PASSAGE RETRIEVAL ON TURKISH LEGAL TEXTS USING BERT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SEDA CIVELEK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2024

Approval of the thesis:

PASSAGE RETRIEVAL ON TURKISH LEGAL TEXTS USING BERT

submitted by **SEDA CIVELEK** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** _______________

Prof. Dr. Nihan Kesim Çiçekli
Supervisor, **Computer Engineering, METU** _______________

Examining Committee Members:

Prof. Dr. Fazlı Can
Computer Engineering, Bilkent University _______________

Prof. Dr. Nihan Kesim Çiçekli
Computer Engineering, METU _______________

Assist. Prof. Dr. Çağrı Toraman
Computer Engineering, METU _______________

Date:04.09.2024

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Seda Civelek

Signature :

ABSTRACT

PASSAGE RETRIEVAL ON TURKISH LEGAL TEXTS USING BERT

Civelek, Seda
M.S., Department of Computer Engineering
Supervisor: Prof. Dr. Nihan Kesim Çiçekli

September 2024, 71 pages

Legal professionals often face challenges in efficiently accessing relevant information from extensive and complex legal documents. This thesis presents a method for passage retrieval in Turkish legal texts utilizing the Bidirectional Encoder Representations from Transformers (BERT) model. The research aims to enhance the retrieval process by leveraging contextual embeddings to understand the nuanced language and terminologies used in legal documents. The study involves the creation of a dataset from Turkish legal books and the application of both BM25 and BERT models for retrieval tasks. Results indicate that the combined use of BM25 and BERT improves the accuracy and relevance of retrieved passages, offering a promising tool for legal research in the Turkish context.

Keywords: Passage Retrieval, Turkish Legal Texts, BM25, BERT, Information Retrieval, Legal Research

ÖZ

BERT KULLANARAK TÜRKÇE HUKUK METİNLERİNDEN PASAJ BULMA

Civelek, Seda
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Nihan Kesim Çiçekli

Eylül 2024 , 71 sayfa

Hukuk profesyonelleri, kapsamlı ve karmaşık hukuk belgelerinden ilgili bilgilere verimli bir şekilde erişme konusunda genellikle zorluklarla karşılaşmaktadır. Bu tez, Bidirectional Encoder Representations from Transformers (BERT) modelini kullanarak Türkçe hukuk metinlerinde pasaj bulma yöntemini sunmaktadır. Araştırma, bağlamsal gömülüleri kullanarak hukuk belgelerinde kullanılan nüanslı dil ve terminolojiyi anlamayı amaçlayarak arama sürecini geliştirmeyi hedeflemektedir. Çalışma, Türkçe hukuk kitaplarından bir veri seti oluşturulmasını ve BM25 ile BERT modellerinin arama görevleri için uygulanmasını içermektedir. Sonuçlar, BM25 ve BERT modellerinin bir arada kullanılmasının, bulunan pasajların doğruluğunu ve alaka düzeyini artırdığını ve Türk hukuk araştırmaları için umut verici bir araç sunduğunu göstermektedir.

Anahtar Kelimeler: Pasaj Bulma, Türkçe Hukuk Metinleri, BM25, BERT, Bilgi Getirme, Hukuk Araştırması

To my beloved family

ACKNOWLEDGMENTS

I am deeply grateful to my supervisor, Prof. Dr. Nihan Kesim Çiçekli. Her exceptional expertise, steadfast support, and boundless patience have greatly enhanced my experience in graduate school.

I would also like to thank my thesis committee members, Prof. Dr. Fazlı Can and Assist. Prof. Dr. Çağrı Toraman, for their precious time and valuable feedback.

I am also thankful to my mother, Songül Civelek, and my father Ahmet Civelek for their never-ending support and love. I always feel their hand on my back. I learned from them to never give up.

I want to thank the members of Seckin Yayincilik, especially Emre Özdemir and Sedat Akel, for their help and valuable contributions during my research process.

Morever, I would like to thank my dearest boyfriend, İrfan Emre Demiray, for always being by my side, supporting me in every moment of my life, and making my life better. I also thank my dearest friends, Yusuf Aydoğdu, Yüksel Pelin Kılıç, Ömer Faruk Özçelik, Melih Çekçi, Oğuz Burak Uyar, Nursena Gültekin, Binnur Nevruz, İsmail Harun Has, for their endless support. They always helped me both physically and mentally.

I am also grateful to METU Computer Engineering faculty members for their patience and guidance.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vi

ACKNOWLEDGMENTS viii

TABLE OF CONTENTS ix

LIST OF TABLES xii

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xv

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation and Problem Definition 1

1.2 Proposed Methods and Models 3

1.3 Contributions 4

1.4 The Outline of the Thesis 4

2 LITERATURE SURVEY 7

2.1 Legal Information Retrieval 7

2.2 Neural Approaches to Document Retrieval 8

2.2.1 Traditional Word Embeddings 8

2.2.2 Large Language Models in IR 10

2.2.2.1 BERT 10

2.2.2.2 BERTurk 11

2.3 BERT in Document Retrieval 12

2.4 BERTurk-Legal 13

2.4.1 Retrieval Augmented Generation 14

3 DATA SET PREPARATION AND PASSAGE RANKING MODELS 15

3.1 Data Set 16

3.1.1 Preparation of Tree Structured Title Paths and Paragraphs . . . 16

3.1.1.1 Title Path Extraction 19

3.1.1.2 Paragraph Extraction 20

3.1.1.3 Post-processing 24

3.2 BERTurk-Seckin: Pre-training BERT with Legal Book Data 24

3.3 BM25 25

3.4 Combination of BM25 and BERT 27

3.5 Generation of Evaluation Data Set 30

3.5.1 Query Generation with OpenAI-GPT 31

3.5.2 Expert Generated Data Set 34

4 EXPERIMENTS AND RESULTS 37

4.1 Evaluation 37

4.2 Experiments with Title Path Query Data Set 42

4.3 Experiments with Generated Query Data Set 43

4.4 Experiments with Seckin Data Set 45

4.5 Impact of Different Query Types on Paragraph Retrieval 47

4.6 Impact of Sentence-Level Inference on Paragraph Retrieval 48

4.6.1 Different Weights of Paragraphs and Sentence-Level Scores . . 49

4.6.2 Different Alpha Weights and Different top-k Sentence Selection 51

5 CONCLUSIONS 57

REFERENCES 59

APPENDICES

APPENDICES 62

A DETAILED DATA SET EXAMPLES 63

LIST OF TABLES

TABLES

Table 3.1 Title Path Data Set Example 22

Table 3.2 Title Path Data Set Example (Table 3.1 continued) 23

Table 3.3 Title Path Data Set Along with Generated Query 33

Table 3.4 Expert Generated Data Set Example 35

Table 3.5 Expert Generated Data Set Example (Table 3.4 continued) 36

Table 4.1 Ranked Paragraphs and Their Relevance For Query 40

Table 4.2 Experiment results with Title Path Query Data Set 42

Table 4.3 Experiment results with Generated Query Data Set 44

Table 4.4 Experiment results with Seckin Data Set 46

Table 4.5 Comparison of Query Types using F1-Score 47

Table 4.6 Comparison of Query Types using Hit Rate 48

Table 4.7 Micro-F1@30 Scores for Different Alpha Values - BERTurk 49

Table 4.8 Micro-F1@30 Scores for Different Alpha Values - BERTurk-Legal . 50

Table 4.9 Micro-F1@30 Scores for Different Alpha Values - BERTurk-Seckin 51

Table 4.10 Micro-F1@30 Scores for Different Alpha and top-k Values - BERTurk 54

Table 4.11 Micro-F1@30 Scores for Different Alpha and top-k Values - BERTurk-Legal 55

Table 4.12 Micro-F1@30 Scores for Different Alpha and top-k Values - BERTurk-Seckin 55

Table A.1 Title Path Data Set Example (Table 3.2 continued) 64

Table A.2 Title Path Data Set Example (Table A.1 continued) 65

Table A.3 Title Path Data Set Example (Table A.2 continued) 66

Table A.4 Generated Query Data Set 67

Table A.5 Generated Query Data Set (Table A.4 continued) 68

Table A.6 Generated Query Data Set (Table A.4 continued) 69

Table A.7 Expert Generated Data Set Example (Table 3.5 continued) 70

Table A.8 Expert Generated Data Set Example (Table A.7 continued) 71

LIST OF FIGURES

FIGURES

Figure 2.1 BERT Input Representation. Sourced by [9] 11

Figure 2.2 BERT Pre-training and Fine-tuning Procedures. Sourced by [9] . 11

Figure 3.1 Method Overview 16

Figure 3.2 An example book data 17

Figure 3.3 An example TOC data 18

Figure 3.4 An example book data in tree structure 19

Figure 3.5 BM25 Paragraph Retrieval Overview 27

Figure 3.6 BERT Re-ranking Overview 28

LIST OF ABBREVIATIONS

ABBREVIATIONS	
TOC	Table of Contents
BERT	Bidirectional Encoder Representations from Transformers
BM25	Best Matching 25
TF-IDF	Term Frequency-Inverse Document Frequency
NLP	Natural Language Processing
IR	Information Retrieval
RNN	Recurrent Neural Network
MLM	Masked Language Modeling

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

The field of law is constantly evolving and updating itself. A vast amount of information needs to be known or examined in various, highly independent, and different areas of expertise. As this information changes and updates over time, the number of related sources increases exponentially, making it significantly more challenging for legal experts to use these resources effectively. For instance, U.S. state courts alone handle around 63 million cases annually [7]. In addition, the rapid pace of legal developments means that professionals must continuously adapt to new regulations, case laws, and legal precedents. This dynamic environment requires legal practitioners to be skilled in their specific areas of law and proficient at quickly assimilating new information.

Consequently, the demand for comprehensive legal research tools and databases has never been higher, as these resources are essential for staying informed and maintaining a high level of competence in the field. The ability to efficiently navigate and utilize these expanding sources of legal information is becoming an increasingly critical skill for legal professionals. Therefore, the use of Natural Language Processing (NLP) and Information Retrieval (IR) techniques has become increasingly important for quickly searching and finding relevant information in legal texts [12].

In the existing literature, efforts to retrieve relevant information from large corpora have predominantly focused on English texts. This focus is largely due to the extensive availability of well-structured and annotated datasets in the English language, which facilitate the development and evaluation of information retrieval systems. Numerous benchmark datasets such as TREC [16], TREC Legal Track [24], CLEF [27], and MSLR-WEB30K [31] have been created to support various tasks in information retrieval, including document and passage ranking. These resources, combined with a wide range of established methodologies and evaluation metrics, have significantly advanced the field of legal text retrieval in English.

In the context of legal text retrieval, one of the primary challenges is the efficient and accurate retrieval of relevant information from long texts and passages [18]. Legal documents are typically lengthy and complex, containing nuanced language and intricate legal terminologies. The problem of long text and passage retrieval involves identifying and extracting the most relevant sections within these extensive documents to meet the user's information needs. Recent advancements in large language models, such as BERT [10], GPT [32], and their variants, have shown promise in addressing this issue. These models are capable of understanding and processing long passages, enabling more effective retrieval of relevant information [23]. However, it is important to note that many language models are predominantly focused on short texts and have limitations when it comes to handling long texts. This limitation is particularly significant in the legal domain, where documents are often extensive and detailed [38].

Research on information retrieval in the Turkish legal domain is relatively limited. While there have been some efforts to develop systems for tasks such as case law retrieval, precedent identification [25], and case outcome prediction [20], these initiatives are insufficient to address the broader needs of legal professionals. A data set containing user queries and related texts is a key point for an information retrieval system. Since Turkish legal books have only recently started to be digitized and there is no system for experts to use these books effectively, there are not enough user queries available. Collecting and processing this type of data set by experts is very time-consuming. Therefore, there is a need to automatically process the existing legal texts and generate relevant queries for these texts.

In this thesis, we study the problem of retrieving relevant passages from long Turkish legal texts. We acquired a sample set of books from Seckin Yayıncılık to create our document corpus. After preprocessing the books to automatically generate queries and relevant passages, we investigated the effectiveness of large language models in retrieving these relevant passages. This research has practical implications for improving the efficiency of legal professionals in Turkish law, where access to relevant legal texts in Turkish is a growing need. By enhancing the retrieval of passages from extensive legal documents, legal practitioners can reduce the time spent manually searching for relevant information, thereby streamlining the legal research process.

1.2 Proposed Methods and Models

In this thesis, our primary goals focus on two key areas. First, we create a system that can perform contextual analysis on long Turkish legal texts and return the most relevant paragraphs for the given queries. Our second objective in this thesis is to create a query-relevant paragraph data set using our Turkish legal book documents.

Since legal texts are highly complex in terms of context, traditional word embedding methods do not perform well. Therefore, we need to use a model that includes contextual embeddings such as Bidirectional Encoder Representations from Transformers (BERT). However, BERT has limitations when it comes to extracting contextual embeddings for long texts. Therefore, we need to follow a different approach for long Turkish legal texts. We attempt to combine both the embeddings of chunks within the size limits acceptable by BERT and the embeddings of the entire passage for each passage to generate query-paragraph relevancy scores. This way, we have the opportunity to examine the performance of contextual embedding methods compared to traditional word embedding methods. Through the achievement of these objectives, this thesis makes significant contributions to domain-specific Turkish texts.

For creating a query-relevant paragraph data set, we try two different methods. First, we divide each paragraph within the books according to the sections and subsections they belong to, and then we determine the relevant queries for these paragraphs by combining the main and subheading texts, considering these combined texts as the queries relevant to the paragraphs. We can think of these queries as keywords that could be related to the context of the paragraph. Secondly, we convert these generated queries into query sentences that users could use in search engines, utilizing a Generative Language Model. This way, we have the opportunity to observe the effects of content richness in queries on relevancy.

1.3 Contributions

Our contributions are as follows:

- Developing a prototype system to rank relevant passages for Turkish legal texts, which returns the most relevant paragraphs for given queries.
- Utilizing BERT to combine embeddings from chunks within its size limits and entire passage embeddings to generate query-paragraph relevancy scores, ensuring relevancy scores are obtained without losing context.
- Introducing different query-relevant paragraph data sets using Turkish legal book documents. Paragraphs are divided according to sections and subsections. Relevant queries for paragraphs are determined by combining main and subheading texts. Queries are transformed into query sentences using a Generative Language Model.
- Comparing and analyzing the performance of different types of queries used to return relevant paragraphs. Observing the impact of different types of queries on understanding the context.

1.4 The Outline of the Thesis

The organization of the sections of this thesis is as follows. Chapter 2 provides an extensive review of the literature and a summary of related studies in the field. Chapter 3 describes the proposed methods and models, detailing the contextualized language models employed in the study, and the data set preparation, including pre-processing steps and a detailed explanation of the generation of query-relevant paragraph data sets. Chapter 4 discusses the experiments conducted, along with their results, analyzing the effectiveness of the proposed methods and models. Finally, Chapter 5 concludes the study, summarizing the findings and suggesting potential directions for future research.

CHAPTER 2

LITERATURE SURVEY

2.1 Legal Information Retrieval

Legal documents are very different from general documents. They are usually long and written in formal, complex language. Simple tasks like finding sentence boundaries or identifying named entities are much harder in legal texts. Also, in many countries, legal documents are unstructured and lack clear headings or patterns. Even though it's easier now to access unlabelled legal texts, it's still hard to find large, annotated datasets for specific tasks in many countries and languages. Existing datasets are not well organized, and manually annotating them requires legal experts, which is both time-consuming and expensive.

Traditional legal information retrieval systems, such as Westlaw [4] and LexisNexis [2], rely heavily on boolean indexing, which necessitates extensive user training. These systems, while widely used, have limitations in terms of scalability and the complexity of legal language, which often includes domain-specific jargon and multilingual elements. Recent research has focused on overcoming these challenges by developing models that incorporate semantic word measures and advanced natural language processing techniques. One prominent approach involves the use of vector space models, which represent legal documents in a multi-dimensional space, allowing for more precise retrieval based on semantic similarity. The study by Sugathadasa et al. (2018) [34] compares three novel models using vector space representations for legal documents and demonstrates that an ensemble model significantly improves retrieval accuracy. This model leverages both raw document references and neural network-based embeddings, illustrating the potential of combining traditional and modern techniques to enhance legal information retrieval systems. The integration of such models into legal IR systems holds promise for more efficient and accurate retrieval of relevant legal documents, thereby addressing the limitations of existing systems.

The study of cross-lingual information retrieval (CLIR), studied by Alishiba Florian D'souza in 2019 [11], in the legal domain addresses the challenges of accessing legal documents in multiple languages, which is increasingly important in our globalized world. Traditional CLIR methods, primarily based on translation, often struggle with issues such as term ambiguity and the quality of translation resources, which significantly impact retrieval performance. In [11], they explore a phrasal dictionary-based query translation approach for legal documents, aiming to minimize term ambiguity by considering context during translation. The research utilizes legal data from the EUR-Lex website, encompassing European Union laws and international agreements, and employs descriptors as queries. The performance of the custom phrasal dictionary is compared with a general-purpose word-to-word dictionary using retrieval models such as Bag of Words (BoW), TF-IDF, BM25, and Word Centroid Distance. Additionally, the study examines the impact of various text preprocessing techniques and evaluates the effectiveness of word embeddings trained on domain-specific and general-purpose data. The results indicate that the phrasal dictionary approach outperforms the standard dictionary method, and the ensemble method of combining multiple retrieval models yields superior precision and recall metrics, highlighting the importance of domain-specific solutions in legal CLIR.

2.2 Neural Approaches to Document Retrieval

2.2.1 Traditional Word Embeddings

Traditional word embedding methods, such as Word2Vec (Mikolov et al., 2013) [19] and GloVe (Pennington et al., 2014) [26], have been foundational in the field of IR. These methods learn vector representations of words based on their co-occurrence in large corpora, capturing semantic similarities between words.

Word2Vec, introduced by Mikolov et al. (2013) [19], is a particularly influential word embedding technique that has played a crucial role in the development of modern natural language processing (NLP) and IR systems. Word2Vec utilizes two main approaches: Continuous Bag of Words (CBOW) and Skip-gram [19]. CBOW predicts the target word based on its surrounding context words within a specified window. This method is computationally efficient and tends to work well with frequent words. Skip-gram, on the other hand, predicts the surrounding context words given a target word. This model focuses on maximizing the probability of the context words appearing in the vicinity of the target word. Skip-gram is particularly effective for representing rare words because it prioritizes learning more about the less frequent words in the corpus.

GloVe, developed by Pennington et al. (2014) [26], is another widely adopted word embedding method. Unlike Word2Vec, which is a predictive model, GloVe is a count-based model that constructs word embeddings by factorizing a word co-occurrence matrix. The key idea behind GloVe is to leverage the global statistical information from a corpus, rather than just local context as in Word2Vec. GloVe works by first constructing a co-occurrence matrix, where each element represents how frequently a word appears in the context of another word within a specified window size. This matrix is then factorized into lower-dimensional word vectors such that the dot product of two-word vectors is proportional to the logarithm of the probability of their co-occurrence [26]. One of the strengths of GloVe is its ability to capture both local and global word semantics. The model ensures that word embeddings reflect not only the immediate context of words but also their broader relationships across the entire corpus. This makes GloVe embeddings particularly useful in IR applications, where understanding the global context can improve the accuracy of document retrieval and ranking.

Despite their advantages, both Word2Vec and GloVe have limitations. They operate on static embeddings, meaning each word has a single representation regardless of its context. This can be problematic in cases where words have multiple meanings. Furthermore, they do not capture subword information, which can be a limitation for languages with rich morphology or for dealing with out-of-vocabulary words.

2.2.2 Large Language Models in IR

Large Language Models like BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) [9] and ELMo (Embeddings from Language Models) (Peters et al., 2018) [28] have marked significant advancements in text understanding. Unlike traditional word embeddings, these models generate contextual representations of words based on their surrounding text, thereby capturing word dependencies and sentence structures. Devlin et al. (2018) [9] introduced BERT, which was pre-trained on large-scale corpora to learn general language patterns and relationships between text segments. This model has been shown to outperform traditional embeddings across various NLP tasks, including passage ranking and document retrieval.

2.2.2.1 BERT

BERT (Bidirectional Encoder Representations from Transformers), proposed by Devlin et al. [9] in 2018, marks substantial progress in the evolution of contextualized language models. BERT pre-trains deep bidirectional representations by simultaneously considering both the left and right context in all layers. This bidirectional methodology enables BERT to grasp a more comprehensive representation of language compared to models that process text in a single direction.

BERT integrates token embeddings, positional embeddings, and segment embeddings as input. The transformer architecture typically struggles with maintaining the order of inputs. To overcome this, BERT employs positional embeddings, which learn and represent the positions of words within a sentence. Furthermore, it learns separate embeddings for the first and second sentences, enhancing the model's capability to differentiate between sentences effectively. The overall input embeddings are the summation of token embeddings, positional embeddings and segment embeddings as seen in Figure 2.1.

BERT's development involves two primary steps. The first step is pre-training, where the model is exposed to extensive amounts of unlabeled text to learn general language representations. This pre-training consists of two methods: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). The second step is fine-tuning,

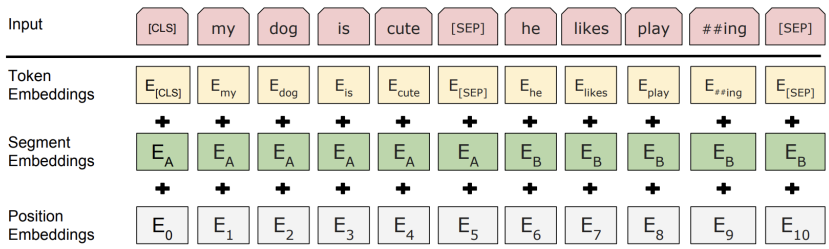

Figure 2.1: BERT Input Representation. Sourced by [9]

which entails training BERT on a smaller, task-specific annotated dataset for applications such as sentiment analysis or question answering. Pre-training and fine-tuning procedures can be seen in Figure 2.2

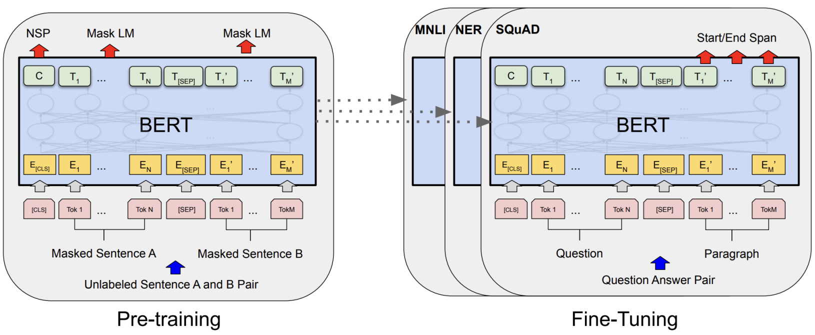

Figure 2.2: BERT Pre-training and Fine-tuning Procedures. Sourced by [9]

2.2.2.2 BERTurk

BERT-BASE Turkish 128k Uncased Model [1], also known as BERTurk, was trained on a curated and sentence-segmented version of the Turkish OSCAR corpus, along with a recent Wikipedia dump, various OPUS corpora, and a special dataset provided by Kemal Oflazer. The final training dataset is 35GB in size, containing 44,049,766,662 tokens. Utilizing Google's TensorFlow Research Cloud (TFRC), the uncased model was trained on a TPU v3-8 for 2 million steps. The model employs a vocabulary size of 128,000. This extensive dataset allowed BERTurk to capture a wide array of linguistic patterns and contextual information, making it highly effective for various natural language processing (NLP) tasks in Turkish.

2.3 BERT in Document Retrieval

BERT's architecture, designed for sentence pair classification, fits well with search tasks that involve understanding the relationships between queries and documents. The model's ability to predict the relevance between two pieces of text and its attention-based architecture, which models local interactions between words, make it particularly effective for IR. Nogueira and Cho (2019) [22] demonstrated the utility of BERT in passage re-ranking tasks, showing significant improvements over traditional IR models.

Dai and Callan (2019) [8] explored the application of BERT for ad-hoc document retrieval, examining its performance on two standard IR data sets: Robust04 and ClueWeb09-B. Their experiments revealed that fine-tuning BERT with limited search data could perform better than strong baselines. This study highlighted the effectiveness of BERT's deeper text understanding in improving search accuracy, especially for queries written in natural language.

Other significant contributions in the field include the work by Masaharu et al. (2021) [37], which employed BERT-based ensemble methods for legal information retrieval and legal textual entailment in the COLIEE competition. Their results demonstrated the effectiveness of combining BERT with other models to improve retrieval accuracy. Additionally, the study by Nguyen et al. (2021) [21] introduced a framework for legal document retrieval using deep neural networks with attention mechanisms, achieving state-of-the-art results by focusing on important text parts and employing hierarchical structures.

The CEDR (Contextualized Embeddings for Document Ranking) model proposed by MacAvaney et al. (2019) [17]. CEDR incorporates contextual embeddings from models like BERT and ELMo into existing neural ranking architectures, addressing practical challenges such as the maximum input length imposed by BERT. Their approach involves using similarity matrices from multiple layers of the language model and combining BERT's classification vector with other neural ranking models, significantly improving ranking performance on large data sets like Robust04 and WebTrack.

Another significant contribution is the work by Yang et al. (2019) [36] on BERTserini, an open-domain question-answering system that combines BERT with the Anserini IR toolkit. This system splits long documents into smaller chunks and processes each chunk independently to handle BERT's input size limitations.

Additionally, Karpukhin et al. (2020) [13] proposed Dense Passage Retrieval (DPR) for open-domain question answering, which also addresses the challenge of long texts by retrieving and ranking passages rather than entire documents. This model splits documents into smaller passages and efficiently retrieves relevant passages using dense vector representations.

Further advancements in this area are demonstrated by the DoSSIER system (2021) [5] utilized dense retrieval models and summarization-based re-ranking for case law retrieval, highlighting the importance of integrating paragraph-level indexing and retrieval to handle long legal documents effectively.

2.4 BERTurk-Legal

BERTurk-Legal is a transformer-based language model designed to retrieve prior legal cases. BERTurk-Legal was introduced by Öztürk [25] and is pre-trained on a dataset from the Turkish legal domain. This dataset does not contain any labels related to the prior court case retrieval task. The dataset is publicly available in [3]. Masked language modeling is used to train BERTurk-Legal in a self-supervised manner. With zero-shot classification, BERTurk-Legal provides state-of-the-art results on a dataset consisting of legal cases from the Court of Cassation of Turkey. The results of the experiments highlight the necessity of developing language models specific to the Turkish law domain. Öztürk's research also delves into the use of dense word embeddings for the Turkish legal domain, such as LawTurk2Vec, which are trained on large corpora of Turkish legal documents. These embeddings, along with recurrent neural network (RNN) autoencoders and their combinations with BM25 algorithms, provided a foundation for more advanced retrieval tasks. The thesis compares various models, including vanilla RNNs, Legal RNN autoencoders, and transformer-based models, demonstrating that the BERTurk-Legal model outperforms these earlier methods significantly.

One of the key innovations in BERTurk-Legal is its ability to leverage the transformer architecture, specifically BERT, to capture the contextual nuances of legal language. This is achieved through masked language modeling, where parts of the input text are masked and the model is trained to predict these masked parts, thereby learning the context of words and phrases in legal texts. The model's architecture includes multiple layers of attention mechanisms that allow it to understand the relationships between different parts of a legal document, making it highly effective for tasks such as document similarity and relevance ranking.

2.4.1 Retrieval Augmented Generation

Retrieval Augmented Generation (RAG) is a hybrid approach that combines retrieval-based models with generative models to enhance the performance of tasks such as question answering, information retrieval, and summarization [15]. Retrieval Augmented Generation (RAG) has recently been applied in legal document retrieval tasks to improve the relevance and generation of responses by combining retrieval-based and generative models [29]. While RAG could enhance the diversity and accuracy of retrieved information, its application in this study was avoided due to efficiency concerns. Since the generative models work better with clear, detailed, and explainable prompts and queries, it is hard to create such a dataset to analyze the performance.

CHAPTER 3

DATA SET PREPARATION AND PASSAGE RANKING MODELS

This thesis is structured into two main phases: data set preparation and the development of a paragraph ranking model. The overall framework of the research is illustrated in Figure 3.1. Initially, the study focused on the detailed process of data set preparation. This phase included several key steps: pre-processing long document texts, extracting title paths from titles and sub-headings, constructing paragraphs under each title path, and implementing query generation. The inputs required for the modeling phase were successfully extracted after this thorough data preparation process.

At the same time, a systematic and thorough approach was adopted to identify and select the most appropriate language models, which is a critical element of this study. This selection process involved several stages, including an extensive review of existing models and their potential applicability to the task. Once the language models were chosen, they were subjected to rigorous and detailed evaluation using the meticulously prepared data set. This evaluation process was designed to ensure a deep and informed understanding of each model's performance and suitability within the specific context of legal passage retrieval tasks. By doing so, the study aimed to identify the models that could provide the most accurate and relevant results, thereby enhancing the overall effectiveness of the research.

The architectural diagram shown in Figure 3.1 highlights the systematic approach of the study, clearly outlining its structure across data set curation and model evaluation. The initial phase involves retrieving candidate paragraphs using the BM25 algorithm, which provides a baseline retrieval set. Following this, the BERT model is employed for re-ranking, utilizing deep learning techniques to refine the initial results based on the semantic similarity between the query and the retrieved paragraphs.

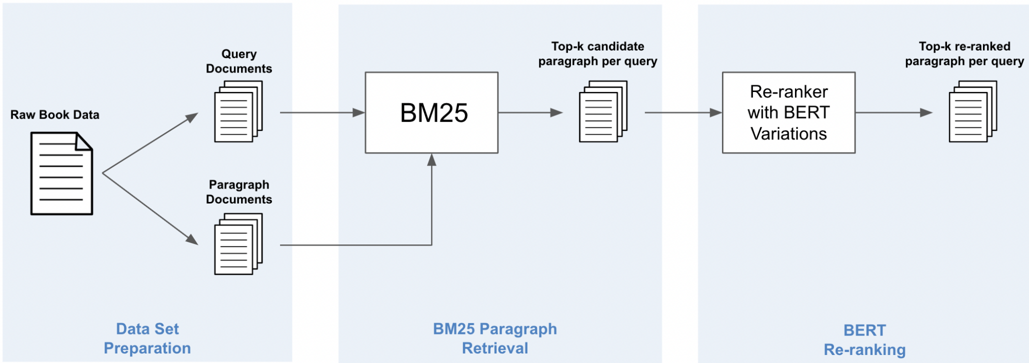

Figure 3.1: Method Overview

3.1 Data Set

Within the scope of this thesis, three different data sets are used, two of them were newly constructed for this study. One was created by combining the titles and subtitles of our legal book in a tree structure, where these combinations serve as queries and the paragraphs under them are determined as relevant paragraphs. The other dataset was created by providing the generated title paths to a generative language model, converting them into more meaningful queries that users are likely to ask. The last one is created using queries and relevant paragraphs identified by experts in the legal domain.

3.1.1 Preparation of Tree Structured Title Paths and Paragraphs

The title path data set is obtained by pre-processing the book Ticari İşletme Hukuku by Poroy and Yasaman, published in 2022, [30] and creating title paths and paragraphs under them. The data set is not publicly available.

The title path data set contains 347 different title paths and 1983 different paragraphs. Each title path can have a varying number of paragraphs under it. The title path with the fewest paragraphs has 1 paragraphs, while the title with the most paragraphs has 89 paragraphs. The raw book data is a JSON file where each element contains two values. The first value represents the index of a paragraph, and the second value contains the paragraph itself. The index information is stored as the paragraph's index - the index of the subsection it belongs to, the index of the higher-level section, ..., and the index of the main section. An example of this data can be found in Figure 3.2.


```
[
  [
    "534-533,532,531",
    "Ticaret Hukuku’nu kısaca tanımlamak istersek, ticari faaliyetleri düzenleyen hukuk koludur diyebiliriz."
  ],
  [
    "535-533,532,531",
    "Belirli bir seviyeye ulaşmış insan toplulukları içinde, üretim-tüketim çarkının dönmesini sağlayan ve esas itibariyle mübadeleye dayanan faaliyetler mevcuttur. Hem iktisat, hem de hukuk ilimlerini ilgilendiren
      bu faaliyetlere “ticari faaliyetler” denir."
  ],
  [
    "536-533,532,531",
    "Ticaret Hukuku’nu, ticari faaliyetlerden hareketle tanımlarken, Ticaret Hukuku’nun tarihsel gelişiminin göz önünde tutulması gerekir. Çok eski zamanlardan beri ticari faaliyetleri düzenleyen hukuk kurallarının
      mevcudiyetine rağmen, ticaret hukuku diye ayrı bir hukuk kolunun ortaya çıkması nispeten yeni olmuştur. Ticaret Hukuku, ortaçağ formalist hukuk sistemine karşı ticaret aleminin bir tepkisi olarak doğmuştur."
  ],
  [
    "2821-2820,2817,2489,2397",
    "Fikir ve sanat eserlerinde, eserin kendisi kadar adının veya eseri tanıtmaya yarayacak olan alametin de korunması gerekir. Eser kamuya belirli bir ad altında sunulur ve kişiler arasında bu adla anılır ve talep
      görür. Esere verilen ad, onun ticari başarısında önemli bir etken olup, eserin bir tür tamamlayıcı parçası olarak görülmektedir. Bu adın başka bir eserde kullanılması alıcıyı yanıltır ve eser sahibinin ticari
      çıkarını zedeler. Bu yüzden adın korunması, “haksız rekabet yasağı” ilkesine dayandırılmaktadır. Ad ve alametlerin haksız rekabet hükümlerine göre korunması için bazı şartların bulunması gerekir."
  ],
  [
    "2822-2820,2817,2489,2397",
    "Bir fikir ve sanat eserinin adı, alametleri ve çoğaltılmış nüshalarının şekli, başka eserlerde ve çoğaltılmış nüshalarda “iltibasa meydan verecek şekilde” kullanılamaz (FSEK m. 83). Aksi halde, TTK’nın haksız
      rekabete ilişkin hükümleri uygulanır. İltibası yaratana (mütecavize) karşı TTK m. 56’da öngörülen tespit, men, haksız rekabetin sonucu olan maddi durumun ortadan kaldırılması, beyanların düzeltilmesi, maddi ve
      manevi tazminat davaları açılabilir. Burada dikkatimizi çeken husus, “iltibas” ibaresidir 1070 . İltibas, haksız rekabette kullanılan bir ibare olup, karıştırılma ihtimalini barındırmaktadır. İltibasın
      şartlarının tespiti, haksız rekabet hükümleri çerçevesinde yapılacaktır 1071 . Ancak iltibasa yol açacak bir şekildeki kullanım haksız rekabet hükümlerinin uygulanmasına sebebiyet verebilir."
  ],
  [
    "2823-2820,2817,2489,2397",
    "FSEK m. 83/2’de belirtildiği gibi, haksız rekabete ilişkin hükümler ve kullanılamama umumen kullanılan ve ayırt edici bir vasfı bulunmayan, ad, alâmet ve dış şekiller hakkında uygulanmaz. Ad ve alametlerin
      haksız rekabet hükümlerine göre korunması, bunların ayırt edici olmalarına bağlıdır 1072 . Herkesin kullanımına açık olan ve ayırt edici nitelik kazanmayan ad ve alametler koruma kapsamına alınmaz. Örneğin
      “Saatleri Ayarlama Enstitüsü” başlığı aklımıza hiç tereddütsüz Ahmet Hamdi Tanpınar, “Uğultulu Tepeler” başlığı aklımıza Emily Bronté’nin eserini getirmektedir. Bu gibi başlıkların başka eserlerde kullanılması,
      Türk Hukuku’na göre FSEK çerçevesinde esere tecavüz olarak kabul edilmemiştir. Bu başlığın veya iltibas yaratacak şekilde benzerinin başkası tarafından kullanılması haksız rekabet çerçevesinde
      değerlendirilmiştir. Ayırt edici olmayan başlıkların ise, haksız rekabet çerçevesinde korunamayacağı açıkça düzenlenmiştir. Nitekim ayırt edici olmayan ve bağımsız bir karakter kazanmayan ad ve alametlerin
      herkesin kullanımına açık olması esastır. Örneğin “İntikam Peşinde” şeklindeki başlık birçok kişi tarafından kullanılmış olup, “umumen kullanılan” bir başlık haline gelmiştir. Nitekim haksız rekabetin temelinde
      , başkasının emeğinden haksız faydalanmanın önlenmesi bulunmaktadır. Eğer bir ad herkes tarafından yaygın olarak kullanılan bir ad ise, bu ad üzerinde bir kimseye haksız rekabet hükümlerine göre korunmaya değer
      bir hak tanınması, haksız rekabetin temel kurallarına da aykırı olacaktır."
  ],
  [
    "2824-2820,2817,2489,2397",
    "FSEK m. 83/3’e göre, bu maddenin uygulanması kanunun 1’inci, 2’nci ve 3’üncü bölümlerindeki şartların tahakkukuna bağlı değildir 1073 . FSEK’in birinci bölümü, Fikir ve Sanat Eserlerini tanımlamakta ve eser
      olmanın şartlarını ve eser türlerini düzenlemektedir. Kanunun ikinci bölümü, eser sahibini, üçüncü bölümü ise eser sahibinin haklarını düzenlemiştir. Belirtmiş olduğumuz gibi, eser adının alametinin korunması
      için bunun tek başına eser olması aranmamıştır. Ancak eserin bir parçası olması aranan şartlardan birisidir. Nitekim, Yargıtay da, öncelikle adın kullanıldığı ürünün eser olup olmadığının belirlenmesi
      gerektiğini kabul etmiştir 1074 . Bir eserin adı üzerindeki hak, o eserin hak sahibi tarafından ileri sürülebilir, ancak adın başkaları tarafından izinsiz olarak kullanılması dolayısıyla FSEK’te belirtilen
      hakları ileri süremez. Eser üzerindeki hak sahibine, ancak haksız rekabet hükümlerine göre bir hak iddia etmesi uygun görülmüştür."
  ],
  [
    "2825-2820,2817,2489,2397",
    "FSEK m. 83/4’e göre, Basın Kanunu’nun 14. maddesinin mevkute adları hakkındaki hükmü mahfuzdur. FSEK’in hazırlandığı tarihteki Basın Kanunu’nun 14. maddesi ile “mevkute sahibin hakları düzenlenmiş olup, bu hükme
      göre, beyannamenin verildiği tarihten bir sene içinde mevkute neşrolunmaz veyahut neşrolunmaya başladıktan sonra neşrine beş yıl müddetle ara verilirse beyanname hükümsüz kalır ve sağladığı haklar düşer. Ancak,
      5680 sayılı Basın Kanunu 2004 yılında yürürlükten kaldırılmış ve yerine 5187 sayılı Basın Kanunu yürürlüğe girmiştir 1075 . Bu Kanunda da süreli yayın sahibinin haklarını kaybetmesi başlıklı madde de süreli
      yayın sahibinin beyanname verdiği tarihten itibaren bir sene içinde süreli yayın yayımlanmaz veya yayımlandıktan sonra yayıma üç yıl müddetle ara verilirse beyanname hükümsüz kalır ve sağladığı hak ortadan
      kalkar. Her iki kanundaki maddelerin başlıkları ve öngörülen “yayıma ara verme süresi” farklı olarak tespit edilmiş olsa da, süreli yayının belli bir süre yayınlanmaması halinde, yayın sahibinin yayın
      üzerindeki ada bağlanan haklarını kaybedeceği belirtilmiştir. Yeni Basın Kanunu’nda bu süre 3 yıl olarak öngörülmüştür 1076 . FSEK m. 83/4’ün Basın Kanunu’na yapmış olduğu atıf, yeni Basın Kanunu m. 9 şeklinde
      anlaşılmalıdır 1077 ."
  ]
]
```


Figure 3.2: An example book data

Another provided data is a JSON file containing the index information of the headings and subheadings found in the table of contents of the book. It contains "headerLevel", "headerText", "tocLineIndex", "tocHeaderRefs", and "contentLineIndex" fields. The field "contentLineIndex" provides the index of the second element in our first dataset. The Table of Contents (TOC) data contains indexes for headings and subheadings, but not for paragraphs. An example of this data can be found in Figure 3.3.

```
[
  {
    "headerLevel": 1,
    "headerText": "GİRİŞ VE GENEL BİLGİLER",
    "tocLineIndex": "88",
    "tocHeaderRefs": [],
    "contentLineIndex": 531
  },
  {
    "headerLevel": 2,
    "headerText": "I. TİCARET HUKUKU KAVRAMI",
    "tocLineIndex": "89",
    "tocHeaderRefs": [
      88
    ],
    "contentLineIndex": 532
  },
  {
    "headerLevel": 3,
    "headerText": "A. Genel Olarak",
    "tocLineIndex": "90",
    "tocHeaderRefs": [
      88,
      89
    ],
    "contentLineIndex": 533
  },
  {
    "headerLevel": 1,
    "headerText": "Onuncu Bölüm HAKSIZ REKABET",
    "tocLineIndex": "405",
    "tocHeaderRefs": [],
    "contentLineIndex": 2397
  },
  {
    "headerLevel": 2,
    "headerText": "II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ",
    "tocLineIndex": "413",
    "tocHeaderRefs": [
      405
    ],
    "contentLineIndex": 2489
  },
  {
    "headerLevel": 3,
    "headerText": "R. FSEK'te Öngörülen Haksız Rekabet Halleri",
    "tocLineIndex": "440",
    "tocHeaderRefs": [
      405,
      413
    ],
    "contentLineIndex": 2817
  },
  {
    "headerLevel": 4,
    "headerText": "1. Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri",
    "tocLineIndex": "441",
    "tocHeaderRefs": [
      405,
      413,
      440
    ],
    "contentLineIndex": 2820
  }
]
```


Figure 3.3: An example TOC data

3.1.1.1 Title Path Extraction

Since we do not have a title path available for the book at hand, we create title paths under which each paragraph is located. The raw data contains index information about which subsections and main sections each paragraph are under. An example of the main section, sub-sections, and paragraph under the sections can be found in Figure 3.4.

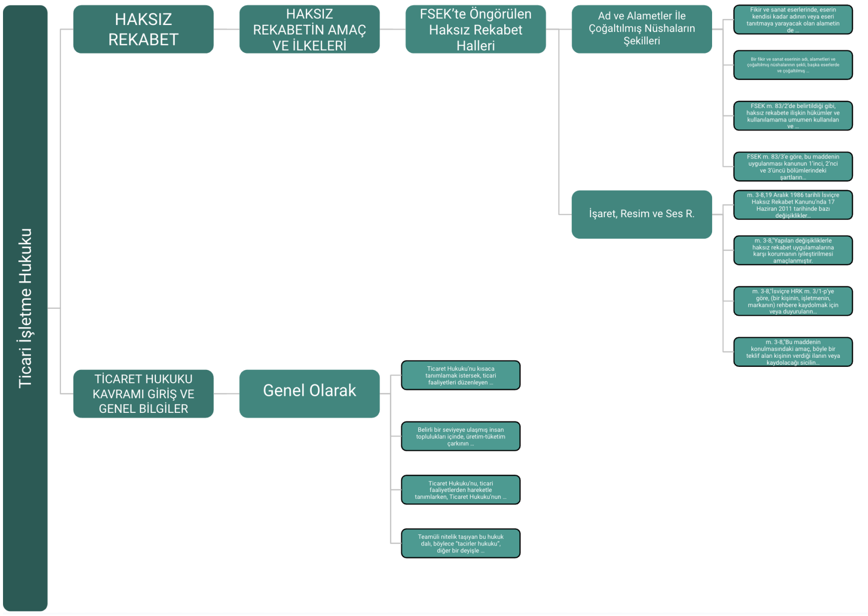

Figure 3.4: An example book data in tree structure

The paragraphs are represented by the leaves in the tree, while the other nodes represent the subsections. Not all paragraphs are at the same hierarchical level; some paragraphs may be at the 5th level while others may be at the 3rd level. Therefore, each title path needs to be created by performing a depth-first search to traverse all the nodes. The title path is created by starting with the highest level node placed first, then concatenating the text of each node with spaces in between, until reaching the lowest level node. It is important to note that the title of the book is not considered a heading text and is not included in any title path. Considering the examples from Figure 3.4:

> The paragraph "Fikir ve sanat eserlerinde, eserin kendisi kadar adının veya eseri tanıtmaya yarayacak olan alametin de..." has a subsection title above it, which

is "Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri" which becomes the first part of the title path. The next higher-level section is "FSEK'te Öngörülen Haksız Rekabet Halleri" Combining these two, the title path becomes "Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri FSEK'te Öngörülen Haksız Rekabet Halleri". Including all the higher-level headings, the final title path is "Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri FSEK'te Öngörülen Haksız Rekabet Halleri HAKSIZ REKABETİN AMAÇ VE İLKELERİ HAKSIZ REKABET".

Similarly, for the paragraph "Ticaret Hukuku'nu kısaca tanımlamak istersek, ticari faaliyetleri düzenleyen..." the subsection title "Genel Olarak" is placed at the beginning of the title path. When combined with the higher-level section "TİCARET HUKUKU KAVRAMI GİRİŞ VE GENEL BİLGİLER" the entire title path becomes "Genel Olarak TİCARET HUKUKU KAVRAMI GİRİŞ VE GENEL BİLGİLER".

As seen in the examples, the lengths of the title paths vary depending on the depth at which the paragraphs are located. Some title paths can be long due to having many levels, while others can be concise. This variation affects the quality of the title paths. While some title paths might be useful as queries for the paragraphs, others might fail due to being too general in context. A generative language model is needed to prevent this and obtain queries that seem more user-generated, thereby increasing the quality of the queries. This aspect is discussed in 3.5.1 Query Generation with OpenAI-GPT.

3.1.1.2 Paragraph Extraction

As stated above our title paths are derived recursively using the depth-first search method of the title paths. Assuming that the paragraphs are located at the highest level of the tree, we can use the same method to create title path-paragraph pairs. We know that the second element of the sub-arrays in the raw data contains text, but this text could be a paragraph or the text of a subsection or main section. To distinguish between these, we need additional information. We can use the TOC (Table of Contents) data for this purpose. If a text is considered a heading or subheading, its index should be present in the TOC data. While traversing the nodes of the tree, we can determine whether the text is a paragraph or a heading/subheading by checking if the index associated with the node is present in the TOC data. So, while traversing all the nodes, we check if the index is in the TOC data. If the node's index is in the TOC data, we consider it a title path element, add it to the title path, and then check the next level. If the index is not in the TOC data, we consider it a paragraph for the title path we have created, and we form a title path-paragraph pair. Example title path-paragraph pairs extracted from the data can be found in Table 3.1. More examples can be found in Table 3.2, A.7, and A.8.

Table 3.1: Title Path Data Set Example

book_code	index	title_path	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	534	A. Genel Olarak I. TİCARET HUKUKU KAVRAMI GİRİŞ VE GENEL BİLGİLER	Ticaret Hukuku'nu kısaca tanımlamak istersek, ticari faaliyetleri düzenleyen hukuk koludur diyebiliriz."
64c09e9d0bd3ea0ee2a71067_240125144019	535	A. Genel Olarak I. TİCARET HUKUKU KAVRAMI GİRİŞ VE GENEL BİLGİLER	Belirli bir seviyeye ulaşmış insan toplulukları içinde, üretim-tüketim çarkının dönmesini sağlayan ve esas itibariyle mübadeleye dayanan faaliyetler mevcuttur. Hem iktisat, hem de hukuk ilimlerini ilgilendiren bu faaliyetlere "ticari faaliyetler" denir."
64c09e9d0bd3ea0ee2a71067_240125144019	536	A. Genel Olarak I. TİCARET HUKUKU KAVRAMI GİRİŞ VE GENEL BİLGİLER	Ticaret Hukuku'nu, ticari faaliyetlerden hareketle tanımlarken, Ticaret Hukuku'nun tarihsel gelişiminin göz önünde tutulması gerekir. Çok eski zamanlardan beri ticari faaliyetleri düzenleyen hukuk kurallarının mevcudiyetine rağmen, ticaret hukuku diye ayrı bir hukuk kolunun ortaya çıkması nispeten yeni olmuştur. Ticaret Hukuku, ortaçağ formalist hukuk sistemine karşı ticaret aleminin bir tepkisi olarak doğmuştur."
64c09e9d0bd3ea0ee2a71067_240125144019	2821	1. Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri R. FSEK'te Öngörülen Haksız Rekabet Halleri II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET	Fikir ve sanat eserlerinde, eserin kendisi kadar adının veya eseri tanıtmaya yarayacak olan alametin de korunması gerekir. Eser kamuya belirli bir ad altında sunulur ve kişiler arasında bu adla anılır ve talep görür. Esere verilen ad, onun ticari başarısında önemli bir etken olup, eserin bir tür tamamlayıcı parçası olarak görülmektedir. Bu adın başka bir eserde kullanılması alıcıyı yanıltır ve eser sahibinin ticari çıkarını zedeler. Bu yüzden adın korunması, "haksız rekabet yasağı" ilkesine dayandırılmaktadır. Ad ve alametlerin haksız rekabet hükümlerine göre korunması için bazı şartların bulunması gerekir.

Table 3.2: Title Path Data Set Example (Table 3.1 continued)

book_code	index	title_path	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	2822	1. Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri R. FSEK'te Öngörülen Haksız Rekabet Halleri II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET	Bir fikir ve sanat eserinin adı, alametleri ve çoğaltılmış nüshalarının şekli, başka eserlerde ve çoğaltılmış nüshalarda "iltibasa meydan verecek şekilde" kullanılamaz (FSEK m. 83). Aksi halde, TTK'nın haksız rekabete ilişkin hükümleri uygulanır. İltibası yaratana (mütecavize) karşı TTK m. 56'da öngörülen tespit, men, haksız rekabetin sonucu olan maddi durumun ortadan kaldırılması, beyanların düzeltilmesi, maddi ve manevi tazminat davaları açılabilir. Burada dikkatimizi çeken husus, "iltibas" ibaresidir 1070 . İltibas, haksız rekabette kullanılan bir ibare olup, karıştırılma ihtimalini barındırmaktadır. İltibasın şartlarının tespiti, haksız rekabet hükümleri çerçevesinde yapılacaktır 1071 . Ancak iltibasa yol açacak bir şekildeki kullanım haksız rekabet hükümlerinin uygulanmasına sebebiyet verebilir.
64c09e9d0bd3ea0ee2a71067_240125144019	2829	1. Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri R. FSEK'te Öngörülen Haksız Rekabet Halleri II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET	Eserin koruma süresi bitmiş olsa dahi, eser sahibi, eser başlığının başkaları tarafından kullanılmasını, karışıklığa yol açacaksa men edebilir. Başlığa tanınan koruma müstakil olduğu için eserin başlığı, eser üzerindeki koruma süresi sona ermiş olsa dahi haksız rekabet hükümlerine göre korunmaya devam edilir 1081 .

The constructed data set has "book_code", "index", "title_path", and "paragraph" columns.

- **book_code:** Represents the unique ID of the book.
- **index:** Represents the index of the paragraph.
- **title_path:** Represents the query that is constructed from titles and sub-titles.

Explained in detail in 3.1.1.1 Title Path Extraction.

- **paragraph:** Represents the paragraph under the given title path. Explained in detail in 3.1.1.2 Paragraph Extraction.

3.1.1.3 Post-processing

As mentioned above, the TOC data is used to determine whether the texts are paragraphs or part of the title path. However, there can be subheadings that are not included in the TOC data. Therefore, there are some irrelevant results in the automatically generated title path-paragraph pairs. To explain this with an example, let's name the headings and subheadings with "H" and the following number to indicate their depth. For instance, H1 represents a heading, H2 represents a subheading under H1, and H3 represents a subheading under H2. The TOC data contains information up to a maximum of H5. However, the book data includes texts from H6 and subheadings below it. Since the index information for these texts is not present in the TOC data, we also treat them as paragraphs. Therefore, we end up creating irrelevant title path-paragraph pairs. Since this can affect the relevancy score and hinder accurate measurement, we need to eliminate these pairs. Out of the 2623 automatically generated paragraphs, 1984 are meaningful pairs, while 639 are irrelevant data due to the mentioned issue. All these irrelevant data are identified and removed manually. Thus, the final version of the Title Path Data Set is obtained.

3.2 BERTurk-Seckin: Pre-training BERT with Legal Book Data

Pre-training BERT models on domain-specific data has been shown to improve performance on downstream tasks within that domain. In this thesis, we pre-train the BERT-BASE Turkish 128k Uncased Model [1] on a dataset of Turkish legal texts to enhance its ability to understand and process legal language. Our approach involves tokenizing the text data, creating a custom dataset, and fine-tuning a pre-trained BERT model using MLM (Masked Language Modeling). The MLM method is a technique that improves the model's language understanding by allowing it to predict specific words. In this method, 15% of the sentences in the training data are randomly selected and masked, meaning they are hidden. The model learns to correctly predict these masked words, thereby understanding the contextual relationships in the language better. This is an effective method to enhance the model's performance, especially in fields with complex and specialized terminology, such as law.

We begin by loading a dataset of Turkish legal texts from the Title Path Data Set. The dataset consists of paragraphs extracted from legal documents. This data is then processed and tokenized for use with the BERT model. The dataset was divided into a train-test split of 80%-20% to ensure robust evaluation. We use the BERT tokenizer to encode the legal paragraphs. The tokenizer converts text into tokens that the BERT model can understand. Each paragraph is truncated or padded to a fixed length of 512 tokens. The model is trained for three epochs, meaning that the entire training dataset is passed through the model three times. Training for multiple epochs helps the model learn better, but the number of epochs must be balanced to avoid over-fitting, where the model performs well on the training data but poorly on unseen data. A batch size of 4 was chosen based on the available GPU memory and the need to balance training speed and memory usage. A smaller batch size may lead to more stable updates, while a larger batch size can speed up training if memory allows. The model checkpoints are saved every 10,000 steps during training. Saving checkpoints periodically ensures that we do not lose significant progress in case of interruptions, such as hardware failures or time limitations on computing resources. It also allows us to evaluate the model's performance at the intermediate stages of training. By keeping only the two most recent checkpoints, we manage disk space efficiently while still maintaining the ability to roll back to recent training states if needed. Older checkpoints are deleted once the limit is exceeded, ensuring that storage is used effectively. Thus, we obtained a BERT variant that we can use to extract embeddings of the queries and paragraphs.

3.3 BM25

BM25 (Best Matching 25) is a state-of-the-art probabilistic information retrieval model that has been widely used for ranking documents based on their relevance to a given search query. It belongs to the family of term frequency-inverse document frequency (TF-IDF) models but introduces several refinements that make it more effective in practical applications. The BM25 algorithm, introduced by Robertson and Walker in 1994 [33], is an enhancement of the earlier Okapi BM11 algorithm. BM25 considers both term frequency (TF) and document length normalization to assess a document's relevance to a given query.

We used an open-source BM25 algorithm developed by Dorian Brown [6]. We use the BM25Okapi variation of the implementation. It uses the ATIRE BM25 [35] variant as the BM25 algorithm. A document's RSV is calculated in ATIRE BM25 as follows:

$$RSV_d = \sum_{t \in q} \left(\log \left(\frac{N - df_t + 0.5}{df_t + 0.5} \right) \times \frac{(k_1 + 1) \times tft_d}{k_1 \times ((1 - b) + b \times \left(\frac{L_d}{L_{\text{avg}}} \right)) + tf_{td}} \right) \tag{3.1}$$

where RSV_d is the relevance score for document d, t represents terms in the query q, N is the total number of documents in the corpus, df_t is the document frequency of term t, tft_d is the term frequency of term t in document d, k_1 and b are free parameters, chosen with $k_1 = 1.5$ and $b = 0.75$, L_d is the length of document d, L_{avg} is the average document length in the corpus [6].

To rank the paragraphs using the queries we have, we primarily use the BM25 algorithm. The process involves several steps to ensure that we calculate and sort the relevancy scores effectively. The overview of the BM25 paragraph retrieval can be found in 3.5. First, we calculate the relevancy score for each paragraph in the corpus against each unique query using the BM25 algorithm. This results in a score that indicates how relevant each paragraph is to a given query. Once we have the relevancy scores, we sort the query-paragraph pairs in descending order, placing the pairs with the highest relevancy scores at the top. This sorting ensures that the most relevant paragraphs for each query are easily identifiable. Next, we take the top 100 paragraphs for each query based on the highest relevancy scores. This step is crucial because it filters out less relevant paragraphs, allowing us to focus on the most relevant ones. By saving these top 100 results for each query, we create a subset of query-paragraph pairs that are considered highly relevant according to BM25. Given that we have 347 unique queries, this process results in a total of 34,700 query-paragraph pairs, with 100 paragraphs selected for each query. This subset provides a robust input for both BM25

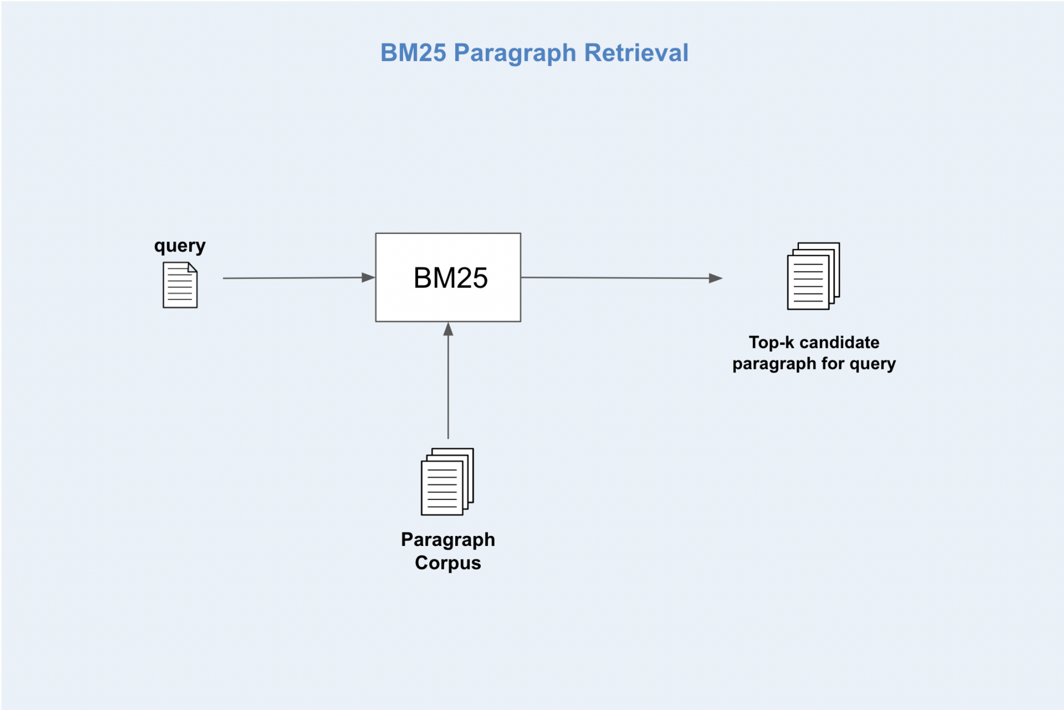

Figure 3.5: BM25 Paragraph Retrieval Overview

ranking results and further re-ranking using BERT. The BM25 scores give us a baseline ranking, and BERT can then perform re-ranking to refine these results further. By combining the strengths of BM25 and BERT, we aim to achieve more accurate and contextually relevant rankings for the given queries and paragraphs.

3.4 Combination of BM25 and BERT

Our approach involves tokenizing queries and paragraphs, performing inference using a pre-trained BERT model, and aggregating scores to rank the paragraphs effectively. We used three different pre-trained BERT models for the same implementation. To enhance the initial ranking, we use BM25 ranking results as inputs for the BERT model, considering the top 100 ranked paragraphs for each query. Before applying BERT for reranking, we utilize the BM25 algorithm to perform an initial ranking of the paragraphs. For each query, we retrieve the top 100 ranked paragraphs using BM25, which serves as the input for our BERT-based re-ranking process. We utilize the BERT tokenizer to encode both the query and the document (paragraph). The

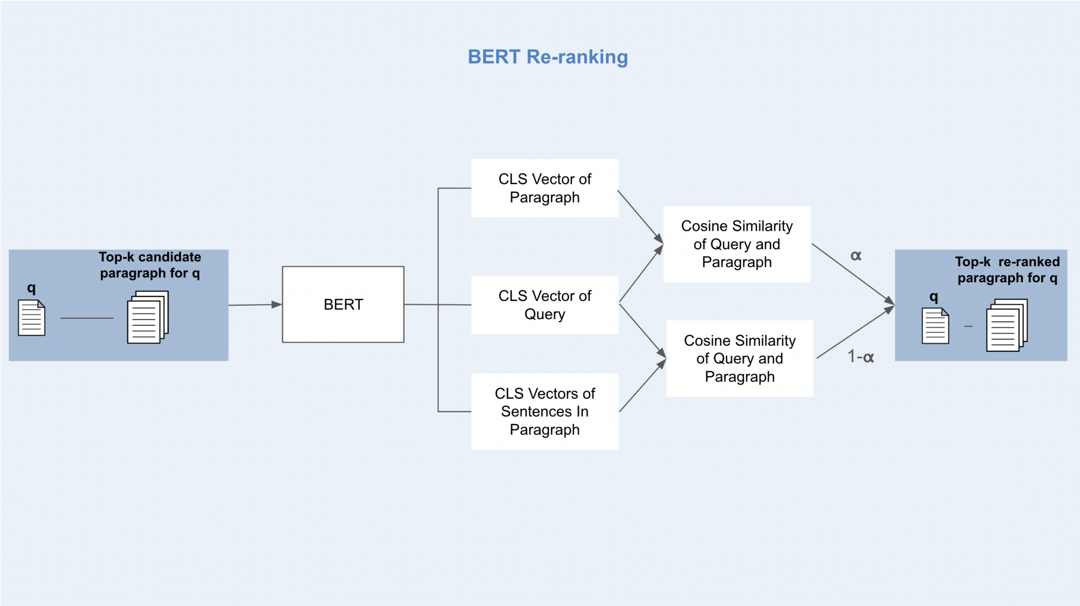

Figure 3.6: BERT Re-ranking Overview

function designed to tokenize the input concatenates the query and paragraph into a single sequence, which is then tokenized and prepared for input into the BERT model. This step ensures that the model can process the input correctly and efficiently, maintaining the context of the query and paragraph.

The initial ranking with BM25 is crucial as it provides a broad filter, narrowing down the vast number of paragraphs to a more manageable subset of the top 100 candidates for each query. This step leverages the strength of BM25 in capturing basic term frequency and inverse document frequency metrics, which are effective for coarse-grained relevance assessment. By doing so, we reduce the computational load on the subsequent BERT-based re-ranking stage.

The process of encoding queries and paragraphs using the BERT tokenizer involves converting the text into a format that the BERT model can process. This typically includes adding special tokens (such as [CLS] and [SEP]), truncating or padding the sequences to a fixed length, and converting the tokens into numerical IDs. This meticulous preparation is necessary to ensure that the BERT model can leverage its pre-trained knowledge to assess the semantic similarity between the query and the paragraph accurately.

Once the input sequences are tokenized, we use three pre-trained BERT models to perform inference. The inference function takes the tokenized inputs and feeds them into the BERT model, which outputs the last hidden state corresponding to the [CLS] token, representing the embeddings of the query and each paragraph. To rank the paragraphs, we first calculate the cosine similarity between these embeddings, providing a direct measure of their semantic similarity.

$$\text{similarity\_score} = \cos(\mathbf{q}, \mathbf{p}) = \frac{\mathbf{q} \cdot \mathbf{p}}{\|\mathbf{q}\| \|\mathbf{p}\|} \tag{3.2}$$

The variable $\mathbf{q}$ represents the embedding of the query. Similarly, the variable $\mathbf{p}$ represents the embedding of the paragraph. The notation $\cos(\mathbf{q}, \mathbf{p})$ denotes the cosine similarity between the query and paragraph embeddings. This similarity is calculated as the dot product of the two embeddings, represented by $\mathbf{q} \cdot \mathbf{p}$, divided by the product of their magnitudes. The magnitudes, or Euclidean norms, of the query and paragraph embeddings are represented by $\|\mathbf{q}\|$ and $\|\mathbf{p}\|$, respectively. This equation essentially measures the cosine of the angle between the two vectors, providing an indication of their semantic similarity.

Additionally, we perform sentence-level inference to capture finer-grained relevance information. Sentences within each paragraph are scored individually, and the top N sentence scores are considered. The final score for each paragraph is an aggregated value, combining the paragraph-level and sentence-level scores, controlled by a weighting parameter alpha.

$$\text{aggregated\_score} = \alpha \cdot \text{paragraph\_bert\_score} + (1 - \alpha) \cdot \frac{\sum_{n=1}^{N} \text{sentence\_score}_n}{\text{N}} \tag{3.3}$$

After computing the scores, paragraphs are ranked based on their aggregated relevance scores. The aggregation involves a weighted sum of the paragraph-level BERT score and the average of the top N sentence scores. This approach ensures that both the overall context and specific relevant sentences contribute to the final ranking.

The incorporation of sentence-level inference is a distinctive feature of our approach. By evaluating the relevance of individual sentences within each paragraph, we can capture more granular relevance signals that might be overlooked when only considering the paragraph as a whole. This is particularly useful in legal documents, where specific sentences can carry significant weight in determining the overall relevance. By aggregating these sentence-level scores, we ensure that our final relevance score reflects both the broader context and the critical details within the paragraph.

The weighting parameter alpha plays a crucial role in balancing the influence of paragraph-level and sentence-level scores. A higher alpha value gives more weight to the paragraph-level score, emphasizing the overall context, while a lower alpha value gives more weight to the sentence-level scores, highlighting the specific relevant details. This flexibility allows us to fine-tune the ranking process according to the specific needs of the application.

In conclusion, our approach combines the strengths of traditional information retrieval techniques and modern deep learning models to deliver highly relevant search results. By integrating BM25 and BERT, and by carefully balancing paragraph-level and sentence-level relevance signals, we achieve a robust and effective system for retrieving and ranking Turkish legal documents. This ensures that both the overall context and specific relevant sentences are considered, leading to more accurate and meaningful rankings.

3.5 Generation of Evaluation Data Set

With the preparation of all these title path-paragraph pairs, the majority of the dataset is completed. The title path list can serve as a query list and the paragraph list can serve as a text for giving the input as BM25 and BERT to get the most relevant paragraphs to a given query.

The title paths and paragraph pairs are divided into 2 separate lists for constructing the queries and paragraph corpus. Since there can be more than one title path due to the variety of paragraphs under it, the query list is constructed by unifying all title paths in the data set.

3.5.1 Query Generation with OpenAI-GPT

Using title paths as queries is a logical approach for systems built from scratch without user input. However, due to the different styles of each book's author, this method carries the risk of generating nonsensical queries. One author might prefer to create headings with comprehensive terms, while another might define them in great detail.

For contextualized language models, the context of a query about the given paragraph is very important. While term occurrence and frequency are important for traditional bag of words approaches, query understanding is more crucial for Contextualized Language Models. Dai and Callan (2019) [8] also mention that description queries yield better results than title queries in Contextualized Language Models.

In light of this information, we realize that the title path data set alone may not be sufficient. We need to enrich the content of the title path queries further. While seeking help from a legal expert could be beneficial, methods reliant on human labor are too time-consuming. Recently, generative language models like OpenAI-GPT or Meta LLama-3 have gained popularity. They possess a wide range of capabilities, from web browsing to image generation, and text document processing to text generation. With the help of well-crafted prompts, we can ask these models to summarize any text, translate a given text into another language, paraphrase the text, extract information from it, and perform many other tasks. These models are proficient in text generation. They provide a much faster and more effective solution for enriching our queries compared to relying on human labor.

In a similar vein, automated text processing systems have proven to be effective in a variety of legal contexts. According to a recent study, the use of information retrieval models to extract and summarize legal documents has been shown to significantly enhance both the speed and accuracy of legal research [14]. By integrating generative models with existing legal text retrieval systems, we can not only automate the extraction of relevant legal information but also improve the overall efficiency of the query enrichment process.

We use the OpenAI-GPT API to generate rich content queries in a short amount of time. With this API, we can easily provide the necessary inputs to the model and save the results efficiently. The inputs we need to provide to the model are as follows:

- **Role:** We can assign two different roles to the model: one as "system" and the other as "user." In the system role content, we specify how the model should generate its responses, actually instructing the model to act as a legal expert and provide responses accordingly. In the user role content, we give the prompt that outlines what the model needs to do.
- **System Role Content:** First, we need to assign a system role content to the model. Since the model knows various domains, it is crucial to generate information specific to a particular domain. As we want the model to generate information in the legal domain, we should provide the following input: "Imagine you are a user of a search engine that contains Turkish books in the field of law. I will give you a query consisting of a paragraph and keywords. This query does not contain sufficient information. Think of the paragraph as an answer to the query you will write. Considering these two texts, what query would you write to get the answer in this paragraph? Provide the answers in Turkish."
- **User Role Content:** In the user role content, we provide our query-paragraph pairs. Since there are multiple paragraphs for a single query, we need to combine all the paragraphs related to the query into a single paragraph when giving it to the model. We provide the user role content prompt as "Given query: {query text}, Given paragraph: {paragraph text}, If this paragraph was the answer from a search engine, what query would you have used to search for it instead of searching with the given query?"

By providing these prompts, we obtained responses for each query-paragraph pair and updated the dataset accordingly. The Title Path Dataset now includes a "generated_query" field as well. An example of the generated queries for title paths can be found in Table 3.3.

Table 3.3: Title Path Data Set Along with Generated Query

title_path	**generated_query**
A. Genel Olarak I. TİCARET HUKUKU KAVRAMI GİRİŞ VE GENEL BİLGİLER	Ticaret Hukuku nedir?
1. Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri R. FSEK'te Öngörülen Haksız Rekabet Halleri II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET	Fikir ve sanat eserlerinde haksız rekabet hükümleri
İ. Tüketici Kredilerinde, Taksitli Satış veya Benzeri Satış Şartlarında Tüketicinin Korunması (m. 55/a-10, 11, 12) II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET	TTK m. 55 haksız rekabet tüketici kredisi reklamlar
İş Hayatı Şartlarına Riayet Etmeme (m. 55/1-e) II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET	iş şartlarına uymamak haksız rekabet
c. Akreditif Türleri 6. Akreditif C. Uluslararası Ticari Satışlarda Ödeme Şekilleri I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR	Akreditif türleri ve özellikleri
b. İnternet Hizmet Sağlayıcıları 2. Davalılar A. Taraflar III. HAKSIZ REKABETTE HUKUK DAVALARI Onuncu Bölüm HAKSIZ REKABET	haksız rekabet internet ortamında TTK 58 ve 5651 sayılı Kanun

3.5.2 Expert Generated Data Set

Although we have found a way to generate queries for a system with no user input, we still need data verified by legal experts to measure the performance of contextualized language models accurately. Additionally, to prove the usability of our data set generation method, we need a different evaluation data set. Due to these needs, a request was made to legal experts at Seckin Yayıncılık for a small but usable evaluation data set.

The provided data set contains 2 different books, 31 different queries, and 84 paragraphs. Each query has at least one relevant paragraph and up to six relevant paragraphs. The dataset is a JSON file containing the fields "query", "answer", "book_name", "isbn", and "paragraph_index". The provided data set can be found in Table 3.4. The descriptions of the fields in the dataset are as follows:

- **query:** The search query provided by the legal experts
- **answer:** The relevant paragraph(s) corresponding to the query
- **book_name:** The name of the book from which the paragraph is taken
- **isbn:** The ISBN number of the book
- **paragraph_index:** The index of the paragraph within the book

The fields that are important for us are "query" and "answer." In this dataset, the "query" field corresponds to our "title_path" and "generated_query" fields, while the "answer" field represents the "paragraph" field.

Table 3.4: Expert Generated Data Set Example

query	answer	book_name	isbn	paragraph-index
İşçinin şikayet nedeniyle işten çıkarılmasının hukuki sonuçları nelerdir?	İşçinin, İş Kanunu'ndan doğan bir hakkını kullanması veya hakkını istemesi nedeniyle iş sözleşmesinin sona erdirilmesi dürüstlük ve objektif iyiniyet kurallarına aykırıdır. Örneğin, bir işçinin ödenmemiş ücretini istemesi ya da işçinin fazla çalışma ücretini talep etmesi nedeniyle işten çıkartılması durumun iş sözleşmesini fesih hakkının kötüye kullanması söz konusu olmaktadır. İşçinin, Sosyal Güvenlik Kurumuna kaydedilme isteği ile işverene başvurması, hafta tatili veya yıllık izin hakkını talep etmesi, iş kazasından kaynaklanan tazminat hakkını istemesi, zorunluluk arz etmeyen fazla süreli çalışmayı reddetmesi gibi nedenlerle iş sözleşmesinin feshedilmesi de objektif ve iyiniyet kurallarına aykırılık teşkil etmekte olup; bu nedenlere dayanarak işçinin iş sözleşmesinin feshedilmesi durumunda fesih hakkının kötüye kullanılması söz konusu olmaktadır.	İş Sözleşmesini Fesih Hakkının Kötüye Kullanılması	978975-0280412	476-468, 439, 395, 359
İşçinin şikayet nedeniyle işten çıkarılmasının hukuki sonuçları nelerdir?	Yargıtay, iş yerinde çalışmakta olan işçinin, ücretlerinin ödenmemesi sebebiyle Bölge Çalışma Müdürlüğüne ve sigortasının yapılmaması sebebiyle de Sosyal Sigortalar Kurumuna şikâyette bulunduğundan dolayı işçinin iş sözleşmesinin işveren tarafından feshedilmesini, işçinin İş Kanunu'ndan doğan yasal haklarını talep etmesinden kaynaklı fesih olduğundan dolayı kötüniyetli fesih olarak kabul etmiştir.	İş Sözleşmesini Fesih Hakkının Kötüye Kullanılması	978975-0280412	477-468, 439, 395, 359

Table 3.5: Expert Generated Data Set Example (Table 3.4 continued)

query	**answer**	**book_name**	**isbn**	**paragraph-index**
Tespit Dava açma Süresinin Niteliği	Dava açma sürenin niteliği konusunda öğretide görüş birliği bulunmamaktadır. Bir kısım yazarlar tarafından Kanunda düzenlenen 5 yıllık sürenin zamanaşımı süresi olduğu ileri sürülmektedir. 161 Diğer bir kısım yazarlar ise bu sürenin hak düşümü süresi olduğu savunmaktadır. 162 Yargıtay'ın görüşüne gelince, Yüksek Mahkeme 5 yıllık dava açma süresinin hak düşümü süresi olduğu yönündeki görüşü istikrar kazanmıştır.	Hizmet Tespiti Davaları	978975-0284458	641-640, 613, 587, 397
Hizmet Tespit Davasının Sosyal Güvenlik Kurumu Yönünden Sonuçları	Hizmet tespiti davasının, sigortalı lehine sonuçlanması ile birlikte, Kurum yönünden prim alacağı doğar ve primlerin Kurum tarafından tahsili gerekir. Bunun yanında, işyeri tescil edilmemiş ise, Kurum tarafından işyerinin resen tescil işlemleri yerine getirilir ve işverene idari para cezası tahakkuk ettirilir. Hizmet tespiti ile ayrıca Kurum yönünden, kısa veya uzun vadeli sigorta kollarından sağlanan yardımları sigortalılara sağlama yükümlülüğü de doğmuş olur.	Hizmet Tespiti Davaları	978975-0284458	1265-1264, 1243, 1242
Hizmet Tespit Davası Prim alacaklarında Zaman aşımı	5510 sayılı Kanunun 93. maddesinin 2. fıkrasında zamanaşımına ilişkin hükme yer verilmiş ve Kurumun prim ve diğer alacaklarının ödeme süresinin dolduğu tarihi takip eden takvim yılı başından başlayarak on yıllık zamanaşımına tabi olduğu belirtilmiştir. 5510 sayılı Kanun zamanaşımının başlangıcı konusunda Borçlar Kanunundaki düzenlemeden ayrılmaktadır.	Hizmet Tespiti Davaları	978975-0284458	472-468, 439, 395, 359

CHAPTER 4

EXPERIMENTS AND RESULTS

In this section, we present the experiments conducted to evaluate the performance of our passage retrieval system using BM25 and BERT on Turkish legal texts. The experiments were designed to test the effectiveness of our proposed hybrid model in retrieving relevant passages from a corpus of legal documents with different evaluation metrics used on ranking models. The experiments were conducted using three different datasets. In all experiments, retrieval was first performed using the BM25 algorithm, and its results were accepted as the baseline. The performance of different BERT models was then measured against this baseline. Also, since our system has sentence-level inference, the different weightage of paragraph-level similarity and sentence-level similarity were conducted and results are discussed.

4.1 Evaluation

In this study, the experiments were evaluated using metrics such as Precision, Recall, F1-Score, and Hit Rate. These metrics are widely utilized in information retrieval to measure the performance of models and systems.

- **Precision@k:** Precision at k is defined as the proportion of relevant items among the total number of items in a list of k results. In essence, it measures the accuracy of the results by indicating how many of the items in the top-k list are actually relevant. The formula is:

$$\text{precision@k} = \frac{\text{number of relevant items at k}}{\text{k}}$$

To compute Precision at k, the number of relevant items within the top-K results is divided by k. The value of k is a threshold that can be selected to limit the evaluation, which is particularly useful when the user is expected to interact with only a subset of the retrieved items.

- **Recall@k:** Recall at k quantifies the proportion of relevant items retrieved in the top k results compared to the total number of relevant items in the entire dataset. In other words, it reflects how many of the relevant items the system was able to identify. The formula is:

  $$\text{recall@k} = \frac{\text{number of relevant items at k}}{\text{total number of relevant items}}$$

  To compute Recall at k, the number of relevant items within the top-K results is divided by the total number of relevant items in the dataset. This metric helps evaluate the system's effectiveness in retrieving all relevant items from the dataset. Similar to Precision, k represents a chosen cut-off rank, limiting the evaluation to the top recommendations.

- **F1-Score@k:** Precision and Recall at k capture different facets of a ranking or recommendation system's performance. Precision evaluates how accurate the system is when making recommendations, while Recall assesses the system's ability to identify all relevant items, even if some irrelevant ones are included. Sometimes, it's beneficial to consider both Precision and Recall simultaneously. The F_β score at k combines these two metrics into a single value, offering a balanced evaluation. The Beta parameter allows you to control the emphasis on Recall versus Precision.The formula is:

  $$F_\beta = \frac{(1+\beta^2) \times \text{precision at k} \times \text{recall at k}}{(\beta^2 \times \text{precision at k}) + \text{recall at k}}$$

  A β value greater than 1 places more weight on Recall, whereas a β value less than 1 gives more importance to Precision. When β equals 1, the formula reduces to the traditional F1 score, which is the harmonic mean of Precision and Recall.

  The F1 score at k ranges from 0 to 1, where a higher score indicates better overall performance by balancing both false positives and false negatives.

- **Hit Rate@k:** Hit-rate at K measures the proportion of queries for which relevant item appears in the top K results. It indicates how often a relevant item is successfully retrieved within the top K results for each query. The formula is:

  $$\text{Hit-rate@k} = \frac{\text{number of queries with at least one relevant item in top k}}{\text{total number of queries}}$$

  To calculate Hit-rate at k, the number of queries with at least one relevant item in the top-k results is divided by the total number of queries. This metric is useful for assessing whether the system can consistently place a relevant item within the top results.

As the values of these metrics increase, the system's performance improves. To explain how these metrics are used in more detail, we can refer to the following example. In Table 4.1, you can find the top 10 paragraphs returned from BM25 retrieval for the query "1. Ad ve Alametler İle Çoğaltılmış Nüshaların Şekilleri R. FSEK'te Öngörülen Haksız Rekabet Halleri II. HAKSIZ REKABETİN AMAÇ VE İLKELERİ Onuncu Bölüm HAKSIZ REKABET", along with information on whether each of these paragraphs is relevant.

Table 4.1: Ranked Paragraphs and Their Relevance For Query

Ranked Paragraphs	**Relevant**
FSEK m. 83'te ad, alamet ve çoğaltılmış nüshaların şeklinden bahsedilmektedir. Yukarıda adlara değindik, ancak aynı hususların eserlerin alametleri ve çoğaltılmış nüshaların şekilleri için geçerli olduğunu da belirtmek isteriz...	1
FSEK m. 83 ve 84, haksız rekabet halleri olarak TTK m. 54 vd. maddelerinde sayılmış olanlara iki hususu daha eklemiştir. FSEK m. 83, bir eserin ad ve alametleriyle çoğaltılmış nüshalarının şeklinin kullanılmasını; FSEK m. 84 ise işaret, resim ve seslerin kullanılmasını yasaklamaktadır.	0
FSEK 83/5 ve 84/2. maddelerde söz konusu olan ad ve alametler ile işaret, resim ve seslerin korunması bakımından haksız rekabet hükümlerinden doğan uyuşmazlıklarda FSEK 76. maddeye göre dava fikri ve sınai haklar hukuk mahkemesinde açılır.	0
FSEK m. 83/2'de belirtildiği gibi, haksız rekabete ilişkin hükümler ve kullanılamama umumen kullanılan ve ayırt edici bir vasfı bulunmayan, ad, alâmet ve dış şekiller hakkında uygulanmaz. Ad ve alametlerin haksız rekabet hükümlerine göre korunması, bunların ayırt edici olmalarına bağlıdır...	1
Bir fikir ve sanat eserinin adı, alametleri ve çoğaltılmış nüshalarının şekli, başka eserlerde ve çoğaltılmış nüshalarda "iltibasa meydan verecek şekilde" kullanılamaz (FSEK m. 83). Aksi halde, TTK'nın haksız rekabete ilişkin hükümleri uygulanır...	1
Arkan 'a göre 861, ticari işlerde ortaya çıkan haksız rekabet halleri TTK hükümlerine, adi işlerdeki haksız rekabet ise TBK m. 57'ye tabi bulunmaktadır. Dolayısıyla tarafların ve özellikle davalı tarafın tacir olmadığı haksız rekabet halleri hakkında TBK m. 57 uygulanacaktır...	0
19 Aralık 1986 tarihli İsviçre Haksız Rekabet Kanunu'nda 17 Haziran 2011 tarihinde bazı değişiklikler yapılmış ve bazı yeni haksız rekabet halleri kanuna ilave edilmiştir. Bu değişiklikler 1 Nisan 2012 tarihinde yürürlüğe girmiştir.	0
Eski TK m. 57 on bent halinde başlıca haksız rekabet hallerini saymıştır. Bu sayma sınırlayıcı değildir 889...	0
ii. Yine TBK'da hizmetlinin rekabet yasağı ile ilgili 444 ve 447. maddelerinden doğan davalar, rekabet yasağı kavramının ticaret mahkemesince daha isabetli olarak değerlendirileceği düşünülerek ticari sayılmıştır.	0
Fikir ve sanat eserlerinde, eserin kendisi kadar adının veya eseri tanıtmaya yarayacak olan alametin de korunması gerekir. Eser kamuya belirli bir ad altında sunulur ve kişiler arasında bu adla anılır ve talep görür...	1

To calculate the Precision, Recall, and F1-score for the given table, we follow these steps: From the table, we have the relevance scores for each paragraph (1 means relevant, 0 means not relevant). The total number of paragraphs is 10.

Relevant Items: There are 4 relevant paragraphs (with a relevance score of 1). To calculate Precision@k, let's assume we are considering all 10 paragraphs (i.e., $k = 10$):

Number of relevant items in top 10: 4
Total number of items in top 10: 10

$$\text{Precision@10} = \frac{4}{10} = 0.4$$

Recall@k requires us to consider the total number of relevant items in the dataset, which is 4.

Number of relevant items in top 10: 4
Total number of relevant items in the dataset: 9

$$\text{Recall@10} = \frac{4}{9} \approx 0.44$$

We calculate the F1-score using the Precision and Recall values:

$$F1\text{-score@10} = \frac{2 \times 0.4 \times 0.44}{0.4 + 0.44} = \frac{0.352}{0.84} \approx 0.42$$

Finally we calculate Hit-rate. In this case, at least one relevant item is found within the top 10 paragraphs, so:

$$\text{Hit-rate@10} = 1.0$$

These results were calculated based on the rankings for only one query. However, we obtained results for multiple queries for evaluation. The results presented in the following sections were calculated by averaging the results computed for each query.

4.2 Experiments with Title Path Query Data Set

The Title Path Query Data Set was processed to remove irrelevant query-paragraph pairs resulting from discrepancies between the book's table of contents (TOC) and its text. The final dataset contained 1984 meaningful pairs from the original 2623 automatically generated pairs. The experiment begins by inputting our queries into the BM25 algorithm, which selects 100 candidate paragraphs from a large document corpus for each query. These candidate paragraphs and the query are then separately fed into BERT models, where the similarity between the embeddings of the query and paragraphs is measured. Based on the similarity scores, the 100 candidate paragraphs are re-ranked. Evaluation metrics are then calculated using the top 30 results returned. Table 4.2 contains the results obtained with the BM25, BERTurk, BERTurk-Legal, and BERTurk-Seckin models. All the results in this table were achieved by setting the sentence-level inference parameter, alpha, to 0.5, and by selecting 3 as the value for the top_k parameter, which is used to rank the sentences within a paragraph based on their similarity and include the top-k results in the calculation. The selection of those hyperparameters is explained in detail in Section 4.6.

Table 4.2: Experiment results with Title Path Query Data Set

Method	Precision@30	Recall@30	F1-Score@30	Hit Rate@30
BM25	**0.055**	**0.412**	**0.086**	**0.726**
BM25 + BERTurk	0.027	0.193	0.042	0.492
BM25 + BERTurk-Legal	0.041	0.292	0.064	0.585
BM25 + BERTurk-Seckin	0.042	0.314	0.066	0.636

BM25 serves as the baseline method in this experiment. It demonstrates moderate performance in terms of Precision@30 and F1-Score@30. However, it shows strong performance in Recall@30 and Hit Rate@30, indicating its effectiveness in retrieving a significant portion of relevant documents. The integration of BERTurk with BM25 results in a noticeable decline in performance across all metrics. Precision@30 and F1-Score@30 are particularly low, suggesting that this combination does not enhance retrieval effectiveness compared to the baseline. The Recall@30 and Hit Rate@30 also decrease significantly, indicating that this method struggles to retrieve a sufficient number of relevant documents. Adding the legal domain-specific BERTurk model to BM25 improves performance compared to BM25 + BERTurk. Precision@30, Recall@30, F1-Score@30, and Hit Rate@30 all show better results, though still not surpassing the baseline BM25 in Recall@30 and Hit Rate@30. This suggests that domain-specific tuning of BERTurk contributes to improved retrieval accuracy and relevance. This method, which combines BM25 with a more specialized version of BERTurk (Seckin), shows further improvement over BM25 + BERTurk-Legal. Precision@30 and Recall@30 increase, as do F1-Score@30 and Hit Rate@30. While it does not outperform the baseline BM25 in all metrics, it indicates that specialization of the language model plays a crucial role in enhancing retrieval performance.

Overall, these results highlight that while BM25 remains robust regarding Recall and Hit Rate, integrating it with specialized BERT models can lead to improvements in certain metrics. However, the general BERTurk model decreases overall performance, underscoring the importance of domain-specific tuning for legal document retrieval tasks. Moreover, BM25 often performs well with short and straightforward queries, typical in many information retrieval tasks. BERT-based models, on the other hand, are designed to understand complex language and context, which might not be as beneficial if the queries are simple and direct. Since title paths are likely to be concise and highly specific, matching closely with the document terms. BM25 is particularly effective in such scenarios where the query terms are expected to appear frequently in relevant documents. BERT-based models, which excel in understanding complex and context-rich queries, may not gain the same advantage with short and specific title path queries.

4.3 Experiments with Generated Query Data Set

Using OpenAI-GPT, we enriched the Title Path queries by generating more user-like queries. This approach aimed to improve the quality of queries and, consequently, the retrieval results. We repeat the same method used to obtain the results for title path queries by applying it to the generated queries. Again, we test the BM25, BERTurk, BERTurk-Legal, and BERTurk-Seckin models individually for all queries in the evaluation set. We input these queries into the BM25 algorithm, select 100 candidates from the same document corpus, and then rerank these candidate paragraphs using the BERT models. Evaluation metrics are calculated for the top 30 results. The results obtained using the generated queries with the BM25, BERTurk, BERTurk-Legal, and BERTurk-Seckin models can be found in Table 4.3. All the results in this table were achieved by setting the sentence-level inference parameter, alpha, to 0.5, and by selecting 3 as the value for the top_k parameter, which is used to rank the sentences within a paragraph based on their similarity and include the top-k results in the calculation.

Table 4.3: Experiment results with Generated Query Data Set

Method	Precision@30	Recall@30	F1-Score@30	Hit Rate@30
BM25	0.084	0.630	0.128	0.933
BM25 + BERTurk	0.035	0.248	0.052	0.559
BM25 + BERTurk-Legal	0.066	0.503	0.100	0.815
BM25 + BERTurk-Seckin	**0.085**	**0.659**	**0.131**	**0.956**

BM25 demonstrates strong overall performance, particularly excelling in Recall@30 and Hit Rate@30. This indicates that BM25 is effective in retrieving a large portion of relevant documents when handling user-like queries, maintaining a solid balance between precision and recall. The combination of BM25 with BERTurk shows a marked decline in performance across all metrics. This suggests that BERTurk, without domain-specific tuning, struggles to enhance the retrieval process for user-like queries and may even detract from the effectiveness of the BM25 model. BERTurk, while more powerful in understanding context and relationships between words, relies on deeper semantic understanding. This can be advantageous in more complex, ambiguous queries. However, for simpler, generated queries that focus on specific terms, BERTurk may overlook some precise term matches due to its emphasis on context. In these cases, the simplicity of BM25's term-matching approach becomes more effective.

Integrating a legal domain-specific version of BERTurk with BM25 results in noticeable improvement compared to the general BERTurk model. This setup shows enhanced precision and recall, indicating that domain-specific adjustments to BERTurk significantly boost its ability to retrieve relevant documents in legal contexts, though it still lags behind the baseline BM25 in overall effectiveness. The BM25 + BERTurk-Seckin method delivers the highest performance among the tested approaches. This combination achieves the best results across all metrics, surpassing even the baseline BM25. The high precision, recall, and hit rate indicate that the specialized BERTurk-Seckin model, tailored for legal document retrieval, effectively enhances BM25's capabilities in handling user-like queries.

While BM25 remains a robust baseline, the experiments highlight the significant potential of combining it with domain-specific BERT models, particularly BERTurk-Seckin, for improved retrieval performance. This demonstrates the importance of model specialization and fine-tuning in achieving superior results in legal document retrieval tasks with user-like queries.

4.4 Experiments with Seckin Data Set

The Seckin Data Set provided a small but expert-verified dataset, which was crucial for evaluating the performance of BERT models. This dataset included detailed queries and relevant paragraphs from legal experts, providing a robust evaluation benchmark.

All queries and paragraphs in this dataset are individually included in a list, forming the query and document corpus provided to the models. We then test the BM25, BERTurk, BERTurk-Legal, and BERTurk-Seckin models one by one for all queries in the evaluation set. These queries are input into the BM25 algorithm, and this time, we select 5 candidates from the document corpus. These candidate paragraphs are reranked using the BERT models, and evaluation metrics are calculated for the top 3 results. The reason for setting the candidate cut-off at 5 instead of 100 is due to the relatively smaller number of paragraphs in this dataset. The results obtained using expert-generated queries with the BM25, BERTurk, BERTurk-Legal, and BERTurk-Seckin models can be found in Table 4.4. All the results in this table were achieved by setting the sentence-level inference parameter, alpha, to 0.5, and by selecting 2 as the value for the top_k parameter, which is used to rank the sentences within a paragraph based on their similarity and include the top-k results in the calculation.

Table 4.4: Experiment results with Seckin Data Set

Method	Precision@3	Recall@3	F1-Score@3	Hit Rate@3
BM25	**0.162**	0.233	0.175	0.365
BM25 + BERTurk	0.130	0.176	0.136	0.341
BM25 + BERTurk-Legal	**0.162**	**0.235**	**0.177**	**0.414**
BM25 + BERTurk-Seckin	0.130	0.178	0.137	0.317

BM25 shows a strong balance between precision and recall, with the highest recall among the models (0.2337). The Micro-F1@3 score of 0.1759 reflects its overall effectiveness in balancing these metrics. However, the Mean Hit-Rate@3 of 0.3659 suggests that while BM25 retrieves relevant documents, it is not as effective in consistently ranking them at the very top. BERTurk, trained on general Turkish text, lacks domain-specific tuning for legal documents, which limits its ability to refine and re-rank results effectively. BERTurk underperforms compared to BM25 and BERTurk-Legal, with lower scores across all metrics. The Micro-Precision@3 and Micro-F1@3 scores indicate that BERTurk struggles with both identifying relevant documents and ensuring they appear in the top 3 results. The Mean Hit-Rate@3 of 0.3415 further reflects its limitations in consistently ranking relevant documents at the top. BERTurk, trained on general Turkish text, lacks domain-specific tuning for legal documents, which limits its ability to refine and re-rank results effectivel BERTurk-Legal outperforms all other models, achieving the highest Micro-Recall@3 (0.2358), Micro-F1@3 (0.1775), and Mean Hit-Rate@3 (0.4146). This suggests that BERTurk-Legal is particularly well-tuned for this retrieval task, effectively balancing precision and recall while ensuring relevant documents are consistently ranked in the top 3. BERTurk-Seckin performs similarly to BERTurk, with slightly higher recall but comparable precision and F1 scores. However, its Mean Hit-Rate@3 is the lowest among the models (0.3171), indicating that while it retrieves relevant documents, it struggles more than the other models in ranking them highly. BERTurk-Seckin, although designed as a specialized model, does not perform as well as BERTurk-Legal. Its lower Mean Hit-Rate@3 suggests that it struggles to prioritize the most relevant documents in the top positions, indicating a potential area for further fine-tuning or adjustment. The lower performance of BERTurk-Seckin, despite its specialization, could be due

Table 4.5: Comparison of Query Types using F1-Score

Method	**F1-Score@30 (Title Path Query)**	**F1-Score@30 (Generated Query)**
BM25	0.086	**0.128**
BM25 + BERTurk	0.042	**0.052**
BM25 + BERTurk-Legal	0.064	**0.100**
BM25 + BERTurk-Seckin	0.066	**0.131**

to the differences in the pre-training corpus or fine-tuning strategies. While BERTurk-Seckin may be effective for specific types of legal documents or queries, it appears to be less versatile than BERTurk-Legal in handling the broader range of queries in the Seckin Data Set. This highlights the importance of further fine-tuning or adjusting the model to better capture the specific nuances of the legal texts used in this dataset.

4.5 Impact of Different Query Types on Paragraph Retrieval

For the same document corpus, as explained in Chapter 3, there are two different types of queries. The title path query type contains keywords formed from more general concepts, while the generated query type consists of sentences that seem to be written by a user. Given that the context of the texts provided to BERT models, as well as the relationships between words and sentences, is crucial, different results were obtained from the same document corpus for these different query types. The results obtained using the BM25, BERTurk, BERTurk-Legal, and BERTurk-Seckin models for these two different query types can be found in Table 4.5 and Table 4.6. F1 score and hit rate results were used to compare the different query types.

BM25 shows a notable improvement in F1-Score and Hit Rate when handling generated queries compared to title path queries. This indicates that BM25 is effective at balancing precision and recall in more natural, user-like query scenarios. The combination of BM25 with BERTurk exhibits low F1-Scores and Hit Rate for both query types, though there is a slight increase for generated queries. This indicates that while the general BERTurk model does not significantly enhance retrieval effectiveness,

Table 4.6: Comparison of Query Types using Hit Rate

Method	**Hit Rate@30 (Title Path Query)**	**Hit Rate@30 (Generated Query)**
BM25	0.726	**0.933**
BM25 + BERTurk	0.492	**0.559**
BM25 + BERTurk-Legal	0.585	**0.815**
BM25 + BERTurk-Seckin	0.636	**0.956**

it performs marginally better with more natural queries. Adding the legal domain-specific BERTurk model improves F1-Score and Hit Rate for both query types, especially for generated queries. This highlights the benefit of domain-specific adjustments to BERTurk for enhancing retrieval performance. The BM25 + BERTurk-Seckin combination delivers the highest F1-Scores and Hit Rate, especially with generated queries, surpassing even the baseline BM25. This indicates that the specialized BERTurk-Seckin model effectively improves retrieval performance for more complex, natural queries.

The results highlight a consistent trend where all methods perform better with user-like queries compared to title path queries. This is particularly evident in the BM25 + BERTurk-Legal and BM25 + BERTurk-Seckin combinations, which show significant improvements in both F1-Score and Hit Rate for user-like queries. BM25 remains a strong baseline, but the specialized models, especially BERTurk-Seckin, demonstrate the highest performance enhancements, underscoring the importance of domain-specific tuning and specialization for achieving superior results in legal document retrieval tasks with user-like queries.

4.6 Impact of Sentence-Level Inference on Paragraph Retrieval

This experiment is conducted using Generated Query Data Set. Since there is a sentence level inference in designed system, the performance of different parameter used for top_k and alpha values. First, top_k parameter is set as 3 and alpha value is set from 0.0 to 1.0. Experiment results can be found in sub-section 4.6.1. Second, both top_k and alpha values are changed. While the top_k values can be 1,2 and 3, alpha values are set from 0.0 to 1.0. Results of this experiment can be found in sub-section 4.6.2.

4.6.1 Different Weights of Paragraphs and Sentence-Level Scores

Table 4.7 shows that the Micro-F1@30 scores vary slightly with different alpha values on BERTurk model. The highest score is observed at alpha = 0.8 (0.0547), indicating that a higher emphasis on sentence-level scores can slightly improve the model's performance. However, the overall changes in performance are minimal, suggesting that BERTurk's effectiveness is relatively stable across different weightings of paragraph and sentence-level scores.

Table 4.7: Micro-F1@30 Scores for Different Alpha Values - BERTurk

Alpha	Micro-F1@30
0.0	0.0533
0.1	0.0531
0.2	0.0532
0.3	0.0540
0.4	0.0533
0.5	0.0522
0.6	0.0530
0.7	0.0539
0.8	**0.0547**
0.9	0.0543
1.0	0.0543

Table 4.8 demonstrates a more pronounced improvement in Micro-F1@30 scores as alpha increases from 0.0 to 0.4, peaking at alpha = 0.4 with a score of 0.1011 on BERTurk-Legal model. This indicates that the BERTurk-Legal model benefits from a balanced approach, where both paragraph-level and sentence-level scores are considered. Beyond alpha = 0.4, the scores tend to decline slightly, suggesting that overly emphasizing either paragraph or sentence-level scores alone is less effective.

Table 4.8: Micro-F1@30 Scores for Different Alpha Values - BERTurk-Legal

Alpha	**Micro-F1@30**
0.0	0.0974
0.1	0.0974
0.2	0.0982
0.3	0.1000
0.4	**0.1011**
0.5	0.1008
0.6	0.0994
0.7	0.0989
0.8	0.0989
0.9	0.0983
1.0	0.0975

Table 4.9 reveals that the highest Micro-F1@30 score occurs at alpha = 0.1 (0.1345), indicating that a slight emphasis on sentence-level scores enhances the BERTurk-Seckin model's performance. As alpha increases beyond 0.1, the scores gradually decline, suggesting that overemphasizing paragraph or sentence-level scores can detract from the model's effectiveness. This trend underscores the importance of a balanced weighting approach to capture both paragraph context and critical sentence-level information.

Table 4.9: Micro-F1@30 Scores for Different Alpha Values - BERTurk-Seckin

Alpha	**Micro-F1@30**
0.0	0.1338
0.1	**0.1345**
0.2	0.1339
0.3	0.1337
0.4	0.1327
0.5	0.1310
0.6	0.1292
0.7	0.1282
0.8	0.1269
0.9	0.1258
1.0	0.1242

These results suggest that while all models benefit from incorporating sentence-level inference, the optimal weighting parameter (alpha) varies depending on the model, reflecting their different capacities to leverage paragraph and sentence-level information for legal document retrieval.

4.6.2 Different Alpha Weights and Different top-k Sentence Selection

This section provides Micro-F1@30 scores for different models (BERTurk, BERTurk-Legal, and BERTurk-Seckin) across varying alpha values and top-k sentence selections. This analysis offers insights into how different weightings between paragraph-level and sentence-level scores (controlled by alpha) and the selection of top sentences (Top-1, Top-2, and Top-3) impact the retrieval performance.

Alpha (α) is the parameter that balances the influence of paragraph-level scores and sentence-level scores in determining the overall relevance of a document. By adjusting alpha, we control the weighting between the two types of inferences:

- **When α is closer to 1**, the model gives more weight to the paragraph-level score, which evaluates the entire paragraph's relevance to the query.
- **When α is closer to 0**, the model gives more weight to sentence-level scores, which focus on the relevance of specific sentences within a paragraph.

Effect of different alpha values:

- If α were set too high (closer to 1), the model would rely heavily on the paragraph-level scores, possibly overlooking specific sentences that may be critical to answering the query. This could reduce the precision of the retrieval since important sentence-level details might be missed.
- If α were set too low (closer to 0), the model would prioritize sentence-level details, which could lead to overly focusing on specific sentences while losing the broader context provided by the entire paragraph. This could negatively impact recall, as the model might fail to capture the overall relevance of the document.

The *top-k* parameter determines how many of the most relevant sentences are considered when calculating the sentence-level score. This allows the model to take into account multiple sentences from a paragraph when determining the final ranking:

- **Top-1** considers only the most relevant sentence from each paragraph.
- **Top-2** and **Top-3** allow the model to include additional relevant sentences, thus incorporating more contextual information from the paragraph.

The choice of *k = 3* was based on initial experiments showing that evaluating three sentences provided a good balance between capturing enough relevant content from a paragraph while avoiding the inclusion of too much irrelevant information. This setting allowed for a more holistic evaluation of the paragraph's relevance without overfitting to individual sentences.

Effect of different top-k values:

- If *k* were set too low (e.g., k = 1), the model would only evaluate the single most relevant sentence, potentially missing out on other relevant sentences that provide additional context. This could lead to lower recall, as important supporting sentences might be ignored.
- If *k* were set too high (e.g., k > 3), the model might include less relevant sentences in the calculation, which could dilute the overall sentence-level score and reduce precision. Too many sentences could introduce noise, especially if they are not directly related to the query.

If different values were chosen for alpha and top-k, the model's ability to balance between capturing detailed sentence-level relevance and broader paragraph-level context would be affected. Lower alpha values would make the model more sensitive to fine-grained sentence-level information, which could be beneficial for very specific queries but detrimental for more general ones. Higher top-k values might increase recall by considering more sentence-level details, but they could also reduce precision by introducing irrelevant information.

Table 4.10 shows BERTurk's best performance when alpha is set around 0.2 to 0.8, indicating a need for balanced sentence and paragraph-level weighting. The Top-1 selection outperforms Top-2 and Top-3, suggesting that focusing on the most relevant sentence can be more effective for this model.

Table 4.10: Micro-F1@30 Scores for Different Alpha and top-k Values - BERTurk

Alpha	Top-1 Micro-F1@30	Top-2 Micro-F1@30	Top-3 Micro-F1@30
0.0	0.0535	0.0523	0.0533
0.1	0.0525	0.0526	0.0531
0.2	**0.0558**	0.0542	0.0532
0.3	0.0546	0.0523	0.0540
0.4	0.0547	0.0525	0.0533
0.5	0.0545	0.0524	0.0522
0.6	0.0545	0.0528	0.0530
0.7	0.0548	0.0539	0.0539
0.8	0.0556	0.0547	0.0547
0.9	0.0540	0.0547	0.0543
1.0	0.0543	0.0543	0.0543

Table 4.11 shows that BERTurk-Legal performs optimally at mid-range alpha values (0.3 to 0.5), with slight differences in Top-1, Top-2, and Top-3 results. This suggests that BERTurk-Legal benefits from a more moderate balance between paragraph and sentence-level relevance, with Top-1 and Top-3 being particularly effective in specific alpha ranges.

Table 4.12 shows that BERTurk-Legal performs optimally at mid-range alpha values (0.3 to 0.5), with slight differences in Top-1, Top-2, and Top-3 results. This suggests that BERTurk-Legal benefits from a more moderate balance between paragraph and sentence-level relevance, with Top-1 and Top-3 being particularly effective in specific alpha ranges.

Table 4.11: Micro-F1@30 Scores for Different Alpha and top-k Values - BERTurk-Legal

Alpha	Top-1 Micro-F1@30	Top-2 Micro-F1@30	Top-3 Micro-F1@30
0.0	0.0998	0.0978	0.0974
0.1	0.1010	0.0987	0.0974
0.2	0.1008	0.0993	0.0982
0.3	0.1017	0.1001	0.1000
0.4	0.1026	0.0998	0.1011
0.5	**0.1033**	0.1001	0.1008
0.6	0.1019	0.0990	0.0994
0.7	0.1014	0.0986	0.0989
0.8	0.1004	0.0995	0.0989
0.9	0.0995	0.0989	0.0983
1.0	0.0975	0.0975	0.0975

Table 4.12: Micro-F1@30 Scores for Different Alpha and top-k Values - BERTurk-Seckin

Alpha	Top-1 Micro-F1@30	Top-2 Micro-F1@30	Top-3 Micro-F1@30
0.0	0.1372	0.1355	0.1338
0.1	0.1376	0.1361	0.1345
0.2	0.1380	0.1358	0.1339
0.3	**0.1382**	0.1356	0.1337
0.4	0.1365	0.1337	0.1327
0.5	0.1348	0.1320	0.1310
0.6	0.1325	0.1300	0.1292
0.7	0.1295	0.1287	0.1282
0.8	0.1271	0.1269	0.1269
0.9	0.1264	0.1257	0.1258
1.0	0.1242	0.1242	0.1242

CHAPTER 5

CONCLUSIONS

In this thesis, we developed a system for retrieving relevant passages from Turkish legal texts using the BERT model. Our approach combined BM25 with BERT to enhance retrieval accuracy by leveraging the strengths of both traditional and contextualized models. The results demonstrated that this hybrid method significantly improves the precision and relevance of retrieved passages, making it a valuable tool for legal research.

The integration of BM25 and BERT models for passage retrieval demonstrated a notable improvement in the accuracy and relevance of retrieved passages. This hybrid approach successfully captures both the lexical and contextual nuances of legal language, making it a valuable tool for legal research in Turkey.

By fine-tuning BERTurk on Turkish legal texts, we developed a specialized model called BERTurk-Seckin, which outperforms general-purpose language models in the context of Turkish legal document retrieval.

We also explored the effectiveness of query expansion using generative language models like OpenAI-GPT. This method helped generate more natural and context-rich queries, further improving retrieval performance in legal text searches.

The findings of this research highlight the transformative potential of deep learning models, particularly BERT, in the domain of legal information retrieval for the Turkish language. Our work addresses the significant challenge of navigating extensive and complex legal documents, offering an enhanced tool for legal professionals seeking relevant information.

Despite the promising results, several limitations were identified in this study. Firstly, the dataset, while robust, may not fully encompass the diverse range of legal documents encountered in practice. Additionally, the models used, although fine-tuned for legal texts, are still based on architectures originally designed for general text processing, which may introduce certain biases and limitations in handling highly specialized legal content.

For future work, we plan to explore the incorporation of learning-to-rank (LTR) techniques to further enhance the retrieval performance. LTR methods can optimize the ranking of passages by learning from the preferences indicated in training data, potentially leading to even more accurate and user-centered retrieval results. Additionally, a more extensive and varied dataset, including different types of legal documents such as case laws, statutes, and legal commentaries, could further enhance the model's generalizability and performance. We are considering using RAG if better and different types of queries can be obtained, thus improving retrieval results. Also, future studies could focus on fine-tuning BERTurk-Seckin for specific legal tasks, such as legal question answering, case law analysis, or legal text summarization, which would further refine the model's utility in practice.

REFERENCES

[1] Bert base Turkish 128k uncased. https://huggingface.co/dbmdz/bert-base-turkish-128k-uncased. Accessed: 2024-07-28.

[2] Lexisnexis. https://www.lexisnexis.com/en-us/gateway.page. Accessed: 2024-07-30.

[3] Turkish prior court case retrieval (tpcr) dataset. https://github.com/koc-lab/yargitay_retrieval_dataset. Accessed: 2024-07-28.

[4] Westlaw. https://www.westlaw.com. Accessed: 2024-07-30.

[5] Sophia Althammer, Arian Askari, Suzan Verberne, and Allan Hanbury. Dossier@coliee 2021: Leveraging dense retrieval and summarization-based re-ranking for case law retrieval, 2021.

[6] Dorian Brown. Rank-BM25: A Collection of BM25 Algorithms in Python, 2020.

[7] Jack Browning. Top court filing statistics from around the country, 2023. Accessed: 2024-09-14.

[8] Zhuyun Dai and Jamie Callan. Deeper text understanding for ir with contextual neural language modeling. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '19. ACM, July 2019.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[11] Alishiba Florian D'souza. *Dictionary-Based Cross-Lingual Information Retrieval for Legal Text*. PhD thesis, University of Magdeburg, 2019.

[12] Debasis Ganguly, Jack G Conrad, Kripabandhu Ghosh, Saptarshi Ghosh, Pawan Goyal, Paheli Bhattacharya, Shubham Kumar Nigam, and Shounak Paul. Legal ir and nlp: the history, challenges, and state-of-the-art. In *European Conference on Information Retrieval*, pages 331–340. Springer, 2023.

[13] Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen tau Yih. Dense passage retrieval for open-domain question answering, 2020.

[14] Dilek Küçük and Fazlı Can. Hukuki metinlerin otomatik işlenmesinde yapay zeka teknolojilerinin kullanımı. *Bilişim Hukuku Dergisi*, 6(1):1–23, 2024.

[15] Patrick S. H. Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. *CoRR*, abs/2005.11401, 2020.

[16] Xin Li and Dan Roth. Learning question classifiers. In *COLING 2002: The 19th International Conference on Computational Linguistics*, 2002.

[17] Sean MacAvaney, Andrew Yates, Arman Cohan, and Nazli Goharian. Cedr: Contextualized embeddings for document ranking. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '19. ACM, July 2019.

[18] Dimitris Mamakas, Petros Tsotsi, Ion Androutsopoulos, and Ilias Chalkidis. Processing long legal documents with pre-trained transformers: Modding legalbert and longformer, 2022.

[19] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[20] Emre Mumcuoğlu, Ceyhun E. Öztürk, Haldun M. Ozaktas, and Aykut Koç. Natural language processing in law: Prediction of outcomes in the higher courts of turkey. *Information Processing Management*, 58(5):102684, 2021.

[21] Ha-Thanh Nguyen, Manh-Kien Phi, Xuan-Bach Ngo, Vu Tran, Le-Minh Nguyen, and Minh-Phuong Tu. Attentive deep neural networks for legal document retrieval. *Artificial Intelligence and Law*, 32(1):57–86, December 2022.

[22] Rodrigo Nogueira and Kyunghyun Cho. Passage re-ranking with bert. *arXiv preprint arXiv:1901.04085*, 2019.

[23] Rodrigo Nogueira and Kyunghyun Cho. Passage re-ranking with bert, 2020.

[24] University of Maryland Institute for Advanced Computer Studies (UMIACS). Trec legal track, 2024. Accessed: 2024-09-14.

[25] Ceyhun Emre Öztürk. Retrieving Turkish prior legal cases with deep learning. Master's thesis, Bilkent University, 2023.

[26] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics.

[27] Carol Peters, Martin Braschler, and Julio Gonzalo. Clef 2010: Conference on multilingual and multimodal information access evaluation. In *Cross-Language Evaluation Forum (CLEF)*. Springer, 2010.

[28] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[29] Nicholas Pipitone and Ghita Houir Alami. Legalbench-rag: A benchmark for retrieval-augmented generation in the legal domain, 2024.

[30] Reha Poroy and Hamdi Yasaman. *Ticari İşletme Hukuku*. Seçkin Yayıncılık, Mustafa Kemal, 2158. Sk. No:13, 06510 Çankaya/Ankara, 2022.

[31] Tao Qin and Tie-Yan Liu. Mslr-web30k: A large-scale dataset for learning to rank, 2010.

[32] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[33] Stephen E Robertson and Steve Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organized by Dublin City University*, pages 232–241. Springer, 1994.

[34] Keet Sugathadasa, Buddhi Ayesha, Nisansa de Silva, Amal Shehan Perera, Vindula Jayawardana, Dimuthu Lakmal, and Madhavi Perera. Legal document retrieval using document vector embeddings and deep learning, 2018.

[35] Andrew Trotman, Xiangfei Jia, and Matt Crane. Towards an efficient and effective search engine. In *OSIR@ SIGIR*, pages 40–47, 2012.

[36] Wei Yang, Yuqing Xie, Aileen Lin, Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin. End-to-end open-domain question answering with. In *Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics, 2019.

[37] Masaharu Yoshioka, Yasuhiro Aoki, and Youta Suzuki. Bert-based ensemble methods with data augmentation for legal textual entailment in coliee statute law task. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, ICAIL '21, page 278–284, New York, NY, USA, 2021. Association for Computing Machinery.

[38] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. How does NLP benefit legal system: A summary of legal artificial intelligence. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5218–5230, Online, July 2020. Association for Computational Linguistics.

APPENDIX A

DETAILED DATA SET EXAMPLES

Additional examples for the three different datasets provided in Chapter 3 can be found in the tables below.

Table A.1: Title Path Data Set Example (Table 3.2 continued)

book_code	index	title_path	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	1617	b. Akreditif İlişkisini Yöneten İlkeler 6. Akreditif C. Uluslararası Ticari Satışlarda Ödeme Şekilleri I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR	Bu aşamada, akreditif ilişkisini yöneten belli ilkelere kısaca değinmek yararlı olacaktır. Akreditifin en önemli özelliği, temelindeki hukuki ilişkiden soyut bir karakter taşımasıdır (500 sayılı Broşür m. 3). Örneğin, bir bankacılık işlemi niteliğinde olan akreditif gereğince yapılacak ödeme, bu ödemeye konu oluşturan alıcı ve satıcı arasındaki satım sözleşmesinin ayıplı ifa sebebiyle ihlalinden etkilenmez. Alıcı, ayıplı ifada bulunan satıcıya karşı haklarını alt ilişkiye dayanarak kullanabilir, ancak bankanın akreditif bedelini ödemekten kaçınmasını talep edemez.
64c09e9d0bd3ea0ee2a71067_240125144019	1619	b. Akreditif İlişkisini Yöneten İlkeler 6. Akreditif C. Uluslararası Ticari Satışlarda Ödeme Şekilleri I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR	Değinilmesi gereken son nokta, bankaların belirli durumlardan ve akreditif işleminde yer alan diğer tarafların hareketlerinden sorumlu tutulamayacaklarıdır. Buna göre bankalar; herhangi bir mesajın, mektubun ya da vesaikin gönderilmesi sırasında kaybolması veya gecikmesi sonucu oluşacak sonuçlardan, teknik terimlerin çeviri ya da yorumlanmasındaki hatalardan sorumlu tutulamazlar (500 sayılı Broşür m. 16). Tabii afet, isyan, ayaklanma, iç karışıklık, savaş hali ya da kendi kontrolleri dışında gelişen diğer nedenlerle veya grev ya da lokavtlar sonucunda faaliyetlerinde meydana gelecek kesintilerden doğacak sonuçlardan da bankalar sorumlu tutulamayacaklardır (500 sayılı Broşür m. 17).

Table A.2: Title Path Data Set Example (Table A.1 continued)

book_code	index	title_path	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	1618	b. Akreditif İlişkisini Yöneten İlkeler 6. Akreditif C. Uluslararası Ticari Satışlarda Ödeme Şekilleri I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR	İkinci ilke, akreditif işlemlerinde bankaların belgeler vasıtasıyla işlem yapmaları ilkesidir. Bununla ifade edilmek istenen şudur; akreditif sebebiyle ödemeyi yapacak olan banka lehdarın edimini gerçekten ve gereği gibi yerine getirip getirmediğini araştırmak yükümlülüğü altında değildir 508 . Akreditif metninde belirtilen belgelerin teslimi ile birlikte amir banka açısından ödeme yükümlülüğü doğacaktır. Üçüncü önemli ilke, akreditif ilişkisinde bankanın yükümlülüğünün belgeleri incelemekle sınırlı olduğudur 509 . Bankalar akreditif metninde bulunan belgeleri makul bir dikkat sarf ederek incelemek yükümlülüğü altındadırlar. Buna karşın ibraz edilen belgelerin gerçeği yansıtıp yansıtmadıkları, sahte olup olmadıkları ya da yasal yönden geçerli bulunup bulunmadıkları bankaları ilgilendirmez (600 sayılı Broşür m. 14). Şu halde, bankaların sorumluluğu, akreditif metninde gösterilen belgelerin varlığını aramakla sınırlandırılmıştır. Ancak bu ilkeye belli hallerde istisna tanımak gerekebilir. Örneğin, sahte olduğu derinlemesine bir inceleme gerekmeksizin açıkça görülebilir konumda bulunan bir belgenin kabulü suretiyle ödemenin yapılması halinde, bankanın TK m. 20/f. 2 çerçevesinde basiretli bir tacir gibi hareket etme yükümünü ihlal ettiği gerekçesiyle sorumluluğuna gidilmesi mümkündür 510 . Belgelerin ibrazı karşılığında ödeme yapmakla yükümlü bulunan bankanın, akreditif amiri tarafından söz konusu belgelerin gerçeği yansıtmadığının likit delillerle ispatlanması halinde de ödeme işlemini gerçekleştirmekten kaçınması gerekir 511.

Table A.3: Title Path Data Set Example (Table A.2 continued)

book_code	index	title_path	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	1424	b. Alıcının Temerrüdü 3. Tacirler Arasındaki Ticari Satış ve Mal Değişimleri (Trampalar) E. İki Tarafı Tacir Sıfatını Haiz Kişiler Olan Ticari İşler Hakkındaki Özel Hükümler IV. TACİR OLMANIN HÜKÜMLERİ Dördüncü Bölüm TACİR	TTK m. 23 bent (b), alıcının temerrüdü halinde, satıcının malın satışına izin verilmesini mahkemeden isteyebileceğini hüküm altına almaktadır 425 . Bendin, bu esas hüküm dışında kalan kısımları satışın teferruatı ile ilgilidir.
64c09e9d0bd3ea0ee2a71067_240125144019	1425	b. Alıcının Temerrüdü 3. Tacirler Arasındaki Ticari Satış ve Mal Değişimleri (Trampalar) E. İki Tarafı Tacir Sıfatını Haiz Kişiler Olan Ticari İşler Hakkındaki Özel Hükümler IV. TACİR OLMANIN HÜKÜMLERİ Dördüncü Bölüm TACİR	Bu hükmü tam olarak değerlendirmek için, kısmi ifada olduğu gibi genel hükümleri incelemek ve TTK'nın getirdiği özelliği belirtmek gerekiyor.
64c09e9d0bd3ea0ee2a71067_240125144019	1426	b. Alıcının Temerrüdü 3. Tacirler Arasındaki Ticari Satış ve Mal Değişimleri (Trampalar) E. İki Tarafı Tacir Sıfatını Haiz Kişiler Olan Ticari İşler Hakkındaki Özel Hükümler IV. TACİR OLMANIN HÜKÜMLERİ Dördüncü Bölüm TACİR	TBK m. 232 alıcının asli borçları olarak, satış bedelini ödemek ve kendisine sunulan satılanı devralmayı saymaktadır 426 . Satış bedelini ödemede temerrüt halini TBK m. 235 ve 236 düzenlemiştir. Bu maddeler hükümlerinden alıcının satış bedelini ödemede temerrüdünün, borçlu temerrüdünün özel bir şekli olarak düzenlendiği görülmektedir. Gerçekten, m. 235'e göre satılanın ancak satış bedeli ödendikten sonra veya ödenme anında devredilmesi gereken durumlarda alıcı temerrüde düşerse satıcı, herhangi bir işleme gerek kalmaksızın satıştan dönebilir. Bu hüküm, hem adi, hem ticari satışlar için öngörülmüştür. TBK m. 236 ise, 235. maddedeki hükmün uygulanması halinde, ticari satışlarda satıcının alıcıdan isteyeceği zarar ve ziyanın nasıl hesap edileceğini düzenlemektedir.

Table A.4: Generated Query Data Set

book_code	index	title_path	generated_query	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	1441	b. Alıcının Temerrüdü 3. Tacirler Arasındaki Ticari Satış ve Mal Değişimleri (Trampalar) E. İki Tarafı Tacir Sıfatını Haiz Kişiler Olan Ticari İşler Hakkındaki Özel Hükümler IV. TACİR OLMANIN HÜKÜMLERİ Dördüncü Bölüm TACİR	TTK m. 23 bent b alıcı temerrüdü	TBK m. 223 f. 1 alıcının, devraldığı satılanın durumunu işlerin olağan akışına göre imkan bulunur bulunmaz gözden geçirmek ve satılanda satıcının sorumluluğunu gerektiren bir ayıp görürse, bunu uygun bir süre içinde ona bildirmek zorunda olduğunu belirtmektedir.
64c09e9d0bd3ea0ee2a71067_240125144019	1443	b. Alıcının Temerrüdü 3. Tacirler Arasındaki Ticari Satış ve Mal Değişimleri (Trampalar) E. İki Tarafı Tacir Sıfatını Haiz Kişiler Olan Ticari İşler Hakkındaki Özel Hükümler IV. TACİR OLMANIN HÜKÜMLERİ Dördüncü Bölüm TACİR	TTK m. 23 bent b alıcı temerrüdü	Ancak, satılan şeyde olağan bir gözden geçirmeyle ortaya çıkarılamayacak bir ayıp var ise, ayıba karşı tekeffül sorumluluğunun devam ettiği süre içinde, (bu süre TBK m. 231'e göre satıcı daha uzun bir süre için üstlenmiş olmadıkça, satılanın alıcıya devrinden başlayarak iki yıldır) meydana çıkar çıkmaz satıcıya hemen bildirmek şartıyla, ayıptan doğan haklarını kullanabilir (TBK m. 223 f. 2) 437 .

Table A.5: Generated Query Data Set (Table A.4 continued)

book_code	index	title_path	generated_query	paragraph
64c09e9d0bd3ea0ee2a71067_ 240125144019	1621	c. Akreditif Türleri 6. Akreditif C. Uluslararası Ticari Satışlarda Ödeme Şekilleri I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR	Akreditif türleri ve özellikleri	Akreditif, dönülebilir-dönülemez ya da teyitli-teyitsiz olarak öngörülebilir. Bu iki ayrımın yanında uygulamada rastlanan farklı akreditif türleri de bulunmaktadır. Son olarak anılan bu akreditif türleri aşağıda tek bir başlık altında incelenecektir.
64c09e9d0bd3ea0ee2a71067_ 240125144019	1622	c. Akreditif Türleri 6. Akreditif C. Uluslararası Ticari Satışlarda Ödeme Şekilleri I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR I. ULUSLARARASI TİCARİ SATIŞLARI DÜZENLEYEN ANLAŞMALAR	Akreditif türleri ve özellikleri	• Dönülebilir (Revocable)-Dönülemez (Irrevocable) Akreditif Ayrımı

Table A.6: Generated Query Data Set (Table A.4 continued)

book_code	index	title_path	generated_query	paragraph
64c09e9d0bd3ea0ee2a71067_240125144019	2129	c. Marka Lisans Sözleşmelerine Uygulanması 6. Acentenin Denkleştirme İsteminin Benzer Sözleşmelere Uygulanması G. Acentelik Sözleşmesinin Sona Ermesinin Hukuki Sonuçları III. ACENTE Sekizinci Bölüm TİCARİ İŞLETMEYE BAĞLI YARDIMCILAR VE ÇEŞİTLİ ARACILIK FAALİYETLERİ	SMK 24. madde marka lisansı	SMK'nın 24. maddesine göre tescilli bir markanın kullanım hakkı, tescil edildiği mal ve hizmetlerin bir kısmı veya tamamı için lisans sözleşmesine konu olabilir. Lisans inhisari ya da inhisarı olmayarak verilir 718 . Marka lisans inhisarı olsun veya olmasın sürekli bir borç ilişkisi yaratır. Markayı lisans ile kullanan, ürünü veya hizmeti tescilli marka altında pazarlar ve satar. Bu şekilde markalı ürünün veya hizmetin satış rakamını yükseltir ve markaya bağlı olan müşteri çevresini genişletir. Bu markanın tanıtılmasında ve müşteri çevresinin genişletilmesinde lisans alanın gayret ve mahareti önemlidir. Marka tanınmış bile olsa, bunun daha çok müşteri kitlesine ulaşması, pazarlama ağının genişletilmesinde lisans alanın emeği vardır 719 . Lisans sözleşmesinin sona ermesinden sonra, lisans veren lisans alanın kurduğu müşteri çevresinden sözleşmenin bitiminden sonra da yararlanır. Lisans veren, bir üçüncü kişiye lisans verirken bile bu gelişmeden faydalanır. Bilhassa, Türkiye'ye ilk girecek markalarda lisans alanın ünü ve gayreti önem kazanır. Büyük bir şirketin veya holdingin yabancı bir markanın lisansını alması, pazardaki gelişmesini hızlandırır. Acentedeki denkleştirme istemine ilişkin şartların burada da aynen aranması gerekir. Bu şartlar var ve hakkaniyet de denkleştirme tazminatı verilmesine uygun ise, bu denkleştirme isteminin kabul edilmesi gerekir 720 .

Table A.7: Expert Generated Data Set Example (Table 3.5 continued)

query	answer	book_name	isbn	paragraph-index
Hizmet Tespit davalarında Temyiz Kanun Yolu	Temyiz incelemesi Yargıtay tarafından yapılır. Hangi kararların temyiz edilebileceği, HMK'nın 361. maddesinde düzenlenmiştir. Bölge Adliye Mahkemeleri tarafından verilen temyizi kabil nihai kararlar ve hakem kararlarının iptali üzerine verilecek olan kararlar için temyiz kanun yoluna başvurulabilecektir. Hukukumuzda atlama yoluyla temyiz kabul edilmemiş olmasına rağmen hakem kararlarının iptali üzerine verilecek olan kararlara karşı istinaf kanun yolu değil, temyiz kanun yoluna başvurulacağı kabul edilmiştir. 592	Hizmet Tespiti Davaları	978975-0284458	1195-1193, 1192, 1152, 708
Hizmet Tespiti davalarında davanın açılmamış sayılması	Yargıtay Hukuk Genel Kurulu'nun konuya ilişkin bir kararında, "… Hal böyle olunca, davadan feragat ile davanın açılmamış sayılması müesseselerinin birbirinden tamamen farklı müesseseler olması, davanın açılmamış sayılmasına karar verilmesi halinde kanunlarda öngörülen süreler içerisinde her zaman yeniden dava açılabilmesinin olanaklı bulunması, HUMK m. 409'da açıklanan kuralın emredici bir kural olması ve kanunlarda açıkça bir istisna getirilmemiş olması nedeniyle iki defadan fazla takipsiz bırakılan davanın açılmamış sayılmasına karar verilmesi gerekirken yargılamaya devam edilerek davanın esastan sonuçlandırılması doğru görülmemiştir." 557 ifadeleri ile iki defadan fazla takipsiz bırakılan davanın açılmamış sayılmasına karar verilmesi gerektiğini hükme bağlamıştır.	Hizmet Tespiti Davaları	978975-0284458	11133-1129, 1121, 1076, 1066, 708

Table A.8: Expert Generated Data Set Example (Table A.7 continued)

query	**answer**	**book_name**	**isbn**	**paragraph-index**
Hizmet Tespiti davasında Davaların birleşmesi	Maddi hukuka göre, bir hakkın birden fazla kimse tarafından birlikte kullanılması veya birden fazla kimseye karşı birlikte ileri sürülmesi ve tamamı hakkında tek bir hüküm verilmesi gereken hallerde, mecburi dava arkadaşlığından söz edilir (HMK m. 59). Yani, birden fazla kişi, bir hak için birlikte dava açacaklarsa davanın hepsi tarafından birlikte açılması, o haktan dolayı kendilerine karşı dava açılacaksa davanın hepsine karşı birlikte açılması zorunludur. 279	Hizmet Tespiti Davaları	978975-0284458	812-811, 808, 739, 708
Hizmet Tespit davasının sona erme şekilleri	Hizmet tespiti davasında yapılan yargılama neticesinde mahkemece davanın kabulüne, kısmen kabulüne veya reddine karar verilir. Bunun yanında bazı durumlarda mahkemece herhangi bir hüküm verilmeksizin davanın sona erdirilmesi de mümkündür. Bu başlık altında taleple bağlılık ilkesi, davanın sona ermesi halleri ve kanun yolları inceleme konusu yapılacaktır.	Hizmet Tespiti Davaları	978975-0284458	1068-1067, 1066, 708
Tespit Davasının çeşitleri	Tespit davaları, olumlu (müspet) tespit davası ve olumsuz (menfi) tespit davası olmak üzere ikiye ayrılır.	Hizmet Tespiti Davaları	978975-0284458	598-597, 593, 587, 397