the Preliminary Search corpus was built and extracted manually, and, although the judgment was done by experts, some problems — such as cases in which the same query presented two different lists of relevant documents — were found. These problems might have decreased this dataset's reliability.

The second problem was the lack of related work using the past RF information to improve the retrieval for future queries, which can be also explained by the aforementioned lack of benchmark datasets. As most researchers use this information to expand their queries, no work that could be replicated and used as comparison was found dealing with the use of similar past queries in a way similar to this study. This issue, alongside the lack of datasets, did not allow us to compare Ulysses-RSFQ with other methods. In this sense, we have performed the experiments comparing the performance achieved by using Ulysses-RFSQ with the performance of the baseline models without RF past information, aiming to confirm that the proposed method can have a positive impact on the results.

Finally, another limitation lied on the need to use legislative data to evaluate the proposed model. Two corpora were used to evaluate Ulysses-RFSQ and only one could be made available, due to privacy issues (Brazilian Chamber of Deputies, 1993). This harms the study's reproducibility.

8.2 FUTURE WORK

As future work, the goal is to evaluate Ulysses-RFSQ with corpora from other domains. This will aim to verify if the improvements observed for the legislative domain can be observed for other scenarios as well. However, the other domains must also present redundancy in the queries in order to Ulysses-RFSQ be effectively applied.

Other approaches and IR methods may also be used as the base algorithms. For instance, LLMs can be used either to retrieve the documents and to search for the similar past queries, in the same way as the LMs used in this study. In addition, the comparison between Ulysses-RFSQ and other re-ranking techniques could also be performed.

REFERENCES

ALBUQUERQUE, H. O.; COSTA, R.; SILVESTRE, G.; SOUZA, E.; SILVA, N. F. F. da; VITÓRIO, D.; MORIYAMA, G.; MARTINS, L.; SOEZIMA, L.; NUNES, A.; SIQUEIRA, F.; TARREGA, J. P.; BEINOTTI, J. V.; DIAS, M.; SILVA, M.; GARDINI, M.; SILVA, V.; CARVALHO, A. C. P. L. F. de; OLIVEIRA, A. L. I. UlyssesNER-Br: A corpus of brazilian legislative documents for Named Entity Recognition. In: *Computational Processing of the Portuguese Language*. [S.l.]: Springer International Publishing, 2022. p. 3–14.

ALBUQUERQUE, H. O.; SOUZA, E.; OLIVEIRA, A. L. I.; MACÊDO, D.; ZANCHETTIN, C.; VITÓRIO, D.; SILVA, N. F. F. da; CARVALHO, A. C. P. L. F. de. On the assessment of deep learning models for named entity recognition of brazilian legal documents. In: *Progress in Artificial Intelligence*. Cham: Springer Nature Switzerland, 2023. p. 93–104.

ALBUQUERQUE, H. O.; SOUZA, E.; SILVA, T.; GOUVEIA, R. P.; JUNIOR, F.; VITÓRIO, D.; SILVA, N. F. F. da; CARVALHO, A. C. de; OLIVEIRA, A. L.; ANDRADE, F. E. de. UlyssesNERQ: Expanding queries from Brazilian Portuguese legislative documents through named entity recognition. In: GAMALLO, P.; CLARO, D.; TEIXEIRA, A.; REAL, L.; GARCIA, M.; OLIVEIRA, H. G.; AMARO, R. (Ed.). *Proceedings of the 16th International Conference on Computational Processing of Portuguese - Vol. 1*. Santiago de Compostela, Galicia/Spain: Association for Computational Lingustics, 2024. p. 342–352. Available at: <https://aclanthology.org/2024.propor-1.35/>.

ALMEIDA, P. G. R. Uma jornada para um Parlamento inteligente: Câmara dos Deputados do Brasil. *Red Información*, v. 24, 2021. Available at: <https://www.redinnovacion.org/revista/red-informaciÃșn-ediciÃșn-nÂř-24-marzo-2021>.

ALY, A. A. Using a query expansion technique to improve document retrieval. *International Journal Information Technologies and Knowledge*, v. 2, n. 4, p. 343–348, 2008.

BEITZEL, S. M.; JENSEN, E. C.; FRIEDER, O. Average R-Precision. In: ______. *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. p. 195.

BEITZEL, S. M.; JENSEN, E. C.; FRIEDER, O. Map. In: ______. *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. p. 1691–1692.

BHATTACHARYA, P.; GHOSH, K.; PAL, A.; GHOSH, S. Methods for computing legal document similarity: A comparative study. *ArXiv*, abs/2004.12307, 2020.

BILLERBECK, B.; SCHOLER, F.; WILLIAMS, H. E.; ZOBEL, J. Query expansion using associated queries. In: *Proceedings of the Twelfth International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2003. (CIKM '03), p. 2–9. ISBN 1581137230.

BRANDT, M. B. Ethical aspects in the organization of legislative information. *KO KNOWLEDGE ORGANIZATION*, Nomos Verlagsgesellschaft mbH & Co. KG, v. 45, n. 1, p. 3–12, 2018.

BRANDT, M. B. *Modelagem da informação legislativa: arquitetura da informação para o processo legislativo brasileiro*. 268 p. Phd Thesis (PhD Thesis) — Faculdade de Filosofia e Ciências da Universidade Estadual Paulista (UNESP), 2020.

Brazilian Chamber of Deputies. *Resolução da Câmara dos Deputados Nº 48, de 1993*. 1993. Available at: <https://www2.camara.leg.br/legin/int/rescad/1993/resolucaodacamaradosdeputados-48-26-agosto-1993-321538-normaatualizada-pl.html>.

BUCKLEY, C.; SALTON, G.; ALLAN, J.; SINGHAL, A. Automatic query expansion using SMART: TREC 3. *NIST special publication sp*, National Institute of Standards & Technology, p. 69–69, 1995.

CANTADOR, I.; SÁNCHEZ, L. Q. Semantic annotation and retrieval of parliamentary content: A case study on the spanish congress of deputies. In: *Proceedings of the First Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020)*. [S.l.: s.n.], 2020. (CEUR Workshop Proceedings, v. 2621).

CARPINETO, C.; ROMANO, G. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys*, v. 44, n. 1, p. 1–50, 2012.

CASELI, H. M.; NUNES, M. G. V. (Ed.). *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. 3. ed. BPLN, 2024. ISBN 978-65-00-95750-1. Available at: <https://brasileiraspln.com/livro-pln/3a-edicao/>.

CETINTAS, S.; SI, L.; YUAN, H. *Using past queries for resource selection in distributed information retrieval*. [S.l.], 2011. Department of Computer Science Technical Reports. Paper 1743. Purdue University.

CHALKIDIS, I.; FERGADIOTIS, M.; MANGINAS, N.; KATAKALOU, E.; MALAKASIOTIS, P. Regulatory compliance through Doc2Doc information retrieval: A case study in EU/UK legislation where text similarity has limitations. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. [S.l.: s.n.], 2021. p. 3498–3511.

CONNEAU, A.; KHANDELWAL, K.; GOYAL, N.; CHAUDHARY, V.; WENZEK, G.; GUZMÁN, F.; GRAVE, E.; OTT, M.; ZETTLEMOYER, L.; STOYANOV, V. Unsupervised cross-lingual representation learning at scale. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. [S.l.]: Association for Computational Linguistics, 2020. p. 8440–8451.

COSTA, R.; ALBUQUERQUE, H. O.; SILVESTRE, G.; SILVA, N. F. F.; SOUZA, E.; VITÓRIO, D.; NUNES, A.; SIQUEIRA, F.; TARREGA, J. P.; BEINOTTI, J. V.; DIAS, M. de S.; PEREIRA, F. S. F.; SILVA, M.; GARDINI, M.; SILVA, V.; CARVALHO, A. C. P. L. F. de; OLIVEIRA, A. L. I. Expanding ulyssesner-br named entity recognition corpus with informal user-generated text. In: *Progress in Artificial Intelligence*. Cham: Springer International Publishing, 2022. p. 767–779.

CRASWELL, N. Mean reciprocal rank. In: ______. *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. p. 1703–1703.

DEERWESTER, S.; DUMAIS, S. T.; FURNAS, G. W.; LANDAUER, T. K.; HARSHMAN, R. Indexing by latent semantic analysis. *Journal of the American society for information science*, Wiley Online Library, v. 41, n. 6, p. 391–407, 1990.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human*

Language Technologies, Volume 1 (Long and Short Papers). [S.l.]: Association for Computational Linguistics, 2019. p. 4171–4186.

EL-GHALI, B.; EL-QADI, A. Context-aware query expansion method using language models and latent semantic analyses. *Knowledge and Information Systems*, Springer, v. 50, p. 751–762, 2017.

FENG, F.; YANG, Y.; CER, D.; ARIVAZHAGAN, N.; WANG, W. Language-agnostic BERT sentence embedding. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. [S.l.]: Association for Computational Linguistics, 2022. p. 878–891.

FILHO, J. A. W.; WILKENS, R.; IDIART, M.; VILLAVICENCIO, A. The brWaC corpus: A new open resource for Brazilian Portuguese. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), 2018.

FITZPATRICK, L.; DENT, M. Automatic feedback using past queries: Social searching? In: *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.]: Association for Computing Machinery, 1997. (SIGIR '97), p. 306–313.

GOMES, T.; LADEIRA, M. A new conceptual framework for enhancing legal information retrieval at the brazilian superior court of justice. In: *Proceedings of the 12th International Conference on Management of Digital EcoSystems*. New York, NY, USA: Association for Computing Machinery, 2020. (MEDES '20), p. 26–29. ISBN 9781450381154.

GOUVEIA, R. P.; CARVALHO, A. C. de; SOUZA, E.; ALBUQUERQUE, H. O.; VITóRIO, D.; SILVA, N. F. da. Applying active learning in named entity recognition corpora expansion in legal domain. In: *Annual International Conference on Digital Government Research*. [S.l.: s.n.], 2025. v. 26.

GUTIÉRREZ-SOTO, C. *Exploring the reuse of past search results in information retrieval*. Phd Thesis (PhD Thesis) — Université de Toulouse, Université Toulouse III-Paul Sabatier, 2016.

GUTIÉRREZ-SOTO, C.; HUBERT, G. Randomized algorithm for information retrieval using past search results. In: *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*. [S.l.: s.n.], 2014. p. 1–9.

GUTIÉRREZ-SOTO, C.; PALOMINO, M. A.; ROA, E.; GALDAMES, P. Improving search engine performance through dynamic caching. In: *2021 40th International Conference of the Chilean Computer Science Society (SCCC)*. [S.l.: s.n.], 2021. p. 1–6.

HUBERT, G.; MOTHE, J. Reusing past queries to facilitate information retrieval. In: *ICSOFT (ISDM/EHST/DC)*. [S.l.: s.n.], 2007. p. 166–171.

HUST, A. Introducing query expansion methods for collaborative information retrieval. In: *Reading and Learning*. [S.l.: s.n.], 2004. p. 252–280.

JANNACH, D.; LERCHE, L.; ZANKER, M. Recommending based on implicit feedback. In: ______. *Social Information Access: Systems and Technologies*. Cham: Springer International Publishing, 2018. p. 510–569. ISBN 978-3-319-90092-6.

JÄRVELIN, K.; KEKÄLÄINEN, J. Discounted cumulated gain. In: _____. *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. p. 849–853.

JOACHIMS, T.; GRANKA, L.; PAN, B.; HEMBROOKE, H.; RADLINSKI, F.; GAY, G. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems*, Association for Computing Machinery, New York, NY, USA, v. 25, n. 2, p. 7–es, apr 2007. ISSN 1046-8188.

KAMPHUIS, C.; VRIES, A. P. de; BOYTSOV, L.; LIN, J. Which BM25 do you mean? A large-scale reproducibility study of scoring variants. In: *Advances in Information Retrieval*. [S.l.: s.n.], 2020. p. 28–34.

LIN, J.; NOGUEIRA, R.; YATES, A. *Pretrained transformers for text ranking: Bert and beyond*. [S.l.]: Springer Nature, 2022.

LV, Y.; ZHAI, C. When documents are very long, BM25 fails! In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Association for Computing Machinery, 2011. (SIGIR '11), p. 1103–1104.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. [S.l.]: Cambridge University Press, 2008.

MAXWELL, K. T.; SCHAFER, B. Concept and context in legal information retrieval. In: *Legal Knowledge and Information Systems - JURIX 2008*. [S.l.]: IOS Press, 2008. (Frontiers in Artificial Intelligence and Applications), p. 63–72. ISBN 9781586039523.

MELO, R.; SANTOS, P. A.; DIAS, J. A semantic search system for the Supremo Tribunal de Justiça. In: *Progress in Artificial Intelligence*. Cham: Springer Nature Switzerland, 2023. p. 142–154.

MOSHFEGHI, Y.; VELINOV, K.; TRIANTAFILLOU, P. Improving search results with prior similar queries. In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2016. (CIKM '16), p. 1985–1988. ISBN 9781450340731.

NEMENYI, P. B. *Distribution-free multiple comparisons*. [S.l.]: Princeton University, 1963.

OKABE, M.; YAMADA, S. Learning filtering rulesets for ranking refinement in relevance feedback. *Knowledge-Based Systems*, v. 18, n. 2, p. 117–124, 2005.

OLIVEIRA, R. A. N. de; JUNIOR, M. C. Experimental analysis of stemming on jurisprudential documents retrieval. *Information*, v. 9, n. 2, p. 28, 2018.

ONODA, T.; MURATA, H.; YAMADA, S. SVM-based interactive document retrieval with active learning. *New Generation Computing*, Springer, v. 26, n. 1, p. 49–61, 2007.

OPIJNEN, M. V.; SANTOS, C. On the concept of relevance in legal information retrieval. *Artificial Intelligence and Law*, Kluwer Academic Publishers, USA, v. 25, n. 1, p. 65–87, mar 2017. ISSN 0924-8463.

REIMERS, N.; GUREVYCH, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. [S.l.]: Association for Computational Linguistics, 2019. p. 3982–3992.

REIMERS, N.; GUREVYCH, I. Making monolingual sentence embeddings multilingual using knowledge distillation. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [S.l.]: Association for Computational Linguistics, 2020. p. 4512–4525.

RISSLAND, E. L.; DANIELS, J. J. The synergistic application of cbr to ir. *Artificial Intelligence Review*, Springer, v. 10, n. 5, p. 441–475, 1996.

ROBERTSON, S. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of Documentation*, Emerald Group Publishing Limited, v. 60, n. 5, p. 503–520, 2004.

ROBERTSON, S.; WALKER, S.; JONES, S.; HANCOCK-BEAULIEU, M.; GATFORD, M. Okapi at TREC-3. In: *TREC*. [S.l.: s.n.], 1994.

ROCHA, F.; SOUZA, E.; VITóRIO, D.; SILVA, N.; CARVALHO, A.; OLIVEIRA, A. Avaliação de frameworks para recuperação de documentos legislativos: um estudo de caso na câmara dos deputados brasileira. In: *Anais do XI Workshop de Computação Aplicada em Governo Eletrônico*. Porto Alegre, RS, Brasil: SBC, 2023. p. 224–231. ISSN 2763-8723.

RUSSELL-ROSE, T.; CHAMBERLAIN, J.; AZZOPARDI, L. Information retrieval in the workplace: A comparison of professional search practices. *Information Processing & Management*, v. 54, n. 6, p. 1042–1057, 2018.

SALTON, G.; BUCKLEY, C. Improving retrieval performance by relevance feedback. *Journal of the American society for information science*, v. 41, n. 4, p. 288–297, 1990.

SALTON, G.; WONG, A.; YANG, C. S. A vector space model for automatic indexing. *Communications of the ACM*, ACM, v. 18, n. 11, p. 613–620, nov 1975.

SANTOS, J. A. dos; SOUZA, E.; FILHO, C. J. A. B.; ALBUQUERQUE, H. O.; VITÓRIO, D.; LUCENA, D. C. G. de; SILVA, N.; CARVALHO, A. de. HIRS: A hybrid information retrieval system for legislative documents. In: *Progress in Artificial Intelligence*. Cham: Springer Nature Switzerland, 2024. p. 320–331.

SAVOY, J. Light stemming approaches for the french, portuguese, german and hungarian languages. In: *Proceedings of the 2006 ACM Symposium on Applied Computing*. New York, NY, USA: Association for Computing Machinery, 2006. (SAC '06), p. 1031–1035.

SCHWEIGHOFER, E.; GEIST, A. Legal query expansion using ontologies and relevance feedback. In: *Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques June 4th, 2007, Stanford University, Stanford, CA, USA*. [S.l.: s.n.], 2007. (CEUR Workshop Proceedings, v. 321), p. 149–160.

SILVEIRA, R.; PONTE, C.; ALMEIDA, V.; PINHEIRO, V.; FURTADO, V. Legalbert-pt: A pretrained language model for the brazilian portuguese legal domain. In: *Intelligent Systems*. Cham: Springer Nature Switzerland, 2023. p. 268–282.

SIQUEIRA, F. A.; VITÓRIO, D.; SOUZA, E.; SANTOS, J. A. P.; ALBUQUERQUE, H. O.; DIAS, M. S.; SILVA, N. F. F.; CARVALHO, A. C. P. L. F. de; OLIVEIRA, A. L. I.; BASTOS-FILHO, C. Ulysses Tesemõ: a new large corpus for Brazilian legal and governmental domain. *Language Resources and Evaluation*, p. 1685–1704, 2025.

SONG, K.; TAN, X.; QIN, T.; LU, J.; LIU, T.-Y. MPNet: masked and permuted pre-training for language understanding. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. [S.l.]: Curran Associates Inc., 2020. (NIPS '20).

SONG, S. kwang; MYAENG, S. H. A novel term weighting scheme based on discrimination power obtained from past retrieval results. *Information Processing & Management*, v. 48, n. 5, p. 919–930, 2012. ISSN 0306-4573. Large-Scale and Distributed Systems for Information Retrieval.

SOUZA, E.; MORIYAMA, G.; VITÓRIO, D.; CARVALHO, A. C. P. d. L. F. de; FÉLIX, N.; ALBUQUERQUE, H.; OLIVEIRA, A. L. I. Assessing the impact of stemming algorithms applied to brazilian legislative documents retrieval. In: *Proceedings of the 13th Brazilian Symposium in Information and Human Language Technology*. Porto Alegre, Brazil: SBC, 2021. p. 227–236.

SOUZA, E.; VITÓRIO, D.; MORIYAMA, G.; SANTOS, L.; MARTINS, L.; SOUZA, M.; FONSECA, M.; FÉLIX, N.; CARVALHO, A. C. P. d. L. F. d.; ALBUQUERQUE, H. O.; OLIVEIRA, A. L. I. An information retrieval pipeline for legislative documents from the brazilian chamber of deputies. In: *Legal Knowledge and Information Systems*. [S.l.]: IOS Press, 2021. p. 119–126.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERTimbau: pretrained BERT models for Brazilian Portuguese. In: *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*. [S.l.: s.n.], 2020.

STUDENT. The probable error of a mean. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 6, n. 1, p. 1–25, 1908. ISSN 00063444, 14643510.

URBANO, J.; LIMA, H.; HANJALIC, A. Statistical significance testing in information retrieval: An empirical analysis of type I, type II and type III errors. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.]: Association for Computing Machinery, 2019. (SIGIR'19), p. 505–514.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. [S.l.]: Curran Associates Inc., 2017. (NIPS'17), p. 6000–6010.

VIEGAS, C. F. O.; COSTA, B. C.; ISHII, R. P. JurisBERT: A new approach that converts a classification corpus into an STS one. In: *Computational Science and Its Applications – ICCSA 2023*. Cham: Springer Nature Switzerland, 2023. p. 349–365.

VITÓRIO, D.; SOUZA, E.; MARTINS, L.; SILVA, N. F. F. d.; CARVALHO, A. C. P. d. L. F. d.; OLIVEIRA, A. L. I. Ulysses-RFSQ: A novel method to improve legal information retrieval based on relevance feedback. In: SPRINGER. *Intelligent Systems: 11th Brazilian Conference, BRACIS 2022, Campinas, Brazil, November 28–December 1, 2022, Proceedings, Part I*. [S.l.], 2022. p. 77–91.

VITÓRIO, D.; SOUZA, E.; MARTINS, L.; SILVA, N. F. F. da; CARVALHO, A. C. P. d. L. de; OLIVEIRA, A. L. I.; ANDRADE, F. E. de. Building a relevance feedback corpus for legal information retrieval in the real-case scenario of the Brazilian Chamber of Deputies. *Language Resources and Evaluation*, p. 1257–1277, 2025.

VITÓRIO, D.; SOUZA, E.; SANTOS, J. A. dos; CARVALHO, A. C. P. d. L. F. de; OLIVEIRA, A. L. I.; SILVA, N. F. F. da. Bm25 x vila sésamo: avaliando modelos sentence-bert para recuperação de informação no cenário legislativo brasileiro. *Linguamática*, v. 17, n. 1, p. 17–33, Jun. 2025. Available at: <https://linguamatica.com/index.php/linguamatica/article/view/474>.

WANG, J.; HUANG, J. X.; TU, X.; WANG, J.; HUANG, A. J.; LASKAR, M. T. R.; BHUIYAN, A. Utilizing BERT for information retrieval: Survey, applications, resources, and challenges. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 56, n. 7, Apr. 2024. ISSN 0360-0300.

WANG, W.; BAO, H.; HUANG, S.; DONG, L.; WEI, F. MiniLMv2: Multi-head self-attention relation distillation for compressing pretrained transformers. In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. [S.l.]: Association for Computational Linguistics, 2021. p. 2140–2151.

WANG, W.; WEI, F.; DONG, L.; BAO, H.; YANG, N.; ZHOU, M. MiniLM: deep self-attention distillation for task-agnostic compression of pre-trained transformers. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2020. (NIPS '20).

WOLF, T.; DEBUT, L.; SANH, V.; CHAUMOND, J.; DELANGUE, C.; MOI, A.; CISTAC, P.; RAULT, T.; LOUF, R.; FUNTOWICZ, M.; DAVISON, J.; SHLEIFER, S.; PLATEN, P. von; MA, C.; JERNITE, Y.; PLU, J.; XU, C.; SCAO, T. L.; GUGGER, S.; DRAME, M.; LHOEST, Q.; RUSH, A. Transformers: State-of-the-art natural language processing. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. [S.l.]: Association for Computational Linguistics, 2020. p. 38–45.

YIN, P.-Y.; BHANU, B.; CHANG, K.-C.; DONG, A. Improving retrieval performance by long-term relevance information. In: *Pattern Recognition, International Conference on*. [S.l.: s.n.], 2002. v. 3, p. 533–536.

ZHANG, E.; ZHANG, Y. Average precision. In: ______. *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. p. 192–193.

ZHANG, H.; CORMACK, G. V.; GROSSMAN, M. R.; SMUCKER, M. D. Evaluating sentence-level relevance feedback for high-recall information retrieval. *Information Retrieval Journal*, Springer, v. 23, n. 1, p. 1–26, 2020.

APPENDIX A – SUMMARY OF THE EXPERIMENTAL CONFIGURATIONS AND HOW THEY ARE REFERENCED

IR algorithm	Preprocessing	Similar queries search	Reference
BM25L	with	preprocessed cosine	BM25L_PRE_PRE
BM25L	with	cosine	BM25L_PRE_NP
BM25L	with	BERTimbau	BM25L_PRE_BERTimbau
BM25L	with	Legal-BERTimbau	BM25L_PRE_LegalBERTimbau
BM25L	with	JurisBERT	BM25L_PRE_JurisBERT
BM25L	with	BERTimbauLaw	BM25L_PRE_BERTimbauLaw
BM25L	with	LegalBert-pt	BM25L_PRE_LegalBertpt
BM25L	with	LaBSE	BM25L_PRE_LaBSE
BM25L	with	Multilingual MPNet	BM25L_PRE_MPNet
BM25L	with	Multilingual MiniLM	BM25L_PRE_MiniLM
BM25L	with	FT BERTimbau	BM25L_PRE_FTBERTimbau
BM25L	with	FT LegalBert-pt	BM25L_PRE_FTLegalBertpt
BM25L	without	preprocessed cosine	BM25L_NP_PRE
BM25L	without	cosine	BM25L_NP_NP
BM25L	without	BERTimbau	BM25L_NP_BERTimbau
BM25L	without	Legal-BERTimbau	BM25L_NP_LegalBERTimbau
BM25L	without	JurisBERT	BM25L_NP_JurisBERT
BM25L	without	BERTimbauLaw	BM25L_NP_BERTimbauLaw
BM25L	without	LegalBert-pt	BM25L_NP_LegalBertpt
BM25L	without	LaBSE	BM25L_NP_LaBSE
BM25L	without	Multilingual MPNet	BM25L_NP_MPNet
BM25L	without	Multilingual MiniLM	BM25L_NP_MiniLM
BM25L	without	FT BERTimbau	BM25L_NP_FTBERTimbau
BM25L	without	FT LegalBert-pt	BM25L_NP_FTLegalBertpt
Okapi BM25	with	preprocessed cosine	OkapiBM25_PRE_PRE
Okapi BM25	with	cosine	OkapiBM25_PRE_NP
Okapi BM25	with	BERTimbau	OkapiBM25_PRE_BERTimbau
Okapi BM25	with	Legal-BERTimbau	OkapiBM25_PRE_LegalBERTimbau

IR algorithm	Preprocessing	Similar queries search	Reference
Okapi BM25	with	JurisBERT	OkapiBM25_PRE_JurisBERT
Okapi BM25	with	BERTimbauLaw	OkapiBM25_PRE_BERTimbauLaw
Okapi BM25	with	LegalBert-pt	OkapiBM25_PRE_LegalBertpt
Okapi BM25	with	LaBSE	OkapiBM25_PRE_LaBSE
Okapi BM25	with	Multilingual MPNet	OkapiBM25_PRE_MPNet
Okapi BM25	with	Multilingual MiniLM	OkapiBM25_PRE_MiniLM
Okapi BM25	with	FT BERTimbau	OkapiBM25_PRE_FTBERTimbau
Okapi BM25	with	FT LegalBert-pt	OkapiBM25_PRE_FTLegalBertpt
Okapi BM25	without	preprocessed cosine	OkapiBM25_NP_PRE
Okapi BM25	without	cosine	OkapiBM25_NP_NP
Okapi BM25	without	BERTimbau	OkapiBM25_NP_BERTimbau
Okapi BM25	without	Legal-BERTimbau	OkapiBM25_NP_LegalBERTimbau
Okapi BM25	without	JurisBERT	OkapiBM25_NP_JurisBERT
Okapi BM25	without	BERTimbauLaw	OkapiBM25_NP_BERTimbauLaw
Okapi BM25	without	LegalBert-pt	OkapiBM25_NP_LegalBertpt
Okapi BM25	without	LaBSE	OkapiBM25_NP_LaBSE
Okapi BM25	without	Multilingual MPNet	OkapiBM25_NP_MPNet
Okapi BM25	without	Multilingual MiniLM	OkapiBM25_NP_MiniLM
Okapi BM25	without	FT BERTimbau	OkapiBM25_NP_FTBERTimbau
Okapi BM25	without	FT LegalBert-pt	OkapiBM25_NP_FTLegalBertpt
BERTimbau	_	preprocessed cosine	BERTimbau_PRE
BERTimbau	_	cosine	BERTimbau_NP
BERTimbau	_	BERTimbau	BERTimbau_BERTimbau
BERTimbau	_	Legal-BERTimbau	BERTimbau_LegalBERTimbau
BERTimbau	_	JurisBERT	BERTimbau_JurisBERT
BERTimbau	_	BERTimbauLaw	BERTimbau_BERTimbauLaw
BERTimbau	_	LegalBert-pt	BERTimbau_LegalBertpt
BERTimbau	_	LaBSE	BERTimbau_LaBSE
BERTimbau	_	Multilingual MPNet	BERTimbau_MPNet
BERTimbau	_	Multilingual MiniLM	BERTimbau_MiniLM
BERTimbau	_	FT BERTimbau	BERTimbau_FTBERTimbau

IR algorithm	Preprocessing	Similar queries search	Reference
BERTimbau	_	FT LegalBert-pt	BERTimbau_FTLegalBertpt
Legal-BERTimbau	_	preprocessed cosine	LegalBERTimbau_PRE
Legal-BERTimbau	_	cosine	LegalBERTimbau_NP
Legal-BERTimbau	_	BERTimbau	LegalBERTimbau_BERTimbau
Legal-BERTimbau	_	Legal-BERTimbau	LegalBERTimbau_LegalBERTimbau
Legal-BERTimbau	_	JurisBERT	LegalBERTimbau_JurisBERT
Legal-BERTimbau	_	BERTimbauLaw	LegalBERTimbau_BERTimbauLaw
Legal-BERTimbau	_	LegalBert-pt	LegalBERTimbau_LegalBertpt
Legal-BERTimbau	_	LaBSE	LegalBERTimbau_LaBSE
Legal-BERTimbau	_	Multilingual MPNet	LegalBERTimbau_MPNet
Legal-BERTimbau	_	Multilingual MiniLM	LegalBERTimbau_MiniLM
Legal-BERTimbau	_	FT BERTimbau	LegalBERTimbau_FTBERTimbau
Legal-BERTimbau	_	FT LegalBert-pt	LegalBERTimbau_FTLegalBertpt
JurisBERT	_	preprocessed cosine	JurisBERT_PRE
JurisBERT	_	cosine	JurisBERT_NP
JurisBERT	_	BERTimbau	JurisBERT_BERTimbau
JurisBERT	_	Legal-BERTimbau	JurisBERT_LegalBERTimbau
JurisBERT	_	JurisBERT	JurisBERT_JurisBERT
JurisBERT	_	BERTimbauLaw	JurisBERT_BERTimbauLaw
JurisBERT	_	LegalBert-pt	JurisBERT_LegalBertpt
JurisBERT	_	LaBSE	JurisBERT_LaBSE
JurisBERT	_	Multilingual MPNet	JurisBERT_MPNet
JurisBERT	_	Multilingual MiniLM	JurisBERT_MiniLM
JurisBERT	_	FT BERTimbau	JurisBERT_FTBERTimbau
JurisBERT	_	FT LegalBert-pt	JurisBERT_FTLegalBertpt
BERTimbauLaw	_	preprocessed cosine	BERTimbauLaw_PRE
BERTimbauLaw	_	cosine	BERTimbauLaw_NP
BERTimbauLaw	_	BERTimbau	BERTimbauLaw_BERTimbau
BERTimbauLaw	_	Legal-BERTimbau	BERTimbauLaw_LegalBERTimbau
BERTimbauLaw	_	JurisBERT	BERTimbauLaw_JurisBERT
BERTimbauLaw	_	BERTimbauLaw	BERTimbauLaw_BERTimbauLaw

IR algorithm	Preprocessing	Similar queries search	Reference
BERTimbauLaw	_	LegalBert-pt	BERTimbauLaw_LegalBertpt
BERTimbauLaw	_	LaBSE	BERTimbauLaw_LaBSE
BERTimbauLaw	_	Multilingual MPNet	BERTimbauLaw_MPNet
BERTimbauLaw	_	Multilingual MiniLM	BERTimbauLaw_MiniLM
BERTimbauLaw	_	FT BERTimbau	BERTimbauLaw_FTBERTimbau
BERTimbauLaw	_	FT LegalBert-pt	BERTimbauLaw_FTLegalBertpt
LegalBert-pt	_	preprocessed cosine	LegalBertpt_PRE
LegalBert-pt	_	cosine	LegalBertpt_NP
LegalBert-pt	_	BERTimbau	LegalBertpt_BERTimbau
LegalBert-pt	_	Legal-BERTimbau	LegalBertpt_LegalBERTimbau
LegalBert-pt	_	JurisBERT	LegalBertpt_JurisBERT
LegalBert-pt	_	BERTimbauLaw	LegalBertpt_BERTimbauLaw
LegalBert-pt	_	LegalBert-pt	LegalBertpt_LegalBertpt
LegalBert-pt	_	LaBSE	LegalBertpt_LaBSE
LegalBert-pt	_	Multilingual MPNet	LegalBertpt_MPNet
LegalBert-pt	_	Multilingual MiniLM	LegalBertpt_MiniLM
LegalBert-pt	_	FT BERTimbau	LegalBertpt_FTBERTimbau
LegalBert-pt	_	FT LegalBert-pt	LegalBertpt_FTLegalBertpt
LaBSE	_	preprocessed cosine	LaBSE_PRE
LaBSE	_	cosine	LaBSE_NP
LaBSE	_	BERTimbau	LaBSE_BERTimbau
LaBSE	_	Legal-BERTimbau	LaBSE_LegalBERTimbau
LaBSE	_	JurisBERT	LaBSE_JurisBERT
LaBSE	_	BERTimbauLaw	LaBSE_BERTimbauLaw
LaBSE	_	LegalBert-pt	LaBSE_LegalBertpt
LaBSE	_	LaBSE	LaBSE_LaBSE
LaBSE	_	Multilingual MPNet	LaBSE_MPNet
LaBSE	_	Multilingual MiniLM	LaBSE_MiniLM
LaBSE	_	FT BERTimbau	LaBSE_FTBERTimbau
LaBSE	_	FT LegalBert-pt	LaBSE_FTLegalBertpt
Multilingual MPNet	_	preprocessed cosine	MPNet_PRE

IR algorithm	Preprocessing	Similar queries search	Reference
Multilingual MPNet	_	cosine	MPNet_NP
Multilingual MPNet	_	BERTimbau	MPNet_BERTimbau
Multilingual MPNet	_	Legal-BERTimbau	MPNet_LegalBERTimbau
Multilingual MPNet	_	JurisBERT	MPNet_JurisBERT
Multilingual MPNet	_	BERTimbauLaw	MPNet_BERTimbauLaw
Multilingual MPNet	_	LegalBert-pt	MPNet_LegalBertpt
Multilingual MPNet	_	LaBSE	MPNet_LaBSE
Multilingual MPNet	_	Multilingual MPNet	MPNet_MPNet
Multilingual MPNet	_	Multilingual MiniLM	MPNet_MiniLM
Multilingual MPNet	_	FT BERTimbau	MPNet_FTBERTimbau
Multilingual MPNet	_	FT LegalBert-pt	MPNet_FTLegalBertpt
Multilingual MiniLM	_	preprocessed cosine	MiniLM_PRE
Multilingual MiniLM	_	cosine	MiniLM_NP
Multilingual MiniLM	_	BERTimbau	MiniLM_BERTimbau
Multilingual MiniLM	_	Legal-BERTimbau	MiniLM_LegalBERTimbau
Multilingual MiniLM	_	JurisBERT	MiniLM_JurisBERT
Multilingual MiniLM	_	BERTimbauLaw	MiniLM_BERTimbauLaw
Multilingual MiniLM	_	LegalBert-pt	MiniLM_LegalBertpt
Multilingual MiniLM	_	LaBSE	MiniLM_LaBSE
Multilingual MiniLM	_	Multilingual MPNet	MiniLM_MPNet
Multilingual MiniLM	_	Multilingual MiniLM	MiniLM_MiniLM
Multilingual MiniLM	_	FT BERTimbau	MiniLM_FTBERTimbau
Multilingual MiniLM	_	FT LegalBert-pt	MiniLM_FTLegalBertpt
FT BERTimbau	_	preprocessed cosine	FTBERTimbau_PRE
FT BERTimbau	_	cosine	FTBERTimbau_NP
FT BERTimbau	_	BERTimbau	FTBERTimbau_BERTimbau
FT BERTimbau	_	Legal-BERTimbau	FTBERTimbau_LegalBERTimbau
FT BERTimbau	_	JurisBERT	FTBERTimbau_JurisBERT
FT BERTimbau	_	BERTimbauLaw	FTBERTimbau_BERTimbauLaw
FT BERTimbau	_	LegalBert-pt	FTBERTimbau_LegalBertpt
FT BERTimbau	_	LaBSE	FTBERTimbau_LaBSE

IR algorithm	Preprocessing	Similar queries search	Reference
FT BERTimbau	_	Multilingual MPNet	FTBERTimbau_MPNet
FT BERTimbau	_	Multilingual MiniLM	FTBERTimbau_MiniLM
FT BERTimbau	_	FT BERTimbau	FTBERTimbau_FTBERTimbau
FT BERTimbau	_	FT LegalBert-pt	FTBERTimbau_FTLegalBertpt
FT LegalBert-pt	_	preprocessed cosine	FTLegalBertpt_PRE
FT LegalBert-pt	_	cosine	FTLegalBertpt_NP
FT LegalBert-pt	_	BERTimbau	FTLegalBertpt_BERTimbau
FT LegalBert-pt	_	Legal-BERTimbau	FTLegalBertpt_LegalBERTimbau
FT LegalBert-pt	_	JurisBERT	FTLegalBertpt_JurisBERT
FT LegalBert-pt	_	BERTimbauLaw	FTLegalBertpt_BERTimbauLaw
FT LegalBert-pt	_	LegalBert-pt	FTLegalBertpt_LegalBertpt
FT LegalBert-pt	_	LaBSE	FTLegalBertpt_LaBSE
FT LegalBert-pt	_	Multilingual MPNet	FTLegalBertpt_MPNet
FT LegalBert-pt	_	Multilingual MiniLM	FTLegalBertpt_MiniLM
FT LegalBert-pt	_	FT BERTimbau	FTLegalBertpt_FTBERTimbau
FT LegalBert-pt	_	FT LegalBert-pt	FTLegalBertpt_FTLegalBertpt

APPENDIX B – PARAMETERS VALUES SELECTED FOR EACH CONFIGURATION USING ULYSSES-RFCORPUS

Configuration	**cut**	δ	**Configuration**	**cut**	δ
BM25L_PRE_PRE	0.5	0.1	BM25L_NP_PRE	0.5	0.5
BM25L_PRE_NP	0.5	0.1	BM25L_NP_NP	0.5	0.5
BM25L_PRE_BERTimbau	0.9	0.1	BM25L_NP_BERTimbau	0.9	0.1
BM25L_PRE_LegalBERTimbau	0.6	0.1	BM25L_NP_LegalBERTimbau	0.4	0.1
BM25L_PRE_JurisBERT	0.8	0.1	BM25L_NP_JurisBERT	0.7	0.1
BM25L_PRE_BERTimbauLaw	0.8	0.1	BM25L_NP_BERTimbauLaw	0.7	0.1
BM25L_PRE_LegalBertpt	0.8	0.1	BM25L_NP_LegalBertpt	0.7	0.1
BM25L_PRE_LaBSE	0.7	0.1	BM25L_NP_LaBSE	0.4	0.1
BM25L_PRE_MPNet	0.8	0.1	BM25L_NP_MPNet	0.8	0.5
BM25L_PRE_MiniLM	0.7	0.1	BM25L_NP_MiniLM	0.8	0.5
BM25L_PRE_FTBERTimbau	0.9	0.1	BM25L_NP_FTBERTimbau	0.8	0.1
BM25L_PRE_FTLegalBertpt	0.9	0.1	BM25L_NP_FTLegalBertpt	0.8	0.1
OkapiBM25_PRE_PRE	0.3	0.1	OkapiBM25_NP_PRE	0.5	0.5
OkapiBM25_PRE_NP	0.5	0.1	OkapiBM25_NP_NP	0.5	0.5
OkapiBM25_PRE_BERTimbau	0.9	0.1	OkapiBM25_NP_BERTimbau	0.9	0.1
OkapiBM25_PRE_LegalBERTimbau	0.6	0.1	OkapiBM25_NP_LegalBERTimbau	0.6	0.1
OkapiBM25_PRE_JurisBERT	0.8	0.1	OkapiBM25_NP_JurisBERT	0.7	0.1
OkapiBM25_PRE_BERTimbauLaw	0.7	0.1	OkapiBM25_NP_BERTimbauLaw	0.8	0.1
OkapiBM25_PRE_LegalBertpt	0.8	0.1	OkapiBM25_NP_LegalBertpt	0.8	0.1
OkapiBM25_PRE_LaBSE	0.7	0.1	OkapiBM25_NP_LaBSE	0.6	0.1
OkapiBM25_PRE_MPNet	0.8	0.1	OkapiBM25_NP_MPNet	0.7	0.1
OkapiBM25_PRE_MiniLM	0.7	0.1	OkapiBM25_NP_MiniLM	0.7	0.1
OkapiBM25_PRE_FTBERTimbau	0.9	0.1	OkapiBM25_NP_FTBERTimbau	0.9	0.1
OkapiBM25_PRE_FTLegalBertpt	0.8	0.1	OkapiBM25_NP_FTLegalBertpt	0.8	0.1
BERTimbau_PRE	0.1	2.0	LegalBERTimbau_PRE	0.2	0.5
BERTimbau_NP	0.2	2.0	LegalBERTimbau_NP	0.2	0.5
BERTimbau_BERTimbau	0.9	2.0	LegalBERTimbau_BERTimbau	0.9	0.1
BERTimbau_LegalBERTimbau	0.6	0.5	LegalBERTimbau_LegalBERTimbau	0.6	0.1

Configuration	cut	δ	Configuration	cut	δ
BERTimbau_JurisBERT	0.8	2.0	LegalBERTimbau_JurisBERT	0.6	0.1
BERTimbau_BERTimbauLaw	0.8	1.0	LegalBERTimbau_BERTimbauLaw	0.7	0.1
BERTimbau_LegalBertpt	0.8	0.5	LegalBERTimbau_LegalBertpt	0.6	0.1
BERTimbau_LaBSE	0.6	0.5	LegalBERTimbau_LaBSE	0.5	0.1
BERTimbau_MPNet	0.7	0.5	LegalBERTimbau_MPNet	0.8	2.0
BERTimbau_MiniLM	0.6	0.5	LegalBERTimbau_MiniLM	0.6	0.1
BERTimbau_FTBERTimbau	0.8	2.0	LegalBERTimbau_FTBERTimbau	0.8	0.1
BERTimbau_FTLegalBertpt	0.8	0.5	LegalBERTimbau_FTLegalBertpt	0.8	0.1
JurisBERT_PRE	0.2	0.5	BERTimbauLaw_PRE	0.1	0.5
JurisBERT_NP	0.2	0.5	BERTimbauLaw_NP	0.3	0.5
JurisBERT_BERTimbau	0.9	0.1	BERTimbauLaw_BERTimbau	0.9	0.1
JurisBERT_LegalBERTimbau	0.5	0.1	BERTimbauLaw_LegalBERTimbau	0.6	0.1
JurisBERT_JurisBERT	0.8	0.1	BERTimbauLaw_JurisBERT	0.7	0.1
JurisBERT_BERTimbauLaw	0.8	0.1	BERTimbauLaw_BERTimbauLaw	0.8	0.1
JurisBERT_LegalBertpt	0.8	0.1	BERTimbauLaw_LegalBertpt	0.8	0.1
JurisBERT_LaBSE	0.7	2.0	BERTimbauLaw_LaBSE	0.7	0.1
JurisBERT_MPNet	0.6	0.1	BERTimbauLaw_MPNet	0.8	0.5
JurisBERT_MiniLM	0.8	0.5	BERTimbauLaw_MiniLM	0.7	0.1
JurisBERT_FTBERTimbau	0.9	1.0	BERTimbauLaw_FTBERTimbau	0.9	0.1
JurisBERT_FTLegalBertpt	0.8	0.1	BERTimbauLaw_FTLegalBertpt	0.8	0.1
LegalBertpt_PRE	0.2	2.0	LaBSE_PRE	0.1	0.5
LegalBertpt_NP	0.3	2.0	LaBSE_NP	0.2	0.5
LegalBertpt_BERTimbau	0.9	0.1	LaBSE_BERTimbau	0.9	0.1
LegalBertpt_LegalBERTimbau	0.6	0.5	LaBSE_LegalBERTimbau	0.6	0.5
LegalBertpt_JurisBERT	0.8	2.0	LaBSE_JurisBERT	0.7	0.1
LegalBertpt_BERTimbauLaw	0.8	1.0	LaBSE_BERTimbauLaw	0.7	0.1
LegalBertpt_LegalBertpt	0.8	0.1	LaBSE_LegalBertpt	0.8	0.1
LegalBertpt_LaBSE	0.7	1.0	LaBSE_LaBSE	0.6	0.1
LegalBertpt_MPNet	0.7	1.0	LaBSE_MPNet	0.7	0.1
LegalBertpt_MiniLM	0.7	0.1	LaBSE_MiniLM	0.6	0.1
LegalBertpt_FTBERTimbau	0.9	1.0	LaBSE_FTBERTimbau	0.8	0.1

Configuration	cut	δ	Configuration	cut	δ
LegalBertpt_FTLegalBertpt	0.8	0.1	LaBSE_FTLegalBertpt	0.8	0.1
MPNet_PRE	0.2	0.5	MiniLM_PRE	0.2	0.5
MPNet_NP	0.3	0.5	MiniLM_NP	0.2	0.5
MPNet_BERTimbau	0.9	0.1	MiniLM_BERTimbau	0.9	0.1
MPNet_LegalBERTimbau	0.6	0.1	MiniLM_LegalBERTimbau	0.7	0.5
MPNet_JurisBERT	0.8	0.1	MiniLM_JurisBERT	0.7	0.1
MPNet_BERTimbauLaw	0.8	0.1	MiniLM_BERTimbauLaw	0.8	0.1
MPNet_LegalBertpt	0.8	0.1	MiniLM_LegalBertpt	0.8	0.1
MPNet_LabSE	0.7	0.1	MiniLM_LaBSE	0.7	0.5
MPNet_MPNet	0.8	0.5	MiniLM_MPNet	0.8	1.0
MPNet_MiniLM	0.8	0.1	MiniLM_MiniLM	0.8	0.5
MPNet_FTBERTimbau	0.9	0.1	MiniLM_FTBERTimbau	0.9	0.1
MPNet_LegalBertpt	0.9	0.1	MiniLM_LegalBertpt	0.9	1.0
FTBERTimbau_PRE	0.1	0.1	FTLegalBerpt_PRE	0.2	1.0
FTBERTimbau_NP	0.2	0.1	FTLegalBerpt_NP	0.3	1.0
FTBERTimbau_BERTimbau	0.9	0.1	FTLegalBerpt_BERTimbau	0.9	0.1
FTBERTimbau_LegalBERTimbau	0.6	0.1	FTLegalBerpt_LegalBERTimbau	0.6	0.1
FTBERTimbau_JurisBERT	0.8	0.1	FTLegalBerpt_JurisBERT	0.7	0.1
FTBERTimbau_BERTimbauLaw	0.8	0.1	FTLegalBerpt_BERTimbauLaw	0.8	0.1
FTBERTimbau_LegalBertpt	0.8	0.1	FTLegalBerpt_LegalBertpt	0.8	0.1
FTBERTimbau_LaBSE	0.7	0.1	FTLegalBerpt_LaBSE	0.7	0.1
FTBERTimbau_MPNet	0.8	0.1	FTLegalBerpt_MPNet	0.7	1.0
FTBERTimbau_MiniLM	0.8	0.1	FTLegalBerpt_MiniLM	0.8	0.1
FTBERTimbau_FTBERTimbau	0.9	0.1	FTLegalBerpt_FTBERTimbau	0.9	0.5
FTBERTimbau_FTLegalBertpt	0.9	0.5	FTLegalBerpt_FTLegalBerpt	0.9	0.5

APPENDIX C – PARAMETERS VALUES SELECTED FOR EACH CONFIGURATION USING THE PRELIMINARY SEARCH CORPUS

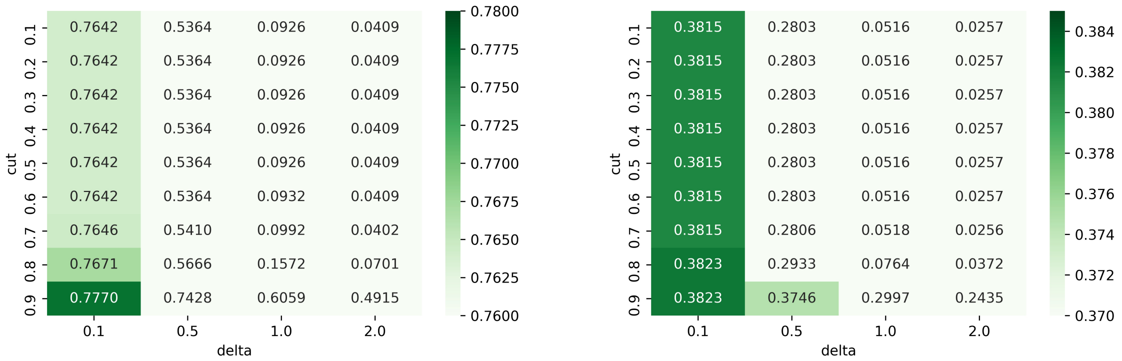
Configuration	cut	δ	Configuration	cut	δ
BM25L_PRE_PRE	0.1	2.0	BM25L_NP_PRE	0.1	2.0
BM25L_PRE_NP	0.1	1.0	BM25L_NP_NP	0.1	2.0
BM25L_PRE_BERTimbau	0.6	0.1	BM25L_NP_BERTimbau	0.9	0.5
BM25L_PRE_LegalBERTimbau	0.6	1.0	BM25L_NP_LegalBERTimbau	0.6	2.0
BM25L_PRE_JurisBERT	0.8	1.0	BM25L_NP_JurisBERT	0.7	0.5
BM25L_PRE_BERTimbauLaw	0.8	1.0	BM25L_NP_BERTimbauLaw	0.7	0.5
BM25L_PRE_LegalBertpt	0.9	0.5	BM25L_NP_LegalBertpt	0.8	0.5
BM25L_PRE_LaBSE	0.6	0.5	BM25L_NP_LaBSE	0.7	2.0
BM25L_PRE_MPNet	0.7	0.5	BM25L_NP_MPNet	0.7	1.0
BM25L_PRE_MiniLM	0.6	0.5	BM25L_NP_MiniLM	0.3	0.5
BM25L_PRE_FTBERTimbau	0.8	0.5	BM25L_NP_FTBERTimbau	0.8	0.5
BM25L_PRE_FTLegalBertpt	0.8	0.5	BM25L_NP_FTLegalBertpt	0.8	0.5
OkapiBM25_PRE_PRE	0.1	2.0	OkapiBM25_NP_PRE	0.1	2.0
OkapiBM25_PRE_NP	0.1	1.0	OkapiBM25_NP_NP	0.3	2.0
OkapiBM25_PRE_BERTimbau	0.7	0.1	OkapiBM25_NP_BERTimbau	0.7	0.1
OkapiBM25_PRE_LegalBERTimbau	0.6	2.0	OkapiBM25_NP_LegalBERTimbau	0.6	1.0
OkapiBM25_PRE_JurisBERT	0.8	1.0	OkapiBM25_NP_JurisBERT	0.8	1.0
OkapiBM25_PRE_BERTimbauLaw	0.8	1.0	OkapiBM25_NP_BERTimbauLaw	0.8	0.5
OkapiBM25_PRE_LegalBertpt	0.9	0.5	OkapiBM25_NP_LegalBertpt	0.5	0.1
OkapiBM25_PRE_LaBSE	0.6	0.5	OkapiBM25_NP_LaBSE	0.7	2.0
OkapiBM25_PRE_MPNet	0.7	0.5	OkapiBM25_NP_MPNet	0.7	0.5
OkapiBM25_PRE_MiniLM	0.6	0.5	OkapiBM25_NP_MiniLM	0.7	0.5
OkapiBM25_PRE_FTBERTimbau	0.8	0.5	OkapiBM25_NP_FTBERTimbau	0.9	0.5
OkapiBM25_PRE_FTLegalBertpt	0.8	0.5	OkapiBM25_NP_FTLegalBertpt	0.9	2.0
BERTimbau_PRE	0.1	2.0	LegalBERTimbau_PRE	0.1	1.0
BERTimbau_NP	0.2	1.0	LegalBERTimbau_NP	0.1	0.5
BERTimbau_BERTimbau	0.9	0.1	LegalBERTimbau_BERTimbau	0.8	0.1
BERTimbau_LegalBERTimbau	0.6	1.0	LegalBERTimbau_LegalBERTimbau	0.6	0.5

Configuration	cut	δ	Configuration	cut	δ
BERTimbau_JurisBERT	0.7	0.5	LegalBERTimbau_JurisBERT	0.6	0.1
BERTimbau_BERTimbauLaw	0.8	2.0	LegalBERTimbau_BERTimbauLaw	0.8	0.5
BERTimbau_LegalBertpt	0.9	2.0	LegalBERTimbau_LegalBertpt	0.7	0.1
BERTimbau_LaBSE	0.6	0.5	LegalBERTimbau_LaBSE	0.6	0.5
BERTimbau_MPNet	0.7	1.0	LegalBERTimbau_MPNet	0.7	0.5
BERTimbau_MiniLM	0.6	0.5	LegalBERTimbau_MiniLM	0.6	0.1
BERTimbau_FTBERTimbau	0.8	2.0	LegalBERTimbau_FTBERTimbau	0.7	0.1
BERTimbau_FTLegalBertpt	0.8	1.0	LegalBERTimbau_FTLegalBertpt	0.8	0.1
JurisBERT_PRE	0.1	2.0	BERTimbauLaw_PRE	0.1	1.0
JurisBERT_NP	0.1	1.0	BERTimbauLaw_NP	0.1	0.5
JurisBERT_BERTimbau	0.8	0.1	BERTimbauLaw_BERTimbau	0.8	0.1
JurisBERT_LegalBERTimbau	0.6	1.0	BERTimbauLaw_LegalBERTimbau	0.6	0.5
JurisBERT_JurisBERT	0.8	1.0	BERTimbauLaw_JurisBERT	0.6	0.1
JurisBERT_BERTimbauLaw	0.8	1.0	BERTimbauLaw_BERTimbauLaw	0.8	0.5
JurisBERT_LegalBertpt	0.9	0.5	BERTimbauLaw_LegalBertpt	0.7	0.1
JurisBERT_LaBSE	0.6	0.5	BERTimbauLaw_LaBSE	0.7	0.5
JurisBERT_MPNet	0.7	0.5	BERTimbauLaw_MPNet	0.7	0.5
JurisBERT_MiniLM	0.6	0.5	BERTimbauLaw_MiniLM	0.6	0.1
JurisBERT_FTBERTimbau	0.9	0.5	BERTimbauLaw_FTBERTimbau	0.9	0.5
JurisBERT_FTLegalBertpt	0.8	1.0	BERTimbauLaw_FTLegalBertpt	0.8	0.1
LegalBertpt_PRE	0.1	0.5	LaBSE_PRE	0.1	1.0
LegalBertpt_NP	0.3	2.0	LaBSE_NP	0.3	2.0
LegalBertpt_BERTimbau	0.9	0.1	LaBSE_BERTimbau	0.9	0.1
LegalBertpt_LegalBERTimbau	0.6	1.0	LaBSE_LegalBERTimbau	0.6	0.5
LegalBertpt_JurisBERT	0.7	0.1	LaBSE_JurisBERT	0.5	0.1
LegalBertpt_BERTimbauLaw	0.8	1.0	LaBSE_BERTimbauLaw	0.8	1.0
LegalBertpt_LegalBertpt	0.9	1.0	LaBSE_LegalBertpt	0.8	0.1
LegalBertpt_LaBSE	0.6	0.1	LaBSE_LaBSE	0.7	0.5
LegalBertpt_MPNet	0.7	1.0	LaBSE_MPNet	0.7	0.5
LegalBertpt_MiniLM	0.6	0.1	LaBSE_MiniLM	0.6	0.1
LegalBertpt_FTBERTimbau	0.9	2.0	LaBSE_FTBERTimbau	0.9	0.5

Configuration	**cut**	δ	**Configuration**	**cut**	δ
LegalBertpt_FTLegalBertpt	0.8	2.0	LaBSE_FTLegalBertpt	0.8	0.1
MPNet_PRE	0.1	1.0	MiniLM_PRE	0.1	2.0
MPNet_NP	0.1	0.5	MiniLM_NP	0.3	2.0
MPNet_BERTimbau	0.6	0.1	MiniLM_BERTimbau	0.8	0.1
MPNet_LegalBERTimbau	0.6	0.5	MiniLM_LegalBERTimbau	0.6	0.5
MPNet_JurisBERT	0.5	0.1	MiniLM_JurisBERT	0.8	1.0
MPNet_BERTimbauLaw	0.8	1.0	MiniLM_BERTimbauLaw	0.8	2.0
MPNet_LegalBertpt	0.5	0.1	MiniLM_LegalBertpt	0.9	0.5
MPNet_LabSE	0.7	0.5	MiniLM_LaBSE	0.6	0.5
MPNet_MPNet	0.8	0.5	MiniLM_MPNet	0.8	1.0
MPNet_MiniLM	0.8	1.0	MiniLM_MiniLM	0.8	0.5
MPNet_FTBERTimbau	0.9	0.5	MiniLM_FTBERTimbau	0.8	0.5
MPNet_FTLegalBertpt	0.9	1.0	MiniLM_FTLegalBertpt	0.9	0.5
FTBERTimbau_PRE	0.1	0.5	FTLegalBertpt_PRE	0.1	1.0
FTBERTimbau_NP	0.1	0.5	FTLegalBertpt_NP	0.1	0.5
FTBERTimbau_BERTimbau	0.8	0.1	FTLegalBertpt_BERTimbau	0.6	0.1
FTBERTimbau_LegalBERTimbau	0.2	0.1	FTLegalBertpt_LegalBERTimbau	0.6	0.5
FTBERTimbau_JurisBERT	0.5	0.1	FTLegalBertpt_JurisBERT	0.5	0.1
FTBERTimbau_BERTimbauLaw	0.3	0.1	FTLegalBertpt_BERTimbauLaw	0.3	0.1
FTBERTimbau_LegalBertpt	0.7	0.1	FTLegalBertpt_LegalBertpt	0.5	0.1
FTBERTimbau_LabSE	0.3	0.1	FTLegalBertpt_LaBSE	0.3	0.1
FTBERTimbau_MPNet	0.4	0.1	FTLegalBertpt_MPNet	0.5	0.1
FTBERTimbau_MiniLM	0.2	0.1	FTLegalBertpt_MiniLM	0.3	0.1
FTBERTimbau_FTBERTimbau	0.7	0.1	FTLegalBertpt_FTBERTimbau	0.5	0.1
FTBERTimbau_FTLegalBertpt	0.8	0.1	FTLegalBertpt_FTLegalBertpt	0.7	0.1

APPENDIX D – PARAMETERS ASSESSMENT FOR BM25L WITH ULYSSES-RFCORPUS AND USING BERT-BASED MODELS TO SEARCH FOR THE SIMILAR QUERIES

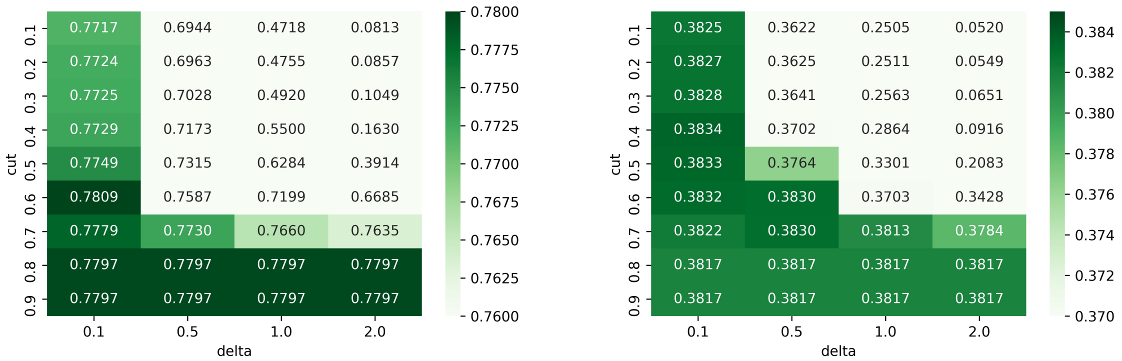
Figure 25 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_BERTimbau (a) and BM25L_NP_BERTimbau (b) with Ulysses-RFCorpus.



(a) BM25L with preprocessing. (b) BM25L without preprocessing.

Source: Created by the author (2025)

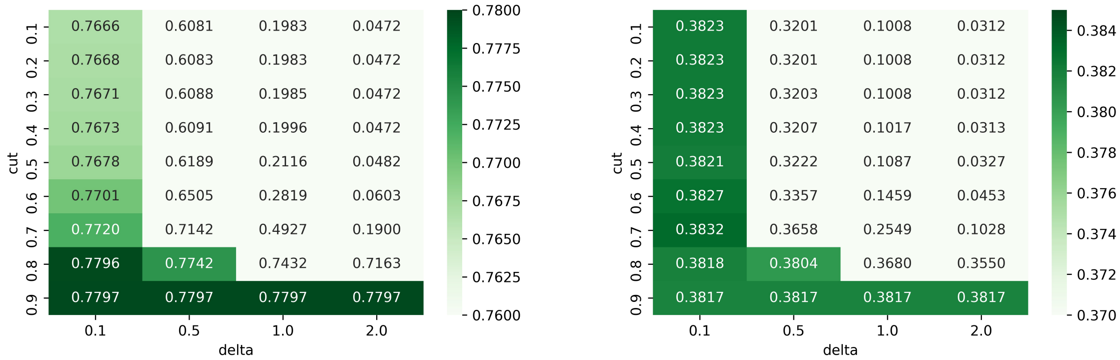
Figure 26 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_LegalBERTimbau (a) and BM25L_NP_LegalBERTimbau (b) with Ulysses-RFCorpus.



(a) BM25L with preprocessing. (b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 27 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_JurisBERT (a) and BM25L_NP_JurisBERT (b) with Ulysses-RFCorpus.

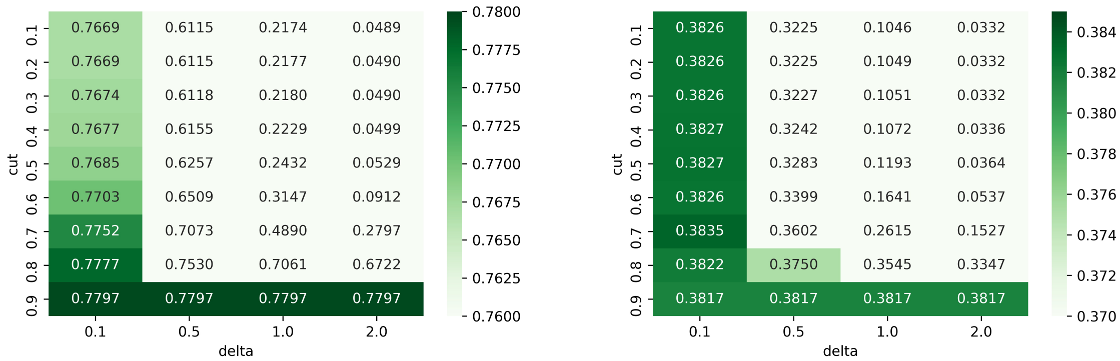


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 28 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_BERTimbauLaw (a) and BM25L_NP_BERTimbauLaw (b) with Ulysses-RFCorpus.

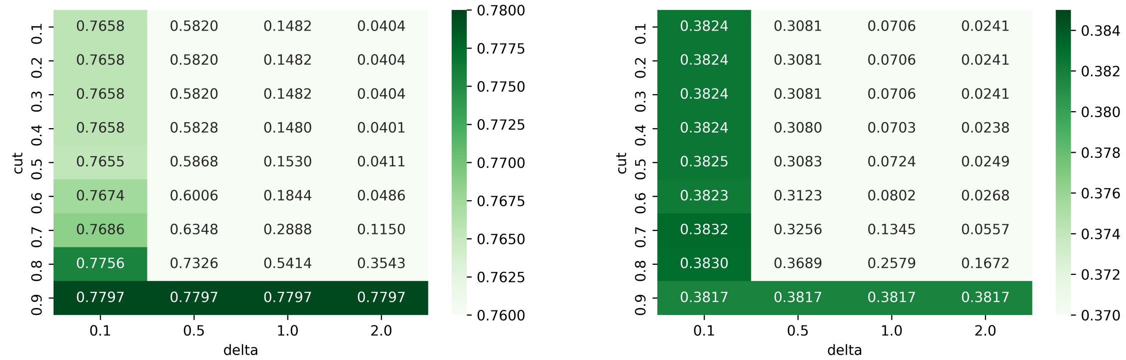


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

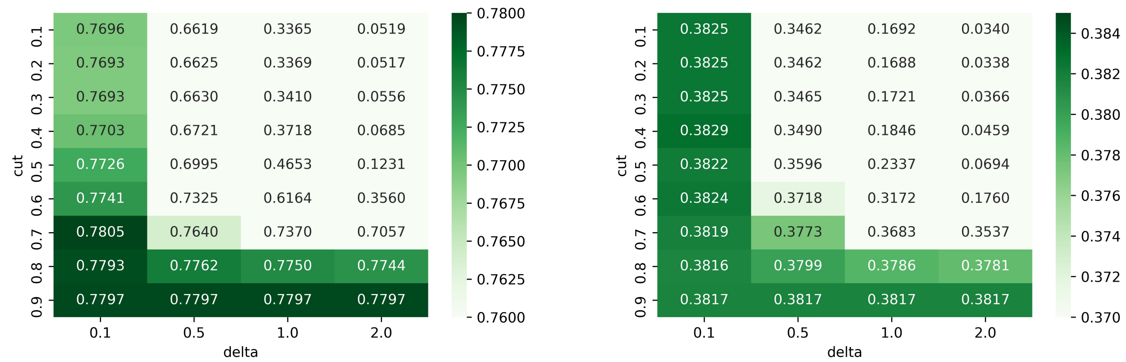
Figure 29 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_LegalBertpt (a) and BM25L_NP_LegalBertpt (b) with Ulysses-RFCorpus.



(a) BM25L with preprocessing. (b) BM25L without preprocessing.

Source: Created by the author (2025)

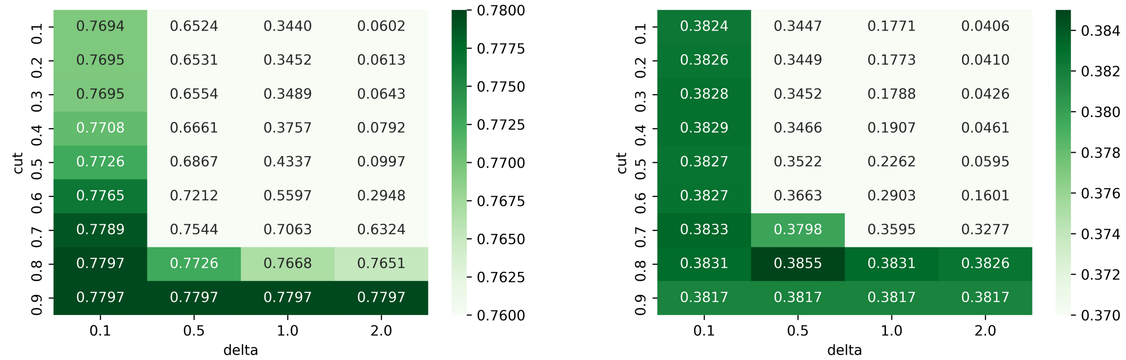
Figure 30 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_LaBSE (a) and BM25L_NP_LaBSE (b) with Ulysses-RFCorpus.



(a) BM25L with preprocessing. (b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 31 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_MPNet (a) and BM25L_NP_MPNet (b) with Ulysses-RFCorpus.

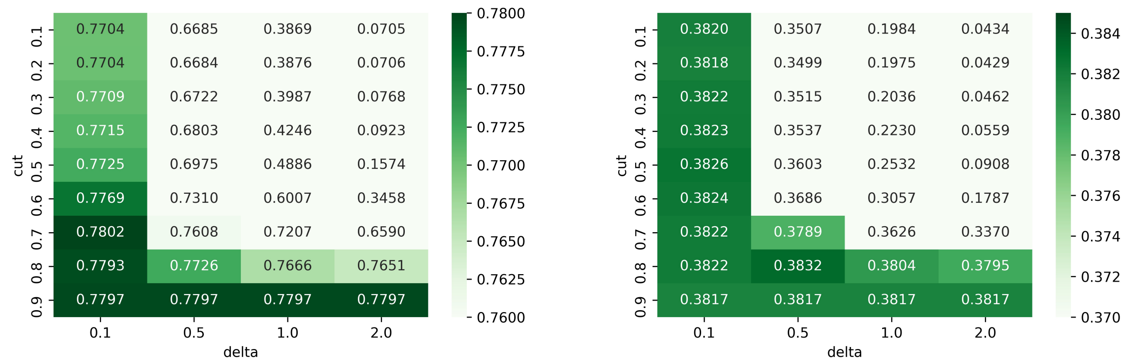


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 32 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_MiniLM (a) and BM25L_NP_MiniLM (b) with Ulysses-RFCorpus.



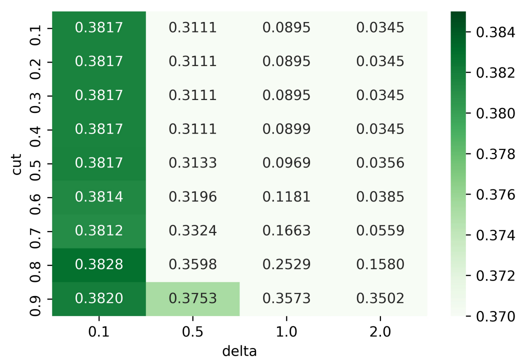
(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 33 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_FTBERTimbau (a) and BM25L_NP_FTBERTimbau (b) with Ulysses-RFCorpus.

(a) BM25L with preprocessing.



(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 34 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_FTLegalBertpt (a) and BM25L_NP_FTLegalBertpt (b) with Ulysses-RFCorpus.

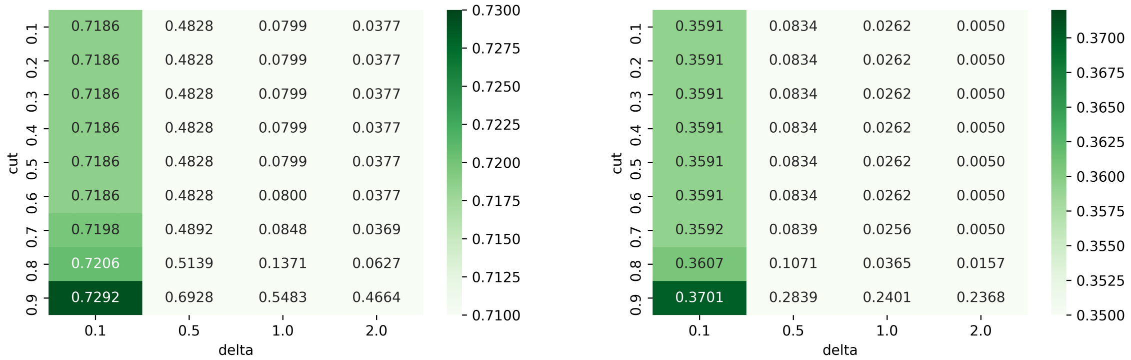
(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

APPENDIX E – PARAMETERS ASSESSMENT FOR OKAPI BM25 WITH ULYSSES-RFCORPUS AND USING BERT-BASED MODELS TO SEARCH FOR THE SIMILAR QUERIES

Figure 35 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_BERTimbau (a) and OkapiBM25_NP_BERTimbau (b) with Ulysses-RFCorpus.

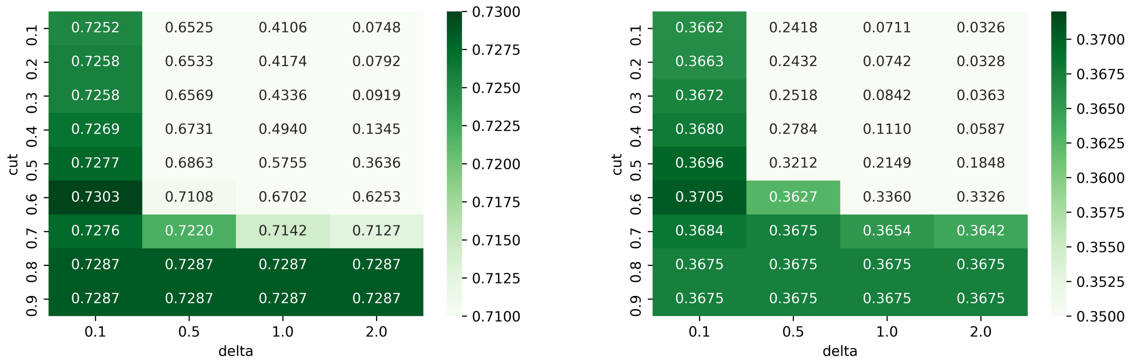


(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 36 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_LegalBERTimbau (a) and OkapiBM25_NP_LegalBERTimbau (b) with Ulysses-RFCorpus.



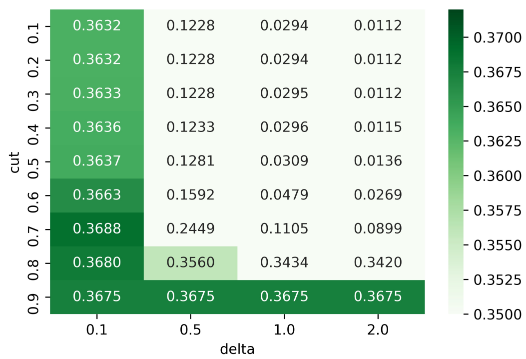
(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 37 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_JurisBERT (a) and OkapiBM25_NP_JurisBERT (b) with Ulysses-RFCorpus.

(a) Okapi BM25 with preprocessing.

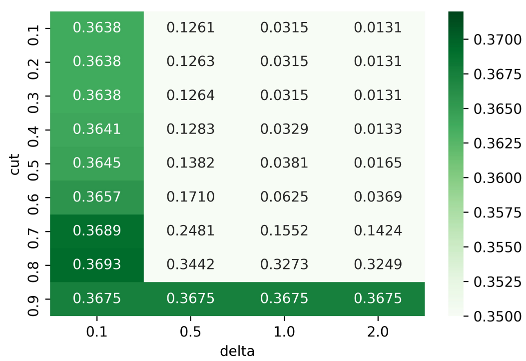


(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 38 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_BERTimbauLaw (a) and OkapiBM25_NP_BERTimbauLaw (b) with Ulysses-RFCorpus.

(a) Okapi BM25 with preprocessing.

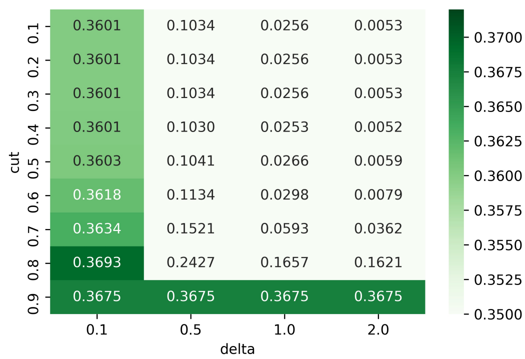


(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 39 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_LegalBertpt (a) and OkapiBM25_NP_LegalBertpt (b) with Ulysses-RFCorpus.

(a) Okapi BM25 with preprocessing.



(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

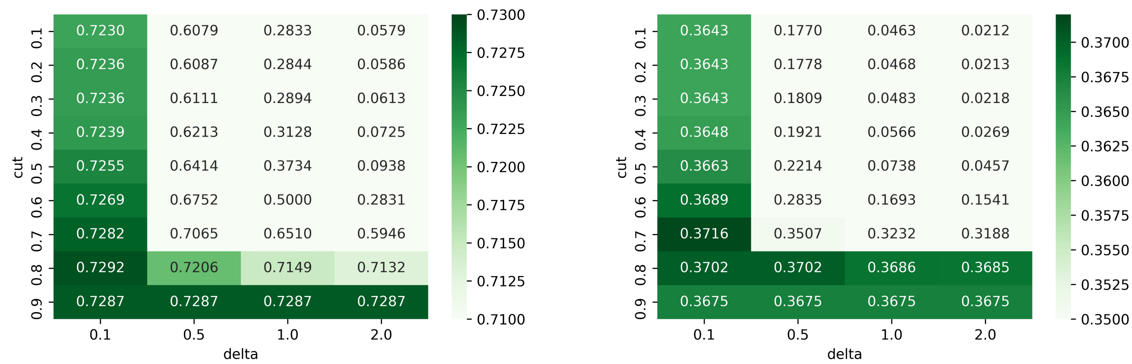
Figure 40 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_LaBSE (a) and OkapiBM25_NP_LaBSE (b) with Ulysses-RFCorpus.

(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

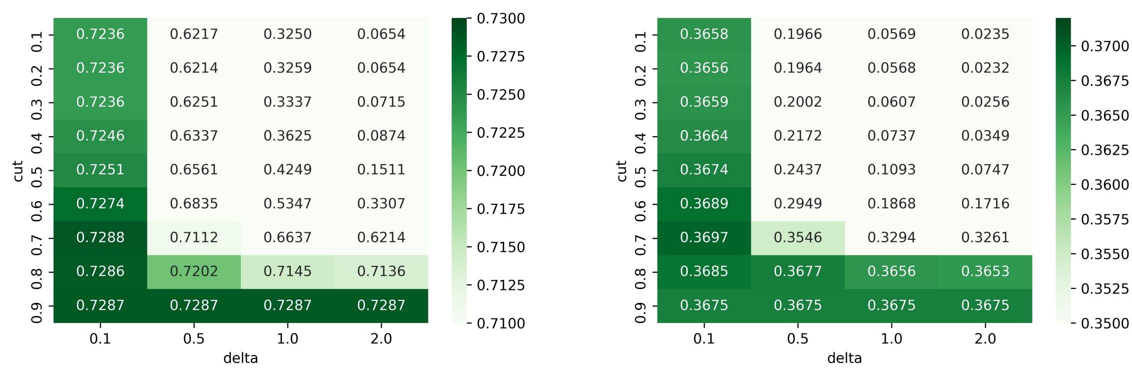
Figure 41 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_MPNet (a) and OkapiBM25_NP_MPNet (b) with Ulysses-RFCorpus.



(a) Okapi BM25 with preprocessing. (b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

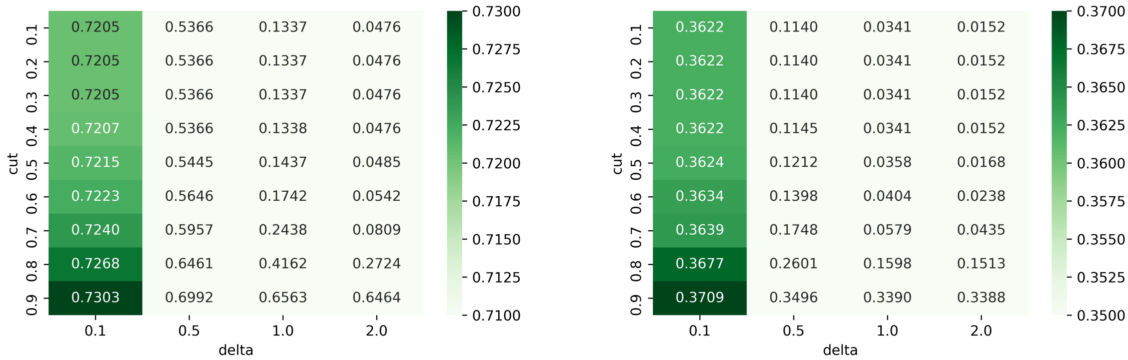
Figure 42 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_MiniLM (a) and OkapiBM25_NP_MiniLM (b) with Ulysses-RFCorpus.



(a) Okapi BM25 with preprocessing. (b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 43 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_FTBERTimbau (a) and OkapiBM25_NP_FTBERTimbau (b) with Ulysses-RFCorpus.

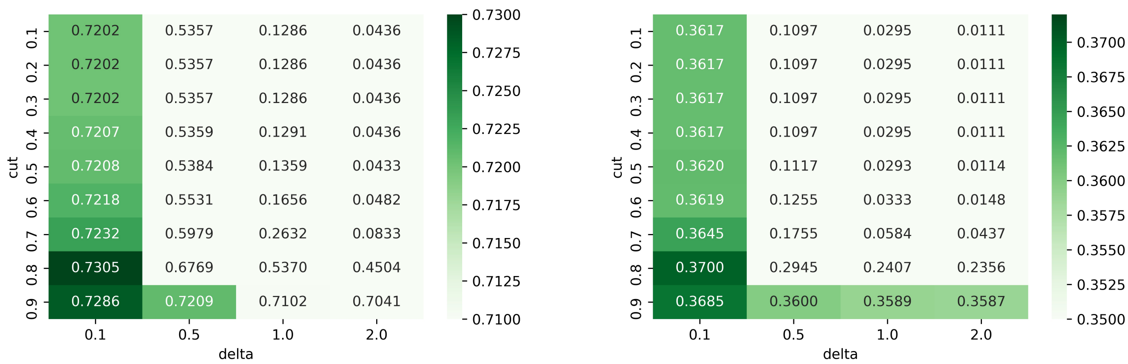


(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 44 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_FTLegalBertpt (a) and OkapiBM25_NP_FTLegalBertpt (b) with Ulysses-RFCorpus.



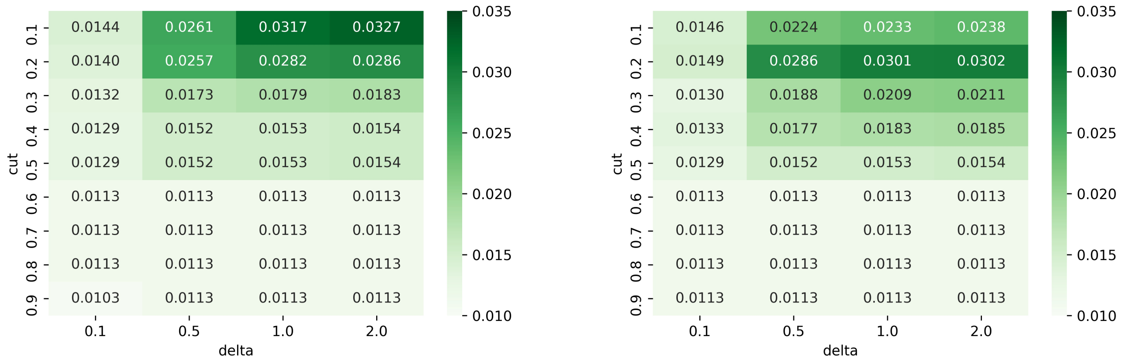
(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

APPENDIX F – PARAMETERS ASSESSMENT FOR BERT-BASED MODELS WITH ULYSSES-RFCORPUS AND USING COSINE TO SEARCH FOR THE SIMILAR QUERIES

Figure 45 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbau_PRE (a) and BERTimbau_NP (b) with Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

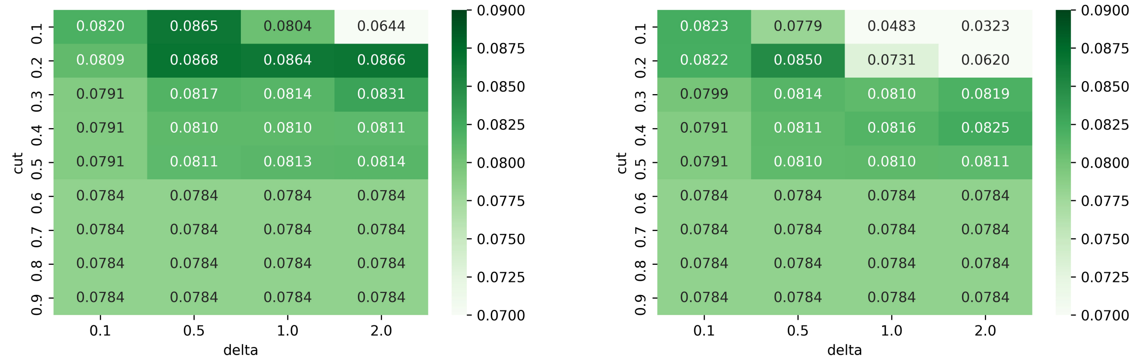
Source: Created by the author (2025)

Figure 46 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBERTimbau_PRE (a) and LegalBERTimbau_NP (b) with Ulysses-RFCorpus.

(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 47 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for JurisBERT_PRE (a) and JurisBERT_NP (b) with Ulysses-RFCorpus.

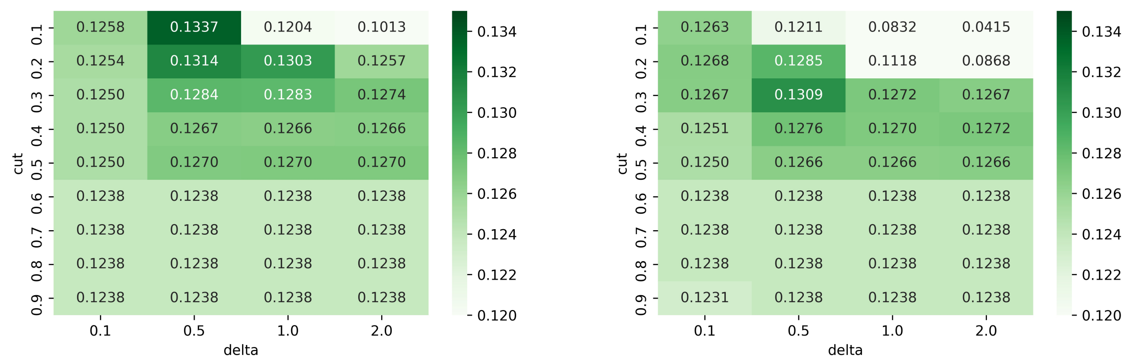


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 48 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbauLaw_PRE (a) and BERTimbauLaw_NP (b) with Ulysses-RFCorpus.

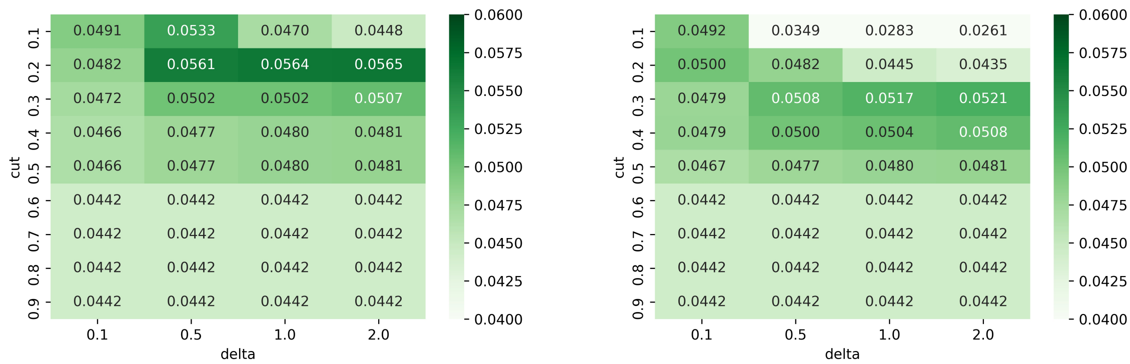


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 49 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBertpt_PRE (a) and LegalBertpt_NP (b) with Ulysses-RFCorpus.

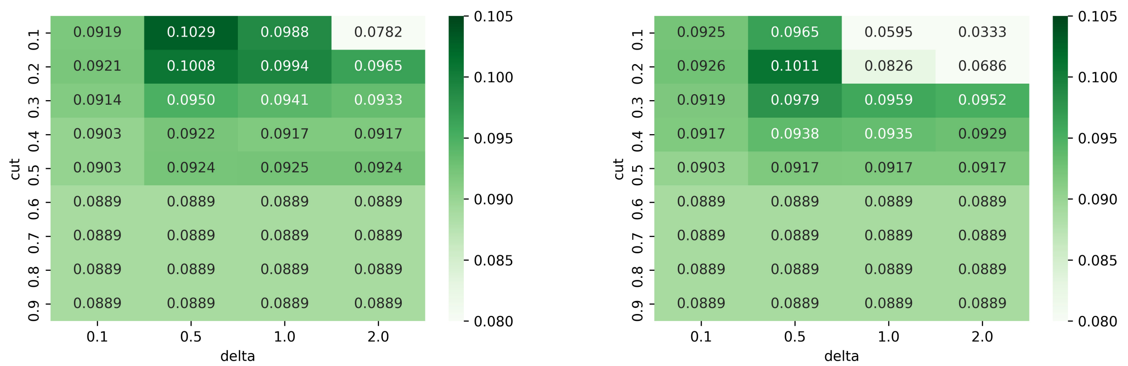


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 50 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LaBSE_PRE (a) and LaBSE_NP (b) with Ulysses-RFCorpus.

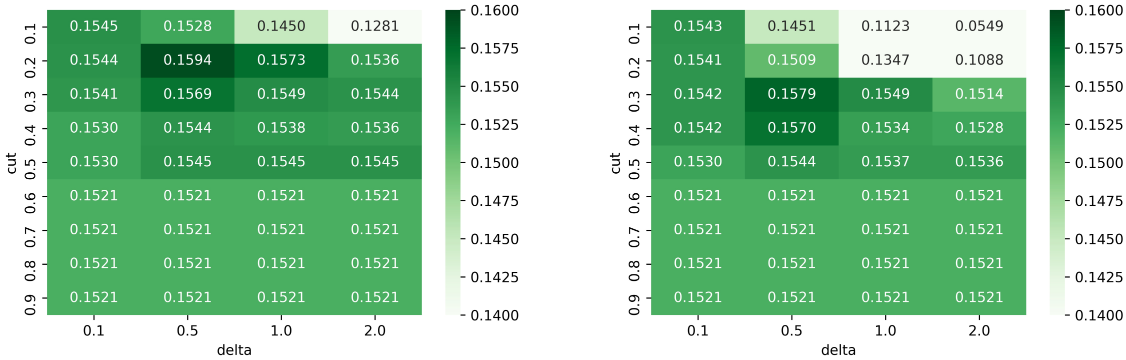


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

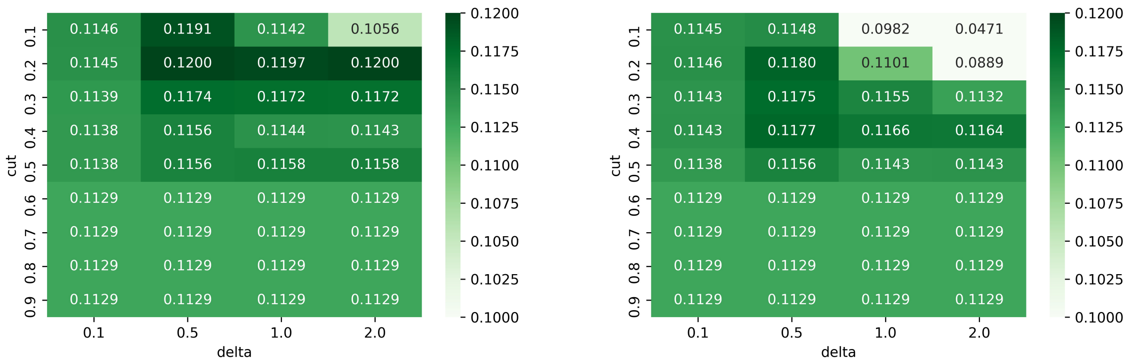
Figure 51 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MPNet_PRE (a) and MPNet_NP (b) with Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

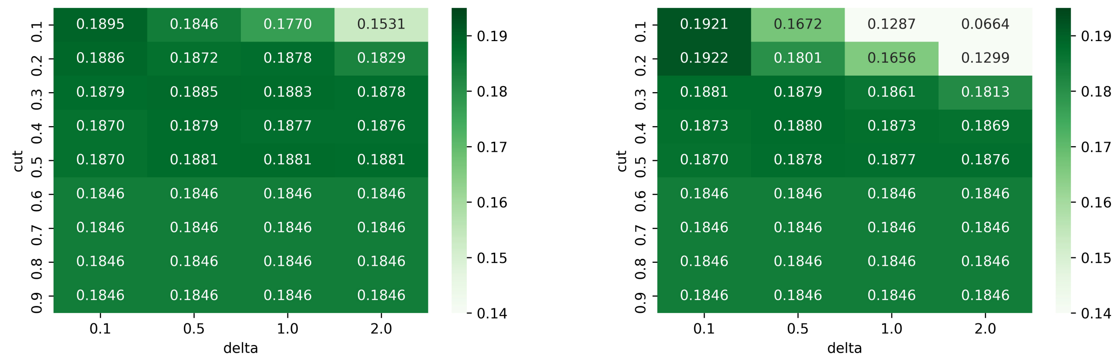
Figure 52 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MiniLM_PRE (a) and MiniLM_NP (b) with Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

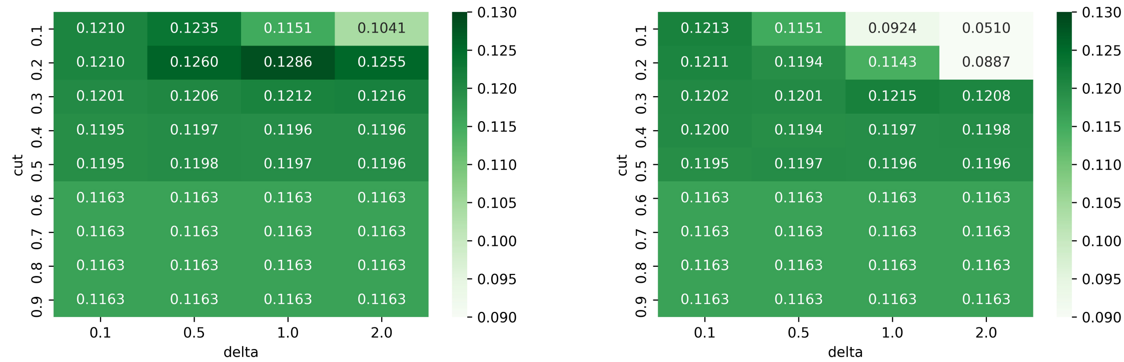
Figure 53 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTBERTimbau_PRE (a) and FTBERTimbau_NP (b) with Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 54 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTLegalBertpt_PRE (a) and FTLegalBertpt_NP (b) with Ulysses-RFCorpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

APPENDIX G – PARAMETERS ASSESSMENT FOR THE BERT-BASED MODELS WITH ULYSSES-RFCORPUS AND USING BERT-BASED MODELS TO SEARCH FOR THE SIMILAR QUERIES

Figure 55 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbau_BERTimbau (a), BERTimbau_LegalBERTimbau (b), BERTimbau_JurisBERT (c), and BERTimbau_BERTimbauLaw (d) with Ulysses-RFCorpus.

(a) BERTimbau.

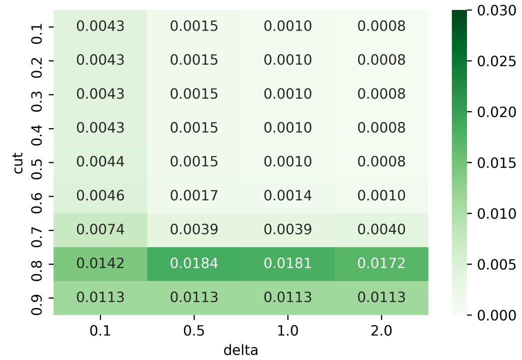
(b) Legal-BERTimbau

(c) JurisBERT.

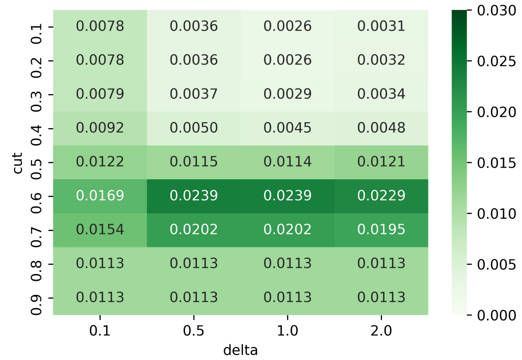
(d) BERTimbauLaw

Source: Created by the author (2025)

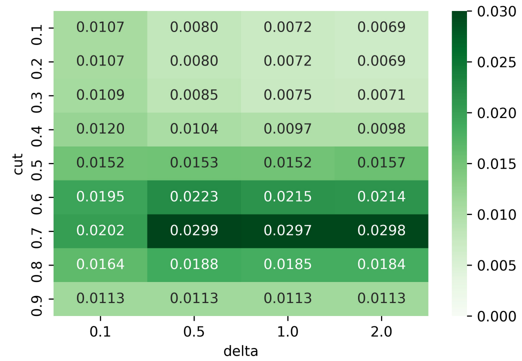
Figure 56 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbau_LegalBertpt (a), BERTimbau_LaBSE (b), BERTimbau_MPNet (c), BERTimbau_MiniLM (d), BERTimbau_FTBERTimbau (e), and BERTimbau_FTLegalBertpt (f) with Ulysses-RFCorpus.



(a) LegalBert-pt.

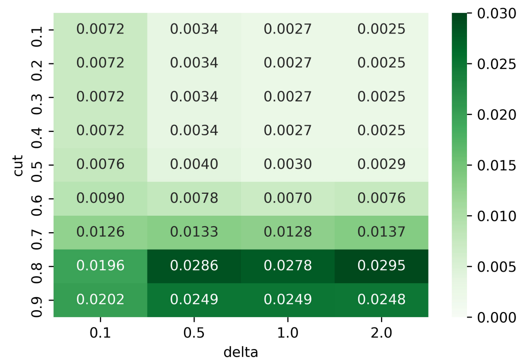


(b) LaBSE

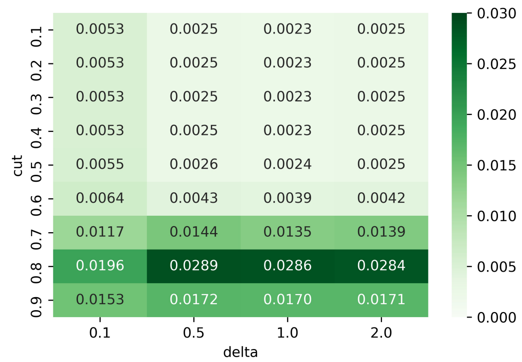


(c) Multilingual MPNet.

(d) Multilingual MiniLM.



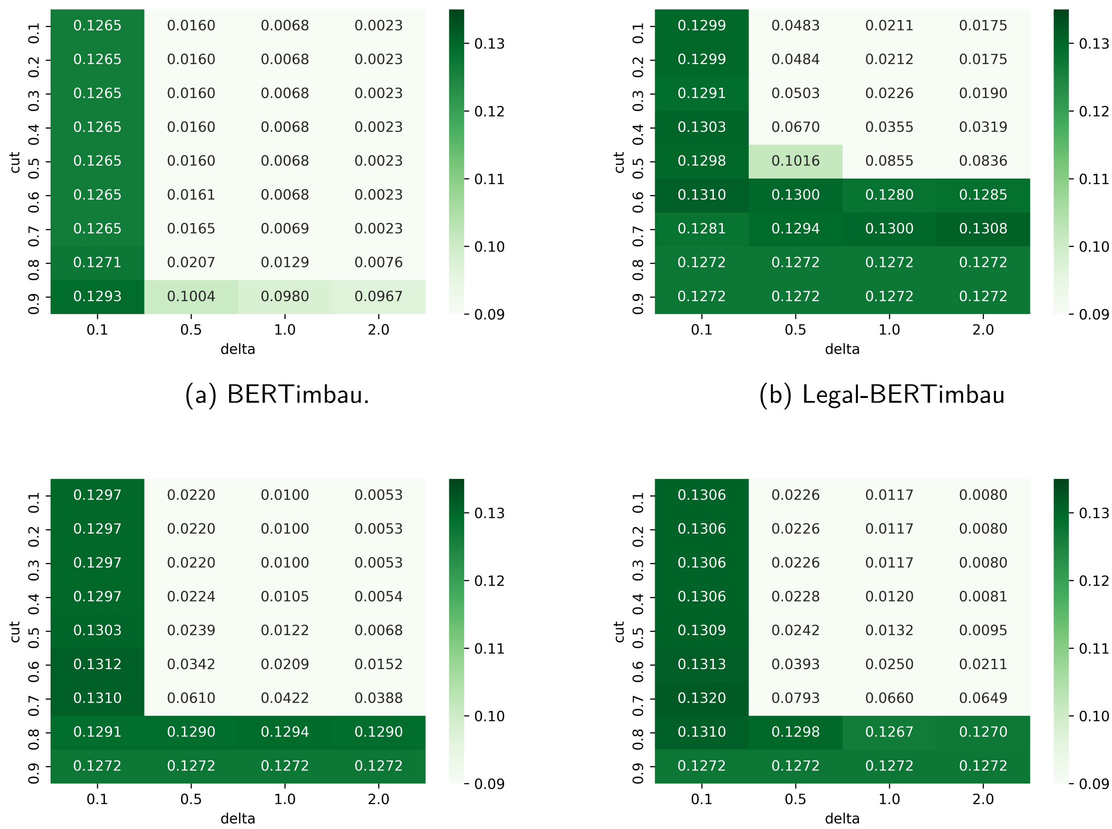
(e) FT BERTimbau.



(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 57 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBERTimbau_BERTimbau (a), LegalBERTimbau_LegalBERTimbau (b), LegalBERTimbau_JurisBERT (c), and LegalBERTimbau_BERTimbauLaw (d) with Ulysses-RFCorpus.



(a) BERTimbau.

(b) Legal-BERTimbau

(c) JurisBERT.

(d) BERTimbauLaw

Source: Created by the author (2025)

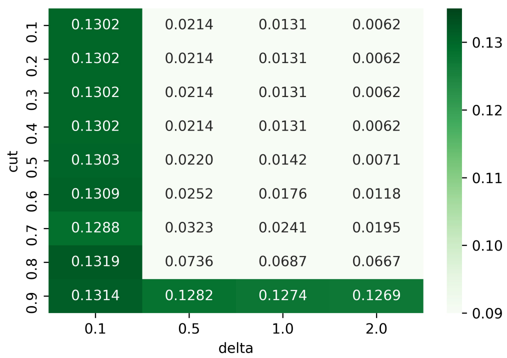
Figure 58 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBERTimbau_LegalBertpt (a), LegalBERTimbau_LaBSE (b), LegalBERTimbau_MPNet (c), LegalBERTimbau_MiniLM (d), LegalBERTimbau_FTBERTimbau (e), and LegalBERTimbau_FTLegalBertpt (f) with Ulysses-RFCorpus.

(a) LegalBert-pt.

(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.



(e) FT BERTimbau.

(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 59 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for JurisBERT_BERTimbau (a), JurisBERT_LegalBERTimbau (b), JurisBERT_JurisBERT (c), and JurisBERT_BERTimbauLaw (d) with Ulysses-RFCorpus.

(a) BERTimbau.

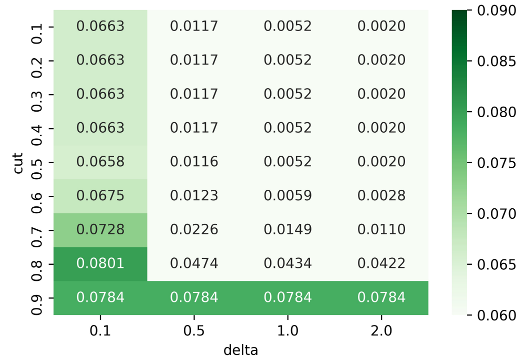
(b) Legal-BERTimbau

(c) JurisBERT.

(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 60 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for JurisBERT_LegalBertpt (a), JurisBERT_LaBSE (b), JurisBERT_MPNet (c), JurisBERT_MiniLM (d), JurisBERT_FTBERTimbau (e), and JurisBERT_FTLegalBertpt (f) with Ulysses-RFCorpus.



(a) LegalBert-pt.

(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.

(e) FT BERTimbau.

(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 61 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbauLaw_BERTimbau (a), BERTimbauLaw_LegalBERTimbau (b), BERTimbauLaw_JurisBERT (c), and BERTimbauLaw_BERTimbauLaw (d) with Ulysses-RFCorpus.

(a) BERTimbau.

(b) Legal-BERTimbau

(c) JurisBERT.

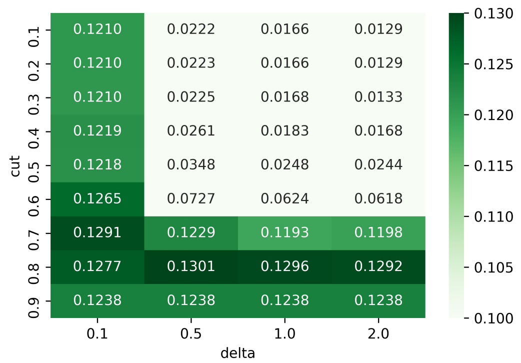
(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 62 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbauLaw_LegalBertpt (a), BERTimbauLaw_LaBSE (b), BERTimbauLaw_MPNet (c), BERTimbauLaw_MiniLM (d), BERTimbauLaw_FTBERTimbau (e), and BERTimbauLaw_FTLegalBertpt (f) with Ulysses-RFCorpus.

(a) LegalBert-pt.

(b) LaBSE



(c) Multilingual MPNet.

(d) Multilingual MiniLM.

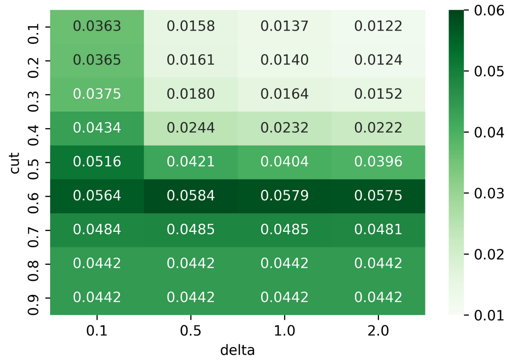
(e) FT BERTimbau.

(f) FT LegalBert-pt.

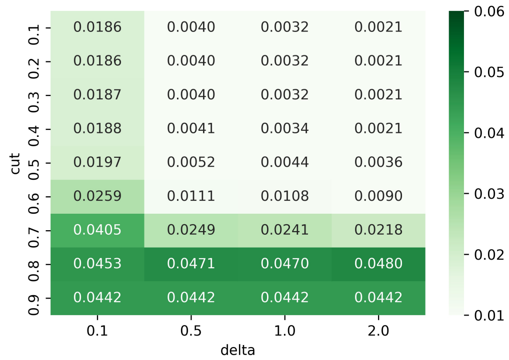
Source: Created by the author (2025)

Figure 63 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBertpt_BERTimbau (a), LegalBertpt_LegalBERTimbau (b), LegalBertpt_JurisBERT (c), and LegalBertpt_BERTimbauLaw (d) with Ulysses-RFCorpus.

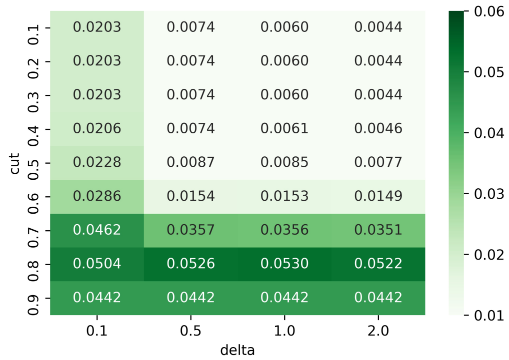
(a) BERTimbau.



(b) Legal-BERTimbau



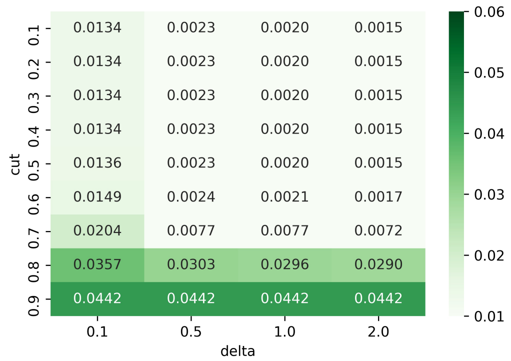
(c) JurisBERT.



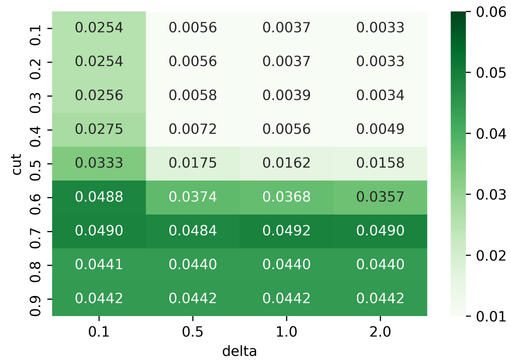
(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 64 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBertpt_LegalBertpt (a), LegalBertpt_LaBSE (b), LegalBertpt_MPNet (c), LegalBertpt_MiniLM (d), LegalBertpt_FTBERTimbau (e), and LegalBertpt_FTLegalBertpt (f) with Ulysses-RFCorpus.



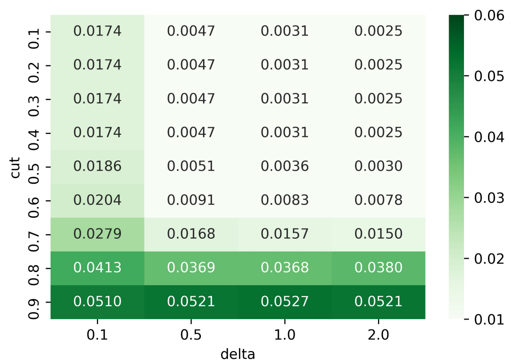
(a) LegalBert-pt.



(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.



(e) FT BERTimbau.

(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 65 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LaBSE_BERTimbau (a), LaBSE_LegalBERTimbau (b), LaBSE_JurisBERT (c), and LaBSE_BERTimbauLaw (d) with Ulysses-RFCorpus.

(a) BERTimbau.

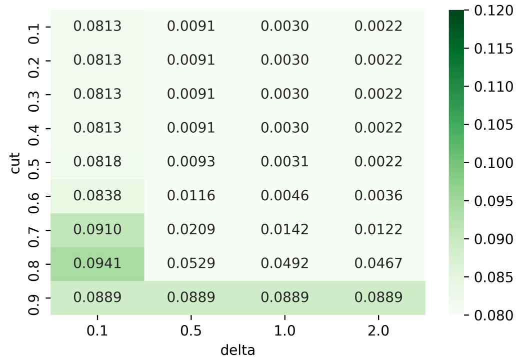
(b) Legal-BERTimbau

(c) JurisBERT.

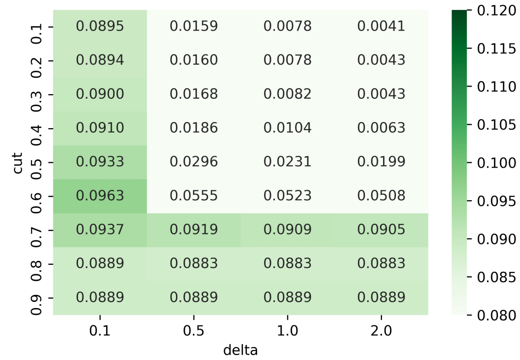
(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 66 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LaBSE_LegalBertpt (a), LaBSE_LaBSE (b), LaBSE_MPNet (c), LaBSE_MiniLM (d), LaBSE_FTBERTimbau (e), and LaBSE_FTLegalBertpt (f) with Ulysses-RFCorpus.

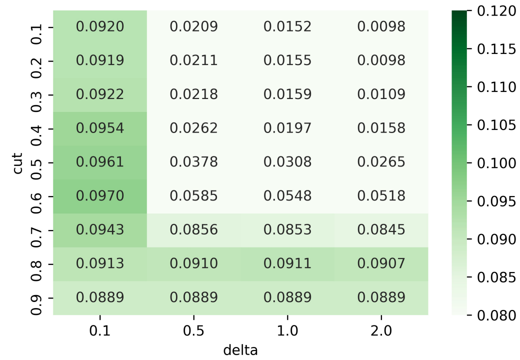


(a) LegalBert-pt.



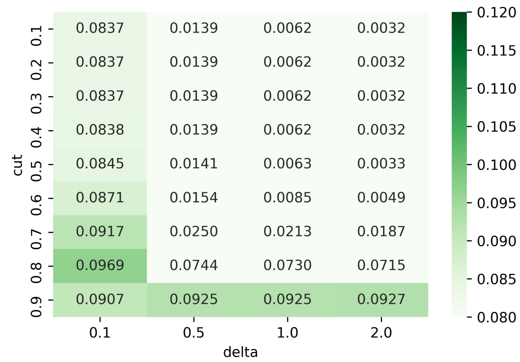
(b) LaBSE

(c) Multilingual MPNet.



(d) Multilingual MiniLM.

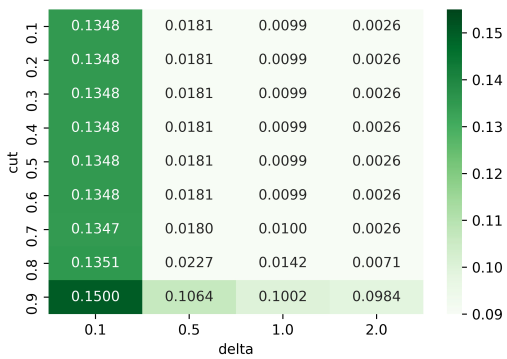
(e) FT BERTimbau.



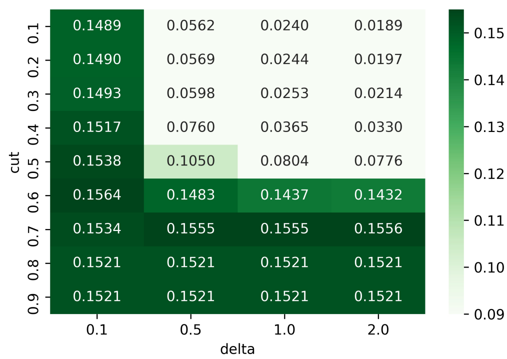
(f) FT LegalBert-pt.

Source: Created by the author (2025)

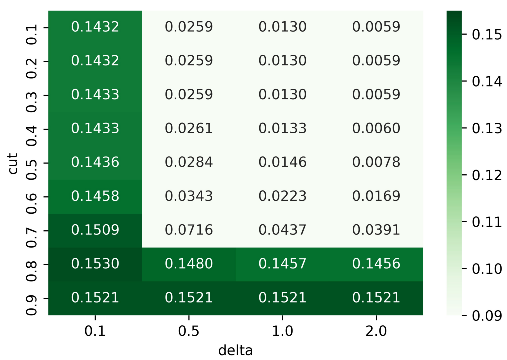
Figure 67 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MPNet_BERTimbau (a), MPNet_LegalBERTimbau (b), MPNet_JurisBERT (c), and MPNet_BERTimbauLaw (d) with Ulysses-RFCorpus.



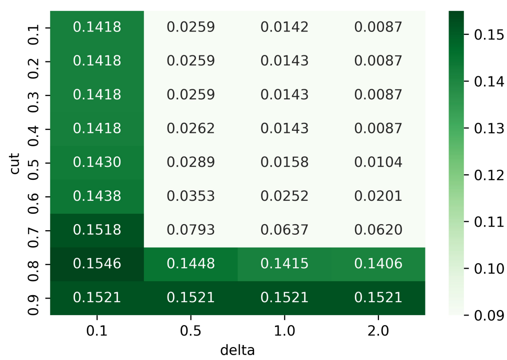
(a) BERTimbau.



(b) Legal-BERTimbau



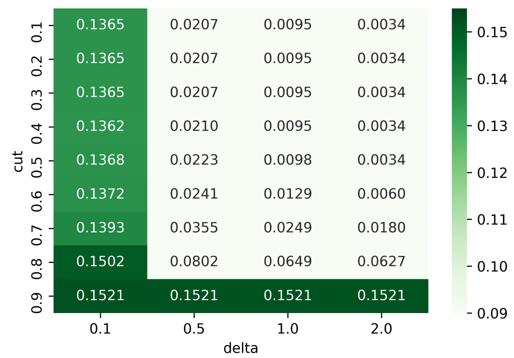
(c) JurisBERT.



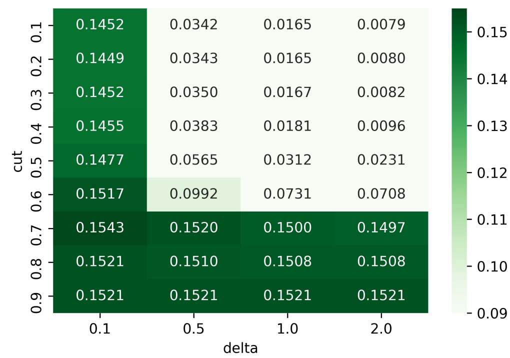
(d) BERTimbauLaw

Source: Created by the author (2025)

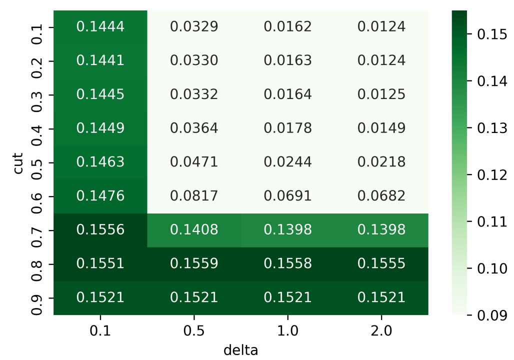
Figure 68 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MPNet_LegalBertpt (a), MPNet_LaBSE (b), MPNet_MPNet (c), MPNet_MiniLM (d), MPNet_FTBERTimbau (e), and MPNet_FTLegalBertpt (f) with Ulysses-RFCorpus.



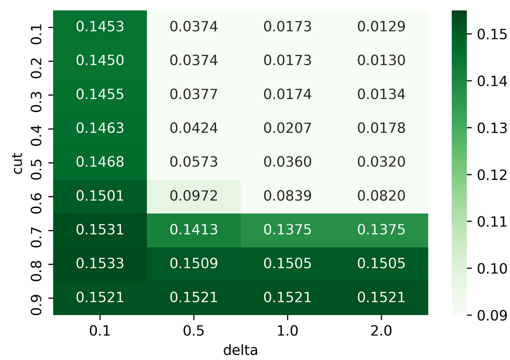
(a) LegalBert-pt.



(b) LaBSE



(c) Multilingual MPNet.



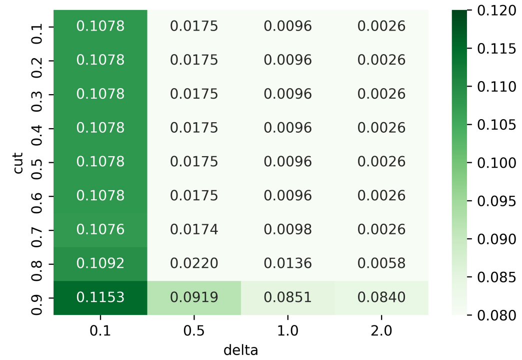
(d) Multilingual MiniLM.

(e) FT BERTimbau.

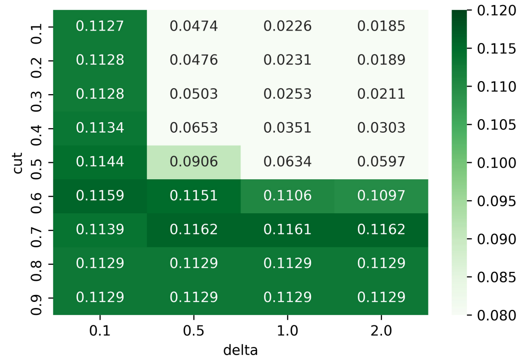
(f) FT LegalBert-pt.

Source: Created by the author (2025)

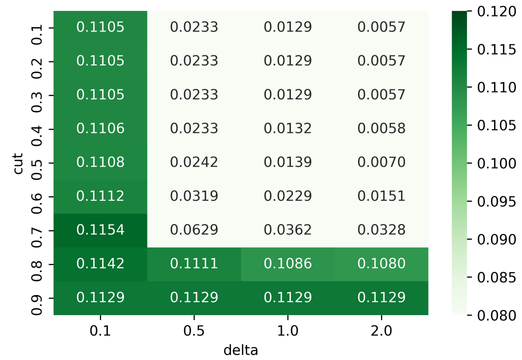
Figure 69 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MiniLM_BERTimbau (a), MiniLM_LegalBERTimbau (b), MiniLM_JurisBERT (c), and MiniLM_BERTimbauLaw (d) with Ulysses-RFCorpus.



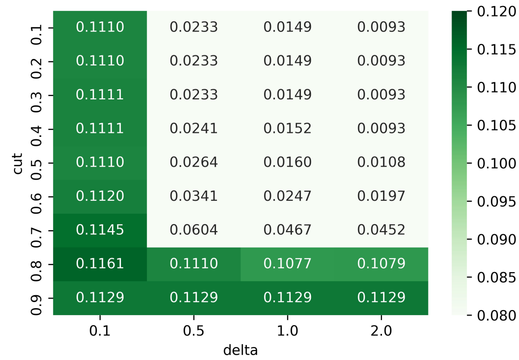
(a) BERTimbau.



(b) Legal-BERTimbau



(c) JurisBERT.

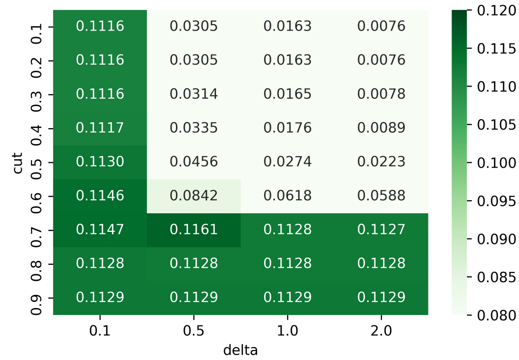


(d) BERTimbauLaw

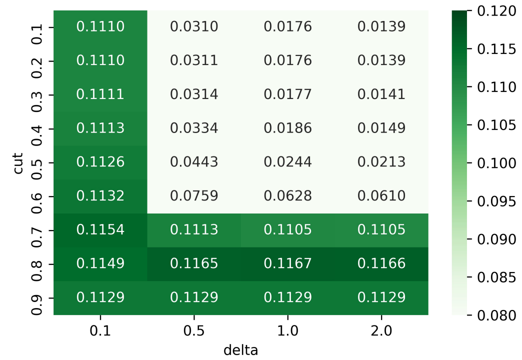
Source: Created by the author (2025)

Figure 70 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MiniLM_LegalBertpt (a), MiniLM_LaBSE (b), MiniLM_MPNet (c), MiniLM_MiniLM (d), MiniLM_FTBERTimbau (e), and MiniLM_FTLegalBertpt (f) with Ulysses-RFCorpus.

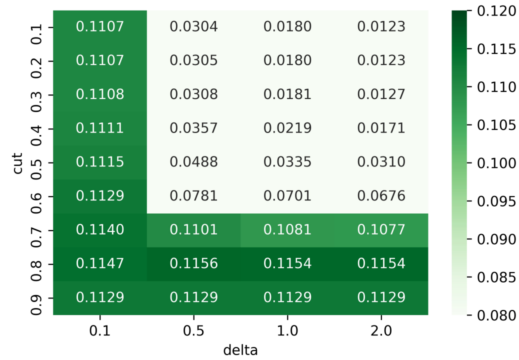
(a) LegalBert-pt.



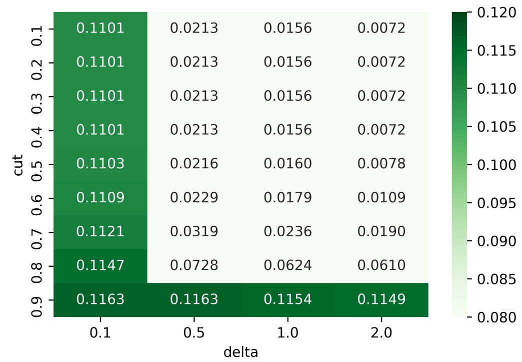
(b) LaBSE



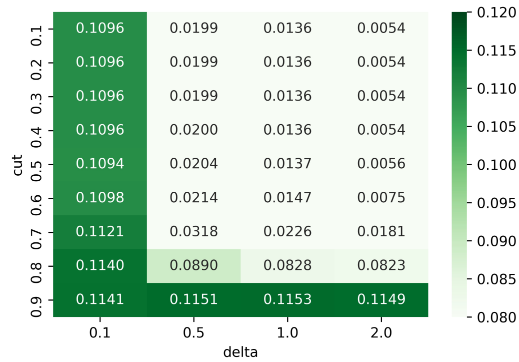
(c) Multilingual MPNet.



(d) Multilingual MiniLM.



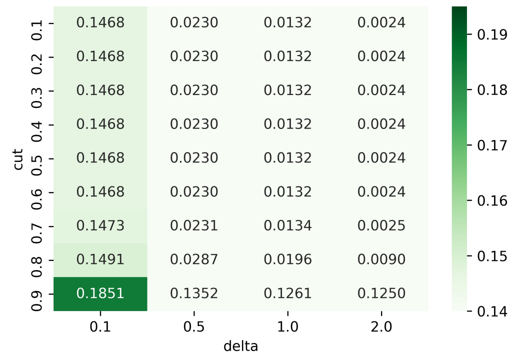
(e) FT BERTimbau.



(f) FT LegalBert-pt.

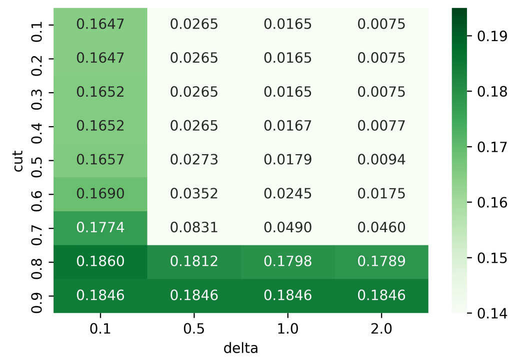
Source: Created by the author (2025)

Figure 71 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTBERTimbau_BERTimbau (a), FTBERTimbau_LegalBERTimbau (b), FTBERTimbau_JurisBERT (c), and FTBERTimbau_BERTimbauLaw (d) with Ulysses-RFCorpus.

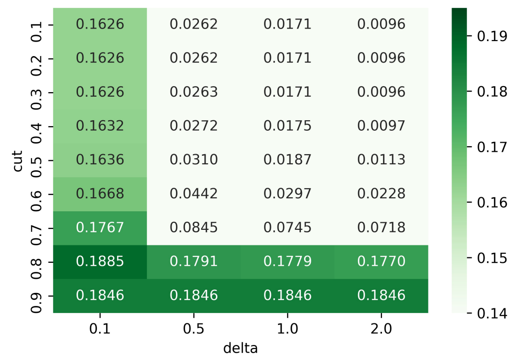


(a) BERTimbau.

(b) Legal-BERTimbau



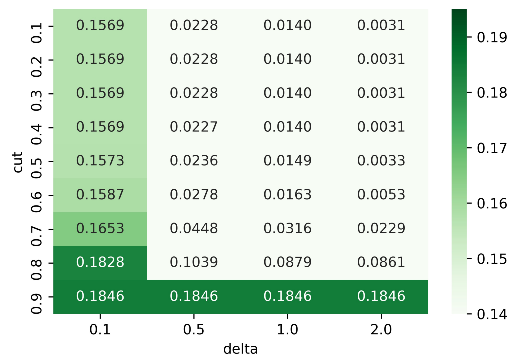
(c) JurisBERT.



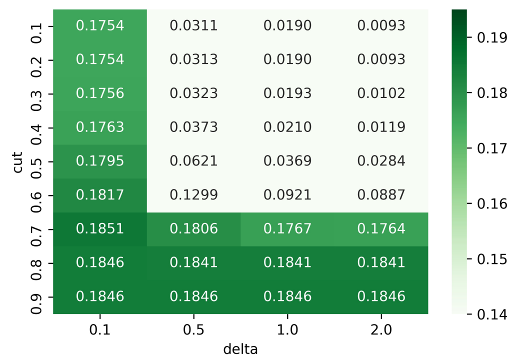
(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 72 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTBERTimbau_LegalBertpt (a), FTBERTimbau_LaBSE (b), FTBERTimbau_MPNet (c), FTBERTimbau_MiniLM (d), FTBERTimbau_FTBERTimbau (e), and FTBERTimbau_FTLegalBertpt (f) with Ulysses-RFCorpus.



(a) LegalBert-pt.

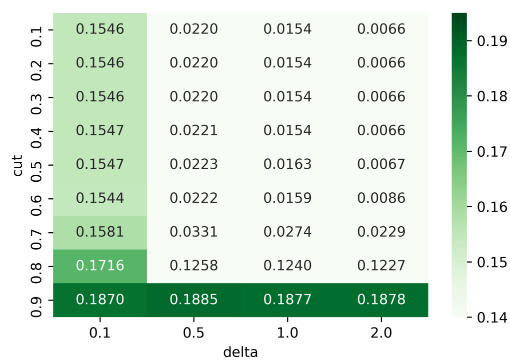


(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.

(e) FT BERTimbau.

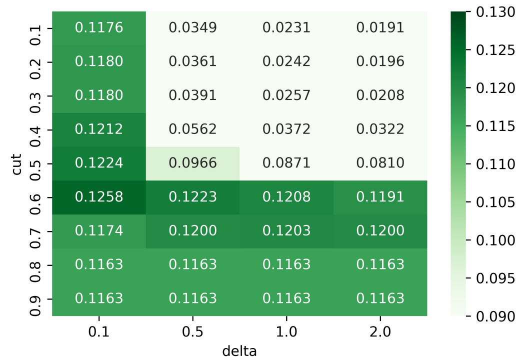


(f) FT LegalBert-pt.

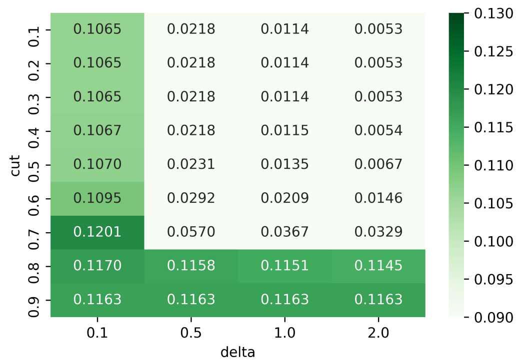
Source: Created by the author (2025)

Figure 73 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTLegalBerpt_BERTimbau (a), FTLegalBerpt_LegalBERTimbau (b), FTLegalBerpt_JurisBERT (c), and FTLegalBerpt_BERTimbauLaw (d) with Ulysses-RFCorpus.

(a) BERTimbau.



(b) Legal-BERTimbau



(c) JurisBERT.

(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 74 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTLegalBerpt_LegalBertpt (a), FTLegalBerpt_LaBSE (b), FTLegalBerpt_MPNet (c), FTLegal-Berpt_MiniLM (d), FTLegalBerpt_FTBERTimbau (e), and FTLegalBerpt_FTLegalBertpt (f) with Ulysses-RFCorpus.

(a) LegalBert-pt.

(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.

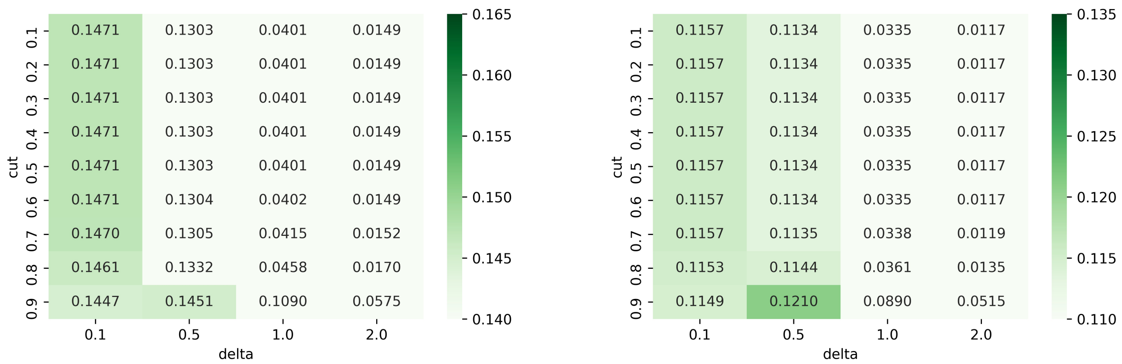
(e) FT BERTimbau.

(f) FT LegalBert-pt.

Source: Created by the author (2025)

APPENDIX H – PARAMETERS ASSESSMENT FOR BM25L WITH THE PRELIMINARY SEARCH CORPUS AND USING BERT-BASED MODELS TO SEARCH FOR THE SIMILAR QUERIES

Figure 75 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_BERTimbau (a) and BM25L_NP_BERTimbau (b) with the Preliminary Search corpus.

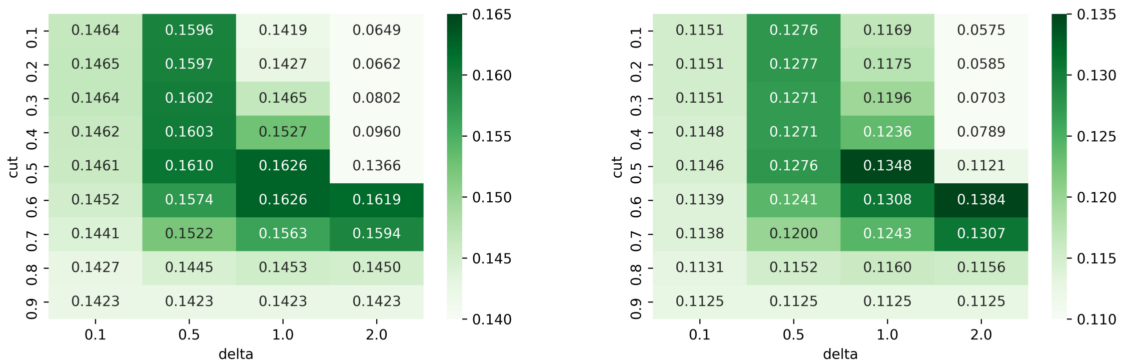


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 76 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_LegalBERTimbau (a) and BM25L_NP_LegalBERTimbau (b) with the Preliminary Search corpus.

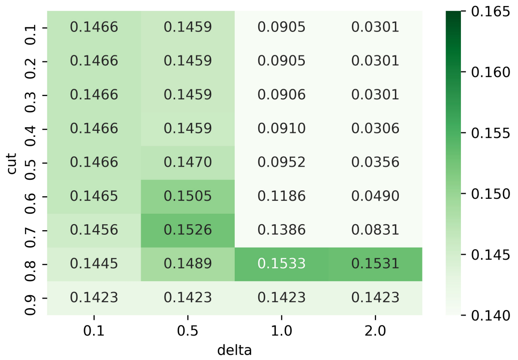


(a) BM25L with preprocessing.

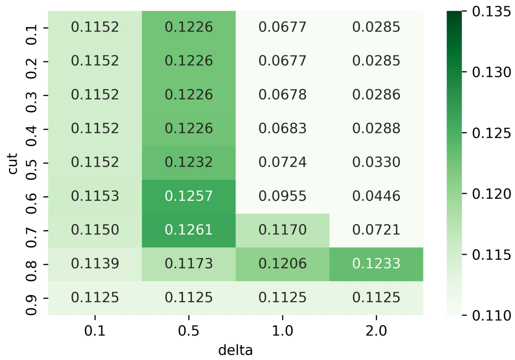
(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 77 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_JurisBERT (a) and BM25L_NP_JurisBERT (b) with the Preliminary Search corpus.



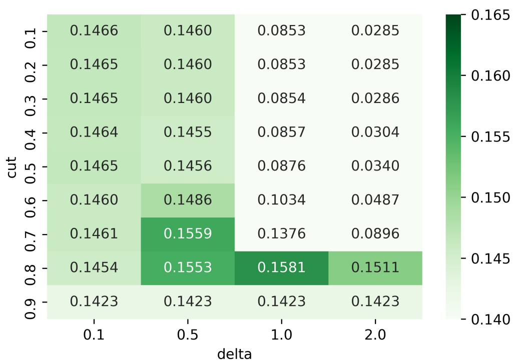
(a) BM25L with preprocessing.



(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 78 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_BERTimbauLaw (a) and BM25L_NP_BERTimbauLaw (b) with the Preliminary Search corpus.

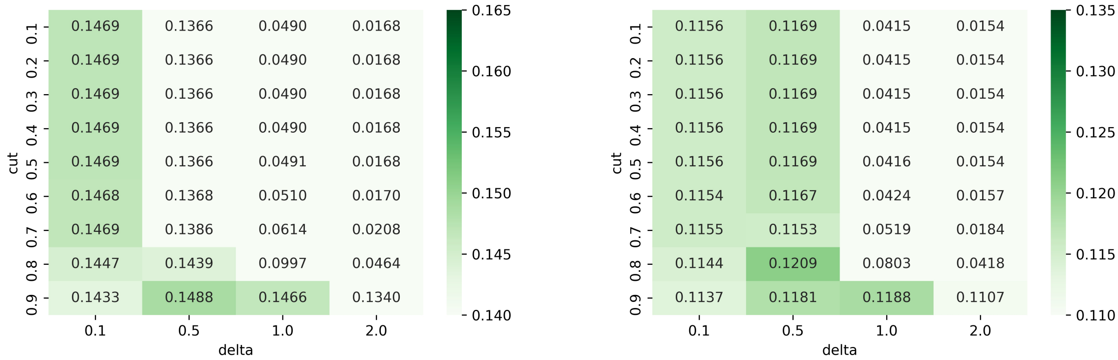


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 79 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_LegalBertpt (a) and BM25L_NP_LegalBertpt (b) with the Preliminary Search corpus.

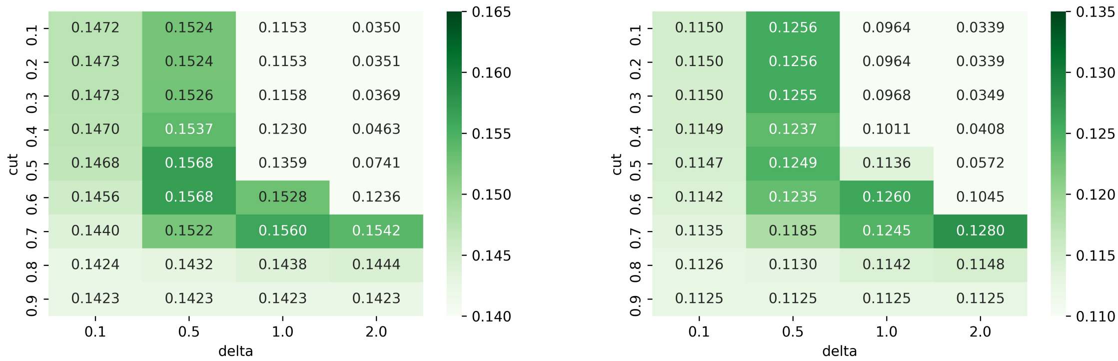


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 80 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_LaBSE (a) and BM25L_NP_LaBSE (b) with the Preliminary Search corpus.

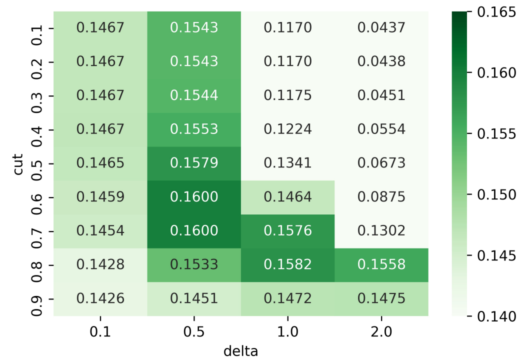


(a) BM25L with preprocessing.

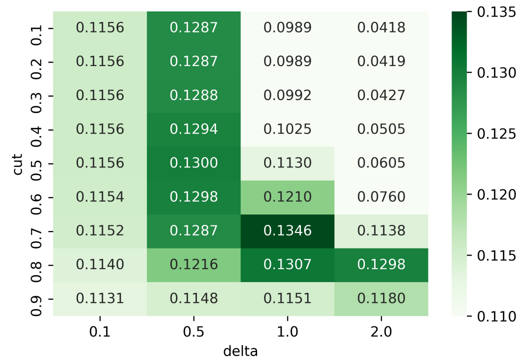
(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 81 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_MPNet (a) and BM25L_NP_MPNet (b) with the Preliminary Search corpus.



(a) BM25L with preprocessing.

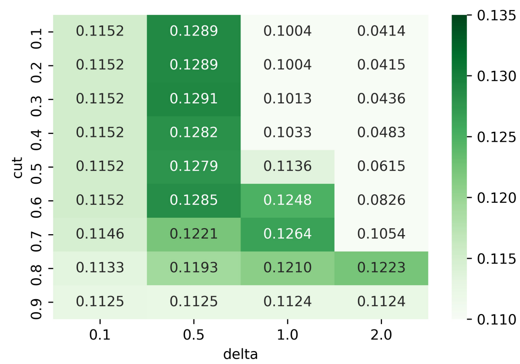


(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 82 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_MiniLM (a) and BM25L_NP_MiniLM (b) with the Preliminary Search corpus.

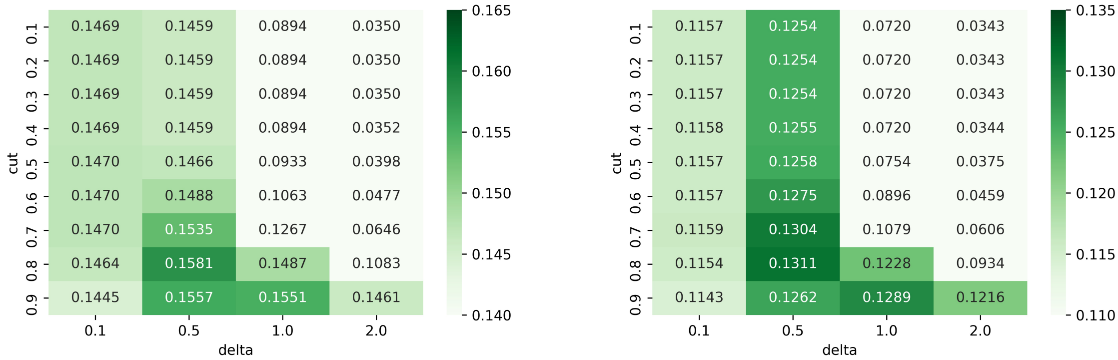
(a) BM25L with preprocessing.



(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 83 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_FTBERTimbau (a) and BM25L_NP_FTBERTimbau (b) with the Preliminary Search corpus.

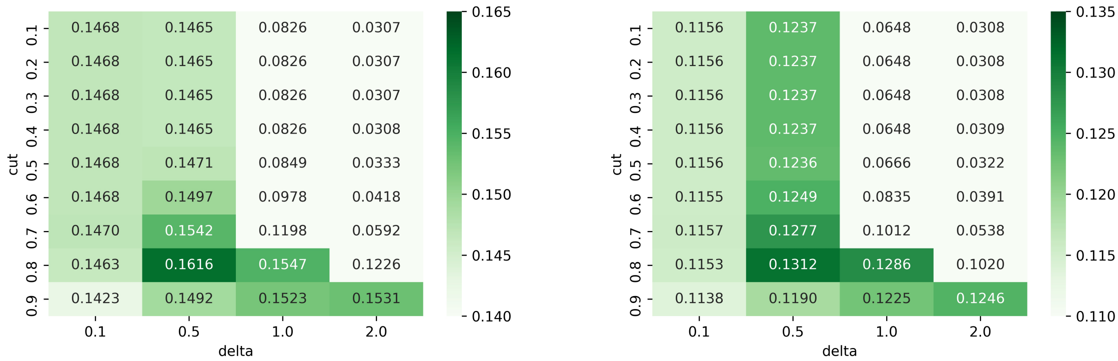


(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

Figure 84 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BM25L_PRE_FTLegalBertpt (a) and BM25L_NP_FTLegalBertpt (b) with the Preliminary Search corpus.



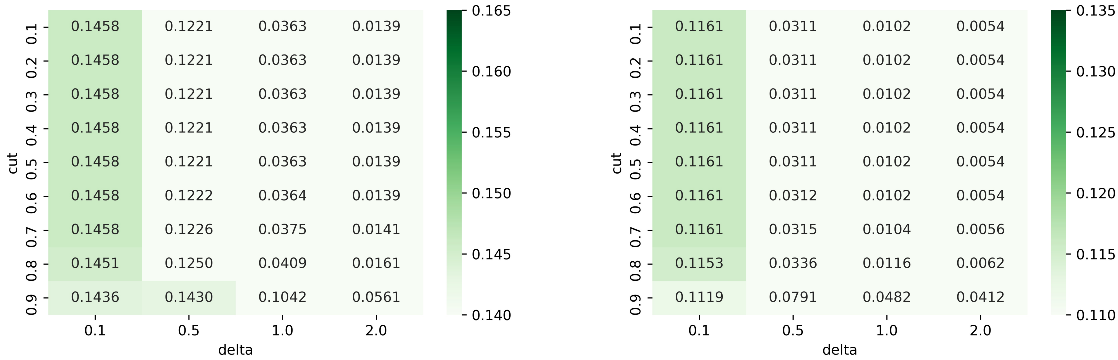
(a) BM25L with preprocessing.

(b) BM25L without preprocessing.

Source: Created by the author (2025)

APPENDIX I – PARAMETERS ASSESSMENT FOR OKAPI BM25 WITH THE PRELIMINARY SEARCH CORPUS AND USING BERT-BASED MODELS TO SEARCH FOR THE SIMILAR QUERIES

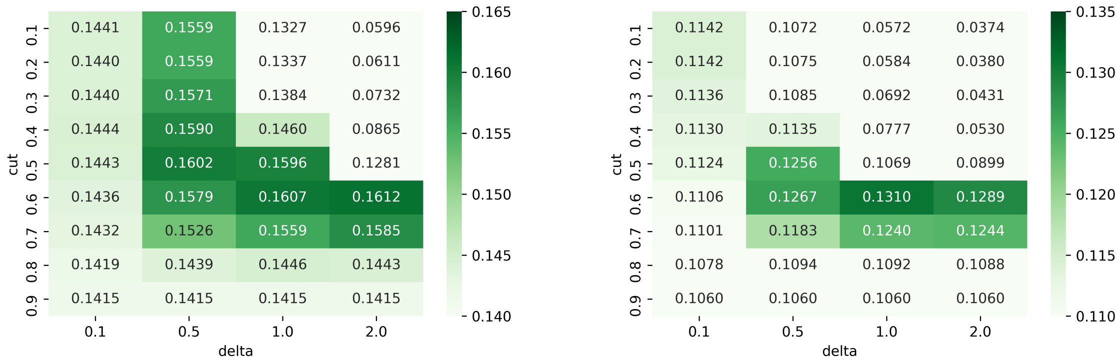
Figure 85 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_BERTimbau (a) and OkapiBM25_NP_BERTimbau (b) with the Preliminary Search corpus.



(a) Okapi BM25 with preprocessing. (b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

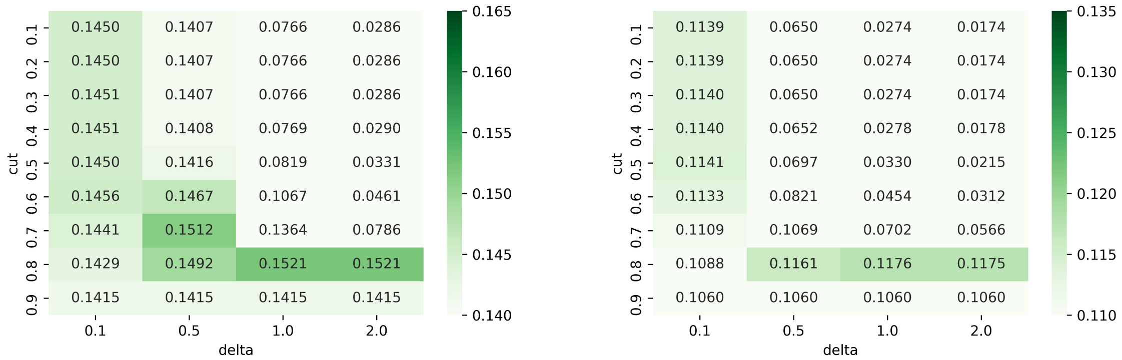
Figure 86 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_LegalBERTimbau (a) and OkapiBM25_NP_LegalBERTimbau (b) with the Preliminary Search corpus.



(a) Okapi BM25 with preprocessing. (b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 87 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_JurisBERT (a) and OkapiBM25_NP_JurisBERT (b) with the Preliminary Search corpus.

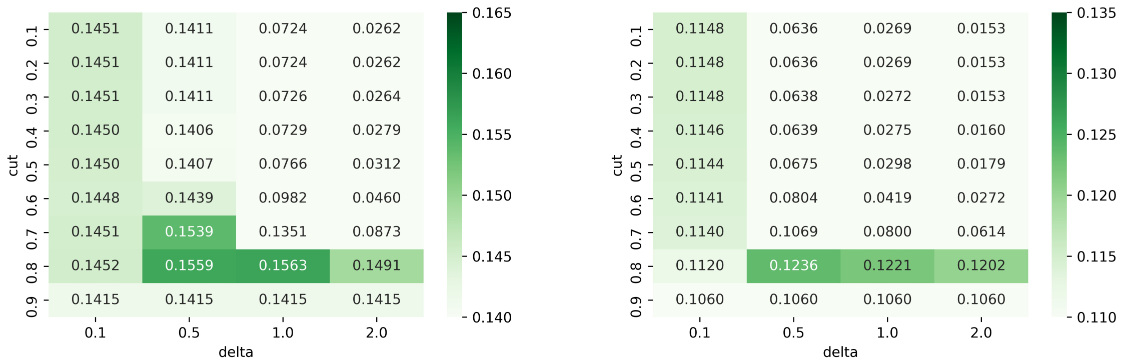


(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 88 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_BERTimbauLaw (a) and OkapiBM25_NP_BERTimbauLaw (b) with the Preliminary Search corpus.

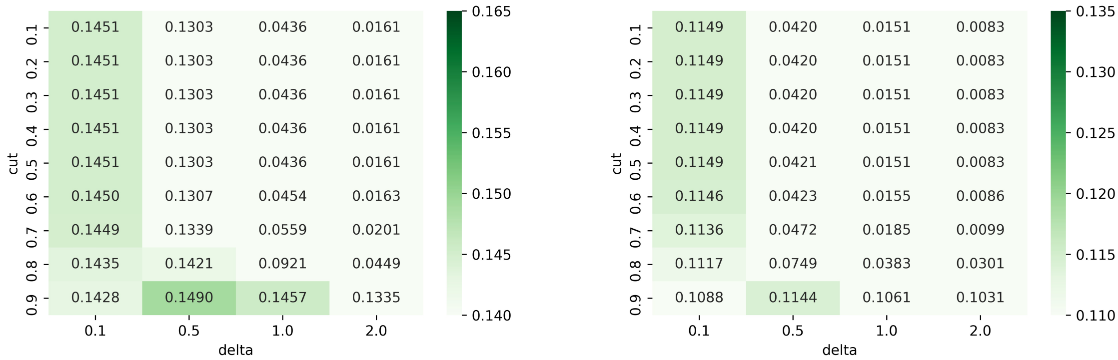


(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 89 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_LegalBertpt (a) and OkapiBM25_NP_LegalBertpt (b) with the Preliminary Search corpus.

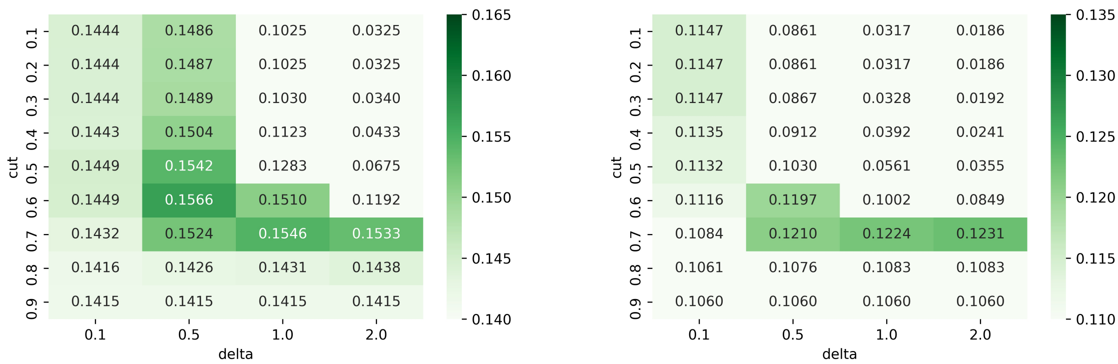


(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 90 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_LaBSE (a) and OkapiBM25_NP_LaBSE (b) with the Preliminary Search corpus.

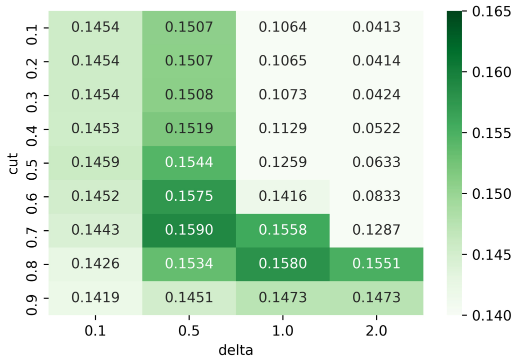


(a) Okapi BM25 with preprocessing.

(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 91 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_MPNet (a) and OkapiBM25_NP_MPNet (b) with the Preliminary Search corpus.

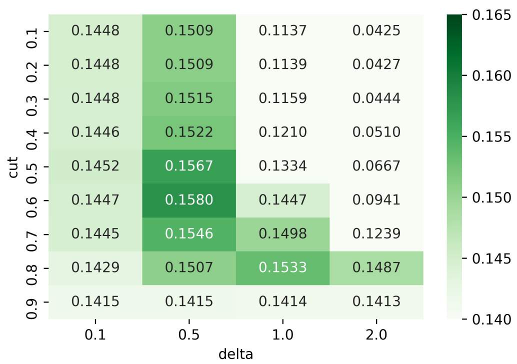


(a) Okapi BM25 with preprocessing.

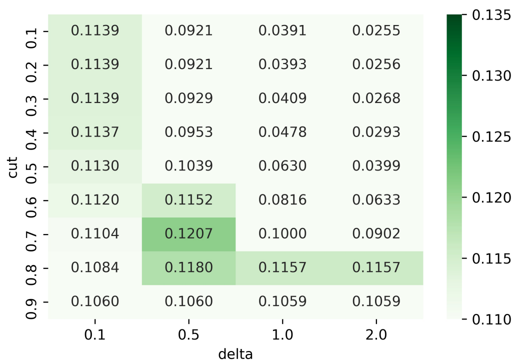
(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 92 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_MiniLM (a) and OkapiBM25_NP_MiniLM (b) with the Preliminary Search corpus.



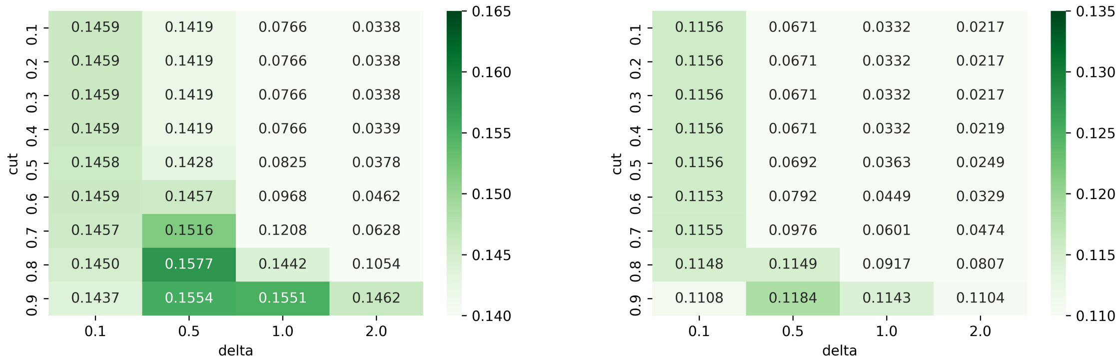
(a) Okapi BM25 with preprocessing.



(b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

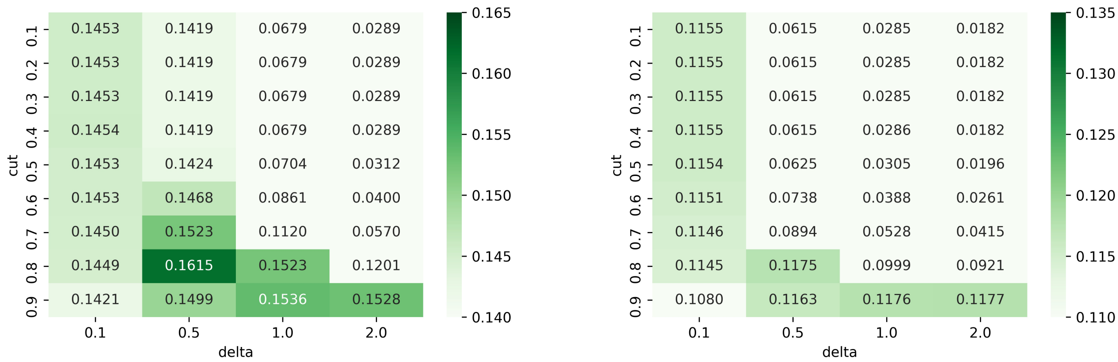
Figure 93 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_FTBERTimbau (a) and OkapiBM25_NP_FTBERTimbau (b) with the Preliminary Search corpus.



(a) Okapi BM25 with preprocessing. (b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

Figure 94 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for OkapiBM25_PRE_FTLegalBertpt (a) and OkapiBM25_NP_FTLegalBertpt (b) with the Preliminary Search corpus.



(a) Okapi BM25 with preprocessing. (b) Okapi BM25 without preprocessing.

Source: Created by the author (2025)

APPENDIX J – PARAMETERS ASSESSMENT FOR BERT-BASED MODELS WITH THE PRELIMINARY SEARCH CORPUS AND USING COSINE TO SEARCH FOR THE SIMILAR QUERIES

Figure 95 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbau_PRE (a) and BERTimbau_NP (b) with the Preliminary Search corpus.

(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 96 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBERTimbau_PRE (a) and LegalBERTimbau_NP (b) with the Preliminary Search corpus.

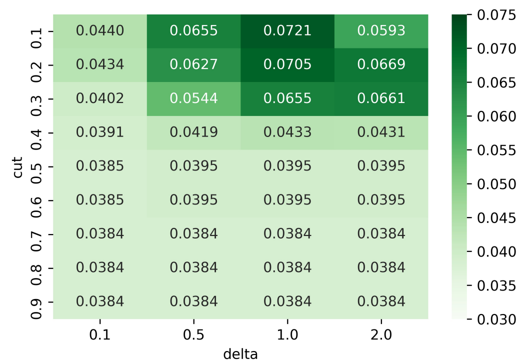
(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 97 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for JurisBERT_PRE (a) and JurisBERT_NP (b) with the Preliminary Search corpus.

(a) Cosine similarity with preprocessing.



(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

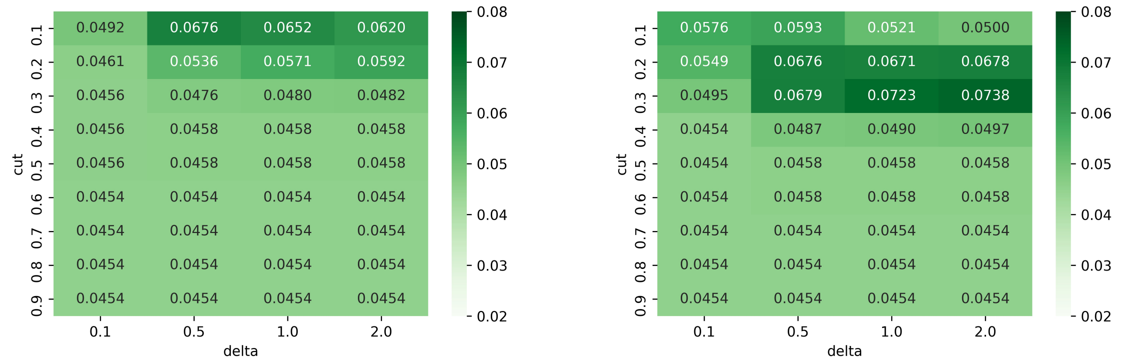
Figure 98 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbauLaw_PRE (a) and BERTimbauLaw_NP (b) with the Preliminary Search corpus.

(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 99 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBertpt_PRE (a) and LegalBertpt_NP (b) with the Preliminary Search corpus.

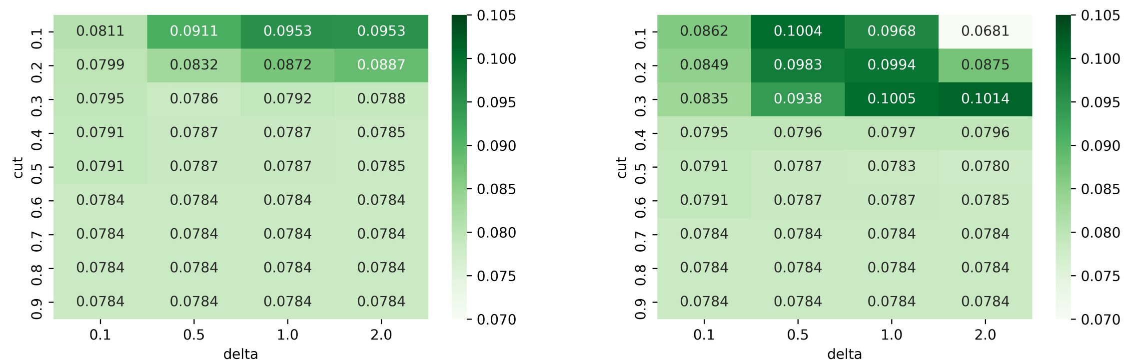


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 100 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LaBSE_PRE (a) and LaBSE_NP (b) with the Preliminary Search corpus.

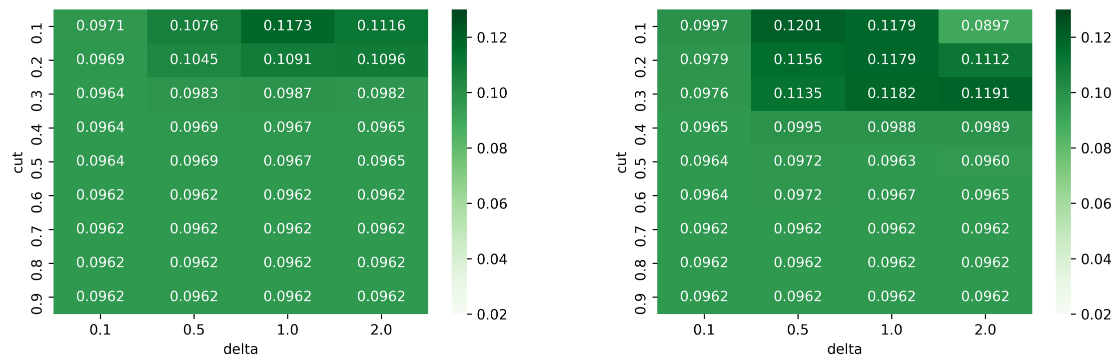


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 101 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MPNet_PRE (a) and MPNet_NP (b) with the Preliminary Search corpus.

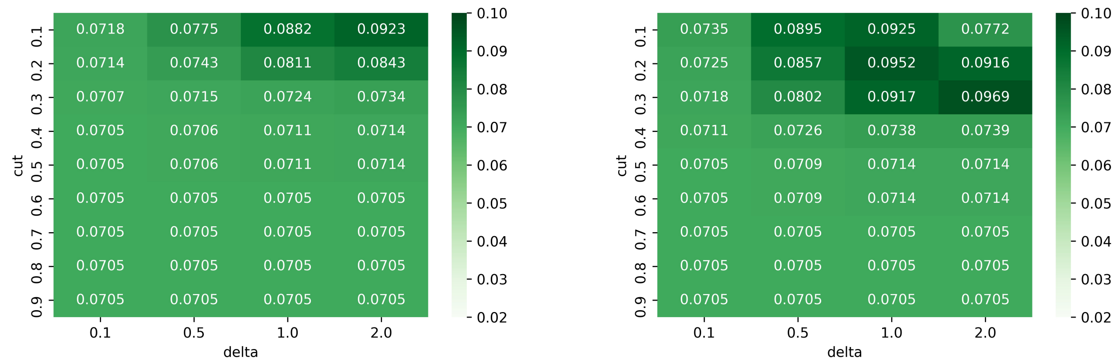


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 102 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MiniLM_PRE (a) and MiniLM_NP (b) with the Preliminary Search corpus.

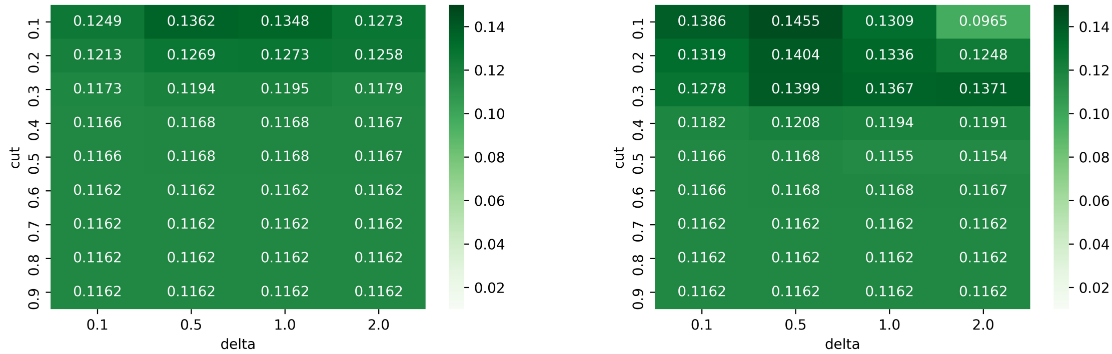


(a) Cosine similarity with preprocessing.

(b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

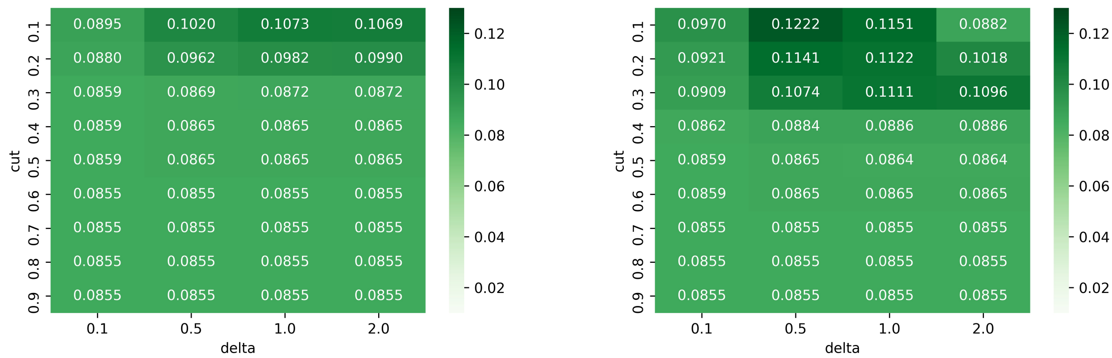
Figure 103 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTBERTimbau_PRE (a) and FTBERTimbau_NP (b) with the Preliminary Search corpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

Figure 104 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTLegalBertpt_PRE (a) and FTLegalBertpt_NP (b) with the Preliminary Search corpus.



(a) Cosine similarity with preprocessing. (b) Cosine similarity without preprocessing.

Source: Created by the author (2025)

APPENDIX K – PARAMETERS ASSESSMENT WITH THE BERT-BASED MODELS WITH THE PRELIMINARY SEARCH CORPUS AND USING BERT-BASED MODELS TO SEARCH WITH THE SIMILAR QUERIES

Figure 105 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbau_BERTimbau (a), BERTimbau_LegalBERTimbau (b), BERTimbau_JurisBERT (c), and BERTimbau_BERTimbauLaw (d) with the Preliminary Search corpus.

(a) BERTimbau.

(b) Legal-BERTimbau

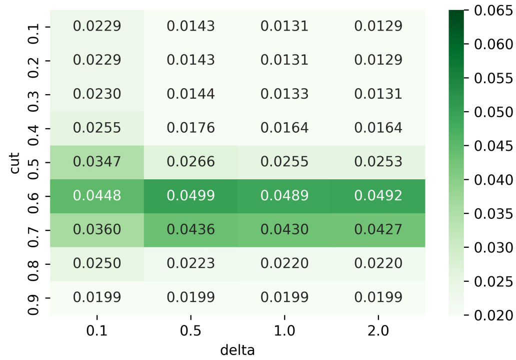
(c) JurisBERT.

(d) BERTimbauLaw

Source: Created by the author (2025)

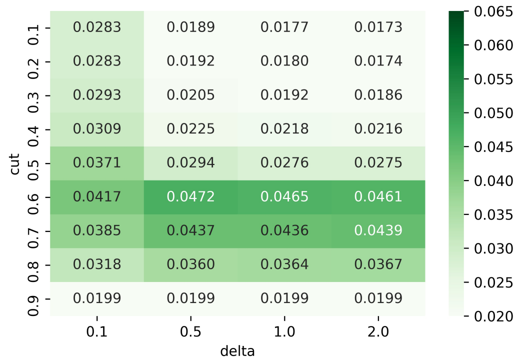
Figure 106 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbau_LegalBertpt (a), BERTimbau_LaBSE (b), BERTimbau_MPNet (c), BERTimbau_MiniLM (d), BERTimbau_FTBERTimbau (e), and BERTimbau_FTLegalBertpt (f) with the Preliminary Search corpus.

(a) LegalBert-pt.



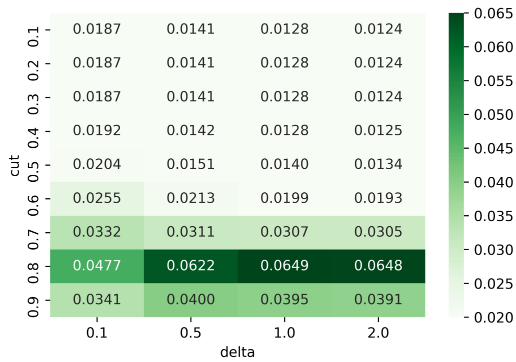
(b) LaBSE

(c) Multilingual MPNet.



(d) Multilingual MiniLM.

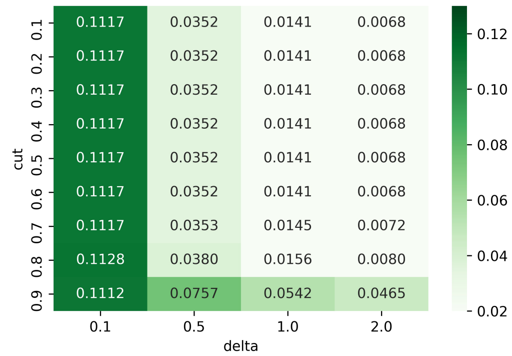
(e) FT BERTimbau.



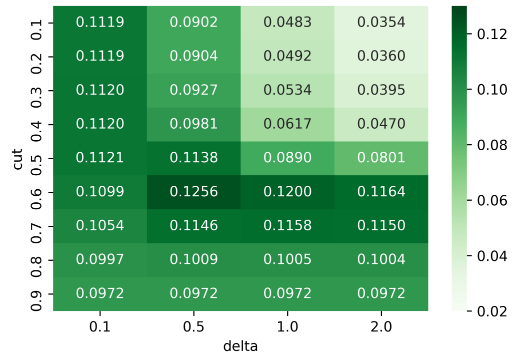
(f) FT LegalBert-pt.

Source: Created by the author (2025)

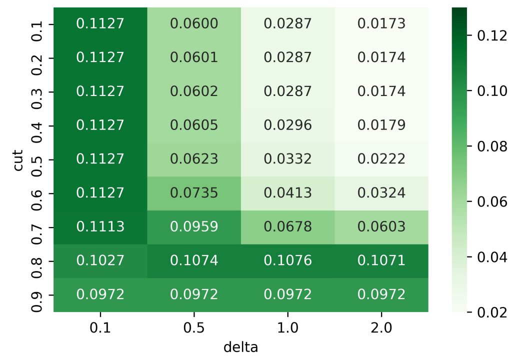
Figure 107 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBERTimbau_BERTimbau (a), LegalBERTimbau_LegalBERTimbau (b), LegalBERTimbau_JurisBERT (c), and LegalBERTimbau_BERTimbauLaw (d) with the Preliminary Search corpus.



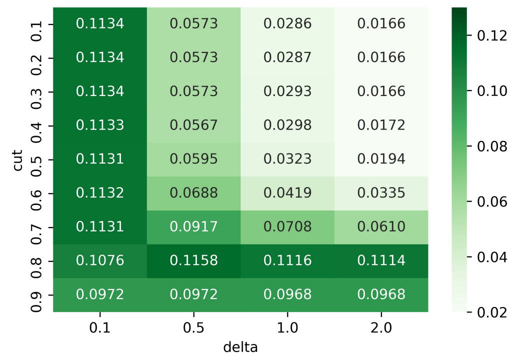
(a) BERTimbau.



(b) Legal-BERTimbau



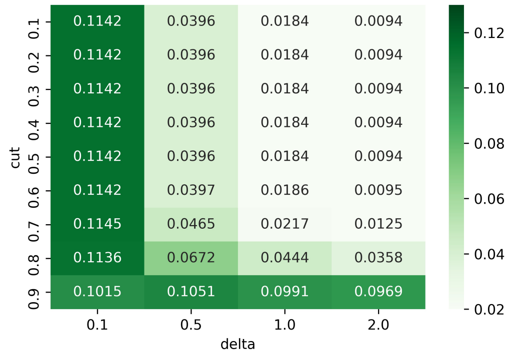
(c) JurisBERT.



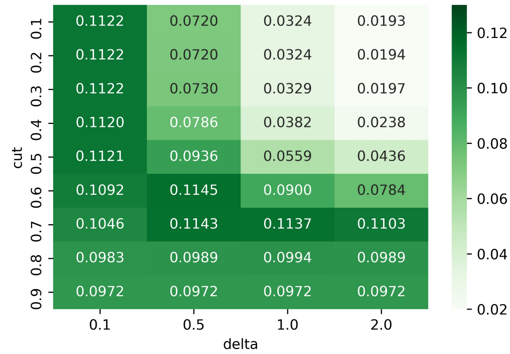
(d) BERTimbauLaw

Source: Created by the author (2025)

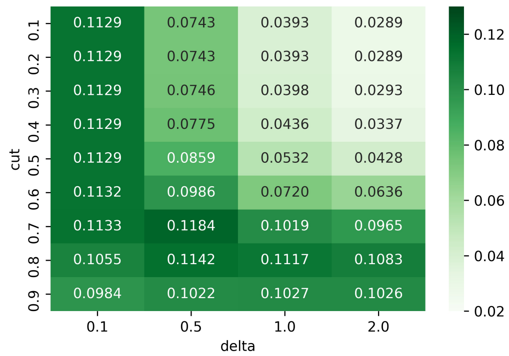
Figure 108 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBERTimbau_LegalBertpt (a), LegalBERTimbau_LaBSE (b), LegalBERTimbau_MPNet (c), LegalBERTimbau_MiniLM (d) LegalBERTimbau_FTBERTimbau (e), and LegalBERTimbau_FTLegalBertpt (f) with the Preliminary Search corpus.



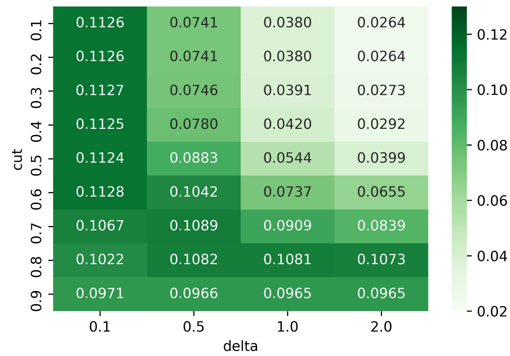
(a) LegalBert-pt.



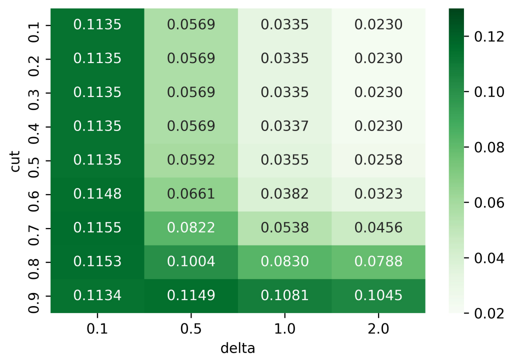
(b) LaBSE



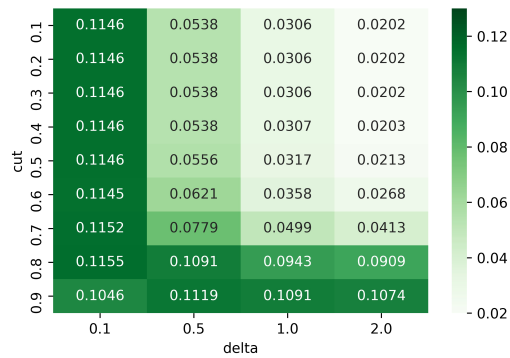
(c) Multilingual MPNet.



(d) Multilingual MiniLM.



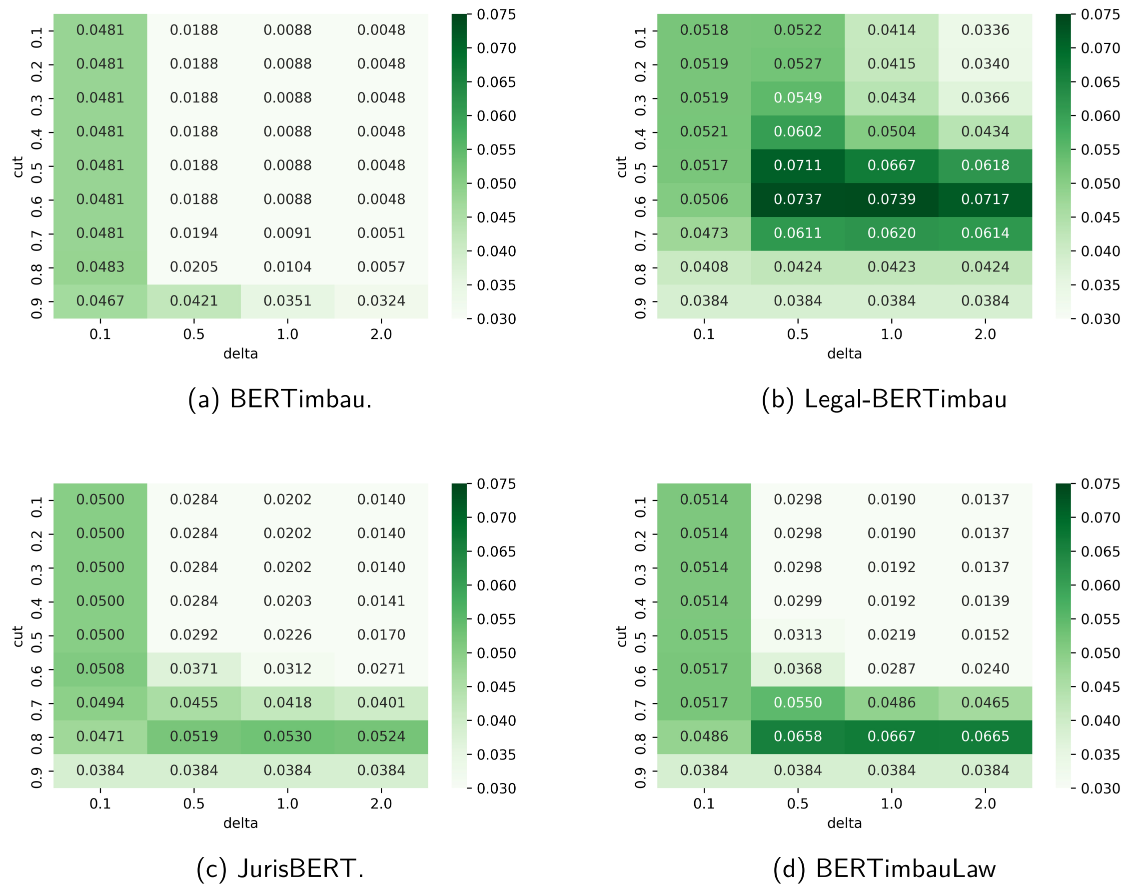
(e) FT BERTimbau.



(f) FT LegalBert-pt.

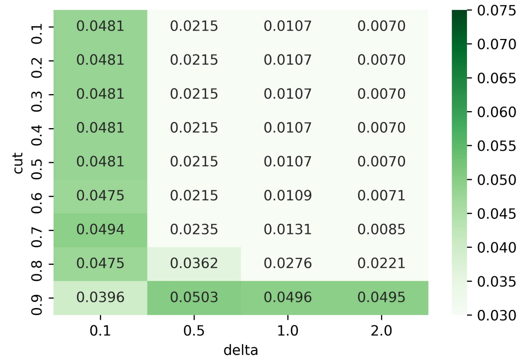
Source: Created by the author (2025)

Figure 109 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for JurisBERT_BERTimbau (a), JurisBERT_LegalBERTimbau (b), JurisBERT_JurisBERT (c), and JurisBERT_BERTimbauLaw (d) with the Preliminary Search corpus.

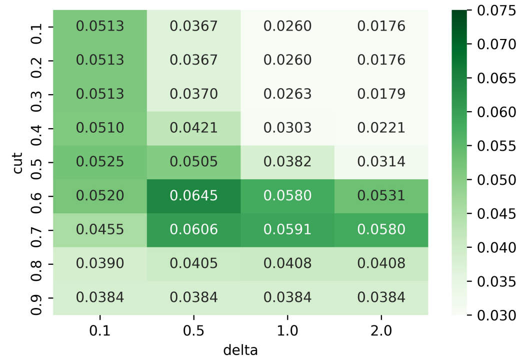


Source: Created by the author (2025)

Figure 110 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for JurisBERT_LegalBertpt (a), JurisBERT_LaBSE (b), JurisBERT_MPNet (c), JurisBERT_MiniLM (d), JurisBERT_FTBERTimbau (e), and JurisBERT_FTLegalBertpt (f) with the Preliminary Search corpus.



(a) LegalBert-pt.



(b) LaBSE

(c) Multilingual MPNet.

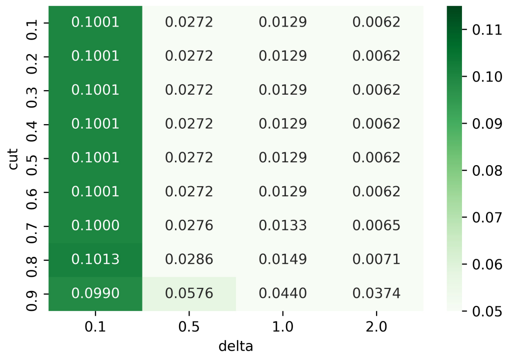
(d) Multilingual MiniLM.

(e) FT BERTimbau.

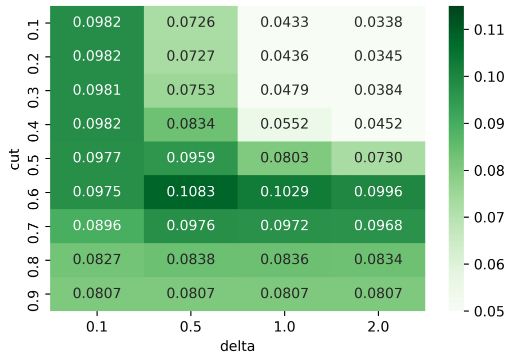
(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 111 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbauLaw_BERTimbau (a), BERTimbauLaw_LegalBERTimbau (b), BERTimbauLaw_JurisBERT (c), and BERTimbauLaw_BERTimbauLaw (d) with the Preliminary Search corpus.



(a) BERTimbau.



(b) Legal-BERTimbau

(c) JurisBERT.

(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 112 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for BERTimbauLaw_LegalBertpt (a), BERTimbauLaw_LaBSE (b), BERTimbauLaw_MPNet (c), BERTimbauLaw_MiniLM (d), BERTimbauLaw_FTBERTimbau (e), BERTimbauLaw_FTLegalBertpt (f) with the Preliminary Search corpus.

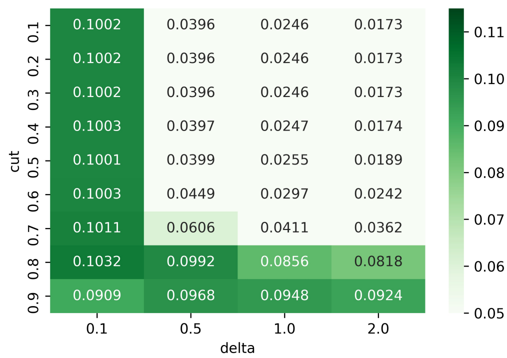
(a) LegalBert-pt.

(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.

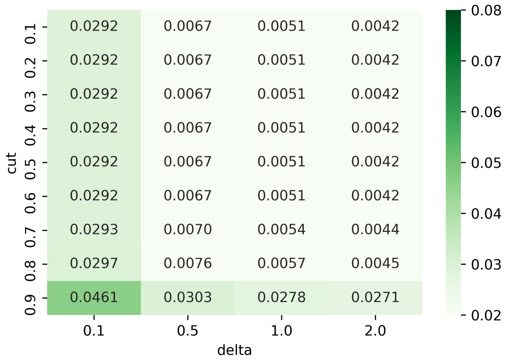
(e) FT BERTimbau.



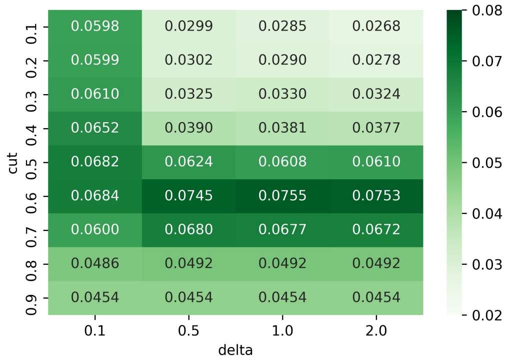
(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 113 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBertpt_BERTimbau (a), LegalBertpt_LegalBERTimbau (b), LegalBertpt_JurisBERT (c), and LegalBertpt_BERTimbauLaw (d) with the Preliminary Search corpus.

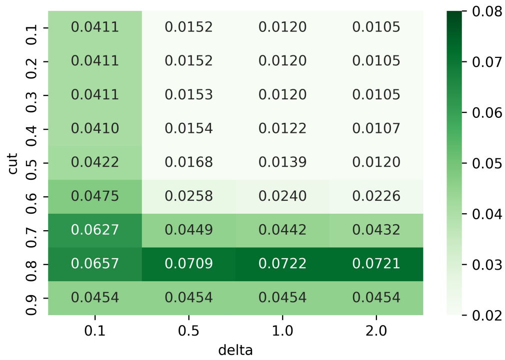


(a) BERTimbau.



(b) Legal-BERTimbau

(c) JurisBERT.

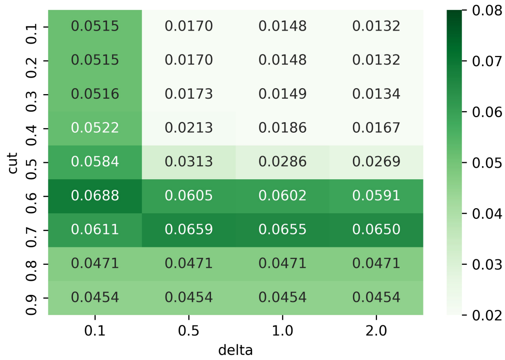


(d) BERTimbauLaw

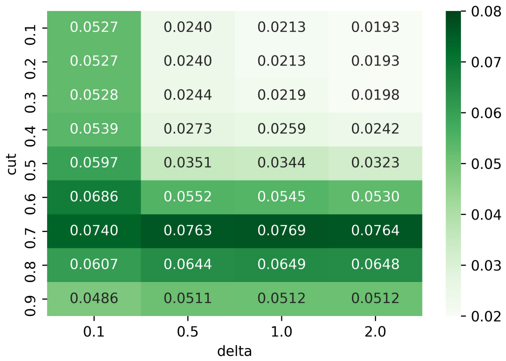
Source: Created by the author (2025)

Figure 114 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LegalBertpt_LegalBertpt (a), LegalBertpt_LaBSE (b), LegalBertpt_MPNet (c), LegalBertpt_MiniLM (d), LegalBertpt_FTBERTimbau (e), LegalBertpt_FTLegalBertpt (f) with the Preliminary Search corpus.

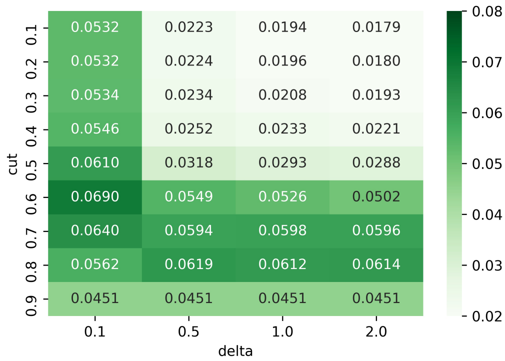
(a) LegalBert-pt.



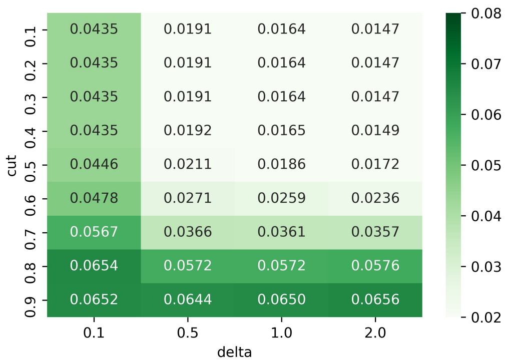
(b) LaBSE



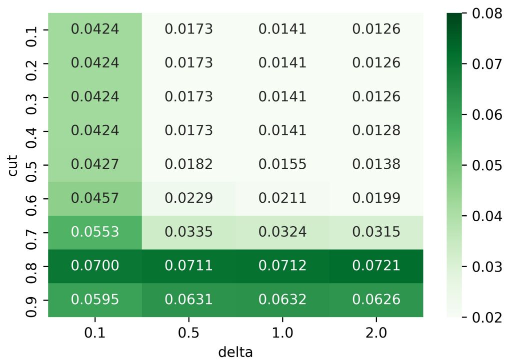
(c) Multilingual MPNet.



(d) Multilingual MiniLM.



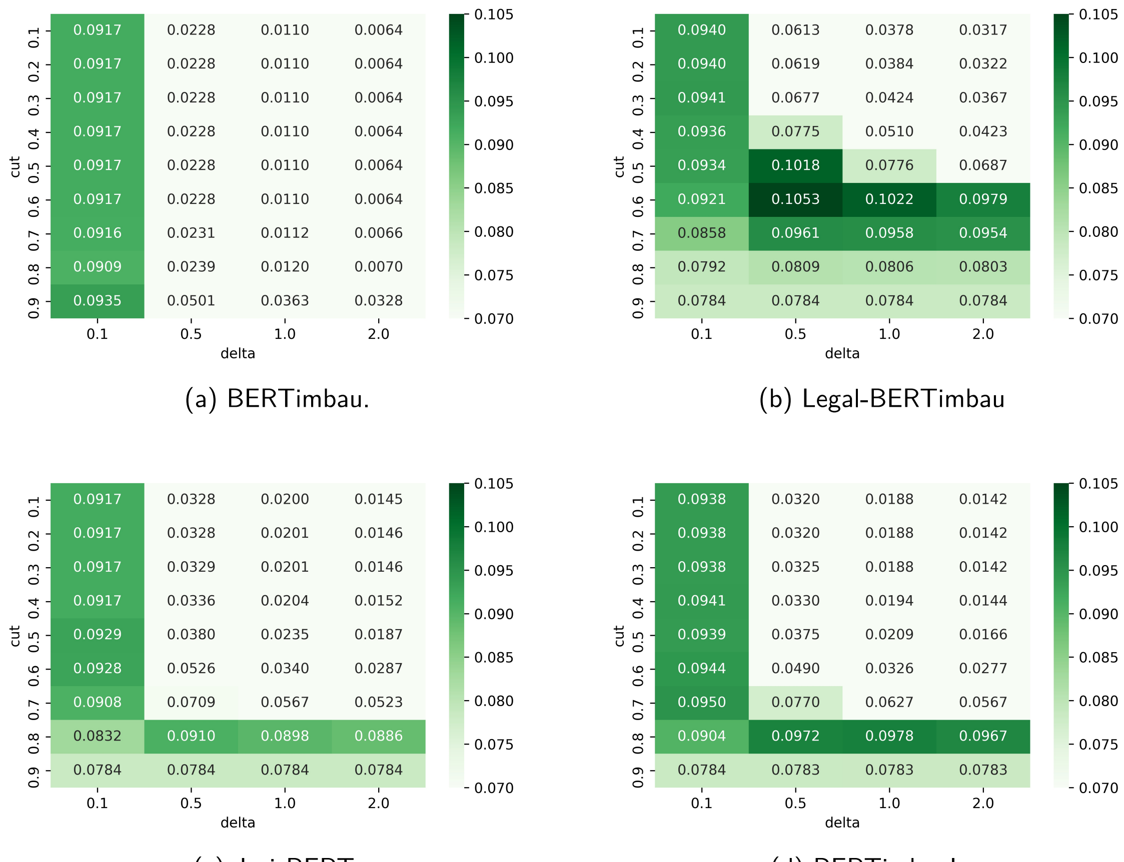
(e) FT BERTimbau.



(f) FT LegalBert-pt.

Source: Created by the author (2025)

Figure 115 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LaBSE_BERTimbau (a), LaBSE_LegalBERTimbau (b), LaBSE_JurisBERT (c), and LaBSE_BERTimbauLaw (d) with the Preliminary Search corpus.



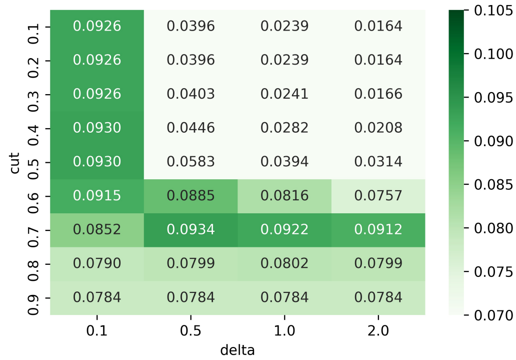
(a) BERTimbau. (b) Legal-BERTimbau

(c) JurisBERT. (d) BERTimbauLaw

Source: Created by the author (2025)

Figure 116 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for LaBSE_LegalBertpt (a), LaBSE_LaBSE (b), LaBSE_MPNet (c), LaBSE_MiniLM (d), LaBSE_FTBERTimbau (e), LaBSE_FTLegalBertpt (f) with the Preliminary Search corpus.

(a) LegalBert-pt.



(b) LaBSE

(c) Multilingual MPNet.

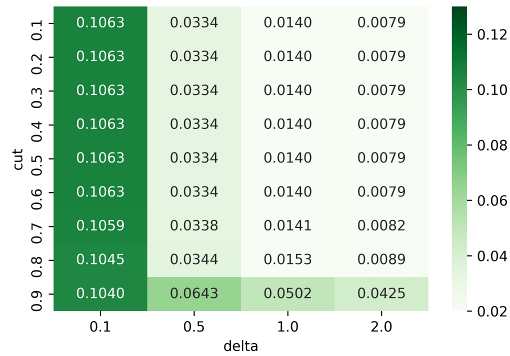
(d) Multilingual MiniLM.

(e) FT BERTimbau.

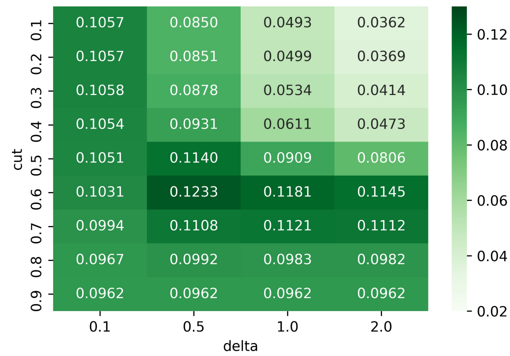
(f) FT LegalBert-pt.

Source: Created by the author (2025)

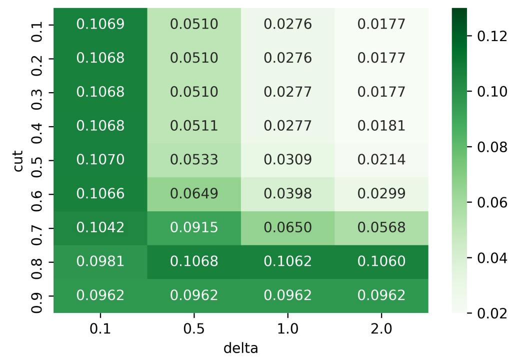
Figure 117 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MPNet_BERTimbau (a), MPNet_LegalBERTimbau (b), MPNet_JurisBERT (c), and MPNet_BERTimbauLaw (d) with the Preliminary Search corpus.



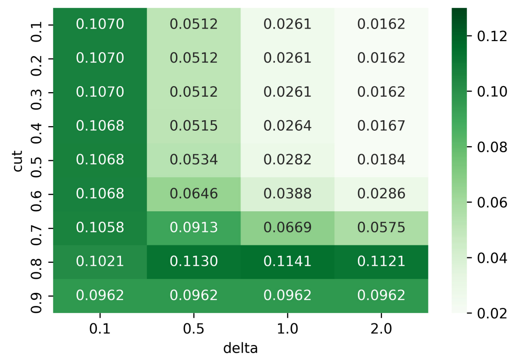
(a) BERTimbau.



(b) Legal-BERTimbau



(c) JurisBERT.

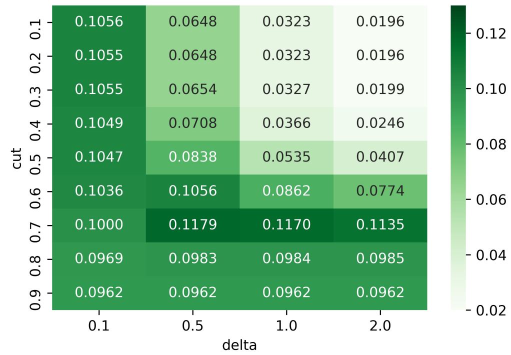


(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 118 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MPNet_LegalBertpt (a), MPNet_LaBSE (b), MPNet_MPNet (c), MPNet_MiniLM (d), MPNet_FTBERTimbau (e), MPNet_FTLegalBertpt (f) with the Preliminary Search corpus.

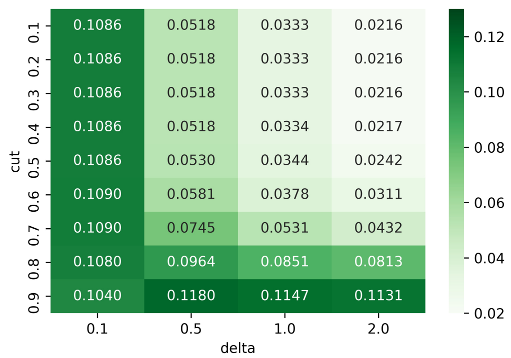
(a) LegalBert-pt.



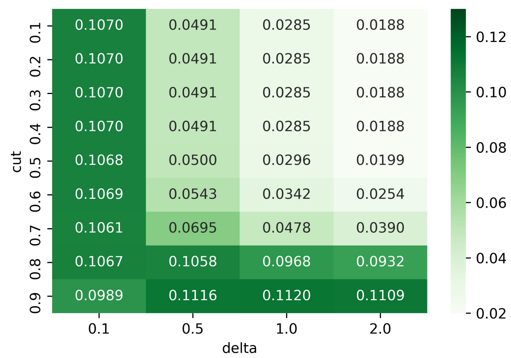
(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.



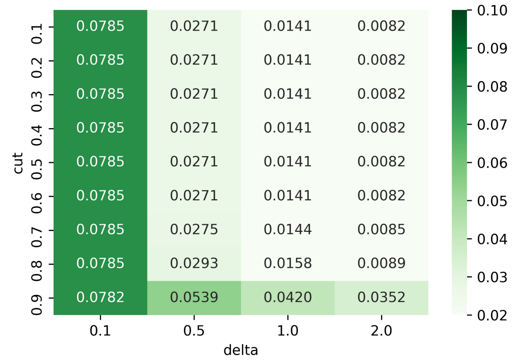
(e) FT BERTimbau.



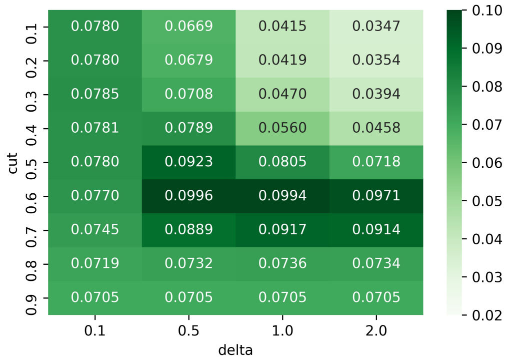
(f) FT LegalBert-pt.

Source: Created by the author (2025)

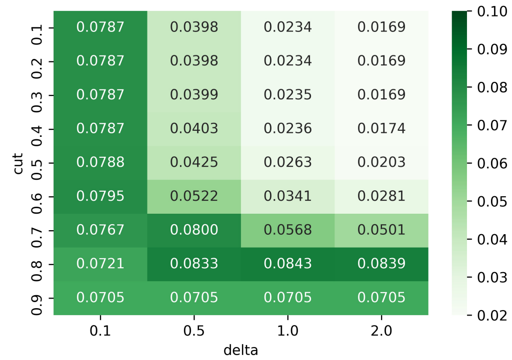
Figure 119 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MiniLM_BERTimbau (a), MiniLM_LegalBERTimbau (b), MiniLM_JurisBERT (c), and MiniLM_BERTimbauLaw (d) with the Preliminary Search corpus.



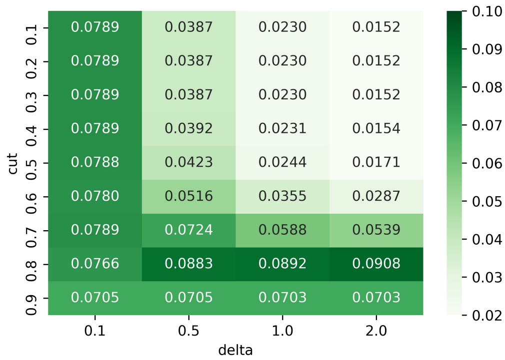
(a) BERTimbau.



(b) Legal-BERTimbau



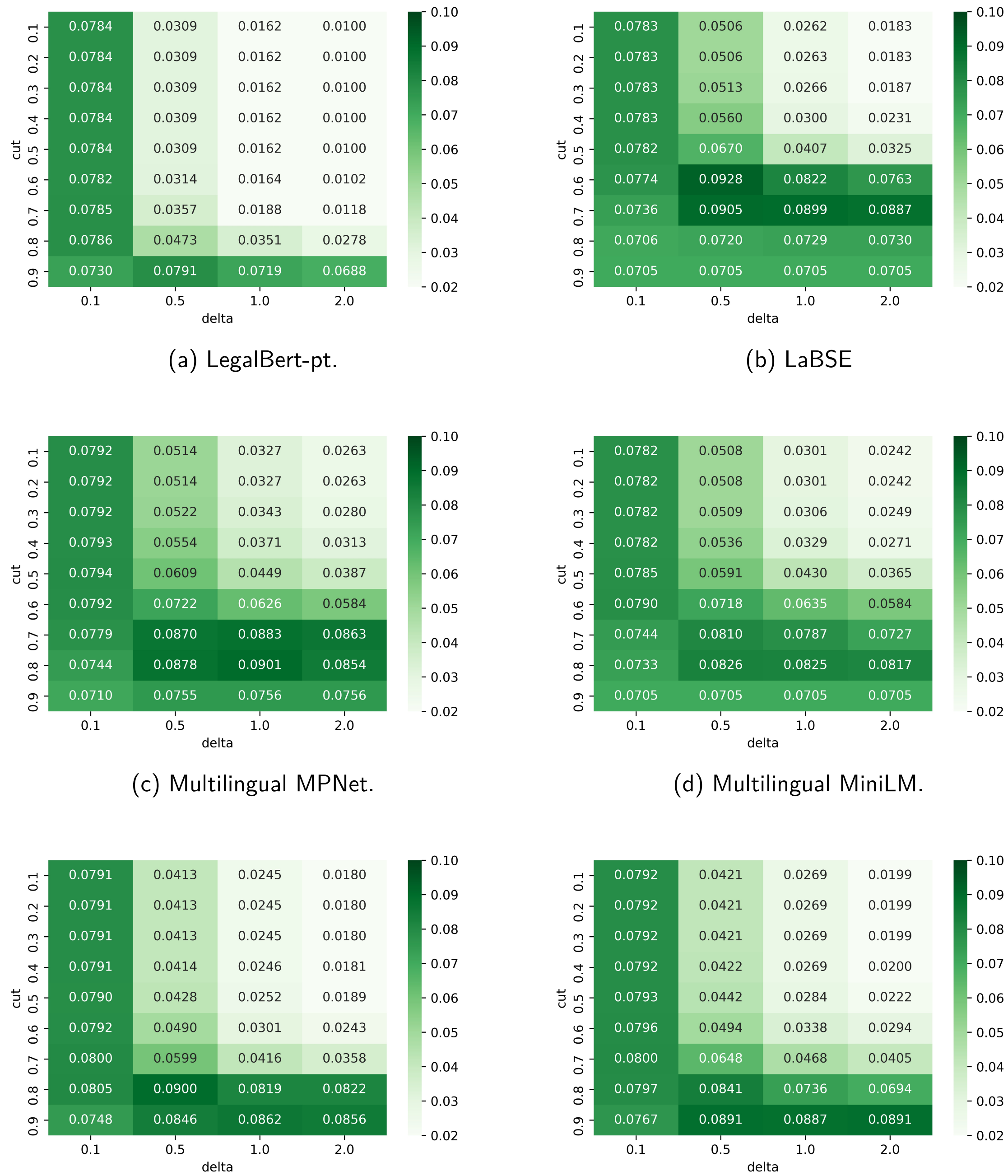
(c) JurisBERT.



(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 120 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for MiniLM_LegalBertpt (a), MiniLM_LaBSE (b), MiniLM_MPNet (c), MiniLM_MiniLM (d), MiniLM_FTBERTimbau (e), MiniLM_FTLegalBertpt (f) with the Preliminary Search corpus.



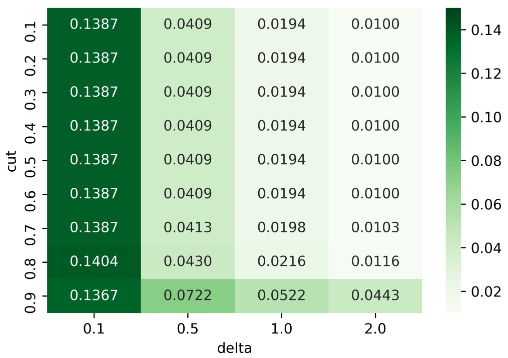
(a) LegalBert-pt. (b) LaBSE

(c) Multilingual MPNet. (d) Multilingual MiniLM.

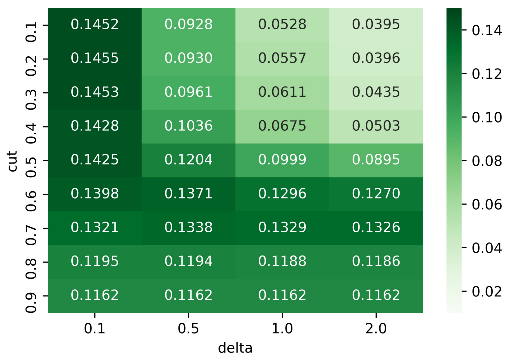
(e) FT BERTimbau. (f) FT LegalBert-pt.

Source: Created by the author (2025)

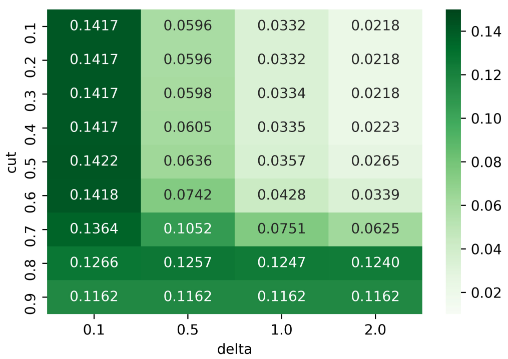
Figure 121 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTBERTimbau_BERTimbau (a), FTBERTimbau_LegalBERTimbau (b), FTBERTimbau_JurisBERT (c), and FTBERTimbau_BERTimbauLaw (d) with the Preliminary Search corpus.



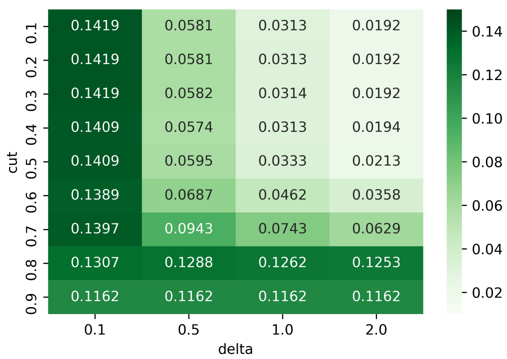
(a) BERTimbau.



(b) Legal-BERTimbau



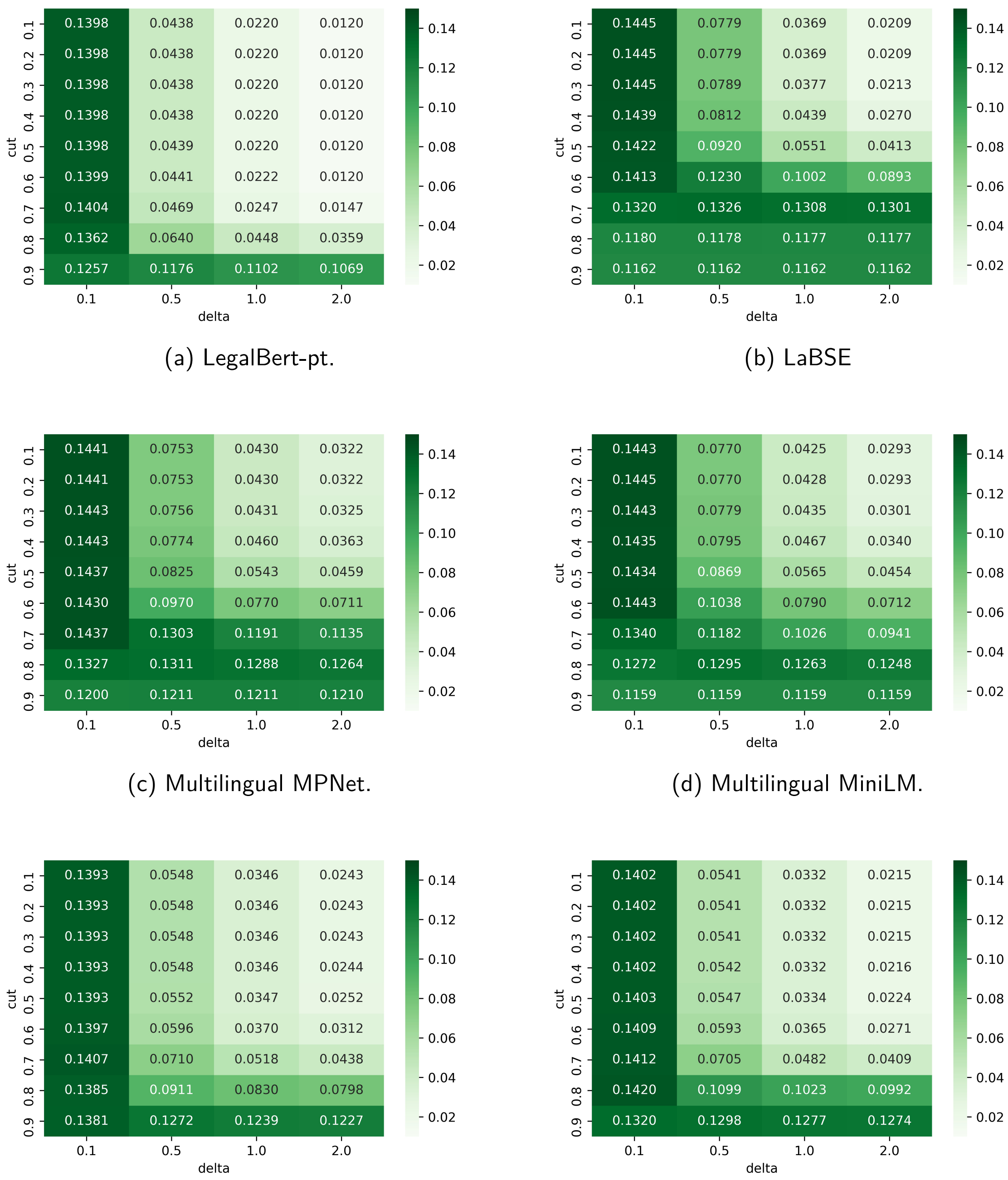
(c) JurisBERT.



(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 122 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTBERTimbau_LegalBertpt (a), FTBERTimbau_LaBSE (b), FTBERTimbau_MPNet (c), FTBERTimbau_MiniLM (d), FTBERTimbau_FTBERTimbau (e), and FTBERTimbau_FTLegalBertpt with the Preliminary Search corpus.



(a) LegalBert-pt.

(b) LaBSE

(c) Multilingual MPNet.

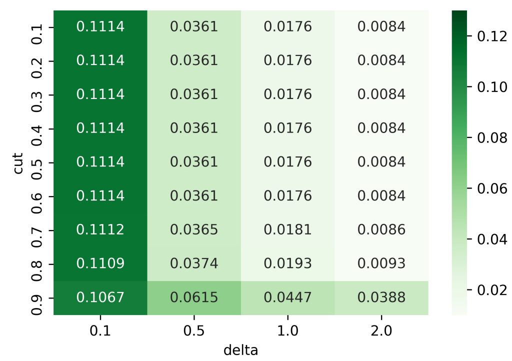
(d) Multilingual MiniLM.

(e) FT BERTimbau.

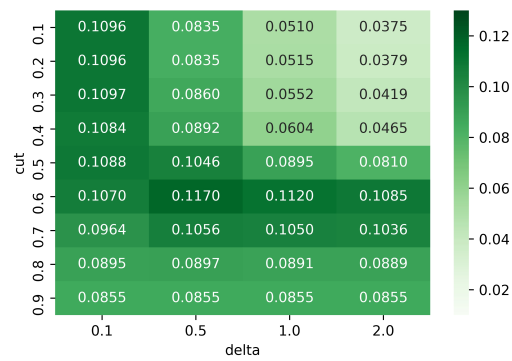
(f) FT LegalBert-pt.

Source: Created by the author (2025)

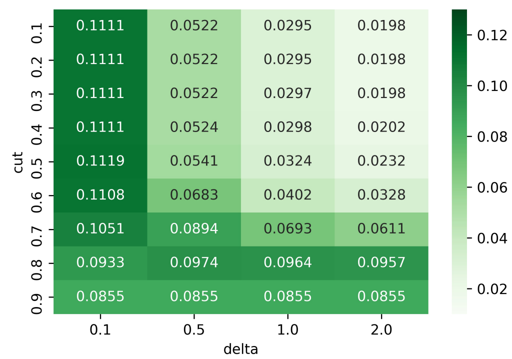
Figure 123 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTLegalBertpt_BERTimbau (a), FTLegalBertpt_LegalBERTimbau (b), FTLegalBertpt_JurisBERT (c), and FTLegalBertpt_BERTimbauLaw (d) with the Preliminary Search corpus.



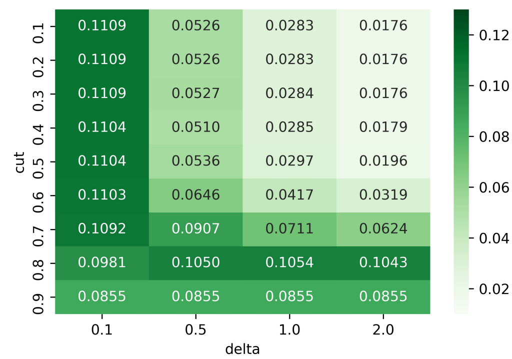
(a) BERTimbau.



(b) Legal-BERTimbau



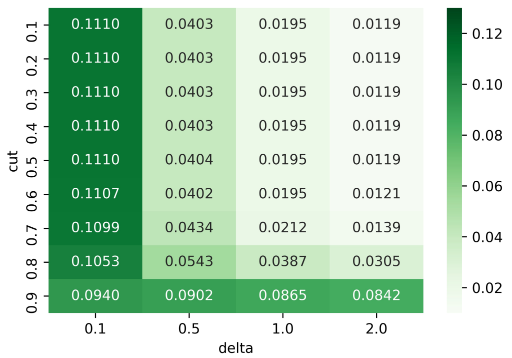
(c) JurisBERT.



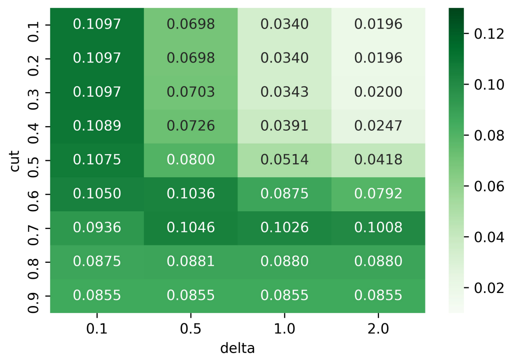
(d) BERTimbauLaw

Source: Created by the author (2025)

Figure 124 – Heatmaps presenting the MAP result achieved by each combination of parameters' values for FTLegalBertpt_LegalBertpt (a), FTLegalBertpt_LaBSE (b), FTLegalBertpt_MPNet (c), FTLegalBertpt_MiniLM (d), FTLegalBertpt_FTBERTimbau (e), and FTLegalBertpt_FTLegalBertpt with the Preliminary Search corpus.



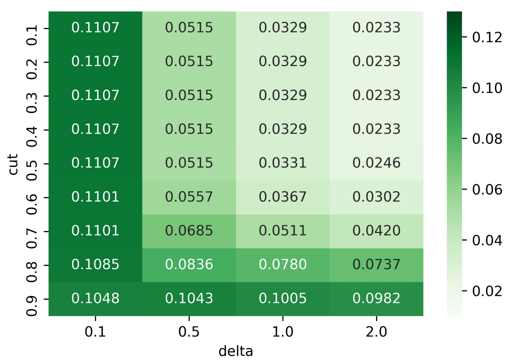
(a) LegalBert-pt.



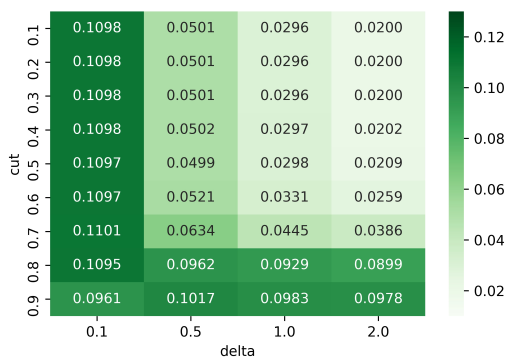
(b) LaBSE

(c) Multilingual MPNet.

(d) Multilingual MiniLM.



(e) FT BERTimbau.



(f) FT LegalBert-pt.

Source: Created by the author (2025)

APPENDIX L – RESULTS ACHIEVED BY THE BASELINES AND ULYSSES-RFSQ-OR

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
BM25L_PRE (baseline)	0.7684	0.6850	0.8597	0.8277	0.1335	0.0988	0.2115
BM25L_PRE_PRE	0.7656	0.6848	0.8583	0.8247	0.1626	0.1243	0.2567
BM25L_PRE_NP	0.7685	0.6863	0.8600	0.8277	0.1698	0.1319	0.2639
BM25L_PRE_BERTimbau	0.7641	0.6861	0.8612	0.8248	0.1485	0.1111	0.2345
BM25L_PRE_LegalBERTimbau	0.7674	**0.6889**	0.8608	0.8261	0.1711	0.1364	0.2774
BM25L_PRE_JurisBERT	0.7688	0.6865	0.8607	0.8279	0.1423	0.1069	0.2262
BM25L_PRE_BERTimbauLaw	0.7647	0.6859	**0.8621**	0.8263	0.1592	0.1243	0.2502
BM25L_PRE_LegalBertpt	0.7628	0.6837	0.8618	0.8242	0.1406	0.1066	0.2239
BM25L_PRE_LaBSE	0.7685	0.6873	0.8613	0.8278	0.1565	0.1229	0.2509
BM25L_PRE_MPNet	**0.7699**	0.6879	0.8613	**0.8287**	0.1670	0.1280	0.2626
BM25L_PRE_MiniLM	0.7683	**0.6889**	0.8615	0.8280	0.1643	0.1285	0.2611
BM25L_PRE_FTBERTimbau	0.7641	0.6854	0.8619	0.8256	**0.1741**	**0.1384**	**0.2777**
BM25L_PRE_FTLegalBertpt	0.7688	0.6869	0.8610	0.8277	0.1704	0.1320	0.2691
BM25L_NP (baseline)	0.3801	0.3913	0.6811	0.5429	0.1107	0.0871	0.1843
BM25L_NP_PRE	0.3847	0.3950	0.6807	0.5485	0.1402	0.1126	0.2319
BM25L_NP_NP	0.3818	0.3916	0.6828	0.5438	0.1442	0.1203	0.2433
BM25L_NP_BERTimbau	0.3836	0.3946	0.6817	0.5465	0.1235	0.1051	0.2132
BM25L_NP_LegalBERTimbau	0.3805	0.3905	0.6836	0.5500	0.1509	0.1230	0.2449
BM25L_NP_JurisBERT	0.3848	0.3938	0.6827	0.5482	0.1374	0.1129	0.2330
BM25L_NP_BERTimbauLaw	0.3839	0.3938	0.6835	0.5465	0.1374	0.1143	0.2307
BM25L_NP_LegalBertpt	0.3843	0.3930	0.6864	0.5476	0.1170	0.0988	0.2064
BM25L_NP_LaBSE	0.3853	0.3952	**0.6866**	0.5486	0.1289	0.1029	0.2101
BM25L_NP_MPNet	0.3853	**0.3969**	0.6807	0.5486	0.1192	0.0940	0.1990
BM25L_NP_MiniLM	0.3783	0.3912	0.6766	0.5434	0.1168	0.0967	0.2055
BM25L_NP_FTBERTimbau	**0.3856**	0.3953	0.6842	0.5487	**0.1522**	**0.1257**	**0.2530**
BM25L_NP_FTLegalBertpt	0.3852	0.3943	0.6853	0.5489	0.1513	0.1247	0.2500
OkapiBM25_PRE (baseline)	0.7183	0.6609	0.8607	0.7942	0.1338	0.0994	0.2126

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
OkapiBM25_PRE_PRE	0.7195	0.6607	0.8611	0.7952	0.1633	0.1254	0.2573
OkapiBM25_PRE_NP	0.7204	0.6622	0.8620	0.7955	0.1713	0.1342	0.2664
OkapiBM25_PRE_BERTimbau	0.7185	0.6610	0.8627	0.7944	0.1494	0.1120	0.2374
OkapiBM25_PRE_LegalBERTimbau	0.7196	0.6612	**0.8632**	0.7951	0.1659	0.1295	0.2618
OkapiBM25_PRE_JurisBERT	0.7201	0.6599	0.8624	0.7957	0.1420	0.1065	0.2264
OkapiBM25_PRE_BERTimbauLaw	0.7149	0.6582	0.8603	0.7938	0.1579	0.1258	0.2487
OkapiBM25_PRE_LegalBertpt	0.7170	0.6587	0.8612	0.7934	0.1411	0.1087	0.2241
OkapiBM25_PRE_LaBSE	0.7201	0.6607	0.8620	0.7957	0.1615	0.1253	0.2534
OkapiBM25_PRE_MPNet	**0.7212**	**0.6624**	0.8630	0.7964	0.1665	0.1285	0.2622
OkapiBM25_PRE_MiniLM	0.7204	0.6603	0.8623	**0.7967**	0.1648	0.1302	0.2623
OkapiBM25_PRE_FTBERTimbau	0.7176	0.6600	0.8621	0.7952	**0.1722**	**0.1352**	**0.2732**
OkapiBM25_PRE_FTLegalBertpt	0.7171	0.6604	0.8613	0.7954	0.1702	0.1343	0.2681
OkapiBM25_NP (baseline)	0.3704	0.3834	0.6718	0.5399	0.1028	0.0840	0.1766
OkapiBM25_NP_PRE	0.3715	0.3846	0.6731	0.5403	0.1396	0.1150	0.2279
OkapiBM25_NP_NP	0.3701	0.3837	0.6691	0.5395	0.1281	0.1053	0.2088
OkapiBM25_NP_BERTimbau	0.3711	0.3833	0.6707	0.5412	0.1171	0.0979	0.2057
OkapiBM25_NP_LegalBERTimbau	0.3748	**0.3885**	0.6737	**0.5452**	0.1170	0.0956	0.1996
OkapiBM25_NP_JurisBERT	**0.3753**	0.3876	0.6723	0.5434	0.1102	0.0888	0.1858
OkapiBM25_NP_BERTimbauLaw	0.3716	0.3840	0.6711	0.5425	0.1311	0.1101	0.2159
OkapiBM25_NP_LegalBertpt	0.3707	0.3825	0.6718	0.5414	0.1180	0.0975	0.2068
OkapiBM25_NP_LaBSE	0.3729	0.3840	0.6686	0.5433	0.1216	0.0979	0.1990
OkapiBM25_NP_MPNet	0.3705	0.3831	0.6602	0.5423	0.1375	0.1162	0.2326
OkapiBM25_NP_MiniLM	0.3735	0.3860	0.6719	0.5445	0.1292	0.1081	0.2223
OkapiBM25_NP_FTBERTimbau	0.3712	0.3856	**0.6738**	0.5436	**0.1502**	**0.1219**	**0.2467**
OkapiBM25_NP_FTLegalBertpt	0.3728	0.3859	0.6731	0.5448	0.1368	0.1102	0.2241
BERTimbau (baseline)	0.0113	0.0209	0.0542	0.0490	0.0188	0.0164	0.0438
BERTimbau_PRE	**0.0504**	**0.0602**	0.1094	**0.1076**	0.0851	0.0776	0.1490
BERTimbau_NP	0.0452	0.0570	**0.1109**	0.1057	0.0902	0.0857	0.1560
BERTimbau_BERTimbau	0.0310	0.0403	0.0802	0.0800	0.0346	0.0330	0.0751
BERTimbau_LegalBERTimbau	0.0419	0.0537	0.1044	0.1047	**0.0961**	**0.0881**	**0.1679**

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
BERTimbau_JurisBERT	0.0227	0.0321	0.0685	0.0639	0.0479	0.0430	0.0893
BERTimbau_BERTimbauLaw	0.0278	0.0403	0.0923	0.0830	0.0716	0.0679	0.1276
BERTimbau_LegalBertpt	0.0238	0.0315	0.0597	0.0604	0.0393	0.0355	0.0689
BERTimbau_LaBSE	0.0310	0.0382	0.0781	0.0779	0.0657	0.0597	0.1183
BERTimbau_MPNet	0.0412	0.0543	0.1040	0.1010	0.0866	0.0794	0.1555
BERTimbau_MiniLM	0.0365	0.0438	0.0933	0.0881	0.0598	0.0543	0.1181
BERTimbau_FTBERTimbau	0.0353	0.0461	0.0983	0.0943	0.0807	0.0732	0.1479
BERTimbau_FTLegalBertpt	0.0430	0.0525	0.1071	0.1029	0.0824	0.0754	0.1511
LegalBERTimbau (baseline)	0.1221	0.1469	0.3320	0.2871	0.0886	0.0720	0.1613
LegalBERTimbau_PRE	0.1378	0.1610	0.3579	0.3121	0.1347	0.1156	0.2292
LegalBERTimbau_NP	**0.1431**	**0.1678**	**0.3641**	**0.3192**	**0.1472**	**0.1244**	**0.2523**
LegalBERTimbau_BERTimbau	0.1283	0.1529	0.3386	0.3004	0.1172	0.0975	0.2137
LegalBERTimbau_LegalBERTimbau	0.1284	0.1499	0.3136	0.3006	0.1398	0.1174	0.2314
LegalBERTimbau_JurisBERT	0.1270	0.1528	0.3362	0.2961	0.1186	0.0992	0.2132
LegalBERTimbau_BERTimbauLaw	0.1321	0.1583	0.3426	0.3033	0.1230	0.1064	0.2099
LegalBERTimbau_LegalBertpt	0.1240	0.1506	0.3270	0.2929	0.1192	0.1003	0.2165
LegalBERTimbau_LaBSE	0.1274	0.1535	0.3362	0.2994	0.1208	0.1021	0.2032
LegalBERTimbau_MPNet	0.1336	0.1560	0.3495	0.3026	0.1358	0.1158	0.2310
LegalBERTimbau_MiniLM	0.1264	0.1509	0.3341	0.2952	0.1202	0.1000	0.2154
LegalBERTimbau_FTBERTimbau	0.1282	0.1532	0.3367	0.3030	0.1273	0.1065	0.2302
LegalBERTimbau_FTLegalBertpt	0.1304	0.1548	0.3426	0.3034	0.1239	0.1025	0.2216
JurisBERT (baseline)	0.0744	0.0924	0.2482	0.1862	0.0425	0.0331	0.0849
JurisBERT_PRE	0.0892	0.1070	0.2482	0.2111	0.0990	0.0858	0.1760
JurisBERT_NP	**0.0905**	**0.1130**	0.2504	**0.2183**	0.1065	0.0930	0.1929
JurisBERT_BERTimbau	0.0822	0.0997	0.2588	0.2017	0.0625	0.0540	0.1286
JurisBERT_LegalBERTimbau	0.0848	0.1041	**0.2671**	0.2040	**0.1133**	**0.1000**	0.1956
JurisBERT_JurisBERT	0.0757	0.0930	0.2389	0.1874	0.0611	0.0492	0.1128
JurisBERT_BERTimbauLaw	0.0798	0.0952	0.2464	0.2007	0.0890	0.0794	0.1569
JurisBERT_LegalBertpt	0.0775	0.0944	0.2443	0.1918	0.0629	0.0528	0.1108
JurisBERT_LaBSE	0.0788	0.0948	0.2517	0.1968	0.0881	0.0757	0.1579

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
JurisBERT_MPNet	0.0798	0.1006	0.2496	0.2022	0.1059	0.0947	0.1865
JurisBERT_MiniLM	0.0783	0.0987	0.2467	0.2015	0.0824	0.0723	0.1520
JurisBERT_FTBERTimbau	0.0856	0.1062	0.2487	0.2078	0.1128	0.0965	**0.1966**
JurisBERT_FTLegalBertpt	0.0816	0.0996	0.2521	0.2066	0.0930	0.0833	0.1653
BERTimbauLaw (baseline)	0.1239	0.1571	0.3481	0.2871	0.0801	0.0659	0.1522
BERTimbauLaw_PRE	**0.1462**	**0.1794**	**0.3720**	**0.3218**	0.1273	0.1089	0.2236
BERTimbauLaw_NP	0.1367	0.1726	0.3584	0.3073	0.1347	0.1165	**0.2335**
BERTimbauLaw_BERTimbau	0.1309	0.1660	0.3543	0.3014	0.0973	0.0832	0.1867
BERTimbauLaw_LegalBERTimbau	0.1337	0.1680	0.3615	0.3012	**0.1367**	**0.1182**	0.2294
BERTimbauLaw_JurisBERT	0.1328	0.1662	0.3447	0.2955	0.1075	0.0941	0.2009
BERTimbauLaw_BERTimbauLaw	0.1301	0.1647	0.3537	0.2981	0.1109	0.0950	0.1907
BERTimbauLaw_LegalBertpt	0.1301	0.1648	0.3474	0.2983	0.1038	0.0891	0.1980
BERTimbauLaw_LaBSE	0.1325	0.1670	0.3560	0.2985	0.1049	0.0876	0.1855
BERTimbauLaw_MPNet	0.1356	0.1681	0.3552	0.3016	0.1230	0.1075	0.2111
BERTimbauLaw_MiniLM	0.1332	0.1670	0.3600	0.2961	0.1198	0.1052	0.2208
BERTimbauLaw_FTBERTimbau	0.1298	0.1642	0.3470	0.2985	0.1355	0.1151	0.2284
BERTimbauLaw_FTLegalBertpt	0.1304	0.1654	0.3490	0.3060	0.1221	0.1060	0.2233
LegalBertpt (baseline)	0.0482	0.0670	0.1792	0.1394	0.0444	0.0397	0.0924
LegalBertpt_PRE	0.0672	**0.0888**	0.1875	0.1667	0.0944	0.0867	0.1718
LegalBertpt_NP	0.0640	0.0854	0.1909	0.1687	0.0843	0.0771	0.1463
LegalBertpt_BERTimbau	0.0551	0.0713	0.1600	0.1516	0.0558	0.0515	0.1094
LegalBertpt_LegalBERTimbau	**0.0706**	0.0887	**0.2009**	**0.1751**	0.1087	0.0955	0.1857
LegalBertpt_JurisBERT	0.0547	0.0734	0.1763	0.1452	0.0738	0.0671	0.1464
LegalBertpt_BERTimbauLaw	0.0572	0.0757	0.1815	0.1593	0.0856	0.0782	0.1502
LegalBertpt_LegalBertpt	0.0473	0.0631	0.1351	0.1328	0.0554	0.0494	0.0985
LegalBertpt_LaBSE	0.0551	0.0714	0.1651	0.1460	0.0825	0.0733	0.1539
LegalBertpt_MPNet	0.0670	0.0847	0.1837	0.1638	0.0963	0.0858	0.1712
LegalBertpt_MiniLM	0.0631	0.0800	0.1934	0.1620	0.0835	0.0748	0.1593
LegalBertpt_FTBERTimbau	0.0626	0.0847	0.1944	0.1661	**0.1136**	**0.1018**	**0.1922**
LegalBertpt_FTLegalBertpt	0.0637	0.0822	0.1763	0.1644	0.0849	0.0774	0.1537

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
LaBSE (baseline)	0.0893	0.1080	0.2910	0.2268	0.0685	0.0549	0.1264
LaBSE_PRE	**0.1169**	**0.1368**	**0.3203**	**0.2655**	0.1159	0.0998	0.2022
LaBSE_NP	0.1114	0.1359	0.2984	0.2555	0.1003	0.0860	0.1685
LaBSE_BERTimbau	0.1010	0.1210	0.3056	0.2440	0.0972	0.0840	0.1813
LaBSE_LegalBERTimbau	0.1018	0.1236	0.2814	0.2434	0.1252	0.1092	0.2090
LaBSE_JurisBERT	0.1016	0.1213	0.3022	0.2485	0.0934	0.0803	0.1747
LaBSE_BERTimbauLaw	0.1029	0.1265	0.3067	0.2502	0.0992	0.0836	0.1674
LaBSE_LegalBertpt	0.0983	0.1193	0.2973	0.2412	0.0973	0.0831	0.1803
LaBSE_LaBSE	0.0997	0.1200	0.2988	0.2397	0.0935	0.0798	0.1607
LaBSE_MPNet	0.1052	0.1244	0.3175	0.2463	0.1148	0.0991	0.1976
LaBSE_MiniLM	0.1052	0.1253	0.3108	0.2483	0.1033	0.0886	0.1867
LaBSE_FTBERTimbau	0.1030	0.1263	0.3107	0.2547	**0.1321**	**0.1116**	**0.2175**
LaBSE_FTLegalBertpt	0.1054	0.1266	0.3119	0.2527	0.1117	0.0953	0.1996
MPNet (baseline)	0.1633	0.1891	0.4204	0.3491	0.0933	0.0752	0.1696
MPNet_PRE	**0.1793**	**0.2053**	**0.4342**	**0.3669**	0.1348	0.1135	0.2368
MPNet_NP	0.1730	0.1997	0.4253	0.3619	0.1443	0.1210	**0.2565**
MPNet_BERTimbau	0.1656	0.1926	0.4161	0.3559	0.1180	0.1003	0.2204
MPNet_LegalBERTimbau	0.1698	0.1975	0.4218	0.3578	0.1452	0.1212	0.2491
MPNet_JurisBERT	0.1690	0.1948	0.4217	0.3560	0.1239	0.1039	0.2298
MPNet_BERTimbauLaw	0.1687	0.1948	0.4309	0.3581	0.1123	0.0948	0.1909
MPNet_LegalBertpt	0.1651	0.1914	0.4158	0.3539	0.1199	0.1030	0.2229
MPNet_LaBSE	0.1675	0.1950	0.4211	0.3537	0.1166	0.0979	0.2047
MPNet_MPNet	0.1722	0.1970	0.4230	0.3582	0.1245	0.1026	0.2170
MPNet_MiniLM	0.1690	0.1957	0.4245	0.3525	0.1144	0.0942	0.1953
MPNet_FTBERTimbau	0.1669	0.1942	0.4201	0.3591	**0.1496**	**0.1274**	0.2478
MPNet_FTLegalBertpt	0.1651	0.1905	0.4215	0.3529	0.1303	0.1071	0.2230
MiniLM (baseline)	0.1208	0.1454	0.3376	0.2692	0.0695	0.0598	0.1378
MiniLM_PRE	0.1282	0.1530	0.3255	0.2789	0.1166	0.1009	0.2054
MiniLM_NP	0.1302	0.1552	**0.3509**	0.2813	0.1028	0.0912	0.1818
MiniLM_BERTimbau	0.1258	0.1502	0.3478	0.2740	0.0904	0.0782	0.1796

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
MiniLM_LegalBERTimbau	0.1245	0.1494	0.3401	0.2755	**0.1241**	**0.1093**	**0.2200**
MiniLM_JurisBERT	0.1272	0.1501	0.3504	0.2785	0.0832	0.0720	0.1515
MiniLM_BERTimbauLaw	0.1254	0.1494	0.3508	0.2768	0.0955	0.0854	0.1672
MiniLM_LegalBertpt	0.1228	0.1486	0.3386	0.2714	0.0812	0.0721	0.1494
MiniLM_LaBSE	0.1195	0.1408	0.3193	0.2697	0.0999	0.0840	0.1778
MiniLM_MPNet	**0.1338**	**0.1591**	0.3471	**0.2825**	0.1037	0.0900	0.1855
MiniLM_MiniLM	0.1206	0.1441	0.3300	0.2682	0.0904	0.0790	0.1651
MiniLM_FTBERTimbau	0.1259	0.1498	0.3497	0.2758	0.1115	0.0997	0.1990
MiniLM_FTLegalBertpt	0.1285	0.1527	0.3433	0.2759	0.1080	0.0930	0.1928
FTBERTimbau (baseline)	0.1902	0.2163	0.4456	0.4070	0.1214	0.0941	0.2024
FTBERTimbau_PRE	**0.2024**	0.2281	0.4664	0.4186	0.1703	0.1380	0.2725
FTBERTimbau_NP	**0.2024**	**0.2310**	**0.4669**	**0.4202**	**0.1839**	**0.1505**	0.2920
FTBERTimbau_BERTimbau	0.1839	0.2109	0.4152	0.4065	0.1561	0.1259	0.2589
FTBERTimbau_LegalBERTimbau	0.1907	0.2180	0.4301	0.4091	0.1792	0.1466	**0.2931**
FTBERTimbau_JurisBERT	0.1927	0.2172	0.4473	0.4098	0.1695	0.1384	0.2788
FTBERTimbau_BERTimbauLaw	0.1900	0.2161	0.4401	0.4111	0.1679	0.1357	0.2774
FTBERTimbau_LegalBertpt	0.1852	0.2132	0.4226	0.4084	0.1638	0.1323	0.2713
FTBERTimbau_LaBSE	0.1949	0.2212	0.4438	0.4134	0.1754	0.1438	0.2869
FTBERTimbau_MPNet	0.1963	0.2222	0.4473	0.4124	0.1743	0.1421	0.2860
FTBERTimbau_MiniLM	0.1883	0.2159	0.4328	0.4070	0.1733	0.1408	0.2849
FTBERTimbau_FTBERTimbau	0.1823	0.2072	0.4146	0.4023	0.1645	0.1350	0.2720
FTBERTimbau_FTLegalBertpt	0.1938	0.2184	0.4424	0.4104	0.1705	0.1387	0.2736
FTLegalBertpt (baseline)	0.1220	0.1523	0.2998	0.3050	0.0890	0.0639	0.1569
FTLegalBertpt_PRE	**0.1367**	**0.1650**	**0.3118**	0.3128	0.1424	0.1163	0.2343
FTLegalBertpt_NP	0.1285	0.1567	0.3024	0.3098	**0.1637**	**0.1338**	**0.2694**
FTLegalBertpt_BERTimbau	0.1286	0.1588	0.3013	0.3123	0.1369	0.1147	0.2357
FTLegalBertpt_LegalBERTimbau	0.1342	0.1639	0.3115	**0.3175**	0.1483	0.1205	0.2395
FTLegalBertpt_JurisBERT	0.1289	0.1567	0.2969	0.3168	0.1473	0.1237	0.2538
FTLegalBertpt_BERTimbauLaw	0.1294	0.1572	0.3066	0.3156	0.1472	0.1240	0.2529
FTLegalBertpt_LegalBertpt	0.1288	0.1591	0.3004	0.3138	0.1412	0.1189	0.2423

	Ulysses-RFCorpus				Preliminary Search		
Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**	**MAP**	**MRP**	**MRR**
FTLegalBertpt_LaBSE	0.1333	0.1625	0.3104	0.3134	0.1511	0.1259	0.2577
FTLegalBertpt_MPNet	0.1222	0.1522	0.2707	0.2843	0.1527	0.1237	0.2619
FTLegalBertpt_MiniLM	0.1232	0.1528	0.2978	0.3032	0.1519	0.1260	0.2605
FTLegalBertpt_FTBERTimbau	0.1292	0.1586	0.3079	0.3058	0.1445	0.1220	0.2475
FTLegalBertpt_FTLegalBertpt	0.1290	0.1594	0.3086	0.3093	0.1449	0.1202	0.2502

APPENDIX M – STUDENT'S T-TEST RESULTS FOR THE COMPARISON BETWEEN THE ALGORITHMS USED TO SEARCH FOR THE SIMILAR QUERIES FROM ULYSSES-RFCORPUS

Figure 125 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using BM25_PRE. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.828	0.917	0.950	0.912	0.812	0.813	0.827	0.839	0.893	0.842	0.751
NP	0.828	1.000	0.748	0.779	0.743	0.983	0.985	0.999	0.674	0.933	0.986	0.921
BERTimbau	0.917	0.748	1.000	0.967	0.995	0.732	0.734	0.746	0.921	0.811	0.761	0.673
Legal-BERTimbau	0.950	0.779	0.967	1.000	0.962	0.763	0.764	0.777	0.888	0.843	0.792	0.703
JurisBERT	0.912	0.743	0.995	0.962	1.000	0.727	0.728	0.741	0.926	0.806	0.756	0.668
BERTimbauLaw	0.812	0.983	0.732	0.763	0.727	1.000	0.998	0.985	0.659	0.917	0.969	0.937
LegalBert-pt	0.813	0.985	0.734	0.764	0.728	0.998	1.000	0.986	0.661	0.918	0.971	0.936
LaBSE	0.827	0.999	0.746	0.777	0.741	0.985	0.986	1.000	0.673	0.932	0.985	0.922
Multilingual MPNet	0.839	0.674	0.921	0.888	0.926	0.659	0.661	0.673	1.000	0.736	0.687	0.603
Multilingual MiniLM	0.893	0.933	0.811	0.843	0.806	0.917	0.918	0.932	0.736	1.000	0.947	0.854
FT BERTimbau	0.842	0.986	0.761	0.792	0.756	0.969	0.971	0.985	0.687	0.947	1.000	0.906
FT LegalBert-pt	0.751	0.921	0.673	0.703	0.668	0.937	0.936	0.922	0.603	0.854	0.906	1.000

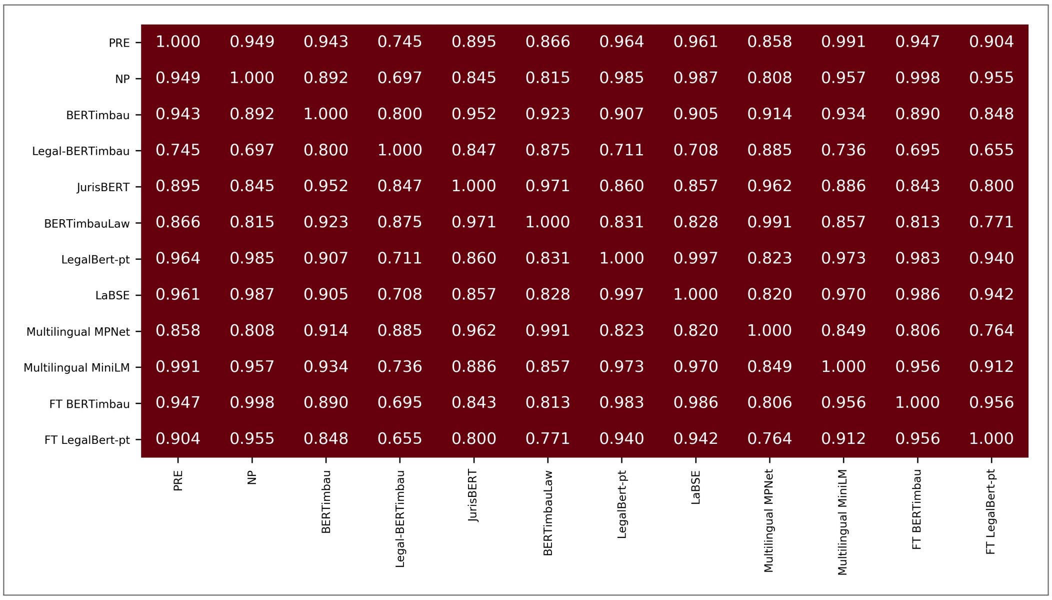
Source: Created by the author (2025)

Figure 126 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using BM25_NP. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.861	0.949	0.960	0.958	0.976	0.993	0.971	0.982	0.799	0.699	0.969
NP	0.861	1.000	0.911	0.900	0.820	0.837	0.853	0.832	0.878	0.937	0.833	0.831
BERTimbau	0.949	0.911	1.000	0.989	0.907	0.925	0.942	0.920	0.967	0.848	0.746	0.919
Legal-BERTimbau	0.960	0.900	0.989	1.000	0.918	0.936	0.953	0.931	0.978	0.837	0.736	0.929
JurisBERT	0.958	0.820	0.907	0.918	1.000	0.982	0.965	0.987	0.940	0.758	0.660	0.989
BERTimbauLaw	0.976	0.837	0.925	0.936	0.982	1.000	0.983	0.995	0.958	0.775	0.676	0.993
LegalBert-pt	0.993	0.853	0.942	0.953	0.965	0.983	1.000	0.978	0.975	0.792	0.692	0.976
LaBSE	0.971	0.832	0.920	0.931	0.987	0.995	0.978	1.000	0.953	0.770	0.671	0.998
Multilingual MPNet	0.982	0.878	0.967	0.978	0.940	0.958	0.975	0.953	1.000	0.815	0.714	0.951
Multilingual MiniLM	0.799	0.937	0.848	0.837	0.758	0.775	0.792	0.770	0.815	1.000	0.895	0.770
FT BERTimbau	0.699	0.833	0.746	0.736	0.660	0.676	0.692	0.671	0.714	0.895	1.000	0.671
FT LegalBert-pt	0.969	0.831	0.919	0.929	0.989	0.993	0.976	0.998	0.951	0.770	0.671	1.000

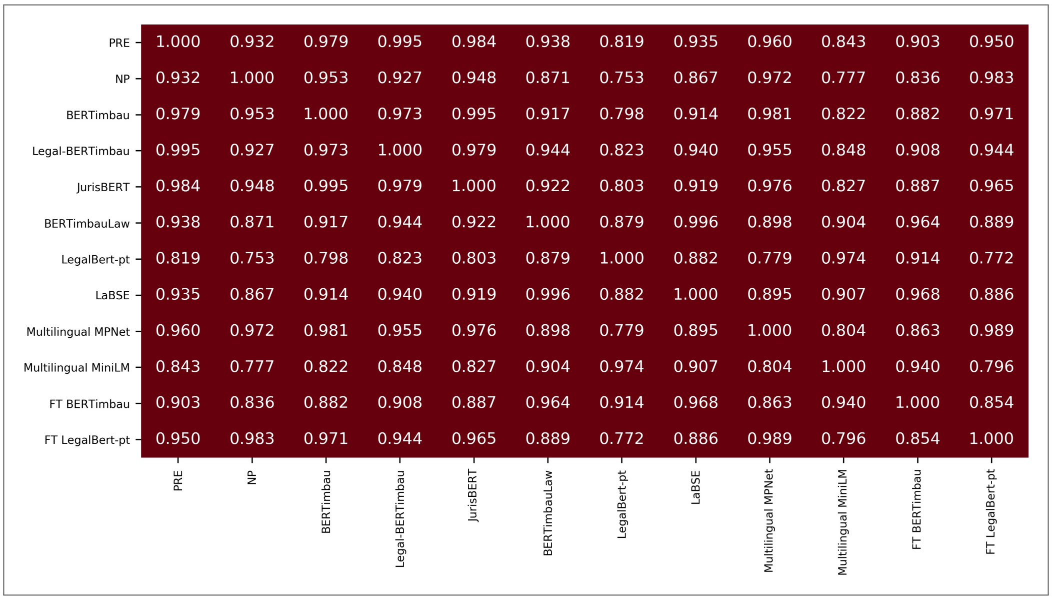
Source: Created by the author (2025)

Figure 127 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using OkapiBM25_PRE. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 128 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using OkapiBM25_NP. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 129 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using BERTimbau. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.537	0.013	0.002	0.057	0.383	0.000	0.014	0.000	0.305	0.085	0.266
NP	0.537	1.000	0.053	0.010	0.182	0.783	0.001	0.053	0.002	0.669	0.250	0.603
BERTimbau	0.013	0.053	1.000	0.592	0.527	0.105	0.170	0.992	0.240	0.135	0.424	0.161
Legal-BERTimbau	0.002	0.010	0.592	1.000	0.223	0.027	0.346	0.603	0.466	0.036	0.166	0.045
JurisBERT	0.057	0.182	0.527	0.223	1.000	0.303	0.041	0.523	0.067	0.372	0.859	0.425
BERTimbauLaw	0.383	0.783	0.105	0.027	0.303	1.000	0.003	0.105	0.006	0.884	0.395	0.813
LegalBert-pt	0.000	0.001	0.170	0.346	0.041	0.003	1.000	0.177	0.849	0.004	0.029	0.006
LaBSE	0.014	0.053	0.992	0.603	0.523	0.105	0.177	1.000	0.248	0.135	0.421	0.161
Multilingual MPNet	0.000	0.002	0.240	0.466	0.067	0.006	0.849	0.248	1.000	0.008	0.047	0.010
Multilingual MiniLM	0.305	0.669	0.135	0.036	0.372	0.884	0.004	0.135	0.008	1.000	0.477	0.927
FT BERTimbau	0.085	0.250	0.424	0.166	0.859	0.395	0.029	0.421	0.047	0.477	1.000	0.536
FT LegalBert-pt	0.266	0.603	0.161	0.045	0.425	0.813	0.006	0.161	0.010	0.927	0.536	1.000

Source: Created by the author (2025)

Figure 130 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using Legal-BERTimbau. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.646	0.398	0.612	0.387	0.511	0.337	0.352	0.217	0.410	0.306	0.710
NP	0.646	1.000	0.193	0.336	0.186	0.266	0.157	0.166	0.091	0.203	0.139	0.409
BERTimbau	0.398	0.193	1.000	0.739	0.988	0.854	0.908	0.932	0.695	0.997	0.859	0.642
Legal-BERTimbau	0.612	0.336	0.739	1.000	0.727	0.882	0.654	0.676	0.470	0.747	0.610	0.894
JurisBERT	0.387	0.186	0.988	0.727	1.000	0.841	0.919	0.944	0.705	0.985	0.870	0.629
BERTimbauLaw	0.511	0.266	0.854	0.882	0.841	1.000	0.765	0.788	0.566	0.860	0.718	0.779
LegalBert-pt	0.337	0.157	0.908	0.654	0.919	0.765	1.000	0.975	0.782	0.907	0.951	0.562
LaBSE	0.352	0.166	0.932	0.676	0.944	0.788	0.975	1.000	0.759	0.931	0.926	0.583
Multilingual MPNet	0.217	0.091	0.695	0.470	0.705	0.566	0.782	0.759	1.000	0.698	0.829	0.394
Multilingual MiniLM	0.410	0.203	0.997	0.747	0.985	0.860	0.907	0.931	0.698	1.000	0.859	0.651
FT BERTimbau	0.306	0.139	0.859	0.610	0.870	0.718	0.951	0.926	0.829	0.859	1.000	0.521
FT LegalBert-pt	0.710	0.409	0.642	0.894	0.629	0.779	0.562	0.583	0.394	0.651	0.521	1.000

Source: Created by the author (2025)

Figure 131 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using JurisBERT. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.896	0.451	0.313	0.709	0.413	0.141	0.257	0.207	0.634	0.236	0.306
NP	0.896	1.000	0.372	0.250	0.613	0.337	0.106	0.202	0.160	0.539	0.184	0.244
BERTimbau	0.451	0.372	1.000	0.790	0.713	0.949	0.454	0.694	0.592	0.768	0.656	0.787
Legal-BERTimbau	0.313	0.250	0.790	1.000	0.532	0.838	0.630	0.901	0.786	0.573	0.859	1.000
JurisBERT	0.709	0.613	0.713	0.532	1.000	0.666	0.277	0.455	0.379	0.931	0.425	0.527
BERTimbauLaw	0.413	0.337	0.949	0.838	0.666	1.000	0.491	0.741	0.635	0.718	0.701	0.836
LegalBert-pt	0.141	0.106	0.454	0.630	0.277	0.491	1.000	0.718	0.837	0.294	0.758	0.624
LaBSE	0.257	0.202	0.694	0.901	0.455	0.741	0.718	1.000	0.881	0.489	0.957	0.900
Multilingual MPNet	0.207	0.160	0.592	0.786	0.379	0.635	0.837	0.881	1.000	0.405	0.922	0.783
Multilingual MiniLM	0.634	0.539	0.768	0.573	0.931	0.718	0.294	0.489	0.405	1.000	0.456	0.567
FT BERTimbau	0.236	0.184	0.656	0.859	0.425	0.701	0.758	0.957	0.922	0.456	1.000	0.857
FT LegalBert-pt	0.306	0.244	0.787	1.000	0.527	0.836	0.624	0.900	0.783	0.567	0.857	1.000

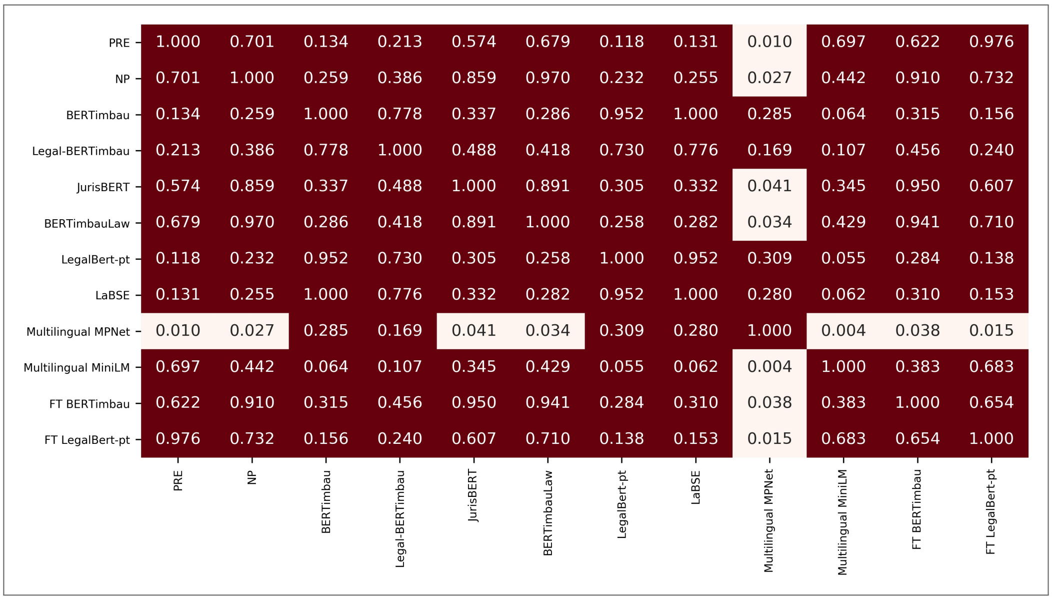
Source: Created by the author (2025)

Figure 132 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using BERTimbauLaw. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.419	0.183	0.158	0.158	0.171	0.257	0.238	0.161	0.281	0.266	0.371
NP	0.419	1.000	0.599	0.545	0.538	0.571	0.733	0.708	0.549	0.787	0.757	0.925
BERTimbau	0.183	0.599	1.000	0.940	0.923	0.966	0.861	0.881	0.940	0.797	0.831	0.670
Legal-BERTimbau	0.158	0.545	0.940	1.000	0.982	0.974	0.802	0.821	1.000	0.738	0.772	0.615
JurisBERT	0.158	0.538	0.923	0.982	1.000	0.956	0.788	0.807	0.982	0.725	0.759	0.605
BERTimbauLaw	0.171	0.571	0.966	0.974	0.956	1.000	0.829	0.849	0.973	0.765	0.799	0.641
LegalBert-pt	0.257	0.733	0.861	0.802	0.788	0.829	1.000	0.977	0.804	0.939	0.972	0.807
LaBSE	0.238	0.708	0.881	0.821	0.807	0.849	0.977	1.000	0.823	0.915	0.949	0.782
Multilingual MPNet	0.161	0.549	0.940	1.000	0.982	0.973	0.804	0.823	1.000	0.740	0.773	0.618
Multilingual MiniLM	0.281	0.787	0.797	0.738	0.725	0.765	0.939	0.915	0.740	1.000	0.967	0.863
FT BERTimbau	0.266	0.757	0.831	0.772	0.759	0.799	0.972	0.949	0.773	0.967	1.000	0.831
FT LegalBert-pt	0.371	0.925	0.670	0.615	0.605	0.641	0.807	0.782	0.618	0.863	0.831	1.000

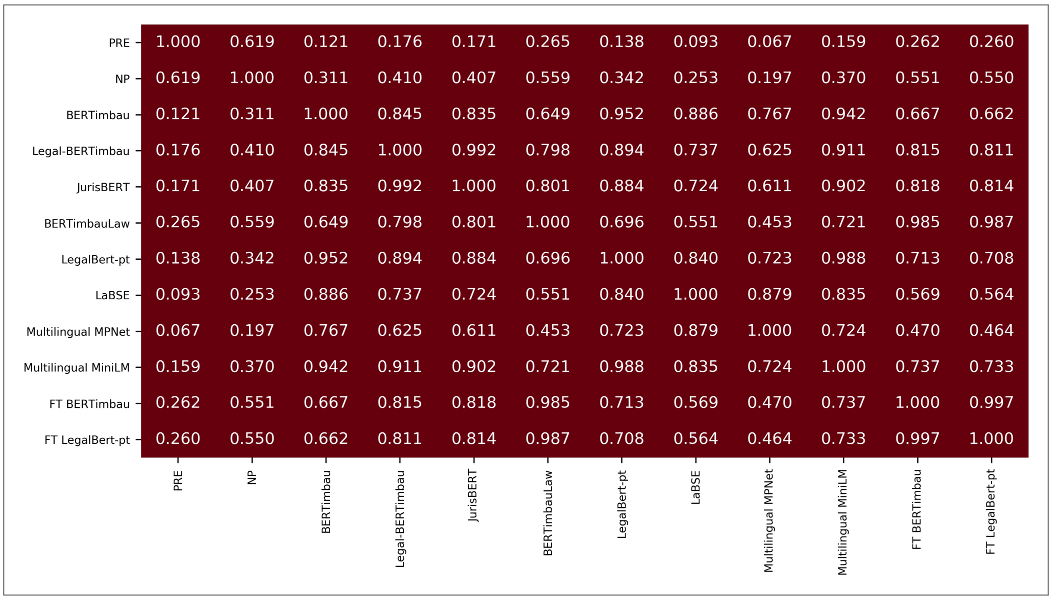
Source: Created by the author (2025)

Figure 133 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using LegalBert-pt. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 134 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using LaBSE. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 135 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using Multilingual MPNet. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.630	0.286	0.413	0.337	0.268	0.428	0.362	0.267	0.466	0.431	0.587
NP	0.630	1.000	0.559	0.738	0.632	0.531	0.756	0.666	0.531	0.805	0.758	0.949
BERTimbau	0.286	0.559	1.000	0.801	0.915	0.967	0.784	0.881	0.968	0.735	0.785	0.606
Legal-BERTimbau	0.413	0.738	0.801	1.000	0.884	0.768	0.982	0.920	0.769	0.931	0.982	0.789
JurisBERT	0.337	0.632	0.915	0.884	1.000	0.882	0.866	0.965	0.883	0.816	0.867	0.681
BERTimbauLaw	0.268	0.531	0.967	0.768	0.882	1.000	0.752	0.848	0.999	0.704	0.753	0.577
LegalBert-pt	0.428	0.756	0.784	0.982	0.866	0.752	1.000	0.902	0.752	0.949	1.000	0.807
LaBSE	0.362	0.666	0.881	0.920	0.965	0.848	0.902	1.000	0.849	0.852	0.902	0.715
Multilingual MPNet	0.267	0.531	0.968	0.769	0.883	0.999	0.752	0.849	1.000	0.704	0.754	0.577
Multilingual MiniLM	0.466	0.805	0.735	0.931	0.816	0.704	0.949	0.852	0.704	1.000	0.950	0.856
FT BERTimbau	0.431	0.758	0.785	0.982	0.867	0.753	1.000	0.902	0.754	0.950	1.000	0.809
FT LegalBert-pt	0.587	0.949	0.606	0.789	0.681	0.577	0.807	0.715	0.577	0.856	0.809	1.000

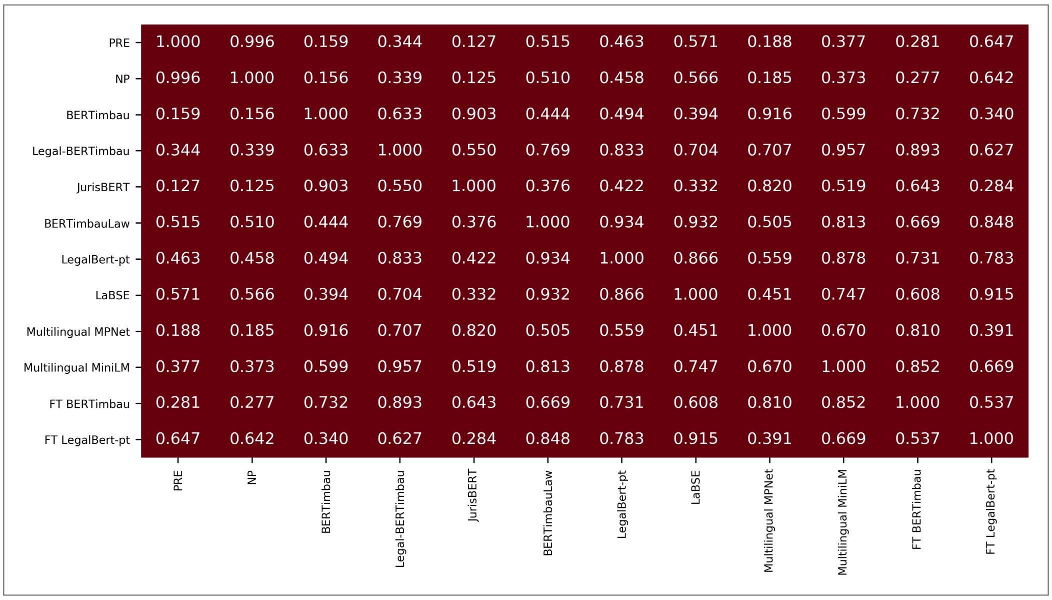
Source: Created by the author (2025)

Figure 136 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using Multilingual MiniLM. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.856	0.831	0.805	0.835	0.981	0.931	0.431	0.630	0.743	0.495	0.630
NP	0.856	1.000	0.689	0.663	0.693	0.875	0.787	0.325	0.501	0.605	0.381	0.758
BERTimbau	0.831	0.689	1.000	0.974	0.996	0.812	0.899	0.561	0.786	0.907	0.634	0.487
Legal-BERTimbau	0.805	0.663	0.974	1.000	0.969	0.786	0.872	0.581	0.811	0.933	0.656	0.464
JurisBERT	0.835	0.693	0.996	0.969	1.000	0.816	0.903	0.557	0.782	0.903	0.630	0.489
BERTimbauLaw	0.981	0.875	0.812	0.786	0.816	1.000	0.912	0.416	0.613	0.724	0.479	0.647
LegalBert-pt	0.931	0.787	0.899	0.872	0.903	0.912	1.000	0.480	0.691	0.807	0.548	0.569
LaBSE	0.431	0.325	0.561	0.581	0.557	0.416	0.480	1.000	0.756	0.642	0.918	0.207
Multilingual MPNet	0.630	0.501	0.786	0.811	0.782	0.613	0.691	0.756	1.000	0.877	0.836	0.337
Multilingual MiniLM	0.743	0.605	0.907	0.933	0.903	0.724	0.807	0.642	0.877	1.000	0.719	0.418
FT BERTimbau	0.495	0.381	0.634	0.656	0.630	0.479	0.548	0.918	0.836	0.719	1.000	0.247
FT LegalBert-pt	0.630	0.758	0.487	0.464	0.489	0.647	0.569	0.207	0.337	0.418	0.247	1.000

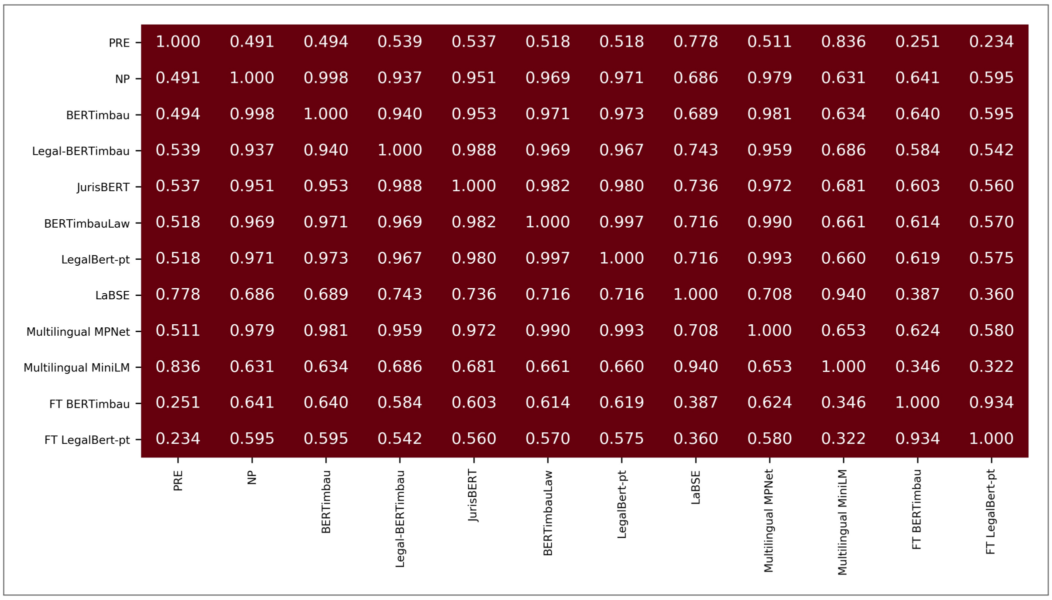
Source: Created by the author (2025)

Figure 137 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using FT BERTimbau. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

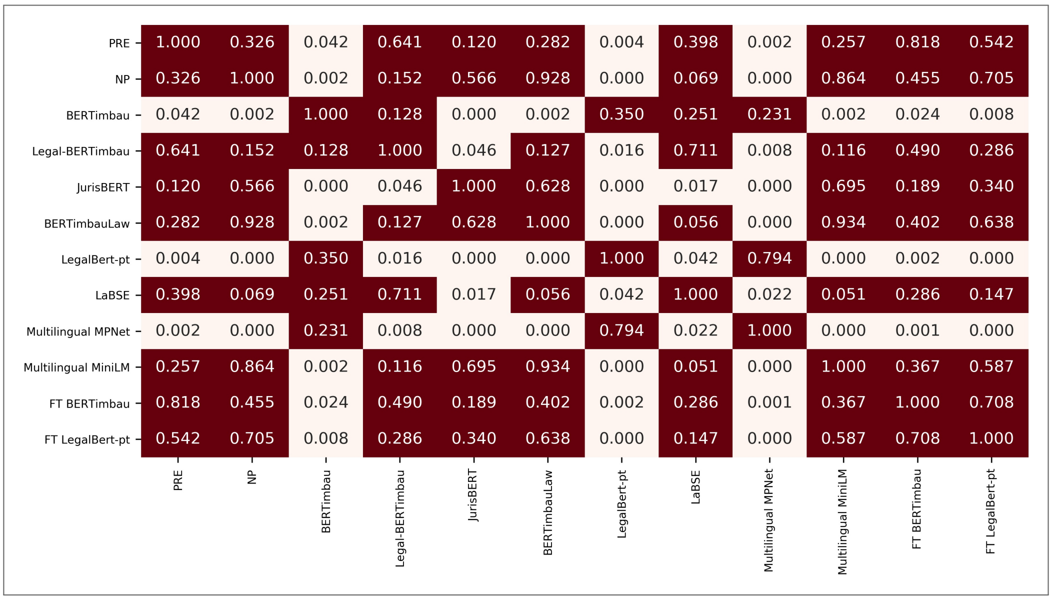
Figure 138 – P-values obtained for the comparison between the algorithms used to search for the similar queries from Ulysses-RFCorpus using FT LegalBert-pt. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

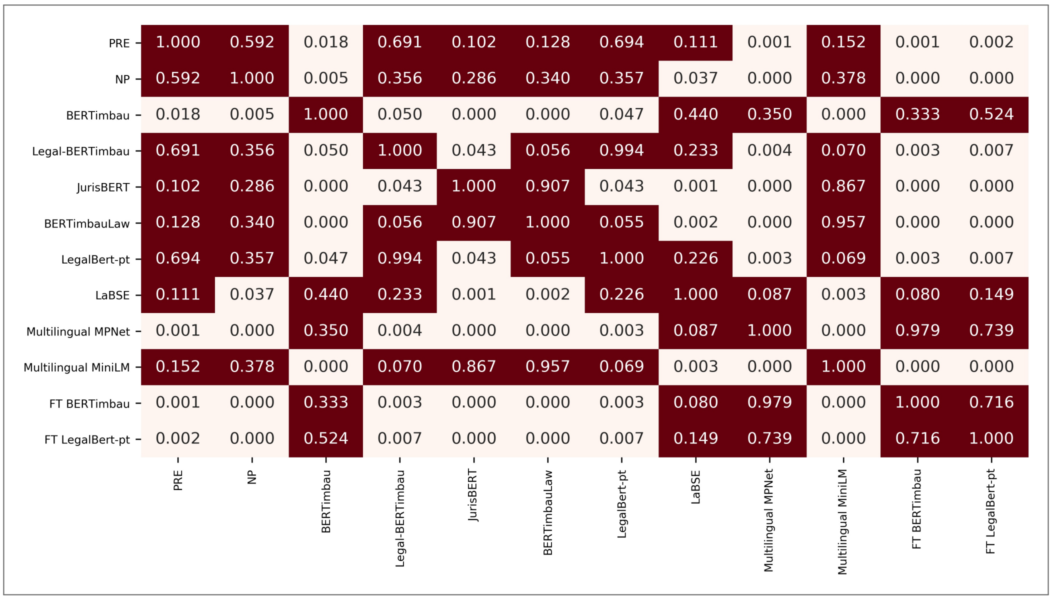
APPENDIX N – STUDENT'S T-TEST RESULTS FOR THE COMPARISON BETWEEN THE ALGORITHMS USED TO SEARCH FOR THE SIMILAR QUERIES FROM THE PRELIMINARY SEARCH CORPUS

Figure 139 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using BM25_PRE. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 140 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using BM25_NP. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 141 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using OkapiBM25_PRE. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.282	0.046	0.461	0.233	0.351	0.002	0.804	0.002	0.730	0.844	0.661
NP	0.282	1.000	0.002	0.073	0.908	0.884	0.000	0.185	0.000	0.482	0.383	0.518
BERTimbau	0.046	0.002	1.000	0.226	0.001	0.003	0.274	0.080	0.216	0.022	0.029	0.014
Legal-BERTimbau	0.461	0.073	0.226	1.000	0.057	0.098	0.025	0.621	0.018	0.291	0.355	0.242
JurisBERT	0.233	0.908	0.001	0.057	1.000	0.793	0.000	0.149	0.000	0.414	0.323	0.446
BERTimbauLaw	0.351	0.884	0.003	0.098	0.793	1.000	0.000	0.237	0.000	0.573	0.465	0.617
LegalBert-pt	0.002	0.000	0.274	0.025	0.000	0.000	1.000	0.005	0.886	0.001	0.001	0.000
LaBSE	0.804	0.185	0.080	0.621	0.149	0.237	0.005	1.000	0.003	0.558	0.658	0.491
Multilingual MPNet	0.002	0.000	0.216	0.018	0.000	0.000	0.886	0.003	1.000	0.001	0.001	0.000
Multilingual MiniLM	0.730	0.482	0.022	0.291	0.414	0.573	0.001	0.558	0.001	1.000	0.880	0.937
FT BERTimbau	0.844	0.383	0.029	0.355	0.323	0.465	0.001	0.658	0.001	0.880	1.000	0.813
FT LegalBert-pt	0.661	0.518	0.014	0.242	0.446	0.617	0.000	0.491	0.000	0.937	0.813	1.000

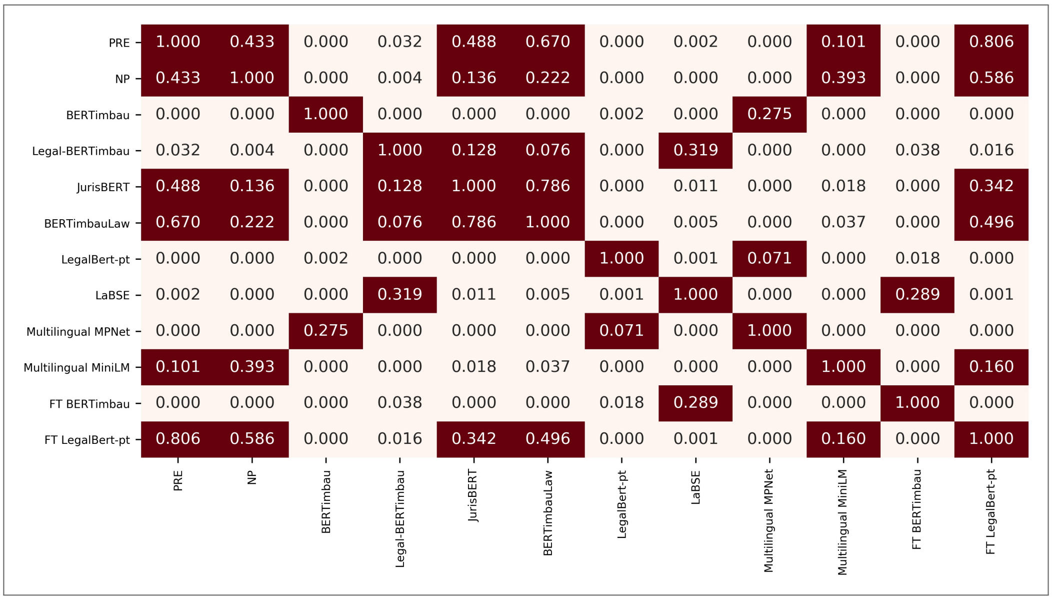
Source: Created by the author (2025)

Figure 142 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using OkapiBM25_NP. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.116	0.002	0.249	0.154	0.708	0.000	0.013	0.002	0.001	0.152	0.777
NP	0.116	1.000	0.114	0.669	0.003	0.236	0.009	0.363	0.147	0.105	0.873	0.195
BERTimbau	0.002	0.114	1.000	0.043	0.000	0.006	0.289	0.511	0.891	0.984	0.078	0.004
Legal-BERTimbau	0.249	0.669	0.043	1.000	0.009	0.442	0.002	0.179	0.058	0.038	0.786	0.381
JurisBERT	0.154	0.003	0.000	0.009	1.000	0.073	0.000	0.000	0.000	0.000	0.004	0.085
BERTimbauLaw	0.708	0.236	0.006	0.442	0.073	1.000	0.000	0.037	0.008	0.005	0.298	0.924
LegalBert-pt	0.000	0.009	0.289	0.002	0.000	0.000	1.000	0.088	0.229	0.290	0.005	0.000
LaBSE	0.013	0.363	0.511	0.179	0.000	0.037	0.088	1.000	0.599	0.492	0.279	0.027
Multilingual MPNet	0.002	0.147	0.891	0.058	0.000	0.008	0.229	0.599	1.000	0.874	0.102	0.005
Multilingual MiniLM	0.001	0.105	0.984	0.038	0.000	0.005	0.290	0.492	0.874	1.000	0.070	0.003
FT BERTimbau	0.152	0.873	0.078	0.786	0.004	0.298	0.005	0.279	0.102	0.070	1.000	0.248
FT LegalBert-pt	0.777	0.195	0.004	0.381	0.085	0.924	0.000	0.027	0.005	0.003	0.248	1.000

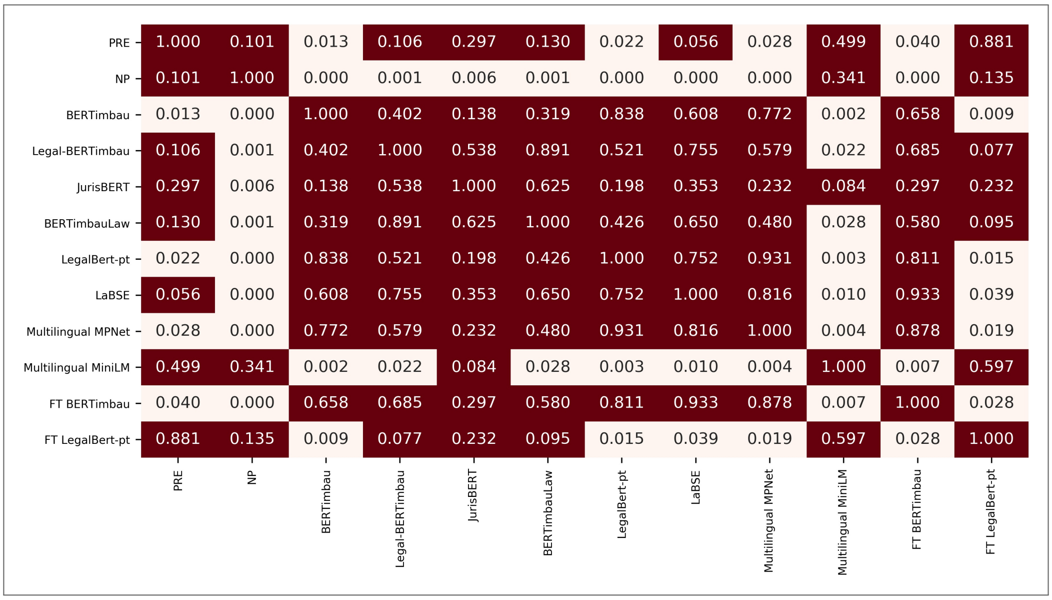
Source: Created by the author (2025)

Figure 143 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using BERTimbau. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 144 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using Legal-BERTimbau. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 145 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using JurisBERT. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.274	0.000	0.130	0.045	0.363	0.000	0.094	0.000	0.042	0.010	0.310
NP	0.274	1.000	0.000	0.010	0.376	0.046	0.000	0.006	0.000	0.349	0.000	0.934
BERTimbau	0.000	0.000	1.000	0.000	0.000	0.000	0.778	0.000	0.951	0.000	0.000	0.000
Legal-BERTimbau	0.130	0.010	0.000	1.000	0.000	0.534	0.000	0.885	0.000	0.000	0.295	0.012
JurisBERT	0.045	0.376	0.000	0.000	1.000	0.004	0.000	0.000	0.000	0.944	0.000	0.331
BERTimbauLaw	0.363	0.046	0.000	0.534	0.004	1.000	0.000	0.439	0.000	0.003	0.090	0.055
LegalBert-pt	0.000	0.000	0.778	0.000	0.000	0.000	1.000	0.000	0.743	0.000	0.000	0.000
LaBSE	0.094	0.006	0.000	0.885	0.000	0.439	0.000	1.000	0.000	0.000	0.364	0.008
Multilingual MPNet	0.000	0.000	0.951	0.000	0.000	0.000	0.743	0.000	1.000	0.000	0.001	0.000
Multilingual MiniLM	0.042	0.349	0.000	0.000	0.944	0.003	0.000	0.000	0.000	1.000	0.000	0.307
FT BERTimbau	0.010	0.000	0.000	0.295	0.000	0.090	0.000	0.364	0.001	0.000	1.000	0.000
FT LegalBert-pt	0.310	0.934	0.000	0.012	0.331	0.055	0.000	0.008	0.000	0.307	0.000	1.000

Source: Created by the author (2025)

Figure 146 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using BERTimbauLaw. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.314	0.000	0.021	0.268	0.466	0.004	0.001	0.001	0.213	0.291	0.551
NP	0.314	1.000	0.000	0.001	0.917	0.079	0.000	0.000	0.000	0.794	0.037	0.110
BERTimbau	0.000	0.000	1.000	0.036	0.000	0.000	0.100	0.232	0.299	0.000	0.001	0.000
Legal-BERTimbau	0.021	0.001	0.036	1.000	0.001	0.103	0.605	0.375	0.276	0.000	0.197	0.090
JurisBERT	0.268	0.917	0.000	0.001	1.000	0.063	0.000	0.000	0.000	0.874	0.029	0.090
BERTimbauLaw	0.466	0.079	0.000	0.103	0.063	1.000	0.027	0.011	0.005	0.047	0.735	0.908
LegalBert-pt	0.004	0.000	0.100	0.605	0.000	0.027	1.000	0.690	0.550	0.000	0.062	0.024
LaBSE	0.001	0.000	0.232	0.375	0.000	0.011	0.690	1.000	0.855	0.000	0.028	0.010
Multilingual MPNet	0.001	0.000	0.299	0.276	0.000	0.005	0.550	0.855	1.000	0.000	0.015	0.005
Multilingual MiniLM	0.213	0.794	0.000	0.000	0.874	0.047	0.000	0.000	0.000	1.000	0.022	0.069
FT BERTimbau	0.291	0.037	0.001	0.197	0.029	0.735	0.062	0.028	0.015	0.022	1.000	0.658
FT LegalBert-pt	0.551	0.110	0.000	0.090	0.090	0.908	0.024	0.010	0.005	0.069	0.658	1.000

Source: Created by the author (2025)

Figure 147 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using LegalBert-pt. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.127	0.000	0.181	0.006	0.141	0.001	0.065	0.000	0.042	0.090	0.781
NP	0.127	1.000	0.000	0.836	0.000	0.921	0.090	0.781	0.000	0.000	0.902	0.071
BERTimbau	0.000	0.000	1.000	0.000	0.000	0.000	0.001	0.000	0.936	0.000	0.000	0.000
Legal-BERTimbau	0.181	0.836	0.000	1.000	0.000	0.909	0.052	0.620	0.000	0.001	0.735	0.105
JurisBERT	0.006	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.497	0.000	0.013
BERTimbauLaw	0.141	0.921	0.000	0.909	0.000	1.000	0.062	0.697	0.000	0.000	0.819	0.078
LegalBert-pt	0.001	0.090	0.001	0.052	0.000	0.062	1.000	0.144	0.001	0.000	0.102	0.000
LaBSE	0.065	0.781	0.000	0.620	0.000	0.697	0.144	1.000	0.000	0.000	0.872	0.033
Multilingual MPNet	0.000	0.000	0.936	0.000	0.000	0.000	0.001	0.000	1.000	0.000	0.000	0.000
Multilingual MiniLM	0.042	0.000	0.000	0.001	0.497	0.000	0.000	0.000	0.000	1.000	0.000	0.077
FT BERTimbau	0.090	0.902	0.000	0.735	0.000	0.819	0.102	0.872	0.000	0.000	1.000	0.047
FT LegalBert-pt	0.781	0.071	0.000	0.105	0.013	0.078	0.000	0.033	0.000	0.077	0.047	1.000

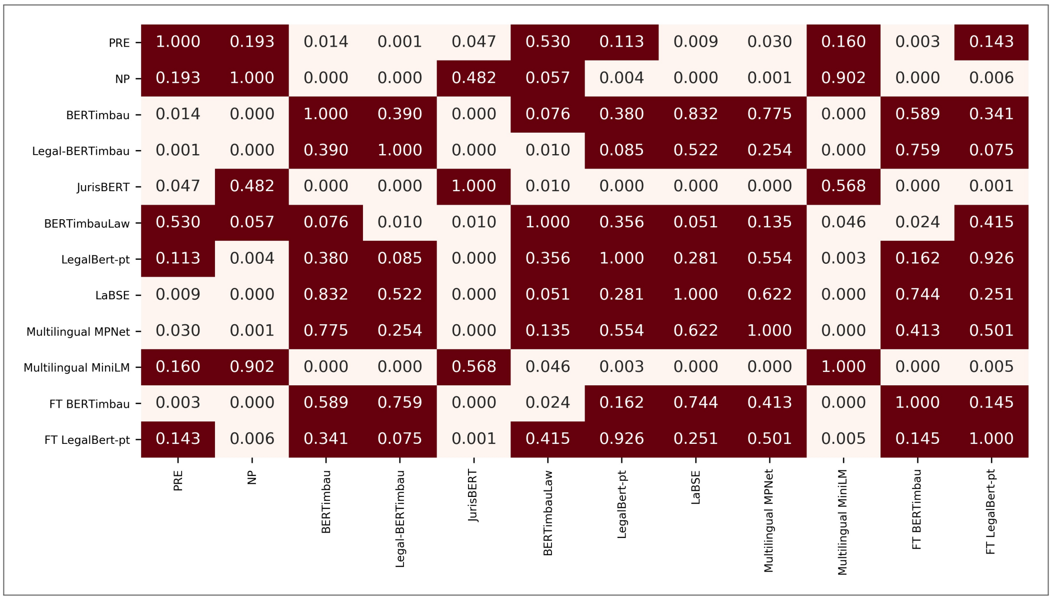
Source: Created by the author (2025)

Figure 148 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using LaBSE. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.024	0.005	0.014	0.028	0.540	0.001	0.001	0.005	0.206	0.065	0.874
NP	0.024	1.000	0.630	0.871	0.000	0.096	0.283	0.302	0.642	0.001	0.652	0.036
BERTimbau	0.005	0.630	1.000	0.748	0.000	0.028	0.542	0.563	0.987	0.000	0.338	0.008
Legal-BERTimbau	0.014	0.871	0.748	1.000	0.000	0.063	0.356	0.376	0.760	0.000	0.532	0.022
JurisBERT	0.028	0.000	0.000	0.000	1.000	0.005	0.000	0.000	0.000	0.364	0.000	0.018
BERTimbauLaw	0.540	0.096	0.028	0.063	0.005	1.000	0.005	0.006	0.029	0.061	0.215	0.651
LegalBert-pt	0.001	0.283	0.542	0.356	0.000	0.005	1.000	0.987	0.531	0.000	0.117	0.001
LaBSE	0.001	0.302	0.563	0.376	0.000	0.006	0.987	1.000	0.552	0.000	0.130	0.002
Multilingual MPNet	0.005	0.642	0.987	0.760	0.000	0.029	0.531	0.552	1.000	0.000	0.347	0.009
Multilingual MiniLM	0.206	0.001	0.000	0.000	0.364	0.061	0.000	0.000	0.000	1.000	0.002	0.155
FT BERTimbau	0.065	0.652	0.338	0.532	0.000	0.215	0.117	0.130	0.347	0.002	1.000	0.092
FT LegalBert-pt	0.874	0.036	0.008	0.022	0.018	0.651	0.001	0.002	0.009	0.155	0.092	1.000

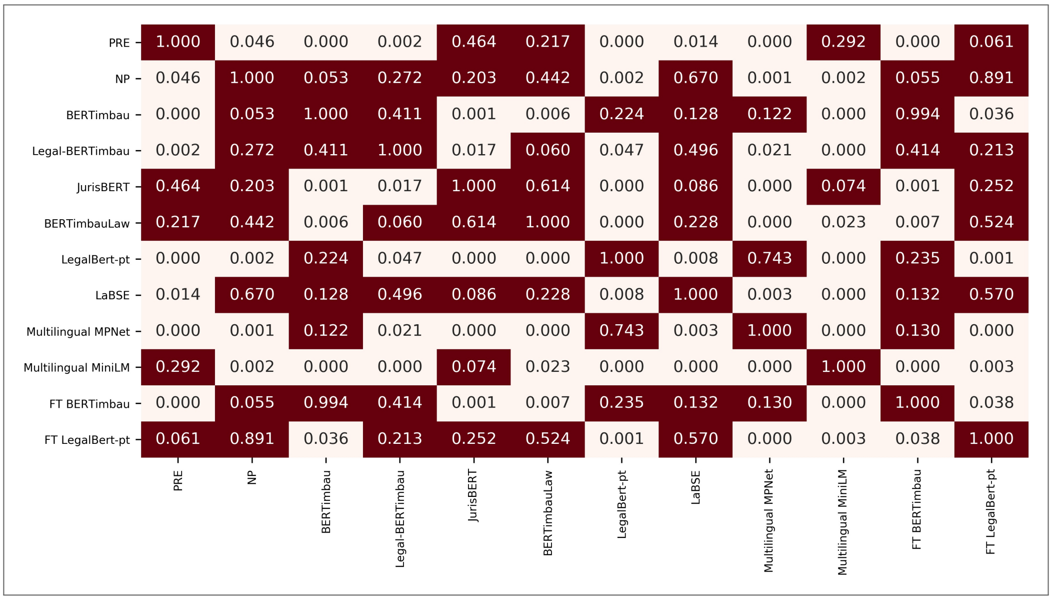
Source: Created by the author (2025)

Figure 149 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using Multilingual MPNet. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 150 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using Multilingual MiniLM. The highlighted values indicate the cases for which there was no statistical significance.



Source: Created by the author (2025)

Figure 151 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using FT BERTimbau. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.098	0.069	0.758	0.466	0.980	0.923	0.527	0.413	0.270	0.708	0.614
NP	0.098	1.000	0.001	0.048	0.017	0.103	0.079	0.300	0.014	0.567	0.195	0.244
BERTimbau	0.069	0.001	1.000	0.125	0.274	0.065	0.083	0.014	0.315	0.003	0.027	0.019
Legal-BERTimbau	0.758	0.048	0.125	1.000	0.669	0.739	0.832	0.343	0.604	0.155	0.490	0.413
JurisBERT	0.466	0.017	0.274	0.669	1.000	0.450	0.525	0.172	0.929	0.066	0.266	0.215
BERTimbauLaw	0.980	0.103	0.065	0.739	0.450	1.000	0.903	0.544	0.398	0.281	0.726	0.632
LegalBert-pt	0.923	0.079	0.083	0.832	0.525	0.903	1.000	0.464	0.468	0.228	0.635	0.546
LaBSE	0.527	0.300	0.014	0.343	0.172	0.544	0.464	1.000	0.145	0.636	0.795	0.897
Multilingual MPNet	0.413	0.014	0.315	0.604	0.929	0.398	0.468	0.145	1.000	0.054	0.229	0.184
Multilingual MiniLM	0.270	0.567	0.003	0.155	0.066	0.281	0.228	0.636	0.054	1.000	0.463	0.547
FT BERTimbau	0.708	0.195	0.027	0.490	0.266	0.726	0.635	0.795	0.229	0.463	1.000	0.896
FT LegalBert-pt	0.614	0.244	0.019	0.413	0.215	0.632	0.546	0.897	0.184	0.547	0.896	1.000

Source: Created by the author (2025)

Figure 152 – P-values obtained for the comparison between the algorithms used to search for the similar queries from the Preliminary Search corpus using FT LegalBert-pt. The highlighted values indicate the cases for which there was no statistical significance.

	PRE	NP	BERTimbau	Legal-BERTimbau	JurisBERT	BERTimbauLaw	LegalBert-pt	LaBSE	Multilingual MPNet	Multilingual MiniLM	FT BERTimbau	FT LegalBert-pt
PRE	1.000	0.006	0.454	0.525	0.780	0.743	0.520	0.248	0.872	0.440	0.210	0.172
NP	0.006	1.000	0.001	0.036	0.014	0.016	0.036	0.112	0.004	0.054	0.138	0.164
BERTimbau	0.454	0.001	1.000	0.166	0.304	0.279	0.163	0.057	0.558	0.130	0.045	0.034
Legal-BERTimbau	0.525	0.036	0.166	1.000	0.722	0.755	0.995	0.605	0.426	0.885	0.536	0.468
JurisBERT	0.780	0.014	0.304	0.722	1.000	0.963	0.716	0.382	0.659	0.619	0.329	0.278
BERTimbauLaw	0.743	0.016	0.279	0.755	0.963	1.000	0.749	0.404	0.624	0.650	0.350	0.296
LegalBert-pt	0.520	0.036	0.163	0.995	0.716	0.749	1.000	0.608	0.421	0.889	0.539	0.470
LaBSE	0.248	0.112	0.057	0.605	0.382	0.404	0.608	1.000	0.188	0.714	0.918	0.836
Multilingual MPNet	0.872	0.004	0.558	0.426	0.659	0.624	0.421	0.188	1.000	0.351	0.157	0.127
Multilingual MiniLM	0.440	0.054	0.130	0.885	0.619	0.650	0.889	0.714	0.351	1.000	0.641	0.567
FT BERTimbau	0.210	0.138	0.045	0.536	0.329	0.350	0.539	0.918	0.157	0.641	1.000	0.918
FT LegalBert-pt	0.172	0.164	0.034	0.468	0.278	0.296	0.470	0.836	0.127	0.567	0.918	1.000

Source: Created by the author (2025)

APPENDIX O – RESULTS ACHIEVED BY THE BASELINES AND ULYSSES-RFSQ-RI

Algorithm	MAP	MRP	MRR	nDCG
BM25L_PRE (baseline)	0.7684	0.6850	0.8597	0.8277
BM25L_PRE_PRE	0.7653	0.6853	0.8592	0.8244
BM25L_PRE_NP	0.7699	0.6872	0.8618	0.8285
BM25L_PRE_BERTimbau	0.7647	0.6869	0.8640	0.8253
BM25L_PRE_LegalBERTimbau	0.7679	**0.6899**	0.8636	0.8272
BM25L_PRE_JurisBERT	0.7688	0.6854	0.8615	0.8283
BM25L_PRE_BERTimbauLaw	0.7649	0.6846	0.8632	0.8266
BM25L_PRE_LegalBertpt	0.7631	0.6860	**0.8659**	0.8243
BM25L_PRE_LaBSE	0.7695	0.6897	0.8641	0.8288
BM25L_PRE_MPNet	**0.7701**	0.6879	0.8632	0.8294
BM25L_PRE_MiniLM	0.7690	0.6888	0.8647	**0.8295**
BM25L_PRE_FTBERTimbau	0.7648	0.6853	0.8640	0.8258
BM25L_PRE_FTLegalBertpt	0.7692	0.6869	0.8628	0.8279
BM25L_NP (baseline)	0.3801	0.3913	0.6811	0.5429
BM25L_NP_PRE	0.3830	0.3923	0.6788	0.5479
BM25L_NP_NP	0.3811	0.3917	0.6799	0.5422
BM25L_NP_BERTimbau	0.3832	0.3937	0.6828	0.5461
BM25L_NP_LegalBERTimbau	0.3799	0.3913	0.6846	**0.5505**
BM25L_NP_JurisBERT	0.3848	0.3938	0.6839	0.5492
BM25L_NP_BERTimbauLaw	0.3838	0.3919	0.6851	0.5483
BM25L_NP_LegalBertpt	0.3832	0.3912	0.6874	0.5467
BM25L_NP_LaBSE	0.3837	0.3920	0.6870	0.5467
BM25L_NP_MPNet	0.3846	**0.3966**	0.6791	0.5471
BM25L_NP_MiniLM	0.3782	0.3912	0.6768	0.5432
BM25L_NP_FTBERTimbau	**0.3855**	0.3947	0.6870	0.5500
BM25L_NP_FTLegalBertpt	0.3849	0.3934	**0.6882**	0.5492
OkapiBM25_PRE (baseline)	0.7183	0.6609	0.8607	0.7942
OkapiBM25_PRE_PRE	0.7212	0.6624	0.8630	0.7966

Algorithm	MAP	MRP	MRR	nDCG
OkapiBM25_PRE_NP	**0.7213**	0.6630	0.8636	0.7962
OkapiBM25_PRE_BERTimbau	0.7183	0.6616	0.8641	0.7951
OkapiBM25_PRE_LegalBERTimbau	0.7195	0.6616	**0.8649**	0.7957
OkapiBM25_PRE_JurisBERT	0.7205	0.6611	0.8642	0.7962
OkapiBM25_PRE_BERTimbauLaw	0.7154	0.6574	0.8645	0.7949
OkapiBM25_PRE_LegalBertpt	0.7175	0.6592	0.8638	0.7943
OkapiBM25_PRE_LaBSE	0.7209	0.6622	0.8646	0.7967
OkapiBM25_PRE_MPNet	**0.7213**	**0.6634**	0.8630	0.7965
OkapiBM25_PRE_MiniLM	0.7204	0.6610	0.8626	**0.7971**
OkapiBM25_PRE_FTBERTimbau	0.7184	0.6611	0.8641	0.7963
OkapiBM25_PRE_FTLegalBertpt	0.7173	0.6608	0.8644	0.7960
OkapiBM25_NP (baseline)	0.3704	0.3834	0.6718	0.5399
OkapiBM25_NP_PRE	0.3709	0.3841	0.6716	0.5395
OkapiBM25_NP_NP	0.3693	0.3832	0.6675	0.5386
OkapiBM25_NP_BERTimbau	0.3696	0.3832	0.6710	0.5397
OkapiBM25_NP_LegalBERTimbau	**0.3746**	**0.3890**	**0.6741**	**0.5451**
OkapiBM25_NP_JurisBERT	0.3741	0.3866	0.6716	0.5435
OkapiBM25_NP_BERTimbauLaw	0.3717	0.3840	0.6715	0.5430
OkapiBM25_NP_LegalBertpt	0.3685	0.3803	0.6715	0.5385
OkapiBM25_NP_LaBSE	0.3711	0.3830	0.6690	0.5422
OkapiBM25_NP_MPNet	0.3696	0.3821	0.6586	0.5410
OkapiBM25_NP_MiniLM	0.3725	0.3854	0.6720	0.5439
OkapiBM25_NP_FTBERTimbau	0.3707	0.3852	0.6730	0.5437
OkapiBM25_NP_FTLegalBertpt	0.3703	0.3845	0.6733	0.5428
BERTimbau (baseline)	0.0113	0.0209	0.0542	0.0490
BERTimbau_PRE	**0.0479**	**0.0585**	0.1095	**0.1057**
BERTimbau_NP	0.0430	0.0555	**0.1106**	0.1036
BERTimbau_BERTimbau	0.0297	0.0393	0.0800	0.0786
BERTimbau_LegalBERTimbau	0.0412	0.0534	0.1041	0.1040
BERTimbau_JurisBERT	0.0226	0.0318	0.0685	0.0638
BERTimbau_BERTimbauLaw	0.0270	0.0391	0.0912	0.0830

Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**
BERTimbau_LegalBertpt	0.0223	0.0301	0.0580	0.0587
BERTimbau_LaBSE	0.0290	0.0372	0.0764	0.0764
BERTimbau_MPNet	0.0406	0.0537	0.1038	0.1010
BERTimbau_MiniLM	0.0345	0.0427	0.0914	0.0866
BERTimbau_FTBERTimbau	0.0343	0.0446	0.0982	0.0930
BERTimbau_FTLegalBertpt	0.0414	0.0504	0.1042	0.0984
LegalBERTimbau (baseline)	0.1221	0.1469	0.3320	0.2871
LegalBERTimbau_PRE	0.1349	0.1581	0.3527	0.3083
LegalBERTimbau_NP	**0.1400**	**0.1644**	**0.3611**	**0.3135**
LegalBERTimbau_BERTimbau	0.1259	0.1502	0.3350	0.2984
LegalBERTimbau_LegalBERTimbau	0.1274	0.1491	0.3129	0.2996
LegalBERTimbau_JurisBERT	0.1233	0.1473	0.3291	0.2930
LegalBERTimbau_BERTimbauLaw	0.1302	0.1558	0.3398	0.3012
LegalBERTimbau_LegalBertpt	0.1191	0.1442	0.3166	0.2877
LegalBERTimbau_LaBSE	0.1233	0.1475	0.3257	0.2938
LegalBERTimbau_MPNet	0.1336	0.1563	0.3495	0.3026
LegalBERTimbau_MiniLM	0.1217	0.1447	0.3219	0.2902
LegalBERTimbau_FTBERTimbau	0.1251	0.1519	0.3312	0.2991
LegalBERTimbau_FTLegalBertpt	0.1273	0.1514	0.3377	0.2999
JurisBERT (baseline)	0.0744	0.0924	0.2482	0.1862
JurisBERT_PRE	0.0870	0.1049	0.2467	0.2085
JurisBERT_NP	**0.0890**	**0.1122**	0.2511	**0.2159**
JurisBERT_BERTimbau	0.0812	0.0992	0.2570	0.2010
JurisBERT_LegalBERTimbau	0.0823	0.1020	**0.2626**	0.2003
JurisBERT_JurisBERT	0.0739	0.0906	0.2340	0.1826
JurisBERT_BERTimbauLaw	0.0773	0.0933	0.2419	0.1966
JurisBERT_LegalBertpt	0.0755	0.0932	0.2400	0.1902
JurisBERT_LaBSE	0.0759	0.0932	0.2453	0.1932
JurisBERT_MPNet	0.0781	0.0998	0.2466	0.1986
JurisBERT_MiniLM	0.0755	0.0969	0.2398	0.1972
JurisBERT_FTBERTimbau	0.0846	0.1047	0.2494	0.2084

Algorithm	MAP	MRP	MRR	nDCG
JurisBERT_FTLegalBertpt	0.0795	0.0992	0.2484	0.2054
BERTimbauLaw (baseline)	0.1239	0.1571	0.3481	0.2871
BERTimbauLaw_PRE	**0.1408**	**0.1729**	**0.3671**	**0.3158**
BERTimbauLaw_NP	0.1347	0.1695	0.3579	0.3039
BERTimbauLaw_BERTimbau	0.1291	0.1644	0.3501	0.2983
BERTimbauLaw_LegalBERTimbau	0.1331	0.1676	0.3594	0.3005
BERTimbauLaw_JurisBERT	0.1311	0.1633	0.3435	0.2932
BERTimbauLaw_BERTimbauLaw	0.1297	0.1635	0.3558	0.2969
BERTimbauLaw_LegalBertpt	0.1287	0.1639	0.3435	0.2950
BERTimbauLaw_LaBSE	0.1324	0.1674	0.3560	0.2980
BERTimbauLaw_MPNet	0.1356	0.1679	0.3552	0.3012
BERTimbauLaw_MiniLM	0.1332	0.1670	0.3588	0.2960
BERTimbauLaw_FTBERTimbau	0.1276	0.1613	0.3468	0.2965
BERTimbauLaw_FTLegalBertpt	0.1272	0.1620	0.3424	0.3013
LegalBertpt (baseline)	0.0482	0.0670	0.1792	0.1394
LegalBertpt_PRE	0.0650	0.0872	0.1859	0.1649
LegalBertpt_NP	0.0625	0.0837	0.1899	0.1667
LegalBertpt_BERTimbau	0.0521	0.0684	0.1556	0.1479
LegalBertpt_LegalBERTimbau	**0.0700**	**0.0879**	**0.2004**	**0.1743**
LegalBertpt_JurisBERT	0.0545	0.0732	0.1764	0.1448
LegalBertpt_BERTimbauLaw	0.0564	0.0752	0.1816	0.1584
LegalBertpt_LegalBertpt	0.0444	0.0602	0.1319	0.1299
LegalBertpt_LaBSE	0.0528	0.0689	0.1610	0.1432
LegalBertpt_MPNet	0.0660	0.0846	0.1838	0.1629
LegalBertpt_MiniLM	0.0626	0.0793	0.1935	0.1615
LegalBertpt_FTBERTimbau	0.0613	0.0843	0.1942	0.1654
LegalBertpt_FTLegalBertpt	0.0606	0.0788	0.1718	0.1596
LaBSE (baseline)	0.0893	0.1080	0.2910	0.2268
LaBSE_PRE	**0.1126**	**0.1337**	0.3166	**0.2613**
LaBSE_NP	0.1075	0.1320	0.2944	0.2517
LaBSE_BERTimbau	0.0980	0.1183	0.3035	0.2398

Algorithm	MAP	MRP	MRR	nDCG
LaBSE_LegalBERTimbau	0.1008	0.1222	0.2811	0.2419
LaBSE_JurisBERT	0.1005	0.1201	0.3005	0.2460
LaBSE_BERTimbauLaw	0.1004	0.1244	0.3025	0.2452
LaBSE_LegalBertpt	0.0963	0.1165	0.2947	0.2377
LaBSE_LaBSE	0.0979	0.1174	0.2967	0.2371
LaBSE_MPNet	0.1047	0.1236	**0.3183**	0.2455
LaBSE_MiniLM	0.1023	0.1236	0.3078	0.2444
LaBSE_FTBERTimbau	0.0995	0.1233	0.3039	0.2507
LaBSE_FTLegalBertpt	0.1021	0.1242	0.3060	0.2473
MPNet (baseline)	0.1633	0.1891	0.4204	0.3491
MPNet_PRE	**0.1763**	**0.2029**	**0.4317**	**0.3645**
MPNet_NP	0.1701	0.1975	0.4225	0.3588
MPNet_BERTimbau	0.1642	0.1911	0.4150	0.3539
MPNet_LegalBERTimbau	0.1703	0.1983	0.4231	0.3578
MPNet_JurisBERT	0.1676	0.1938	0.4206	0.3534
MPNet_BERTimbauLaw	0.1677	0.1944	0.4291	0.3571
MPNet_LegalBertpt	0.1640	0.1909	0.4152	0.3525
MPNet_LaBSE	0.1671	0.1948	0.4205	0.3531
MPNet_MPNet	0.1721	0.1968	0.4230	0.3582
MPNet_MiniLM	0.1686	0.1957	0.4244	0.3520
MPNet_FTBERTimbau	0.1660	0.1942	0.4194	0.3574
MPNet_FTLegalBertpt	0.1649	0.1907	0.4217	0.3527
MiniLM (baseline)	0.1208	0.1454	0.3376	0.2692
MiniLM_PRE	0.1258	0.1505	0.3248	0.2768
MiniLM_NP	0.1286	0.1535	0.3492	0.2790
MiniLM_BERTimbau	0.1239	0.1489	0.3446	0.2727
MiniLM_LegalBERTimbau	0.1240	0.1491	0.3399	0.2747
MiniLM_JurisBERT	0.1255	0.1482	0.3470	0.2764
MiniLM_BERTimbauLaw	0.1250	0.1490	**0.3496**	0.2771
MiniLM_LegalBertpt	0.1206	0.1467	0.3345	0.2674
MiniLM_LaBSE	0.1184	0.1402	0.3178	0.2687

Algorithm	MAP	MRP	MRR	nDCG
MiniLM_MPNet	**0.1337**	**0.1590**	0.3473	**0.2825**
MiniLM_MiniLM	0.1205	0.1437	0.3295	0.2673
MiniLM_FTBERTimbau	0.1248	0.1493	0.3493	0.2747
MiniLM_FTLegalBertpt	0.1283	0.1526	0.3434	0.2755
FTBERTimbau (baseline)	0.1902	0.2163	0.4456	0.4070
FTBERTimbau_PRE	0.1985	0.2235	0.4608	0.4164
FTBERTimbau_NP	**0.1996**	**0.2284**	**0.4635**	**0.4175**
FTBERTimbau_BERTimbau	0.1809	0.2080	0.4128	0.4029
FTBERTimbau_LegalBERTimbau	0.1898	0.2176	0.4297	0.4081
FTBERTimbau_JurisBERT	0.1915	0.2156	0.4442	0.4082
FTBERTimbau_BERTimbauLaw	0.1889	0.2145	0.4393	0.4106
FTBERTimbau_LegalBertpt	0.1801	0.2097	0.4149	0.4034
FTBERTimbau_LaBSE	0.1927	0.2196	0.4410	0.4106
FTBERTimbau_MPNet	0.1959	0.2215	0.4468	0.4116
FTBERTimbau_MiniLM	0.1880	0.2155	0.4328	0.4064
FTBERTimbau_FTBERTimbau	0.1811	0.2062	0.4140	0.4006
FTBERTimbau_FTLegalBertpt	0.1930	0.2182	0.4423	0.4095
FTLegalBertpt (baseline)	0.1220	0.1523	0.2998	0.3050
FTLegalBertpt_PRE	**0.1340**	0.1624	0.3087	0.3114
FTLegalBertpt_NP	0.1266	0.1546	0.3013	0.3080
FTLegalBertpt_BERTimbau	0.1261	0.1572	0.2999	0.3074
FTLegalBertpt_LegalBERTimbau	0.1336	**0.1628**	**0.3135**	**0.3166**
FTLegalBertpt_JurisBERT	0.1277	0.1561	0.2982	0.3133
FTLegalBertpt_BERTimbauLaw	0.1282	0.1555	0.3077	0.3144
FTLegalBertpt_LegalBertpt	0.1250	0.1544	0.2955	0.3093
FTLegalBertpt_LaBSE	0.1318	0.1617	0.3096	0.3115
FTLegalBertpt_MPNet	0.1220	0.1516	0.2709	0.2850
FTLegalBertpt_MiniLM	0.1230	0.1526	0.2971	0.3034
FTLegalBertpt_FTBERTimbau	0.1277	0.1576	0.3075	0.3054
FTLegalBertpt_FTLegalBertpt	0.1289	0.1592	0.3087	0.3089

APPENDIX P – RESULTS ACHIEVED BY THE BASELINES AND ULYSSES-RFSQ-DRL

Algorithm	MAP	MRP	MRR	nDCG
BM25L_PRE (baseline)	0.7684	0.6850	0.8597	0.8277
BM25L_PRE_PRE	0.7649	0.6832	0.8573	0.8243
BM25L_PRE_NP	0.7677	0.6849	0.8590	0.8271
BM25L_PRE_BERTimbau	0.7643	0.6851	0.8592	0.8248
BM25L_PRE_LegalBERTimbau	0.7671	0.6874	0.8603	0.8262
BM25L_PRE_JurisBERT	0.7684	0.6855	0.8597	0.8274
BM25L_PRE_BERTimbauLaw	0.7647	0.6846	0.8601	0.8260
BM25L_PRE_LegalBertpt	0.7624	0.6828	0.8591	0.8237
BM25L_PRE_LaBSE	0.7685	0.6865	0.8603	0.8276
BM25L_PRE_MPNet	**0.7690**	0.6867	0.8601	0.8279
BM25L_PRE_MiniLM	0.7683	**0.6879**	**0.8606**	**0.8281**
BM25L_PRE_FTBERTimbau	0.7645	0.6847	0.8604	0.8257
BM25L_PRE_FTLegalBertpt	0.7680	0.6856	0.8599	0.8270
BM25L_NP (baseline)	0.3801	0.3913	0.6811	0.5429
BM25L_NP_PRE	0.3856	0.3941	0.6850	0.5485
BM25L_NP_NP	0.3817	0.3918	0.6827	0.5436
BM25L_NP_BERTimbau	0.3839	0.3947	0.6827	0.5476
BM25L_NP_LegalBERTimbau	0.3815	0.3919	0.6812	**0.5517**
BM25L_NP_JurisBERT	0.3855	0.3939	0.6840	0.5490
BM25L_NP_BERTimbauLaw	0.3845	0.3931	0.6855	0.5479
BM25L_NP_LegalBertpt	0.3850	0.3923	0.6873	0.5493
BM25L_NP_LaBSE	**0.3859**	0.3946	**0.6875**	0.5494
BM25L_NP_MPNet	0.3857	**0.3969**	0.6809	0.5490
BM25L_NP_MiniLM	0.3793	0.3916	0.6779	0.5440
BM25L_NP_FTBERTimbau	0.3858	0.3948	0.6850	0.5505
BM25L_NP_FTLegalBertpt	0.3854	0.3943	0.6861	0.5506
OkapiBM25_PRE (baseline)	0.7183	0.6609	0.8607	0.7942
OkapiBM25_PRE_PRE	0.7191	0.6609	0.8609	0.7952

Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**
OkapiBM25_PRE_NP	0.7191	0.6607	0.8609	0.7946
OkapiBM25_PRE_BERTimbau	0.7176	0.6602	0.8609	0.7936
OkapiBM25_PRE_LegalBERTimbau	0.7184	0.6597	0.8611	0.7946
OkapiBM25_PRE_JurisBERT	0.7193	0.6594	0.8615	0.7950
OkapiBM25_PRE_BERTimbauLaw	0.7156	0.6601	0.8609	0.7943
OkapiBM25_PRE_LegalBertpt	0.7168	0.6600	0.8610	0.7934
OkapiBM25_PRE_LaBSE	0.7192	0.6601	0.8613	0.7951
OkapiBM25_PRE_MPNet	**0.7201**	**0.6614**	0.8619	0.7955
OkapiBM25_PRE_MiniLM	0.7197	0.6600	**0.8621**	**0.7962**
OkapiBM25_PRE_FTBERTimbau	0.7167	0.6585	0.8610	0.7947
OkapiBM25_PRE_FTLegalBertpt	0.7178	0.6612	0.8609	0.7958
OkapiBM25_NP (baseline)	0.3704	0.3834	0.6718	0.5399
OkapiBM25_NP_PRE	0.3714	0.3843	0.6726	0.5401
OkapiBM25_NP_NP	0.3702	0.3833	0.6690	0.5395
OkapiBM25_NP_BERTimbau	0.3719	0.3840	0.6727	0.5421
OkapiBM25_NP_LegalBERTimbau	0.3744	**0.3869**	0.6762	**0.5455**
OkapiBM25_NP_JurisBERT	**0.3760**	0.3865	0.6766	0.5440
OkapiBM25_NP_BERTimbauLaw	0.3725	0.3839	0.6740	0.5430
OkapiBM25_NP_LegalBertpt	0.3718	0.3838	0.6747	0.5422
OkapiBM25_NP_LaBSE	0.3732	0.3846	0.6720	0.5437
OkapiBM25_NP_MPNet	0.3718	0.3849	0.6617	0.5434
OkapiBM25_NP_MiniLM	0.3740	0.3858	0.6741	0.5450
OkapiBM25_NP_FTBERTimbau	0.3715	0.3858	0.6744	0.5432
OkapiBM25_NP_FTLegalBertpt	0.3738	0.3866	**0.6775**	0.5452
BERTimbau (baseline)	0.0113	0.0209	0.0542	0.0490
BERTimbau_PRE	**0.0483**	**0.0571**	0.1082	**0.1090**
BERTimbau_NP	0.0431	0.0538	**0.1103**	0.1016
BERTimbau_BERTimbau	0.0317	0.0403	0.0852	0.0783
BERTimbau_LegalBERTimbau	0.0422	0.0560	0.1059	0.1055
BERTimbau_JurisBERT	0.0235	0.0322	0.0724	0.0644
BERTimbau_BERTimbauLaw	0.0287	0.0409	0.0945	0.0831

Algorithm	MAP	MRP	MRR	nDCG
BERTimbau_LegalBertpt	0.0230	0.0297	0.0592	0.0610
BERTimbau_LaBSE	0.0298	0.0389	0.0742	0.0762
BERTimbau_MPNet	0.0414	0.0558	0.1033	0.1033
BERTimbau_MiniLM	0.0352	0.0420	0.0852	0.0853
BERTimbau_FTBERTimbau	0.0345	0.0428	0.0915	0.0956
BERTimbau_FTLegalBertpt	0.0423	0.0522	0.1034	0.1038
LegalBERTimbau (baseline)	0.1221	0.1469	0.3320	0.2871
LegalBERTimbau_PRE	0.1343	0.1596	0.3499	0.3076
LegalBERTimbau_NP	**0.1397**	**0.1644**	**0.3572**	**0.3156**
LegalBERTimbau_BERTimbau	0.1277	0.1522	0.3389	0.2979
LegalBERTimbau_LegalBERTimbau	0.1314	0.1531	0.3238	0.3031
LegalBERTimbau_JurisBERT	0.1280	0.1517	0.3382	0.2973
LegalBERTimbau_BERTimbauLaw	0.1318	0.1567	0.3435	0.3030
LegalBERTimbau_LegalBertpt	0.1265	0.1505	0.3333	0.2966
LegalBERTimbau_LaBSE	0.1284	0.1514	0.3391	0.2992
LegalBERTimbau_MPNet	0.1346	0.1568	0.3520	0.3025
LegalBERTimbau_MiniLM	0.1281	0.1502	0.3396	0.2977
LegalBERTimbau_FTBERTimbau	0.1282	0.1512	0.3389	0.3024
LegalBERTimbau_FTLegalBertpt	0.1308	0.1540	0.3453	0.3023
JurisBERT (baseline)	0.0744	0.0924	0.2482	0.1862
JurisBERT_PRE	**0.0885**	0.1063	0.2497	0.2110
JurisBERT_NP	0.0884	**0.1098**	0.2432	**0.2116**
JurisBERT_BERTimbau	0.0823	0.1010	0.2573	0.2008
JurisBERT_LegalBERTimbau	0.0844	0.1036	**0.2656**	0.2036
JurisBERT_JurisBERT	0.0756	0.0941	0.2394	0.1869
JurisBERT_BERTimbauLaw	0.0801	0.0975	0.2466	0.1995
JurisBERT_LegalBertpt	0.0775	0.0960	0.2433	0.1910
JurisBERT_LaBSE	0.0803	0.0986	0.2521	0.1969
JurisBERT_MPNet	0.0807	0.1010	0.2505	0.2031
JurisBERT_MiniLM	0.0795	0.0975	0.2504	0.1995
JurisBERT_FTBERTimbau	0.0864	0.1058	0.2506	0.2077

Algorithm	MAP	MRP	MRR	nDCG
JurisBERT_FTLegalBertpt	0.0823	0.1017	0.2555	0.2038
BERTimbauLaw (baseline)	0.1239	0.1571	0.3481	0.2871
BERTimbauLaw_PRE	**0.1436**	**0.1793**	**0.3656**	**0.3181**
BERTimbauLaw_NP	0.1364	0.1720	0.3565	0.3068
BERTimbauLaw_BERTimbau	0.1310	0.1667	0.3527	0.3006
BERTimbauLaw_LegalBERTimbau	0.1330	0.1666	0.3603	0.2987
BERTimbauLaw_JurisBERT	0.1327	0.1658	0.3474	0.2967
BERTimbauLaw_BERTimbauLaw	0.1300	0.1642	0.3543	0.2975
BERTimbauLaw_LegalBertpt	0.1302	0.1647	0.3525	0.2975
BERTimbauLaw_LaBSE	0.1315	0.1659	0.3551	0.2965
BERTimbauLaw_MPNet	0.1362	0.1681	0.3562	0.3017
BERTimbauLaw_MiniLM	0.1332	0.1673	0.3596	0.2967
BERTimbauLaw_FTBERTimbau	0.1304	0.1650	0.3499	0.2978
BERTimbauLaw_FTLegalBertpt	0.1313	0.1656	0.3550	0.3062
LegalBertpt (baseline)	0.0482	0.0670	0.1792	0.1394
LegalBertpt_PRE	0.0667	**0.0886**	0.1880	0.1680
LegalBertpt_NP	0.0645	0.0859	0.1916	0.1679
LegalBertpt_BERTimbau	0.0558	0.0729	0.1699	0.1536
LegalBertpt_LegalBERTimbau	**0.0702**	0.0869	**0.2010**	**0.1736**
LegalBertpt_JurisBERT	0.0555	0.0734	0.1798	0.1453
LegalBertpt_BERTimbauLaw	0.0576	0.0766	0.1823	0.1596
LegalBertpt_LegalBertpt	0.0468	0.0630	0.1400	0.1364
LegalBertpt_LaBSE	0.0550	0.0707	0.1684	0.1470
LegalBertpt_MPNet	0.0667	0.0856	0.1811	0.1652
LegalBertpt_MiniLM	0.0619	0.0779	0.1944	0.1606
LegalBertpt_FTBERTimbau	0.0632	0.0837	0.1961	0.1664
LegalBertpt_FTLegalBertpt	0.0634	0.0830	0.1806	0.1650
LaBSE (baseline)	0.0893	0.1080	0.2910	0.2268
LaBSE_PRE	**0.1143**	**0.1330**	**0.3201**	**0.2621**
LaBSE_NP	0.1115	0.1326	0.3050	0.2555
LaBSE_BERTimbau	0.1003	0.1205	0.3073	0.2428

Algorithm	MAP	MRP	MRR	nDCG
LaBSE_LegalBERTimbau	0.1031	0.1242	0.2860	0.2453
LaBSE_JurisBERT	0.1013	0.1196	0.3047	0.2497
LaBSE_BERTimbauLaw	0.1043	0.1268	0.3087	0.2515
LaBSE_LegalBertpt	0.0983	0.1185	0.3014	0.2409
LaBSE_LaBSE	0.0995	0.1192	0.3004	0.2391
LaBSE_MPNet	0.1049	0.1228	0.3186	0.2473
LaBSE_MiniLM	0.1057	0.1257	0.3130	0.2480
LaBSE_FTBERTimbau	0.1047	0.1267	0.3160	0.2550
LaBSE_FTLegalBertpt	0.1044	0.1251	0.3153	0.2504
MPNet (baseline)	0.1633	0.1891	0.4204	0.3491
MPNet_PRE	**0.1771**	**0.2030**	**0.4311**	**0.3645**
MPNet_NP	0.1724	0.1982	0.4244	0.3614
MPNet_BERTimbau	0.1658	0.1932	0.4179	0.3556
MPNet_LegalBERTimbau	0.1696	0.1969	0.4238	0.3580
MPNet_JurisBERT	0.1701	0.1958	0.4240	0.3575
MPNet_BERTimbauLaw	0.1677	0.1945	0.4283	0.3584
MPNet_LegalBertpt	0.1657	0.1922	0.4194	0.3542
MPNet_LaBSE	0.1678	0.1951	0.4233	0.3528
MPNet_MPNet	0.1713	0.1970	0.4238	0.3572
MPNet_MiniLM	0.1687	0.1955	0.4242	0.3521
MPNet_FTBERTimbau	0.1664	0.1939	0.4202	0.3594
MPNet_FTLegalBertpt	0.1647	0.1902	0.4215	0.3526
MiniLM (baseline)	0.1208	0.1454	0.3376	0.2692
MiniLM_PRE	0.1266	0.1502	0.3254	0.2796
MiniLM_NP	0.1279	0.1521	0.3487	0.2804
MiniLM_BERTimbau	0.1260	0.1508	0.3458	0.2742
MiniLM_LegalBERTimbau	0.1245	0.1492	0.3402	0.2767
MiniLM_JurisBERT	0.1271	0.1505	0.3513	0.2787
MiniLM_BERTimbauLaw	0.1253	0.1494	**0.3514**	0.2763
MiniLM_LegalBertpt	0.1233	0.1486	0.3391	0.2710
MiniLM_LaBSE	0.1190	0.1411	0.3229	0.2731

Algorithm	MAP	MRP	MRR	nDCG
MiniLM_MPNet	**0.1340**	**0.1591**	0.3486	**0.2824**
MiniLM_MiniLM	0.1209	0.1449	0.3306	0.2684
MiniLM_FTBERTimbau	0.1214	0.1462	0.3293	0.2810
MiniLM_FTLegalBertpt	0.1278	0.1521	0.3426	0.2759
FTBERTimbau (baseline)	0.1902	0.2163	0.4456	0.4070
FTBERTimbau_PRE	0.2012	0.2270	0.4633	0.4178
FTBERTimbau_NP	**0.2018**	**0.2291**	**0.4657**	**0.4214**
FTBERTimbau_BERTimbau	0.1861	0.2138	0.4176	0.4111
FTBERTimbau_LegalBERTimbau	0.1919	0.2172	0.4324	0.4099
FTBERTimbau_JurisBERT	0.1927	0.2183	0.4476	0.4103
FTBERTimbau_BERTimbauLaw	0.1905	0.2177	0.4368	0.4148
FTBERTimbau_LegalBertpt	0.1881	0.2161	0.4253	0.4095
FTBERTimbau_LaBSE	0.1962	0.2222	0.4451	0.4134
FTBERTimbau_MPNet	0.1973	0.2224	0.4520	0.4128
FTBERTimbau_MiniLM	0.1894	0.2163	0.4360	0.4074
FTBERTimbau_FTBERTimbau	0.1826	0.2094	0.4134	0.4049
FTBERTimbau_FTLegalBertpt	0.1938	0.2194	0.4424	0.4105
FTLegalBertpt (baseline)	0.1220	0.1523	0.2998	0.3050
FTLegalBertpt_PRE	0.1326	**0.1625**	0.3067	0.3104
FTLegalBertpt_NP	0.1285	0.1560	0.3028	0.3107
FTLegalBertpt_BERTimbau	0.1293	0.1594	0.3046	0.3137
FTLegalBertpt_LegalBERTimbau	0.1331	0.1620	0.3097	0.3164
FTLegalBertpt_JurisBERT	0.1280	0.1567	0.2970	**0.3175**
FTLegalBertpt_BERTimbauLaw	0.1298	0.1579	0.3082	0.3166
FTLegalBertpt_LegalBertpt	0.1296	0.1576	0.3040	0.3160
FTLegalBertpt_LaBSE	**0.1333**	0.1611	**0.3144**	0.3136
FTLegalBertpt_MPNet	0.1222	0.1515	0.2725	0.2862
FTLegalBertpt_MiniLM	0.1238	0.1528	0.2997	0.3044
FTLegalBertpt_FTBERTimbau	0.1283	0.1575	0.3067	0.3054
FTLegalBertpt_FTLegalBertpt	0.1282	0.1589	0.3075	0.3080

APPENDIX Q – RESULTS ACHIEVED BY THE BASELINES AND ULYSSES-RFSQ-ALL

Algorithm	MAP	MRP	MRR	nDCG
BM25L_PRE (baseline)	0.7684	0.6850	0.8597	0.8277
BM25L_PRE_PRE	0.7655	0.6852	0.8592	0.8246
BM25L_PRE_NP	0.7697	0.6873	0.8610	0.8284
BM25L_PRE_BERTimbau	0.7656	0.6876	0.8639	0.8259
BM25L_PRE_LegalBERTimbau	0.7686	0.6901	0.8641	0.8280
BM25L_PRE_JurisBERT	0.7691	0.6858	0.8615	0.8284
BM25L_PRE_BERTimbauLaw	0.7657	0.6853	0.8631	0.8274
BM25L_PRE_LegalBertpt	0.7640	0.6866	**0.8651**	0.8246
BM25L_PRE_LaBSE	**0.7704**	**0.6904**	0.8643	0.8296
BM25L_PRE_MPNet	0.7703	0.6883	0.8630	0.8294
BM25L_PRE_MiniLM	0.7696	0.6895	0.8648	**0.8299**
BM25L_PRE_FTBERTimbau	0.7657	0.6860	0.8633	0.8262
BM25L_PRE_FTLegalBertpt	0.7694	0.6872	0.8628	0.8280
BM25L_NP (baseline)	0.3801	0.3913	0.6811	0.5429
BM25L_NP_PRE	0.3839	0.3921	0.6834	0.5482
BM25L_NP_NP	0.3810	0.3914	0.6796	0.5421
BM25L_NP_BERTimbau	0.3831	0.3936	0.6829	0.5469
BM25L_NP_LegalBERTimbau	0.3813	0.3918	0.6821	**0.5515**
BM25L_NP_JurisBERT	0.3851	0.3946	0.6844	0.5494
BM25L_NP_BERTimbauLaw	0.3840	0.3918	0.6861	0.5488
BM25L_NP_LegalBertpt	0.3834	0.3907	0.6874	0.5477
BM25L_NP_LaBSE	0.3840	0.3926	**0.6877**	0.5470
BM25L_NP_MPNet	0.3849	**0.3965**	0.6793	0.5475
BM25L_NP_MiniLM	0.3790	0.3916	0.6781	0.5437
BM25L_NP_FTBERTimbau	**0.3853**	0.3937	0.6862	0.5508
BM25L_NP_FTLegalBertpt	0.3847	0.3932	0.6873	0.5503
OkapiBM25_PRE (baseline)	0.7183	0.6609	0.8607	0.7942
OkapiBM25_PRE_PRE	0.7212	0.6620	0.8630	0.7968

Algorithm	MAP	MRP	MRR	nDCG
OkapiBM25_PRE_NP	0.7212	0.6623	0.8636	0.7962
OkapiBM25_PRE_BERTimbau	0.7178	0.6612	0.8626	0.7944
OkapiBM25_PRE_LegalBERTimbau	0.7197	0.6615	0.8647	0.7961
OkapiBM25_PRE_JurisBERT	0.7210	0.6619	0.8643	0.7964
OkapiBM25_PRE_BERTimbauLaw	0.7164	0.6594	0.8638	0.7953
OkapiBM25_PRE_LegalBertpt	0.7175	0.6608	0.8631	0.7942
OkapiBM25_PRE_LaBSE	0.7214	0.6630	**0.8648**	0.7969
OkapiBM25_PRE_MPNet	**0.7217**	**0.6639**	0.8629	0.7966
OkapiBM25_PRE_MiniLM	0.7211	0.6621	0.8634	**0.7975**
OkapiBM25_PRE_FTBERTimbau	0.7175	0.6593	0.8626	0.7956
OkapiBM25_PRE_FTLegalBertpt	0.7183	0.6616	0.8636	0.7967
OkapiBM25_NP (baseline)	0.3704	0.3834	0.6718	0.5399
OkapiBM25_NP_PRE	0.3708	0.3838	0.6711	0.5394
OkapiBM25_NP_NP	0.3696	0.3826	0.6674	0.5386
OkapiBM25_NP_BERTimbau	0.3704	0.3843	0.6728	0.5407
OkapiBM25_NP_LegalBERTimbau	0.3742	**0.3878**	0.6765	**0.5451**
OkapiBM25_NP_JurisBERT	**0.3747**	0.3862	0.6743	0.5444
OkapiBM25_NP_BERTimbauLaw	0.3719	0.3841	0.6744	0.5427
OkapiBM25_NP_LegalBertpt	0.3692	0.3819	0.6744	0.5386
OkapiBM25_NP_LaBSE	0.3712	0.3836	0.6723	0.5421
OkapiBM25_NP_MPNet	0.3710	0.3839	0.6602	0.5424
OkapiBM25_NP_MiniLM	0.3729	0.3850	0.6742	0.5444
OkapiBM25_NP_FTBERTimbau	0.3709	0.3858	0.6736	0.5433
OkapiBM25_NP_FTLegalBertpt	0.3712	0.3850	**0.6777**	0.5434
BERTimbau (baseline)	0.0113	0.0209	0.0542	0.0490
BERTimbau_PRE	**0.0461**	0.0545	0.1079	**0.1071**
BERTimbau_NP	0.0416	0.0522	**0.1106**	0.1003
BERTimbau_BERTimbau	0.0308	0.0399	0.0851	0.0779
BERTimbau_LegalBERTimbau	0.0421	**0.0559**	0.1060	0.1051
BERTimbau_JurisBERT	0.0234	0.0319	0.0724	0.0642
BERTimbau_BERTimbauLaw	0.0282	0.0399	0.0934	0.0830

Algorithm	MAP	MRP	MRR	nDCG
BERTimbau_LegalBertpt	0.0215	0.0280	0.0576	0.0592
BERTimbau_LaBSE	0.0274	0.0368	0.0726	0.0745
BERTimbau_MPNet	0.0411	0.0552	0.1035	0.1029
BERTimbau_MiniLM	0.0332	0.0399	0.0830	0.0834
BERTimbau_FTBERTimbau	0.0333	0.0415	0.0908	0.0921
BERTimbau_FTLegalBertpt	0.0405	0.0502	0.1002	0.0997
LegalBERTimbau (baseline)	0.1221	0.1469	0.3320	0.2871
LegalBERTimbau_PRE	0.1312	0.1551	0.3451	0.3031
LegalBERTimbau_NP	**0.1375**	**0.1625**	**0.3557**	**0.3093**
LegalBERTimbau_BERTimbau	0.1254	0.1495	0.3351	0.2959
LegalBERTimbau_LegalBERTimbau	0.1308	0.1530	0.3238	0.3017
LegalBERTimbau_JurisBERT	0.1242	0.1463	0.3314	0.2939
LegalBERTimbau_BERTimbauLaw	0.1297	0.1538	0.3405	0.2992
LegalBERTimbau_LegalBertpt	0.1213	0.1434	0.3222	0.2901
LegalBERTimbau_LaBSE	0.1237	0.1466	0.3279	0.2944
LegalBERTimbau_MPNet	0.1346	0.1571	0.3520	0.3025
LegalBERTimbau_MiniLM	0.1230	0.1449	0.3259	0.2924
LegalBERTimbau_FTBERTimbau	0.1253	0.1505	0.3322	0.2965
LegalBERTimbau_FTLegalBertpt	0.1275	0.1511	0.3396	0.2985
JurisBERT (baseline)	0.0744	0.0924	0.2482	0.1862
JurisBERT_PRE	0.0860	0.1040	0.2489	0.2087
JurisBERT_NP	**0.0864**	**0.1078**	0.2441	**0.2090**
JurisBERT_BERTimbau	0.0809	0.1007	0.2554	0.2002
JurisBERT_LegalBERTimbau	0.0821	0.1020	**0.2615**	0.1989
JurisBERT_JurisBERT	0.0740	0.0912	0.2339	0.1838
JurisBERT_BERTimbauLaw	0.0775	0.0945	0.2413	0.1952
JurisBERT_LegalBertpt	0.0761	0.0948	0.2405	0.1917
JurisBERT_LaBSE	0.0774	0.0959	0.2452	0.1932
JurisBERT_MPNet	0.0792	0.1000	0.2487	0.1994
JurisBERT_MiniLM	0.0769	0.0960	0.2443	0.1947
JurisBERT_FTBERTimbau	0.0853	0.1043	0.2511	0.2081

Algorithm	MAP	MRP	MRR	nDCG
JurisBERT_FTLegalBertpt	0.0799	0.1006	0.2518	0.2019
BERTimbauLaw (baseline)	0.1239	0.1571	0.3481	0.2871
BERTimbauLaw_PRE	**0.1386**	**0.1728**	**0.3628**	**0.3125**
BERTimbauLaw_NP	0.1340	0.1687	0.3563	0.3037
BERTimbauLaw_BERTimbau	0.1281	0.1633	0.3462	0.2974
BERTimbauLaw_LegalBERTimbau	0.1327	0.1659	0.3584	0.2978
BERTimbauLaw_JurisBERT	0.1310	0.1630	0.3453	0.2932
BERTimbauLaw_BERTimbauLaw	0.1298	0.1631	0.3540	0.2960
BERTimbauLaw_LegalBertpt	0.1282	0.1625	0.3456	0.2960
BERTimbauLaw_LaBSE	0.1315	0.1662	0.3552	0.2960
BERTimbauLaw_MPNet	0.1361	0.1679	0.3562	0.3014
BERTimbauLaw_MiniLM	0.1332	0.1675	0.3584	0.2967
BERTimbauLaw_FTBERTimbau	0.1282	0.1622	0.3478	0.2966
BERTimbauLaw_FTLegalBertpt	0.1282	0.1623	0.3468	0.3021
LegalBertpt (baseline)	0.0482	0.0670	0.1792	0.1394
LegalBertpt_PRE	0.0645	**0.0867**	0.1864	0.1656
LegalBertpt_NP	0.0628	0.0841	0.1916	0.1657
LegalBertpt_BERTimbau	0.0524	0.0697	0.1657	0.1498
LegalBertpt_LegalBERTimbau	**0.0700**	0.0866	**0.2009**	**0.1731**
LegalBertpt_JurisBERT	0.0553	0.0731	0.1798	0.1448
LegalBertpt_BERTimbauLaw	0.0570	0.0763	0.1817	0.1589
LegalBertpt_LegalBertpt	0.0436	0.0602	0.1357	0.1331
LegalBertpt_LaBSE	0.0525	0.0689	0.1644	0.1448
LegalBertpt_MPNet	0.0659	0.0851	0.1813	0.1641
LegalBertpt_MiniLM	0.0615	0.0783	0.1946	0.1603
LegalBertpt_FTBERTimbau	0.0620	0.0832	0.1963	0.1651
LegalBertpt_FTLegalBertpt	0.0595	0.0789	0.1749	0.1604
LaBSE (baseline)	0.0893	0.1080	0.2910	0.2268
LaBSE_PRE	**0.1096**	**0.1293**	0.3147	**0.2567**
LaBSE_NP	0.1080	0.1291	0.3034	0.2514
LaBSE_BERTimbau	0.0977	0.1183	0.3053	0.2390

Algorithm	MAP	MRP	MRR	nDCG
LaBSE_LegalBERTimbau	0.1021	0.1233	0.2858	0.2438
LaBSE_JurisBERT	0.0999	0.1186	0.3006	0.2465
LaBSE_BERTimbauLaw	0.1014	0.1244	0.3022	0.2462
LaBSE_LegalBertpt	0.0962	0.1157	0.2981	0.2369
LaBSE_LaBSE	0.0977	0.1173	0.2985	0.2374
LaBSE_MPNet	0.1044	0.1221	**0.3194**	0.2465
LaBSE_MiniLM	0.1025	0.1231	0.3112	0.2440
LaBSE_FTBERTimbau	0.1011	0.1246	0.3092	0.2484
LaBSE_FTLegalBertpt	0.1012	0.1227	0.3116	0.2433
MPNet (baseline)	0.1633	0.1891	0.4204	0.3491
MPNet_PRE	**0.1741**	**0.2004**	**0.4281**	**0.3620**
MPNet_NP	0.1692	0.1950	0.4220	0.3591
MPNet_BERTimbau	0.1643	0.1916	0.4165	0.3532
MPNet_LegalBERTimbau	0.1700	0.1977	0.4252	0.3578
MPNet_JurisBERT	0.1680	0.1949	0.4207	0.3544
MPNet_BERTimbauLaw	0.1668	0.1941	0.4280	0.3573
MPNet_LegalBertpt	0.1640	0.1913	0.4181	0.3519
MPNet_LaBSE	0.1674	0.1948	0.4227	0.3521
MPNet_MPNet	0.1712	0.1968	0.4238	0.3572
MPNet_MiniLM	0.1683	0.1956	0.4241	0.3516
MPNet_FTBERTimbau	0.1655	0.1937	0.4195	0.3577
MPNet_FTLegalBertpt	0.1645	0.1903	0.4217	0.3523
MiniLM (baseline)	0.1208	0.1454	0.3376	0.2692
MiniLM_PRE	0.1241	0.1482	0.3247	0.2765
MiniLM_NP	0.1263	0.1506	0.3470	0.2780
MiniLM_BERTimbau	0.1240	0.1498	0.3427	0.2726
MiniLM_LegalBERTimbau	0.1241	0.1489	0.3402	0.2764
MiniLM_JurisBERT	0.1253	0.1491	0.3480	0.2760
MiniLM_BERTimbauLaw	0.1247	0.1488	**0.3503**	0.2762
MiniLM_LegalBertpt	0.1207	0.1466	0.3353	0.2680
MiniLM_LaBSE	0.1180	0.1405	0.3213	0.2722

Algorithm	**MAP**	**MRP**	**MRR**	**nDCG**
MiniLM_MPNet	**0.1339**	**0.1589**	0.3487	**0.2824**
MiniLM_MiniLM	0.1208	0.1445	0.3300	0.2675
MiniLM_FTBERTimbau	0.1203	0.1456	0.3291	0.2787
MiniLM_FTLegalBertpt	0.1278	0.1520	0.3428	0.2757
FTBERTimbau (baseline)	0.1902	0.2163	0.4456	0.4070
FTBERTimbau_PRE	0.1971	0.2224	0.4576	0.4156
FTBERTimbau_NP	**0.1987**	**0.2262**	**0.4623**	**0.4191**
FTBERTimbau_BERTimbau	0.1823	0.2098	0.4143	0.4067
FTBERTimbau_LegalBERTimbau	0.1909	0.2167	0.4329	0.4091
FTBERTimbau_JurisBERT	0.1915	0.2166	0.4445	0.4086
FTBERTimbau_BERTimbauLaw	0.1890	0.2163	0.4357	0.4137
FTBERTimbau_LegalBertpt	0.1824	0.2105	0.4172	0.4038
FTBERTimbau_LaBSE	0.1941	0.2198	0.4431	0.4110
FTBERTimbau_MPNet	0.1969	0.2214	0.4515	0.4120
FTBERTimbau_MiniLM	0.1890	0.2156	0.4360	0.4067
FTBERTimbau_FTBERTimbau	0.1810	0.2082	0.4132	0.4029
FTBERTimbau_FTLegalBertpt	0.1930	0.2182	0.4419	0.4098
FTLegalBertpt (baseline)	0.1220	0.1523	0.2998	0.3050
FTLegalBertpt_PRE	0.1303	0.1605	0.3044	0.3086
FTLegalBertpt_NP	0.1270	0.1542	0.3028	0.3085
FTLegalBertpt_BERTimbau	0.1269	0.1571	0.3039	0.3083
FTLegalBertpt_LegalBERTimbau	**0.1329**	**0.1614**	0.3121	0.3154
FTLegalBertpt_JurisBERT	0.1280	0.1549	0.3017	0.3142
FTLegalBertpt_BERTimbauLaw	0.1292	0.1561	0.3109	**0.3158**
FTLegalBertpt_LegalBertpt	0.1260	0.1541	0.3008	0.3108
FTLegalBertpt_LaBSE	0.1319	0.1602	**0.3135**	0.3119
FTLegalBertpt_MPNet	0.1223	0.1512	0.2727	0.2868
FTLegalBertpt_MiniLM	0.1237	0.1528	0.2998	0.3046
FTLegalBertpt_FTBERTimbau	0.1270	0.1557	0.3062	0.3050
FTLegalBertpt_FTLegalBertpt	0.1281	0.1587	0.3075	0.3079

APPENDIX R – COMPARISON BETWEEN THE MAP AND NDCG RESULTS OF THE FOUR ULYSSES-RFSQ VERSIONS

	OR		RI		DRL		ALL	
Algorithm	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**
BM25L_PRE_PRE	**0.7656**	**0.8247**	0.7653	0.8244	0.7649	0.8243	0.7655	0.8246
BM25L_PRE_NP	0.7685	0.8277	**0.7699**	**0.8285**	0.7677	0.8271	0.7697	0.8284
BM25L_PRE_BERTimbau	0.7641	0.8248	0.7647	0.8253	0.7643	0.8248	**0.7656**	**0.8259**
BM25L_PRE_LegalBERTimbau	0.7674	0.8261	0.7679	0.8272	0.7671	0.8262	**0.7686**	**0.8280**
BM25L_PRE_JurisBERT	0.7688	0.8279	0.7688	0.8283	0.7684	0.8274	**0.7691**	**0.8284**
BM25L_PRE_BERTimbauLaw	0.7647	0.8263	0.7649	0.8266	0.7647	0.8260	**0.7657**	**0.8274**
BM25L_PRE_LegalBertpt	0.7628	0.8242	0.7631	0.8243	0.7624	0.8237	**0.7640**	**0.8246**
BM25L_PRE_LaBSE	0.7685	0.8278	0.7695	0.8288	0.7685	0.8276	**0.7704**	**0.8296**
BM25L_PRE_MPNet	0.7699	0.8287	0.7701	**0.8294**	0.7690	0.8279	**0.7703**	**0.8294**
BM25L_PRE_MiniLM	0.7683	0.8280	0.7690	0.8295	0.7683	0.8281	**0.7696**	**0.8299**
BM25L_PRE_FTBERTimbau	0.7641	0.8256	0.7648	0.8258	0.7645	0.8257	**0.7657**	**0.8262**
BM25L_PRE_FTLegalBertpt	0.7688	0.8277	0.7692	0.8279	0.7680	0.8270	**0.7694**	**0.8280**
BM25L_NP_PRE	0.3847	**0.5485**	0.3830	0.5479	**0.3856**	**0.5485**	0.3839	0.5482
BM25L_NP_NP	**0.3818**	**0.5438**	0.3811	0.5422	0.3817	0.5436	0.3810	0.5421
BM25L_NP_BERTimbau	0.3836	0.5465	0.3832	0.5461	**0.3839**	**0.5476**	0.3831	0.5469
BM25L_NP_LegalBERTimbau	0.3805	0.5500	0.3799	0.5505	**0.3815**	**0.5517**	0.3813	0.5515
BM25L_NP_JurisBERT	0.3848	0.5482	0.3848	0.5492	**0.3855**	0.5490	0.3851	**0.5494**
BM25L_NP_BERTimbauLaw	0.3839	0.5465	0.3838	0.5483	0.3845	0.5479	0.3840	0.5488
BM25L_NP_LegalBertpt	0.3843	0.5476	0.3832	0.5467	**0.3850**	**0.5493**	0.3834	0.5477
BM25L_NP_LaBSE	0.3853	0.5486	0.3837	0.5467	**0.3859**	**0.5494**	0.3840	0.5470
BM25L_NP_MPNet	0.3853	0.5486	0.3846	0.5471	**0.3857**	**0.5490**	0.3849	0.5475
BM25L_NP_MiniLM	0.3783	0.5434	0.3782	0.5432	**0.3793**	**0.5440**	0.3790	0.5437
BM25L_NP_FTBERTimbau	0.3856	0.5487	0.3855	0.5500	**0.3858**	0.5505	0.3853	**0.5508**
BM25L_NP_FTLegalBertpt	0.3852	0.5489	0.3849	0.5492	**0.3854**	**0.5506**	0.3847	0.5503
OkapiBM25_PRE_PRE	0.7195	0.7952	**0.7212**	0.7966	0.7191	0.7952	**0.7212**	**0.7968**
OkapiBM25_PRE_NP	0.7204	0.7955	**0.7213**	**0.7962**	0.7191	0.7946	0.7212	**0.7962**

	OR		RI		DRL		ALL	
Algorithm	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**
OkapiBM25_PRE_BERTimbau	**0.7185**	0.7944	0.7183	**0.7951**	0.7176	0.7936	0.7178	0.7944
OkapiBM25_PRE_LegalBERTimbau	0.7196	0.7951	0.7195	0.7957	0.7184	0.7946	**0.7197**	**0.7961**
OkapiBM25_PRE_JurisBERT	0.7201	0.7957	0.7205	0.7962	0.7193	0.7950	**0.7210**	**0.7964**
OkapiBM25_PRE_BERTimbauLaw	0.7149	0.7938	0.7154	0.7949	0.7156	0.7943	**0.7164**	**0.7953**
OkapiBM25_PRE_LegalBertpt	0.7170	0.7934	**0.7175**	**0.7943**	0.7168	0.7934	**0.7175**	0.7942
OkapiBM25_PRE_LaBSE	0.7201	0.7957	0.7209	0.7967	0.7192	0.7951	**0.7214**	**0.7969**
OkapiBM25_PRE_MPNet	0.7212	0.7964	0.7213	0.7965	0.7201	0.7955	**0.7217**	**0.7966**
OkapiBM25_PRE_MiniLM	0.7204	0.7967	0.7204	0.7971	0.7197	0.7962	**0.7211**	**0.7975**
OkapiBM25_PRE_FTBERTimbau	0.7176	0.7952	**0.7184**	**0.7963**	0.7167	0.7947	0.7175	0.7956
OkapiBM25_PRE_FTLegalBertpt	0.7171	0.7954	0.7173	0.7960	0.7178	0.7958	**0.7183**	**0.7967**
OkapiBM25_NP_PRE	**0.3715**	**0.5403**	0.3709	0.5395	0.3714	0.5401	0.3708	0.5394
OkapiBM25_NP_NP	0.3701	**0.5395**	0.3693	0.5386	**0.3702**	**0.5395**	0.3696	0.5386
OkapiBM25_NP_BERTimbau	0.3711	0.5412	0.3696	0.5397	**0.3719**	**0.5421**	0.3704	0.5407
OkapiBM25_NP_LegalBERTimbau	**0.3748**	0.5452	0.3746	0.5451	0.3744	**0.5455**	0.3742	0.5451
OkapiBM25_NP_JurisBERT	0.3753	0.5434	0.3741	0.5435	**0.3760**	0.5440	0.3747	**0.5444**
OkapiBM25_NP_BERTimbauLaw	0.3716	0.5425	0.3717	**0.5430**	**0.3725**	**0.5430**	0.3719	0.5427
OkapiBM25_NP_LegalBertpt	0.3707	0.5414	0.3685	0.5385	**0.3718**	**0.5422**	0.3692	0.5386
OkapiBM25_NP_LaBSE	0.3729	0.5433	0.3711	0.5422	**0.3732**	**0.5437**	0.3712	0.5421
OkapiBM25_NP_MPNet	0.3705	0.5423	0.3696	0.5410	**0.3718**	**0.5434**	0.3710	0.5424
OkapiBM25_NP_MiniLM	0.3735	0.5445	0.3725	0.5439	**0.3740**	**0.5450**	0.3729	0.5444
OkapiBM25_NP_FTBERTimbau	0.3712	0.5436	0.3707	**0.5437**	**0.3715**	0.5432	0.3709	0.5433
OkapiBM25_NP_FTLegalBertpt	0.3728	0.5448	0.3703	0.5428	**0.3738**	**0.5452**	0.3712	0.5434
BERTimbau_PRE	**0.0504**	0.1076	0.0479	0.1057	0.0483	**0.1090**	0.0461	0.1071
BERTimbau_NP	**0.0452**	**0.1057**	0.0430	0.1036	0.0431	0.1016	0.0416	0.1003
BERTimbau_BERTimbau	0.0310	**0.0800**	0.0297	0.0786	**0.0317**	0.0783	0.0308	0.0779
BERTimbau_LegalBERTimbau	0.0419	0.1047	0.0412	0.1040	**0.0422**	**0.1055**	0.0421	0.1051
BERTimbau_JurisBERT	0.0227	0.0639	0.0226	0.0638	**0.0235**	**0.0644**	0.0234	0.0642
BERTimbau_BERTimbauLaw	0.0278	0.0830	0.0270	0.0830	**0.0287**	**0.0831**	0.0282	0.0830
BERTimbau_LegalBertpt	**0.0238**	0.0604	0.0223	0.0587	0.0230	**0.0610**	0.0215	0.0592
BERTimbau_LaBSE	**0.0310**	**0.0779**	0.0290	0.0764	0.0298	0.0762	0.0274	0.0745

	OR		RI		DRL		ALL	
Algorithm	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**
BERTimbau_MPNet	0.0412	0.1010	0.0406	0.1010	**0.0414**	**0.1033**	0.0411	0.1029
BERTimbau_MiniLM	**0.0365**	**0.0881**	0.0345	0.0866	0.0352	0.0853	0.0332	0.0834
BERTimbau_FTBERTimbau	**0.0353**	0.0943	0.0343	0.0930	0.0345	**0.0956**	0.0333	0.0921
BERTimbau_FTLegalBertpt	**0.0430**	0.1029	0.0414	0.0984	0.0423	**0.1038**	0.0405	0.0997
LegalBERTimbau_PRE	**0.1378**	**0.3121**	0.1349	0.3083	0.1343	0.3076	0.1312	0.3031
LegalBERTimbau_NP	**0.1431**	**0.3192**	0.1400	0.3135	0.1397	0.3156	0.1375	0.3093
LegalBERTimbau_BERTimbau	**0.1283**	**0.3004**	0.1259	0.2984	0.1277	0.2979	0.1254	0.2959
LegalBERTimbau_LegalBERTimbau	0.1284	0.3006	0.1274	0.2996	**0.1314**	**0.3031**	0.1308	0.3017
LegalBERTimbau_JurisBERT	0.1270	0.2961	0.1233	0.2930	**0.1280**	**0.2973**	0.1242	0.2939
LegalBERTimbau_BERTimbauLaw	**0.1321**	**0.3033**	0.1302	0.3012	0.1318	0.3030	0.1297	0.2992
LegalBERTimbau_LegalBertpt	0.1240	0.2929	0.1191	0.2877	**0.1265**	**0.2966**	0.1213	0.2901
LegalBERTimbau_LaBSE	0.1274	**0.2994**	0.1233	0.2938	**0.1284**	0.2992	0.1237	0.2944
LegalBERTimbau_MPNet	0.1336	**0.3026**	0.1336	**0.3026**	**0.1346**	0.3025	**0.1346**	0.3025
LegalBERTimbau_MiniLM	0.1264	0.2952	0.1217	0.2902	**0.1281**	**0.2977**	0.1230	0.2924
LegalBERTimbau_FTBERTimbau	**0.1282**	**0.3030**	0.1251	0.2991	**0.1282**	0.3024	0.1253	0.2965
LegalBERTimbau_FTLegalBertpt	0.1304	**0.3034**	0.1273	0.2999	**0.1308**	0.3023	0.1275	0.2985
JurisBERT_PRE	**0.0892**	**0.2111**	0.0870	0.2085	0.0885	0.2110	0.0860	0.2087
JurisBERT_NP	**0.0905**	**0.2183**	0.0890	0.2159	0.0884	0.2116	0.0864	0.2090
JurisBERT_BERTimbau	0.0822	**0.2017**	0.0812	0.2010	**0.0823**	0.2008	0.0809	0.2002
JurisBERT_LegalBERTimbau	**0.0848**	**0.2040**	0.0823	0.2003	0.0844	0.2036	0.0821	0.1989
JurisBERT_JurisBERT	**0.0757**	**0.1874**	0.0739	0.1826	0.0756	0.1869	0.0740	0.1838
JurisBERT_BERTimbauLaw	0.0798	**0.2007**	0.0773	0.1966	**0.0801**	0.1995	0.0775	0.1952
JurisBERT_LegalBertpt	**0.0775**	**0.1918**	0.0755	0.1902	**0.0775**	0.1910	0.0761	0.1917
JurisBERT_LaBSE	0.0788	0.1968	0.0759	0.1932	**0.0803**	**0.1969**	0.0774	0.1932
JurisBERT_MPNet	0.0798	0.2022	0.0781	0.1986	**0.0807**	**0.2031**	0.0792	0.1994
JurisBERT_MiniLM	0.0783	**0.2015**	0.0755	0.1972	**0.0795**	0.1995	0.0769	0.1947
JurisBERT_FTBERTimbau	**0.0856**	0.2078	0.0846	**0.2084**	0.0864	0.2077	0.0853	0.2081
JurisBERT_FTLegalBertpt	0.0816	**0.2066**	0.0795	0.2054	**0.0823**	0.2038	0.0799	0.2019
BERTimbauLaw_PRE	**0.1462**	**0.3218**	0.1408	0.3158	0.1436	0.3181	0.1386	0.3125
BERTimbauLaw_NP	**0.1367**	**0.3073**	0.1347	0.3039	0.1364	0.3068	0.1340	0.3037

	OR		RI		DRL		ALL	
Algorithm	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**
BERTimbauLaw_BERTimbau	0.1309	**0.3014**	0.1291	0.2983	**0.1310**	0.3006	0.1281	0.2974
BERTimbauLaw_LegalBERTimbau	**0.1337**	**0.3012**	0.1331	0.3005	0.1330	0.2987	0.1327	0.2978
BERTimbauLaw_JurisBERT	**0.1328**	0.2955	0.1311	0.2932	0.1327	**0.2967**	0.1310	0.2932
BERTimbauLaw_BERTimbauLaw	**0.1301**	**0.2981**	0.1297	0.2969	0.1300	0.2975	0.1298	0.2960
BERTimbauLaw_LegalBertpt	0.1301	**0.2983**	0.1287	0.2950	**0.1302**	0.2975	0.1282	0.2960
BERTimbauLaw_LaBSE	**0.1325**	**0.2985**	0.1324	0.2980	0.1315	0.2965	0.1315	0.2960
BERTimbauLaw_MPNet	0.1356	0.3016	0.1356	0.3012	**0.1362**	**0.3017**	0.1361	0.3014
BERTimbauLaw_MiniLM	**0.1332**	0.2961	**0.1332**	0.2960	**0.1332**	**0.2967**	**0.1332**	**0.2967**
BERTimbauLaw_FTBERTimbau	0.1298	**0.2985**	0.1276	0.2965	**0.1304**	0.2978	0.1282	0.2966
BERTimbauLaw_FTLegalBertpt	0.1304	0.3060	0.1272	0.3013	**0.1313**	**0.3062**	0.1282	0.3021
LegalBertpt_PRE	**0.0672**	**0.1667**	0.0650	0.1649	0.0667	0.1680	0.0645	0.1656
LegalBertpt_NP	0.0640	**0.1687**	0.0625	0.1667	**0.0645**	0.1679	0.0628	0.1657
LegalBertpt_BERTimbau	0.0551	0.1516	0.0521	0.1479	**0.0558**	**0.1536**	0.0524	0.1498
LegalBertpt_LegalBERTimbau	**0.0706**	**0.1751**	0.0700	0.1743	0.0702	0.1736	0.0700	0.1731
LegalBertpt_JurisBERT	0.0547	0.1452	0.0545	0.1448	**0.0555**	**0.1453**	0.0553	0.1448
LegalBertpt_BERTimbauLaw	0.0572	0.1593	0.0564	0.1584	**0.0576**	**0.1596**	0.0570	0.1589
LegalBertpt_LegalBertpt	**0.0473**	0.1328	0.0444	0.1299	0.0468	**0.1364**	0.0436	0.1331
LegalBertpt_LaBSE	**0.0551**	0.1460	0.0528	0.1432	0.0550	**0.1470**	0.0525	0.1448
LegalBertpt_MPNet	**0.0670**	0.1638	0.0660	0.1629	0.0667	**0.1652**	0.0659	0.1641
LegalBertpt_MiniLM	**0.0631**	**0.1620**	0.0626	0.1615	0.0619	0.1606	0.0615	0.1603
LegalBertpt_FTBERTimbau	0.0626	0.1661	0.0613	0.1654	**0.0632**	**0.1664**	0.0620	0.1651
LegalBertpt_FTLegalBertpt	**0.0637**	0.1644	0.0606	0.1596	0.0634	**0.1650**	0.0595	0.1604
LaBSE_PRE	**0.1169**	**0.2655**	0.1126	0.2613	0.1143	0.2621	0.1096	0.2567
LaBSE_NP	0.1114	**0.2555**	0.1075	0.2517	**0.1115**	**0.2555**	0.1080	0.2514
LaBSE_BERTimbau	**0.1010**	0.2440	0.0980	0.2398	0.1003	**0.2428**	0.0977	0.2390
LaBSE_LegalBERTimbau	0.1018	0.2434	0.1008	0.2419	**0.1031**	**0.2453**	0.1021	0.2438
LaBSE_JurisBERT	**0.1016**	0.2485	0.1005	0.2460	0.1013	**0.2497**	0.0999	0.2465
LaBSE_BERTimbauLaw	0.1029	0.2502	0.1004	0.2452	**0.1043**	**0.2515**	0.1014	0.2462
LaBSE_LegalBertpt	**0.0983**	**0.2412**	0.0963	0.2377	**0.0983**	0.2409	0.0962	0.2369
LaBSE_LaBSE	**0.0997**	**0.2397**	0.0979	0.2371	0.0995	0.2391	0.0977	0.2374

	OR		RI		DRL		ALL	
Algorithm	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**
LaBSE_MPNet	**0.1052**	0.2463	0.1047	0.2455	0.1049	**0.2473**	0.1044	0.2465
LaBSE_MiniLM	0.1052	**0.2483**	0.1023	0.2444	**0.1057**	0.2480	0.1025	0.2440
LaBSE_FTBERTimbau	0.1030	0.2547	0.0995	0.2507	**0.1047**	**0.2550**	0.1011	0.2484
LaBSE_FTLegalBertpt	**0.1054**	**0.2527**	0.1021	0.2473	0.1044	0.2504	0.1012	0.2433
MPNet_PRE	**0.1793**	**0.3669**	0.1763	0.3645	0.1771	0.3645	0.1741	0.3620
MPNet_NP	**0.1730**	**0.3619**	0.1701	0.3588	0.1724	0.3614	0.1692	0.3591
MPNet_BERTimbau	0.1656	**0.3559**	0.1642	0.3539	**0.1658**	0.3556	0.1643	0.3532
MPNet_LegalBERTimbau	0.1698	0.3578	**0.1703**	0.3578	0.1696	**0.3580**	0.1700	0.3578
MPNet_JurisBERT	0.1690	0.3560	0.1676	0.3534	**0.1701**	**0.3575**	0.1680	0.3544
MPNet_BERTimbauLaw	**0.1687**	0.3581	0.1677	0.3571	0.1677	**0.3584**	0.1668	0.3573
MPNet_LegalBertpt	0.1651	0.3539	0.1640	0.3525	**0.1657**	**0.3542**	0.1640	0.3519
MPNet_LaBSE	0.1675	**0.3537**	0.1671	0.3531	**0.1678**	0.3528	0.1674	0.3521
MPNet_MPNet	**0.1722**	**0.3582**	0.1721	**0.3582**	0.1713	0.3572	0.1712	0.3572
MPNet_MiniLM	**0.1690**	**0.3525**	0.1686	0.3520	0.1687	0.3521	0.1683	0.3516
MPNet_FTBERTimbau	**0.1669**	0.3591	0.1660	0.3574	0.1664	**0.3594**	0.1655	0.3577
MPNet_FTLegalBertpt	0.**1651**	**0.3529**	0.1649	0.3527	0.1647	0.3526	0.1645	0.3523
MiniLM_PRE	**0.1282**	0.2789	0.1258	0.2768	0.1266	**0.2796**	0.1241	0.2765
MiniLM_NP	**0.1302**	**0.2813**	0.1286	0.2790	0.1279	0.2804	0.1263	0.2780
MiniLM_BERTimbau	0.1258	0.2740	0.1239	0.2727	**0.1260**	**0.2742**	0.1240	0.2726
MiniLM_LegalBERTimbau	**0.1245**	0.2755	0.1240	0.2747	**0.1245**	0.2767	0.1241	**0.2764**
MiniLM_JurisBERT	**0.1272**	0.2785	0.1255	0.2764	0.1271	**0.2787**	0.1253	0.2760
MiniLM_BERTimbauLaw	**0.1254**	0.2768	0.1250	**0.2771**	0.1253	0.2763	0.1247	0.2762
MiniLM_LegalBertpt	0.1228	**0.2714**	0.1206	0.2674	**0.1233**	0.2710	0.1207	0.2680
MiniLM_LaBSE	**0.1195**	0.2697	0.1184	0.2687	0.1190	**0.2731**	0.1180	0.2722
MiniLM_MPNet	0.1338	**0.2825**	0.1337	**0.2825**	**0.1340**	0.2824	0.1339	0.2824
MiniLM_MiniLM	0.1206	0.2682	0.1205	0.2673	**0.1209**	**0.2684**	0.1208	0.2675
MiniLM_FTBERTimbau	**0.1259**	0.2758	0.1248	0.2747	0.1214	**0.2810**	0.1203	0.2787
MiniLM_FTLegalBertpt	0.1285	0.2759	0.1283	0.2755	0.1278	0.2759	0.1278	0.2757
FTBERTimbau_PRE	**0.2024**	**0.4186**	0.1985	0.4164	0.2012	0.4178	0.1971	0.4156
FTBERTimbau_NP	**0.2024**	0.4202	0.1996	0.4175	0.2018	**0.4214**	0.1987	0.4191

	OR		RI		DRL		ALL	
Algorithm	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**	**MAP**	**nDCG**
FTBERTimbau_BERTimbau	0.1839	0.4065	0.1809	0.4029	**0.1861**	**0.4111**	0.1823	0.4067
FTBERTimbau_LegalBERTimbau	0.1907	0.4091	0.1898	0.4081	**0.1919**	**0.4099**	0.1909	0.4091
FTBERTimbau_JurisBERT	0.1927	0.4098	0.1915	0.4082	0.1927	0.4103	0.1915	0.4086
FTBERTimbau_BERTimbauLaw	0.1900	0.4111	0.1889	0.4106	0.1905	0.4148	0.1890	0.4137
FTBERTimbau_LegalBertpt	0.1852	0.4084	0.1801	0.4034	0.1881	0.4095	0.1824	0.4038
FTBERTimbau_LaBSE	0.1949	0.4134	0.1927	0.4106	0.1962	0.4134	0.1941	0.4110
FTBERTimbau_MPNet	0.1963	0.4124	0.1959	0.4116	0.1973	0.4128	0.1969	0.4120
FTBERTimbau_MiniLM	0.1883	0.4070	0.1880	0.4064	0.1894	0.4074	0.1890	0.4067
FTBERTimbau_FTBERTimbau	0.1823	0.4023	0.1811	0.4006	0.1826	0.4049	0.1810	0.4029
FTBERTimbau_FTLegalBertpt	0.1938	0.4104	0.1930	0.4095	0.1938	0.4105	0.1930	0.4098
FTLegalBertpt_PRE	**0.1367**	**0.3128**	0.1340	0.3114	0.1326	0.3104	0.1303	0.3086
FTLegalBertpt_NP	**0.1285**	0.3098	0.1266	0.3080	**0.1285**	**0.3107**	0.1270	0.3085
FTLegalBertpt_BERTimbau	0.1286	0.3123	0.1261	0.3074	**0.1293**	**0.3137**	0.1269	0.3083
FTLegalBertpt_LegalBERTimbau	**0.1342**	**0.3175**	0.1336	0.3166	0.1331	0.3164	0.1329	0.3154
FTLegalBertpt_JurisBERT	**0.1289**	0.3168	0.1277	0.3133	0.1280	**0.3175**	0.1280	0.3142
FTLegalBertpt_BERTimbauLaw	0.1294	0.3156	0.1282	0.3144	**0.1298**	**0.3166**	0.1292	0.3158
FTLegalBertpt_LegalBertpt	0.1288	0.3138	0.1250	0.3093	**0.1296**	**0.3160**	0.1260	0.3108
FTLegalBertpt_LaBSE	**0.1333**	0.3134	0.1318	0.3115	**0.1333**	**0.3136**	0.1319	0.3119
FTLegalBertpt_MPNet	0.1222	0.2843	0.1220	0.2850	0.1222	0.2862	**0.1223**	**0.2868**
FTLegalBertpt_MiniLM	0.1232	0.3032	0.1230	0.3034	**0.1238**	0.3044	0.1237	**0.3046**
FTLegalBertpt_FTBERTimbau	**0.1292**	**0.3058**	0.1277	0.3054	0.1283	0.3054	0.1270	0.3050
FTLegalBertpt_FTLegalBertpt	**0.1290**	**0.3093**	0.1289	0.3089	0.1282	0.3080	0.1281	0.3079