

Sistem Temu Kembali untuk Berita Berkaitan Pasal - Pasal KUHP pada Hukum Indonesia

Ferdiansyah Dwi Nurcahyo
Dep. Ilmu Komputer dan Elektronika
Universitas Gadjah Mada

Muny Safitri
Dep. Ilmu Komputer dan Elektronika
Universitas Gadjah Mada

Raden Mas Garda Zhafir
Dep. Ilmu Komputer dan Elektronika
Universitas Gadjah M

I. INTRODUCTION

A. Latar Belakang

Munculnya Hukum dimaksudkan untuk menimbulkan dampak jera kepada orang-orang yang melakukan tindakan kejahatan. Hal ini bertujuan untuk mencegah timbulnya suatu kejahatan dalam kehidupan bermasyarakat. Untuk melindungi masyarakat dari kejahatan maka diperlukan sanksi berupa pidana. Adanya hukum yang diterapkan tidak hanya melindungi masyarakat yang dirugikan atau korban, melainkan hukum juga harus melindungi hak-hak tersangka dan memberikan rasa keadilan bagi masyarakat. Berhukum memang dimulai dari teks (undang-undang). Teks hukum yang bersifat umum itu memerlukan akurasi atau penajaman yang kreatif saat diterapkan pada kejadian nyata di masyarakat.

Berhukum dengan teks baru merupakan awal perjalanan panjang untuk mewujudkan tujuan agar hukum dapat mewujudkan keadilan dan kemanfaatan bagi kemanusiaan (Ardiansyah, 2018). Tegak tidaknya kaidah-kaidah hukum sangat ditentukan oleh tingkat ketaatan warga masyarakat pada perintah-perintah yang terkandung didalamnya, dan ketaatan ini ditentukan oleh kekuatan sanksi-sanksi yang terkandung dalam kaidah-kaidah tersebut. Diketahui bahwa sekalipun sanksi itu tidaklah mampu menjamin terealisasinya ketaatan warga masyarakat dan tegaknya kaidah-kaidah hukum (Ucuk, 2013).

Teknologi telah menjangkau berbagai sektor industri seperti finansial, transportasi, imigrasi, pembayaran, dan lain sebagainya. Kini saatnya teknologi menyentuh industri hukum. Dengan adanya bantuan teknologi yang mumpuni, hukum bukan lagi menjadi suatu yang kompleks dan dihindari melainkan menjadi sesuatu yang ingin dipahami dan dicari tahu masyarakat (Syahrin, 2017). Kebanyakan orang lebih suka mendapatkan informasi dari orang lain daripada dari sistem temu kembali informasi. Namun, selama beberapa tahun terakhir, optimasi tanpa henti dari efektivitas pencarian informasi telah mendorong mesin pencari web ke tingkat kualitas baru yang membuat *user* pemakai puas sepanjang waktu. Pencarian web telah menjadi standar dan sumber pencarian informasi yang sering disukai (Christopher D., 2009).

Beberapa tahun belakangan ini, *Legal Information*

Retrieval(LIR) menjadi salah satu topik utama dalam domain hukum yang bertujuan untuk memodelkan pencarian informasi

seperti yang dilakukan oleh praktisi hukum untuk mengidentifikasi informasi yang berguna untuk pekerjaan mereka. Secara khusus, LIR mencakup tugas yang berbeda seperti pemodelan dokumen elektronik (*Discovery of Electronically Stored Information*), Pengambilan paten dan/atau pemulihan kasus sebelumnya (Christopher D., 2009).

Dengan demikian, sistem temu kembali informasi dapat mencakup mendukung pengguna dalam menelusuri atau memproses lebih lanjut sekumpulan dokumen yang diambil, menghasilkan pengelompokan dokumen berdasarkan isinya, dapat mengklasifikasi dokumen secara otomatis (Christopher D., 2009).

B. Related Work

Menurut Van Opijnen et al. (2017) seperti dikutip oleh Sansone dan Sperli (2021), informasi produk hukum memiliki sifat sebagai berikut:

1. *Volume*: jumlah informasi hukum mencakup yang ada pada penyimpanan publik dan penyimpanan pribadi (kegiatan hukum yang pernah dilakukan)
2. *Document size*: ukuran dokumen pada informasi yang berdomain hukum umumnya lebih besar dari domain lainnya
3. *Structure*: Tiap dokumen mempunyai struktur yang spesifik
4. *Heterogenity*: tipe dokumen dapat memiliki variasi yang besar, seperti kasus hukum, berkas perkara, paten, hingga pasal
5. *Self-contained document*: Masing-masing dokumen memiliki otoritasnya sendiri karena domain hukum didefinisikan oleh dokumen-dokumen tersebut
6. *Legal hierarchy*: Masing-masing dokumen memiliki hierarki otoritasnya sendiri
7. *Temporal aspect*: Mengedepankan unsur waktu
8. *Citations' relevance*: Masing-masing teks dan dokumen bisa saling terhubung satu sama lain
9. *Audience*: Kemampuan dan keahlian legal memiliki diversitas tinggi
10. *Personal data*: *Legal memory* dibentuk berdasarkan

pemahaman orang-orang yang

berbeda dan berada di tempat yang berbeda-beda

11. *Multilingualism and multi-jurisdictionality*: produk hukum dibentuk berdasarkan bahasa yurisdiksi yang berbeda-beda

12. *Sparseness of legal resources*: Terdapat kejaran dalam menemukan sumber produk hukum atau sumber produk hukum tidak tersebar merata di internet.

Mitra dan Craswell (2018) seperti dikutip oleh Sansone dan Sperli (2021) menyatakan bahwa setidaknya, sebuah sistem TKI secara umum harus memiliki karakteristik seperti (i) pemahaman semantik; (ii) ketahanan terhadap masukan yang bermacam-macam; (iii) ketahanan terhadap corpus yang bermacam-macam; (iv) ketahanan terhadap kenaikan jumlah dokumen; (v) ketahanan terhadap kesalahan masukan; dan (vi) sensitif terhadap konteks. Pada sistem TKI untuk domain hukum, terdapat beberapa ketentuan yang harus dipenuhi yang mencakup (i) tahan terhadap macam-macam cara seorang pengguna mendefinisikan kebutuhan informasinya melalui *query*; (ii) mendapatkan makna semantik dari dokumen yang ditelaah; (iii) menemu-kembalikan dokumen yang sesuai dengan keinginan praktisi hukum.

Untuk mengatasi permasalahan seperti yang tertera pada latar belakang dan rumusan masalah serta mengacu pada tinjauan pustaka, kami menawarkan solusi berupa sebuah sistem TKI yang berfokus pada domain hukum, dalam hal ini adalah pasal-pasal pada Kitab Umum Hukum Pidana (KUHP). Hal ini disebabkan informasi produk hukum di Indonesia memenuhi 13 karakteristik yang didefinisikan oleh Van Opijnen, terutama pada karakteristik *structure* dan *legal hierarchy*.

C. Proposed approach

Untuk mengatasi permasalahan seperti yang tertera pada latar belakang dan rumusan masalah serta mengacu pada tinjauan pustaka, kami menawarkan solusi berupa sebuah sistem TKI yang berfokus pada domain hukum, dalam hal ini adalah pasal-pasal pada Kitab Umum Hukum Pidana (KUHP). Hal ini disebabkan informasi produk hukum di Indonesia memenuhi 13 karakteristik yang didefinisikan oleh Van Opijnen, terutama pada karakteristik *structure* dan *legal hierarchy*.

Kami memilih solusi diatas adalah karena bagi para praktisi hukum, sistem temu kembali informasi ini dapat membantu mengambil keputusan yang lebih akurat dapat dilakukan dengan metode statistika. Selain itu, dengan solusi yang kami buat ini, maka akan menumbuhkan sikap dan etika yang sesuai dengan standar praktisi hukum. Solusi yang kami buat juga dapat dijadikan sebagai fungsi recheck bagi praktisi hukum.

II. METHODS

A. Akuisisi Data

Akuisisi data dilakukan dengan cara web Scraping dengan tools Python, yaitu dengan libraries BeautifulSoup. Web scrapping dilakukan pada laman <https://www.hukumonline.com/klinik/>. Scraping dilakukan pada tiap halaman yang kemudian di ekspor dengan format CSV.

B. Transformasi Data

Transformasi data yang dilakukan menggunakan pandas untuk mengubah data mentah dari akuisisi data menjadi data yang dapat digunakan dengan lebih mudah.

Transformasi pada Sistem Temu Kembali untuk Pasal-Pasal KUHP pada Hukum Indonesia berfokus pada pengkategorian dan penyelarasan format-format data kuantitatif, dalam hal ini adalah tanggal dan pemasukan kategori tiap data hasil scraping. Kategori berada pada array dengan isi sebagai berikut: {'bisnis', 'hak-asasi-manusia', 'ilmu-hukum', 'kekayaan-intelektual', 'keluarga', 'ketenagakerjaan', 'perdata', 'perlindungan-konsumen', 'pertanahan-properti', 'pidana', 'profesi-hukum', 'start-up-umkm', 'teknologi'}. Fungsi kategori ini adalah untuk memungkinkan adanya pengelompokkan hasil retrieval nanti. Proses penetapan dan penyimpanan kategori ini dapat dilakukan ketika scrapping ulang di mana kategori tiap dokumen hasil scraping terdapat pada url-nya.

Selain itu, terdapat penyelarasan format penanggalan artikel pada tanggal hasil scraping di mana sekarang hanya tahun yang disimpan. Hal ini merujuk pada asumsi bahwa makin terbaru suatu dokumen, dapat menjadi rujukan yang lebih baik karena memiliki dasar hukum yang lebih kuat dan lebih aktual daripada yang lebih lama.

C. Pembuatan Indeks

Pembuatan indeks adalah mengubah data hasil scraping yang sudah di transformasi tadi ke dalam bentuk indeks. Indeks dibuat dalam bentuk Inverted Index.

1) *Tokenisasi*: Tahap tokenisasi adalah mengubah data menjadi bentuk term. Hal ini dilakukan untuk mempermudah proses lainnya.

2) *Inverted Index*: Inverted index adalah bentuk akhir dari pembuatan indeks. Frekuensi banyaknya term dan juga letak termnya dapat dilihat pada Inverted Index.

D. Kompresi Indeks

Berdasarkan informasi dari Techopedia [5], kompresi data adalah proses modifikasi, encode, dan konversi suatu informasi dari sisi struktur digit biner sedemikian rupa sehingga suatu informasi tersebut dapat mengonsumsi tempat penyimpanan yang lebih sedikit. Proses kompresi data dibutuhkan dalam pembuatan dictionary karena pembuatan dan penyimpanan dictionary dilakukan di *main memory* yang memiliki kapasitas terbatas dan harus berbagi kapasitas dengan program lain sehingga kompresi harus dilakukan untuk mengurangi penggunaan kapasitas *memory* sebanyak mungkin.

Main memory memiliki metode random access sehingga proses pencarian dictionary dapat dilakukan lebih cepat. daripada sequential access yang terjadi pada disk memory. Ini juga disebabkan oleh dictionary yang mengonsumsi memori lebih sedikit daripada postings list. Postings list mengonsumsi memori lebih banyak daripada dictionary, biasanya dalam faktor 10 dari konsumsi memori dictionary (Introduction to Information Retrieval, 2009). Oleh karena itu, postings list biasanya langsung ditempatkan pada disk memory.

Postings list harus dikompresi sepadu mungkin karena

[illegible]

4) *Kompresi Indeks*: Kompresi indeks dilakukan untuk mengurangi ukuran file yang disimpan dalam main memory. Pada bagian ini, akan menghasilkan indeks terkompresi yang berbentuk string dictionary (*dictionary-as-a-string*).

Fig. 4. Dictionary as String

1) *Akuisisi Data*: Metode akuisisi data dengan web scraping menghasilkan beberapa dokumen data kejadian yang berkaitan dengan KUHP di Indonesia. Data yang didapat masih sedikit karena belum mengambil dari semua kategori dan juga masih berantakan sehingga sulit dipakai untuk *indexing*

Fig. 1. Dataset Hasil Scraping

No	Judul_Artikel	Update	Nitra	Category	Author	Tahun	
1	Langkah Jaka Mantar Karyawan Membocorkan Rahas...	09 Sep 2022		Si Pokrol	Bismadeth Aurelia Oktavia, S.H.	2022	
2	Prijaman Melebihi Modat Dasar PT, Penuhak Per...	08 Sep 2022	Tjorlat Tandjung & Associates	Bisnis	Syafidha Tandjung, S.H., LL.M	2022	
3	Persamaan Asosiasi dengan Pengumpulan dan Pros...	06 Sep 2022		Si Pokrol	Bisnis	David Christian, S.H.	2022
4	7 Alasan Mengapa Investor Asing Bernvestasi di...	01 Sep 2022		Si Pokrol	Bisnis	David Christian, S.H.	2022
5	Fungsi Lembaga Pengelola Investasi dan Bedanya...	31 Agt 2022	Pusat Kajian Hukum Bisnis FH Unar	Bisnis	Dr. Widhayani Dian Praewati S.H., M.H.	2022	
...	
1177	Hukumnya Bebohong dan Menyebarkan Fitnah di Medsos	26 Feb 2019		Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1178	Hukum Menyebarkan Lagu yang Bermuatan Peng...	26 Feb 2019		Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1179	Mengunggah Screenshot SMS Mengganggu ke Med...	22 Feb 2019		Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1180	Apakah Mengirim Link Berita tentang Seks Term...	19 Feb 2019		Si Pokrol	Teknologi	Tri Jata Ayu Pramesti, S.H.	2019
1181	Recovery File Foto Pribadi di HP dan Masalah H...	15 Feb 2019	Indonesia Cyber Law Community (ICLC)	Teknologi	Teguh Anjidy, S.H., M.H.	2019	

3) *Pembuatan Indeks*: Metode pembuatan indeks menghasilkan data inverted index yang berisi term, frekuensi dari term tersebut, dan posting list (Posisi term pada dokumen asli). Struktur data seperti ini dapat mempermudah proses pencarian dan mudah untuk dikompresi.

IV. DISCUSSION

Metode kompresi dengan menggunakan *dictionary-as-a-string* bukanlah metode kompresi paling efisien. Menurut *Introduction to Information Retrieval* (2008), metode paling efektif adalah metode *front-code*. Selain itu, metode kompresi *postings list* juga dapat dilakukan dengan *delta encoding*. Perlu dilakukan pembuatan sistem temu kembali informasi berfokus pada konten hukum dengan melakukan pendekatan-pendekatan metode kompresi tersebut. Indikator uji dapat dilakukan dengan mengecek jumlah penyimpanan terpakai serta *responds time* sejak dikirimkan *query* hingga didapatkan dokumen ke pengguna.

Dari sisi konsep, penulis pikir bahwa sistem temu kembali yang *re-retrieve* berita berkaitan dengan pasal hukum ini mendukung salah satu sifat hukum, yaitu yurisprudensi. Selain itu, adanya sistem temu kembali informasi ini memberikan pengguna perbandingan dengan kasus hukum yang dirasa relevan dengan kasus hukum yang sedang dialami pengguna sehingga diharapkan, dapat melakukan tindakan hukum yang lebih sesuai dan tepat sasaran.

REFERENCES

- [1] U. Agiyanto “Eksplorasi Konsep Keadilan Berdimensi Ketuhanan”, Publikasi Ilmiah Universitas Muhammadiyah Surakarta, , Jan-2018, [Online]. Available: <https://publikasiilmiah.ums.ac.id/handle/11617/9722>. [Accessed: 21-Aug-2022].
- [2] Ardiansyah, “Penegakan Hukum di Indonesia: Polemik dan Tantangan Progresif dalam Sistem Hukum Indonesia”, Jurnal de June, 2018, [Online]. Available: <https://jurnal.law.uniba-bpn.ac.id/index.php/jurnaldejure/article/view/16>. [Accessed: 21-Aug-2022].
- [3] M. Lubis, W. Syahran, “Demi Masyarakat Melek Hukum, Asosiasi Regtech dan Legaltech Diluncurkan”. Bisnis.com, 25-Sep-2017, [Online]. Available: <https://kabar24.bisnis.com/read/20170925/16/692948/demi-masyarakat-melek-hukum-asosiasi-regtech-legaltech-diluncurkan>.
- [4] C. D. Manning, “An Introduction to Information Retrieval”, Higher Education from Cambridge University Press, 7-Jul-2008, [Online]. Available: <https://www.cambridge.org/highereducation/books/introduction-to-information-retrieval/669D108D20F556C5C30957D63B5AB65C#overview>. [Accessed: 21-Aug-2022].
- [5] “Data Compression”, Techopedia, 28-Mei-2019, [Online]. Available: <https://www.techopedia.com/definition/884/data-compression#:~:text=D ata%20compression%20is%20the%20process,coding%20or%20bit%2D rate%20reduction>. [Accessed: 3-Okt-2022].
- [6] Sansone, C. Sansone ,G. Sperli, “Legal Information Retrieval systems: State-of-the-art and open issues”, ScienceDirect. May-2022, [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0306437921001551> [Accessed: 22-Aug-2022].