**PROPOSAL SKRIPSI**

**SISTEM PENCARIAN DOKUMEN PUTUSAN PENGADILAN DENGAN RERANKER BERBASIS TRANSFORMER**



**Oleh:**

**M. SHOKHIBUL ANWAR**

**200411100130**

**Dosen Pembimbing 1 : Firdaus Solihin, S.Kom., M.Kom.**

**Dosen Pembimbing 2 :**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2023**

# ABSTRAK

Jumlah dokumen putusan pengadilan yang dipublikasikan pada situs web resmi Mahkamah Agung Republik Indonesia terus bertambah dengan rata-rata 900.000 kasus per-tahun. Karenanya, penerapan sistem pencarian yang efektif sangat diperlukan untuk mempermudah dan mempercepat pencarian informasi relevan dari kasus terdahulu. Kendati demikian metode pencarian yang digunakan situs resmi Mahkamah Agung masih menggunakan metode pencarian tradisional *boolean search* yang menuntut pengguna untuk memiliki keahlian teknis yang tinggi agar dapat menggunakan sistem dengan maksimal. Penelitian ini bermaksud untuk menerapkan metode pencarian berbasis model probabilistik (BM25) yang digabung dengan *reranker* berbasis *Bidirectional Encoder Representations from Transformers (BERT).* Teknik BM25 mempertimbangkan frekuensi *term* dan karakteristik dokumen sehingga dapat meningkatkan fleksibilitas query serta menangani kompleksitas dan panjang dokumen. Sedangkan komponen *reranker* berfungsi untuk meningkatkan relevansi dokumen hasil BM25 dengan mempertimbangkan pemahaman semantik menggunakan model *pre-trained* berbasis transformer. Aplikasi metode ini diharapkan dapat meningkatkan akurasi dan fleksibilitas pencarian dokumen putusan pengadilan sehingga dapat meningkatkan efektifitas dan efisiensi pekerjaan para praktisi hukum.

Kata kunci : Information Retrieval, Legal Case Retrieval, BM25, BERT

# DAFTAR ISI

[ABSTRAK i](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756936)

[DAFTAR ISI ii](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756937)

[DAFTAR GAMBAR iv](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756938)

[DAFTAR TABEL v](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756939)

[BAB 1 1](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756940)

[1.1 Latar Belakang 1](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756941)

[1.2 Permasalahan 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756942)

[1.3 Metode Usulan 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756943)

[1.4 Pertanyaan penelitian 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756944)

[1.5 Tujuan 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756945)

[1.6 Manfaat 2](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756946)

[1.7 Batasan masalah 3](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756947)

[1.8 Sistematika Penulisan 3](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756948)

[BAB 2 4](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756949)

[2.1 Leukimia Limfoblastik akut (ALL) 4](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756950)

[2.2 Klasifikasi Citra 5](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756951)

[2.3 Pengolahan Citra Digital (PCD) 6](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756952)

[2.4 Augmentasi Data 7](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756953)

[2.5 Convolutional Neural Network (CNN) 8](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756954)

[2.5.1 *Convolutional Layer* 9](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756955)

[*2.5.2 Pooling Layer* 10](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756956)

[*2.5.3 Rectified Linear Unit (ReLU)* 11](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756957)

[*2.5.4 Fully Connected Layer (FCL)* 12](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756958)

[*2.5.5 Squeeze and Excitation (SE) Layer* 12](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756959)

[2.5.6 *Softmax* 14](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756960)

[2.6 HSV (Hue Saturation Value) 15](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756961)

[2.7 Fungsi aktivasi 16](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756962)

[2.7.1 Sigmoid 16](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756963)

[2.7.2 Swish 17](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756964)

[2.7.3 H – Swish 18](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756965)

[2.8 MobileNet 19](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756966)

[2.9 MobileNet V3 21](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756967)

[2.10 Confusion Matrix 22](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756968)

[2.11 Penelitian Terkait 24](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756969)

[BAB 3 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756970)

[3.1 Arsitektur Sistem 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756971)

[3.2 Dataset 27](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756972)

[3.3 Prepocessing 28](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756973)

[3.4 HSV Coloring 29](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756974)

[3.5 Mobile – Net V3 30](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756975)

[3.7 Skenario Pengujian 32](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756976)

[3.8 Evaluasi 33](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756977)

[3.9 Jadwal Penelitian 33](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756978)

[DAFTAR PUSTAKA 34](file:///C:\Users\user\AppData\Local\Microsoft\Windows\INetCache\IE\M3HREP7E\Proposal_Skripsi(190411100082).pdf%5b1%5d.docx#_Toc147756979)

# DAFTAR GAMBAR

# DAFTAR TABEL

# 

**PENDAHULUAN**

## 1.1 Latar Belakang

Negara Kesatuan Republik Indonesia (NKRI) adalah sebuah negara hukum yang secara eksplisit ditegaskan melalui Undang-Undang Dasar Negara Republik Indonesia Pasal 1 ayat (3)[1]. Dalam penegakan hukum, kekuasaan kehakiman khususnya Mahkamah Agung memiliki peranan penting dalam menyelenggarakan peradilan sebagaimana diatur dalam Undang-Undang no. 48 tahun 2009[2]. Melalui serangkaian proses persidangan dan permusyawaratan hakim, suatu kasus akan dinyatakan selesai setelah sahnya putusan hakim berkekuatan hukum tetap dan tidak ada upaya hukum lanjutan[3]. Putusan pengadilan yang sudah berkekuatan hukum tetap memiliki kekuatan mengikat, kekuatan pembuktian dan kekuatan eksekutorial (pelaksanaan) untuk terdakwa atau pihak-pihak yang bersengketa [4].

Putusan pengadilan atau putusan hakim adalah produk dari pemeriksaan perkara di persidangan yang berisikan pernyataan hakim untuk menyelesaikan atau memperlancar pemeriksaan suatu perkara [5]. Putusan hakim terdiri dari banyak macam yang berbeda-beda berdasarkan tiap jenis perkara hukum. Setiap jenis perkara hukum juga memiliki banyak macam putusan yang dibagi berdasarkan beberapa faktor seperti amar putusan, tingkat putusan, tingkat pengakhiran perkara, kehadiran para pihak, dll [5]. Perbedaan-perbedaan jenis ini mempengaruhi karakteristik isi putusan yang mengakibatkan perbedaan fungsi putusan terhadap pihak-pihak yang terlibat. Selain ditujukan bagi pihak-pihak yang terlibat, putusan pengadilan juga dapat berguna untuk dipelajari bagi publik misalnya materi pembelajaran akademisi bidang hukum, atau kalangan aparat dan praktisi hukum.

Data putusan pengadilan yang diunggah pada situs Direktori Putusan MA pada Februari 2024 mencapai lebih dari 8,5 juta dokumen dengan penambahan rata-rata 65 ribu dokumen setiap bulan [6]. Agar bisa mengimbangi banyaknya jumlah dokumen, diperlukan sistem pencarian otomatis yang efektif, efisien dan memiliki skalabilitas baik. Sistem pencarian *(Information Retrieval)* pada bidang hukum khususnya dokumen konten panjang seperti putusan pengadilan memiliki tantangan tersendiri yang berbeda dengan bidang lain. Tantangan ini umumnya berupa pengoptimalan performa terkait besarnya basis data, pemahaman semantik dokumen, pengolahan panjang dokumen, dan pemahaman jargon atau istilah-istilah khusus [7].

Metode pencarian dokumen yang digunakan pada situs resmi Direktori Putusan MA saat penelitian ini dibuat adalah metode pencarian leksikal berbasis model pencarian *boolean*. *Boolean search* didasarkan pada logika *Boolean* dan teori himpunan di mana dokumen dan kueri dianggap sebagai *bag-of-words* atau sekumpulan istilah *(terms)* [8]. Cara kerjanya adalah mencari berdasarkan ada atau tidaknya istilah-istilah pada teks kueri pada dokumen, serta terpenuhinya operator logika *boolean* yang ada pada teks kueri. Keunggulan Boolean search ada pada konsepnyayaitu sama persis setiap kata *(hard match)* yang membuatnya mudah diaplikasikan, mudah dipahami dan membuat pengguna memegang kendali penuh atas hasil pencarian [9].

Disisi lain, karakteristik lugas dari metode *boolean retrieval* ini tidak selalu sesuai dengan apa yang dibutuhkan oleh pengguna. Keterbatasan model antara lain adalah pada segi fleksibilitas, sensitivitas terhadap pembuatan kueri, dan tidak adanya pemeringkatan relevansi sehingga menghambat pencarian informasi dengan topik yang luas [10]. Hal ini menuntut pengguna untuk memiliki pengetahuan teknis seperti pemahaman sintaksis dan istilah-istilah pada topik terkait agar dapat menemukan hasil yang tepat secara efektif. Solusi permasalahan-permasalahan tersebut telah banyak diteliti seiring dengan berjalannya waktu dan perkembangan teknologi. Diantaranya dengan implementasi algoritma pemeringkatan *(ranking)*[11], penanganan sinonim dan variasi kata [12], implementasi pemahaman semantik [13], serta pengaplikasian berbagai teknik pemrosesan bahasa alami *(NLP)* [14].

Penerapan NLP pada sistem pencarian telah banyak dibuktikan dapat meningkatkan akurasi pencarian secara signifikan pada berbagai bidang [15]. Salah satu teknik mutakhir untuk sistem pencarian yang akurat pada dokumen dengan karakteristik khusus *(domain-specific)* adalah teknik , poin unggul, sekilas bukti keunggulan

Penelitian ini, hipotesis singkat, harapan

## 1.2 Permasalahan

Bagaimana menerapkan sebuah sistem pencarian informasi yang efektif dan efisien untuk mencari dokumen relevan dari basis data dokumen putusan pengadilan yang beragam serta, mengekstrak informasi spesifik dalam dokumen yang panjang dan memiliki struktur yang kompleks.

## 1.3 Metode Usulan

Implementasi model pencarian dengan algoritma pemeringkatan probabilistik berbasis *term-frequency* yaitu BM25 dengan komponen *reranker* berbasis *transfer learning* menggunakan model *Pre-Trained* *Bidirectional Encoder Representations from Transformers (BERT)*  yang mampu menangkap makna semantik dari dokumen sehingga dapat meningkatkan akurasi pencarian.

## Pertanyaan penelitian

1. Seberapa akurat algoritma BM25 dalam menemukan dokumen relevan pada sistem pencarian dokumen putusan pengadilan?
2. Seberapa efektif penerapan BERT reranker dalam meningkatkan tingkat akurasi pencarian dokumen putusan pengadilan berbasis BM25?

## 1.5 Tujuan

Menerapkan metode pencarian BM25 dengan BERT reranker pada sistem pencarian dokumen putusan pengadilan Mahkamah Agung Republik Indonesia

## 1.6 Manfaat

Penelitian ini bertujuan memberikan kontribusi dalam bidang hukum khususnya terkait dokumen putusan pengadilan dengan menciptakan sistem pencarian yang lebih fleksibel dari segi query masukan dan memberikan hasil yang lebih akurat.

## 1.7 Batasan masalah

Penelitian ini berfokus pada :

1. Data yang digunakan berasal dari <https://github.com/ir-nlp-csui/indo-law> berisi 22630 dokumen putusan yang sudah dilakukan *parsing*, *cleaning* dan anotasi segmen oleh [16].
2. Jenis putusan yang digunakan hanya jenis putusan perkara pidana umum yang sudah berstatus berkekuatan hukum tetap *(inkracht)*.

## 1.8 Sistematika Penulisan

Struktur penulisan proposal disusun sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi 5 subbab yaitu: 1) Latar Belakang, 2) Rumusan Masalah, 3) Tujuan Penelitian, 4) Batasan Masalah, dan 5) Sistematika Penulisan. Menjelaskan konteks penelitian seperti pengenalan bidang penelitian, masalah yang ada dan kontribusi penelitian ini pada bidang tersebut.

BAB 2 LANDASAN TEORI

Dalam bab ini dijelaskan tentang teori-teori yang menjadi landasan bagi penelitian ini, yang mana berisikan: 1) Penjelasan *Information Retrieval*, 2.) *Legal Information Retrieval*, 3) *Best Matching 25* *(BM25)*, 4) *BERT*, dan 5) *BERT Reranker*.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang akan diterapkan, mencakup analisis rancangan keseluruhan sistem, Pemrosesan dataset, alur penelitian, dan jadwal penelitian.

# 

**LANDASAN TEORI**

# DAFTAR PUSTAKA

[1] Indonesia, *Undang-Undang Dasar 1945, Pasal 1 Ayat (3)*. 1945.

[2] Indonesia, Pemerintah Pusat, *Undang-Undang No. 48 Tahun 2009*. 2009.

[3] M. Y. Harahap, *Hukum Acara Perdata: Tentang Gugatan, Persidangan, Penyitaan, Pembuktian, dan Putusan Pengadilan*. Sinar Grafika, 2017.

[4] S. Mertokusumo, “Hukum acara perdata Indonesia,” 2009, Accessed: Feb. 26, 2024. [Online]. Available: http://senayan.iain-palangkaraya.ac.id/index.php?p=show\_detail&id=11461&keywords=

[5] Y. Yulia, *Hukum Acara Perdata*. Lhokseumawe: Unimal Press, 2018. Accessed: Feb. 27, 2024. [Online]. Available: https://repository.unimal.ac.id/5938/

[6] “Direktori Putusan.” Accessed: Feb. 29, 2024. [Online]. Available: https://putusan3.mahkamahagung.go.id/

[7] T. Maxwell and B. Schafer, “Natural language processing and query expansion in legal information retrieval: Challenges and a response,” *Int. Rev. Law Comput. Technol.*, vol. 24, no. 1, pp. 63–72, Mar. 2010, doi: 10.1080/13600860903570194.

[8] F. W. Lancaster and E. Gallup, “Information retrieval on-line,” 1973. Accessed: Mar. 02, 2024. [Online]. Available: https://trid.trb.org/view/9772

[9] M. S. Oghli and M. M. Almustafa, “Comparison of basic Information Retrieval Models,” *Int. J. Eng. Res. Technol.*, vol. 10, no. 09, 2021, Accessed: Mar. 02, 2024. [Online]. Available: https://www.researchgate.net/profile/Muhammad-Almustafa/publication/358509603\_Comparison\_of\_basic\_Information\_Retrieval\_Models/links/620563ee7b05f82592deda67/Comparison-of-basic-Information-Retrieval-Models.pdf

[10] G. Salton, *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[11] R. Baeza-Yates, “Modern Information Retrieval,” *Addison Wesley Google Sch.*, vol. 2, pp. 127–136, 1999.

[12] J. Ullman, “Principles of database systems Computer Science Press,” *Rockvillc Matyland*, 1980.

[13] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *J. Am. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, Sep. 1990, doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9.

[14] G. G. Chowdhury, *Introduction to modern information retrieval*. Facet publishing, 2010. Accessed: Mar. 03, 2024. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=cN4qDgAAQBAJ&oi=fnd&pg=PR11&dq=modern+Information+Retrieval&ots=rFffnm52gp&sig=toOlODeUnqeCo\_uBXKeuXXugv4k

[15] D. D. Lewis and K. S. Jones, “Natural language processing for information retrieval,” *Commun. ACM*, vol. 39, no. 1, pp. 92–101, Jan. 1996, doi: 10.1145/234173.234210.

[16] E. Q. Nuranti, E. Yulianti, and H. S. Husin, “Predicting the Category and the Length of Punishment in Indonesian Courts Based on Previous Court Decision Documents,” *Computers*, vol. 11, no. 6, Art. no. 6, Jun. 2022, doi: 10.3390/computers11060088.