



IMPLEMENTASI KONSEP *INFORMATION RETRIEVAL* DENGAN METODE *CASE INSENSITIVE SEARCH* PADA MESIN PENCARI DOKUMEN *QUALITY*

Agung Brotokuncoro^{1*}, Hasanul Fahmi², Wiranto Herry Utomo³

Program Pascasarjana Teknologi Informasi President University

*Email Corresponding: agung.brotokuncoro@student.president.ac.id

Abstrak

Information Retrieval adalah bidang yang berkaitan dengan struktur, analisis, organisasi, penyimpanan, pencarian, dan pengambilan informasi. Mesin pencari dokumen yang menerapkan konsep *information retrieval* ini telah tersedia di PT Chemco Harapan Nusantara namun belum menunjukkan performa yang optimal. Mesin pencari dokumen yang tersedia belum dapat menentukan informasi mana yang terbaik relevansinya sesuai dengan permintaan pengguna dikarenakan masih terikat pada nama dokumen dengan penempatan huruf kapital dan huruf kecil yang sama persis dengan dokumen yang tersedia, sehingga saat pengguna berbeda menempatkan huruf capital dan huruf kecil saat memasukkan *query*, maka mesin pencari dokumen tidak menunjukkan hasil pencarian. Mesin pencari dokumen tersebut belum dapat memprediksi apa yang sebenarnya diinginkan pengguna sehingga memerlukan waktu yang lama untuk menemukan dokumen. Konsep *Information Retrieval* memiliki banyak teknik diantaranya adalah metode *case insensitive search* yaitu teknik memperlakukan huruf besar dan kecil sebagai identik saat melakukan pencarian teks. Tujuan penelitian ini adalah menerapkan metode *case insensitive search* pada mesin pencarian dokumen untuk mencapai *efficient search* dengan meningkatkan relevansi dan akurasi hasil pencarian melalui penelitian eksperimen kuantitatif. Hasil penelitian menunjukkan dengan penerapan metode *case insensitive search* tercapai *efficient search* yang meliputi nilai *Recall* 0.85, nilai Presisi 1.00, dan nilai *F1-Score* 0.91. Pada hasil tertera bahwa nilai *F1-Score* mendekati 1.0 maka model memiliki kinerja yang baik dalam mencapai keseimbangan antara Presisi dan *Recall*.

Kata kunci: *efficient search*, *information retrieval*, mesin pencari dokumen, metode *case insensitive search*

PENDAHULUAN

Mesin pencari dokumen dengan penerapan konsep *information retrieval* ini telah tersedia di PT Chemco Harapan Nusantara terutama pada penyimpanan digital dokumen prosedur, namun masih ada beberapa hal yang perlu dikembangkan dalam rangka meningkatkan performa mesin pencari dokumen tersebut untuk mencapai *efficient search*. Hal yang umum terjadi pada mesin pencari dokumen di PT Chemco Harapan Nusantara adalah masalah pencarian informasi yang ada kalanya belum dapat menentukan informasi mana yang terbaik relevansinya sesuai dengan permintaan pengguna tertentu. Mesin pencari dokumen tersebut masih terikat pada nama dokumen dengan penempatan huruf kapital dan huruf kecil yang sama persis dengan dokumen yang tersedia, sehingga saat pengguna berbeda menempatkan huruf capital dan huruf kecil saat memasukkan *query* maka mesin pencari tidak menunjukkan hasil pencarian. Sistem *information retrieval* tersebut belum dapat memprediksi apa yang sebenarnya diinginkan pengguna dan memerlukan waktu yang lama untuk menemukan dokumen.

Penelitian pendahuluan penulis melalui wawancara pada para pengguna didapatkan hasil bahwa kepuasan pengguna umumnya terkait dengan bagaimana pengguna menemukan sesuatu yang berhubungan dengan apa yang mereka cari, dan menemukannya dengan cepat bahkan hanya dengan menginput kata kunci yang berbeda penempatan huruf capital dan huruf kecilnya pada suatu

kalimat. Dengan demikian diperlukan penggunaan metode yang tepat untuk meningkatkan *efficient search* dalam hal meningkatnya relevansi serta akurasi hasil pencarian dokumen.

Information Retrieval memiliki beberapa metode salah satunya adalah dengan metode *case-insensitive search*. Menurut Roshdi & Roohparvaar (2015) pencarian yang tidak peka terhadap huruf besar atau kecil (*case insensitive search*) adalah suatu metode dalam *Information Retrieval* yang memperlakukan huruf kapital dan huruf kecil sebagai identik saat melakukan pencarian teks. Ini berarti bahwa tidak ada perbedaan antara huruf kapital dan kecil saat mencocokkan kata atau frasa selama proses pencarian. Menurut Baeza-Yates & Riberiro-Neto (2011) metode ini umumnya diterapkan dalam sistem pencarian, basis data, dan aplikasi lainnya untuk meningkatkan kemudahan penggunaan dan memastikan bahwa hasil pencarian mencakup semua kemungkinan variasi penulisan huruf (*uppercase atau lowercase*). Pencarian yang tidak peka terhadap huruf besar dan kecil memungkinkan sistem IR untuk memproses kueri pencarian dan dokumen yang diindeks tanpa membedakan antara huruf besar dan kecil. Menurut Ibrich dkk. (2022) tujuan memproses *query* adalah untuk mengurangi kompleksitas pencarian bagi pengguna dan memastikan bahwa variasi dalam penulisan tidak menghambat proses pencarian informasi yang efektif. Dalam model spasial IR, dokumen dan kueri direpresentasikan sebagai vektor dalam ruang multidimensi.

Case insensitive search membantu dalam menyederhanakan representasi ini dengan mengurangi dimensi vektor (yaitu, mengurangi variabilitas token berdasarkan kapitalisasi), sehingga mempercepat pencocokan kueri dengan dokumen. Salah satu teknik utama dalam pencarian yang tidak peka terhadap huruf adalah normalisasi teks, di mana semua teks diubah menjadi huruf kecil (atau huruf besar) baik pada saat indeksasi maupun pada saat pencarian. Teknik ini memastikan bahwa pencarian tidak dipengaruhi oleh perbedaan kapitalisasi. Penelitian relevan terkait dengan penerapan metode *case insensitive* untuk meningkatkan performa mesin pencari dokumen adalah yang dilakukan oleh Sivarajkumar dkk. (2024) & Amrullah (2018) bahwa pencarian *case insensitive* yang diimplementasikan dalam database relasional secara signifikan meningkatkan kinerja pencarian dibandingkan dengan metode pencarian *case sensitive* tradisional, pendekatan baru ini menawarkan waktu respons yang lebih cepat dan hasil yang lebih relevan, terutama dalam skenario penggunaan dengan volume data yang besar dan variasi penulisan huruf besar/kecil.

Tujuan penelitian eksperimen ini adalah menerapkan metode *case insensitive search* pada mesin pencarian dokumen dan menganalisis performa mesin pencari dokumen tersebut dalam hal *efficient search* yang meliputi relevansi dan akurasi hasil pencarian. Dengan mempelajari kajian teori yang ada dan hasil penelitian terdahulu maka diharapkan metode *case insensitive search* dapat meningkatkan performa mesin pencari dokumen di PT Chemco Harapan Nusantara dalam hal *efficient search* yang meliputi relevansi dan akurasi hasil pencarian (Mourão dkk., 2020). Işık & Dağ Isik (2020) menyelidiki bagaimana penggunaan huruf kecil meningkatkan kinerja dalam klasifikasi teks. Penggunaan huruf kecil merupakan salah satu teknik preprocessing yang paling sederhana namun efektif dalam penambahan teks, meningkatkan kinerja klasifier dengan mengurangi sensitivitas huruf besar-kecil dan memastikan konsistensi dalam data tekstual.

Sharma dkk. (2017) menguji dampak preprocessing pada empat set data teks Twitter yang berbeda, termasuk olahraga, politik, hiburan, dan keuangan. Hasil mereka menunjukkan bahwa mengonversi teks menjadi huruf kecil meningkatkan akurasi klasifikasi data sampel Twitter. Camacho & Pilehvar (2018) melakukan studi tentang peran teknik preprocessing sederhana, termasuk penggunaan huruf kecil, terhadap kinerja klasifier teks neural. Hickman dkk. (2022) menyimpulkan bahwa semua teks sebaiknya dikonversi menjadi huruf kecil, karena hal ini mengurangi dimensi data sambil mempertahankan informasi semantik. Sebagian besar studi organisasi melaporkan penggunaan konversi huruf kecil dalam penambahan teks dengan kosakata terbuka. Kobayashi dkk. (2018) menekankan bahwa konversi huruf kecil cenderung bermanfaat

karena mengurangi dimensi data, sehingga meningkatkan kekuatan statistik, dan biasanya tidak mengurangi validitas. HaCohen-Kerner dkk. (2020) menemukan bahwa mengonversi huruf besar menjadi huruf kecil adalah satu-satunya metode preprocessing tunggal yang menunjukkan peningkatan signifikan dibandingkan dengan metode preprocessing lainnya. Hasil terbaik untuk dataset mereka diperoleh ketika mereka melakukan koreksi ejaan dan konversi menjadi huruf kecil.

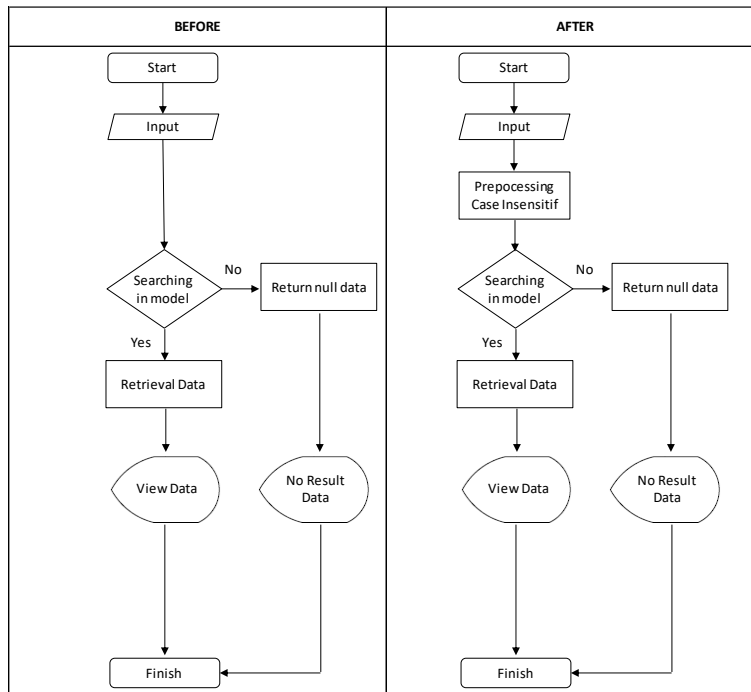
Rosid dkk. (2020) mempelajari dokumen-dokumen yang mengandung huruf besar atau kesalahan penulisan serupa. Mereka menyoroti bahwa dalam preprocessing teks, proses penyusunan huruf bertujuan untuk mengubah semua huruf dalam dokumen teks menjadi huruf kecil, seperti mengubah "Wifi" menjadi "wifi". Tabassum dkk. (2020) mengidentifikasi bahwa teknik huruf kecil merupakan kontributor utama dalam meningkatkan akurasi dari algoritma machine learning berbasis teks apa pun. Garcia & Smith (2023) mengemukakan bahwa pendekatan yang tidak peka terhadap huruf besar-kecil dalam sistem pengambilan dokumen telah mengalami kemajuan signifikan, yang dijelaskan dalam penelitian mereka tentang sistem tersebut. Kim dkk. (2018) menyoroti dampak dari kueri yang tidak peka terhadap huruf besar-kecil dalam pencarian dokumen yang ramah pengguna.

Rodriguez dkk. (2020) meneliti dampak kemajuan dalam pencarian yang tidak peka terhadap huruf besar-kecil terhadap efisiensi pengambilan dokumen. Mereka menyoroti peningkatan signifikan dalam aksesibilitas informasi dan keterbacaan dengan penggunaan kueri yang tidak peka terhadap huruf besar-kecil. Temuan mereka menunjukkan bahwa pendekatan ini meningkatkan kemudahan pengguna dalam menemukan dan mengakses informasi yang relevan. Thompson dkk. (2022) melakukan studi perbandingan untuk meningkatkan pengalaman pengguna dalam pengambilan dokumen menggunakan pencarian yang tidak peka terhadap huruf besar-kecil. Mereka menyelidiki efek dari pendekatan ini terhadap efektivitas dan kepuasan pengguna dalam menemukan informasi yang relevan. Hasil studi menunjukkan bahwa penggunaan pencarian yang tidak peka terhadap huruf besar-kecil secara signifikan meningkatkan pengalaman pengguna dalam pengambilan dokumen.

METODE

Penelitian ini merupakan penelitian kuantitatif dengan metode eksperimen pada mesin pencari dokumen inspeksi PT Chemco Harapan Nusantara untuk mencapai *efficient search* yang ditunjukkan oleh tingkat relevansi dan akurasi hasil pencarian dengan perhitungan nilai Presisi, *Recall*, dan nilai *F1-Score*. Gambar 1 merupakan gambaran alur proses penerapan metode *case insensitive research* yang diterapkan pada mesin pencari dokumen di PT Chemco Harapan Nusantara.

Gambar 1. Alur proses penerapan metode *case insensitive search*



Sebelum penerapan metode *case insensitive search*, proses dimulai dengan input data, yang dapat berupa dokumen dalam format teks atau data yang perlu diindeks oleh mesin pencari selanjutnya. Pada tahap ini, mesin pencari mungkin mempertimbangkan perbedaan huruf besar dan kecil, sehingga kata-kata yang sama dengan penulisan yang berbeda dianggap sebagai entitas yang berbeda. Ketika pengguna melakukan pencarian, mesin pencari akan mencocokkan kueri pencarian dengan indeks yang telah dibuat sebelumnya. Jika pencarian dilakukan tanpa mempertimbangkan sensitivitas huruf, hanya kata-kata dengan penulisan huruf yang sama dengan kueri yang akan dihasilkan.

Setelah penerapan metode *case insensitive search*, proses dimulai dengan input data yang sama seperti sebelumnya, yaitu dokumen dalam format teks atau data yang perlu diindeks oleh mesin pencari. Proses ditambah dengan metode *case insensitive*. Ini berarti kata-kata diindeks tanpa memperhatikan perbedaan antara huruf besar dan kecil. Ketika pengguna melakukan pencarian, mesin pencari akan mencocokkan kueri pencarian dengan indeks yang telah diperbarui dengan metode *case insensitive*. Dalam hal ini, kata-kata dengan penulisan huruf yang berbeda tetapi memiliki makna yang sama akan dianggap sebagai hasil yang relevan. Dengan demikian, penerapan metode *case insensitive search* pada mesin pencari dokumen memungkinkan pengguna untuk melakukan pencarian tanpa memperhatikan perbedaan huruf besar dan kecil, sehingga meningkatkan kemungkinan untuk menemukan dokumen yang relevan dalam hasil pencarian.

Analisis performa mesin pencari dokumen dalam rangka mencapai *efficient search* dilakukan perhitungan Presisi (*Precision*) dan *Recall* yang merupakan dua metrik evaluasi kinerja umum yang digunakan dalam sistem information retrieval informasi. Berikut adalah persamaan untuk menghitung Presisi dan *Recall*, beserta keterangan tiap komponennya:

Presisi (Precision)

$$\text{Presisi} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall (Recall/Sensitivitas)

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Keterangan:

1. Jumlah dokumen yang relevan dan ditemukan (*True Positive, TP*)
2. Jumlah dokumen yang tidak relevan tetapi keliru diidentifikasi sebagai relevan (*False Positive, FP*)
3. Jumlah dokumen yang sebenarnya relevan tetapi tidak ditemukan (*False Negative, FN*)
4. Jumlah dokumen yang tidak relevan dan diabaikan (*True Negative, TN*)

F1-Score adalah metrik evaluasi yang menggabungkan Presisi (*Precision*) dan Recall untuk memberikan gambaran keseluruhan tentang kinerja suatu sistem atau model dalam konteks *information retrieval*. F1-Score adalah salah satu metrik evaluasi kinerja yang umum digunakan dalam bidang klasifikasi, terutama dalam konteks *machine learning* dan *data mining*.

Persamaan F1-Score

$$\text{F1-Score} = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{(\text{presisi} + \text{Recall})}$$

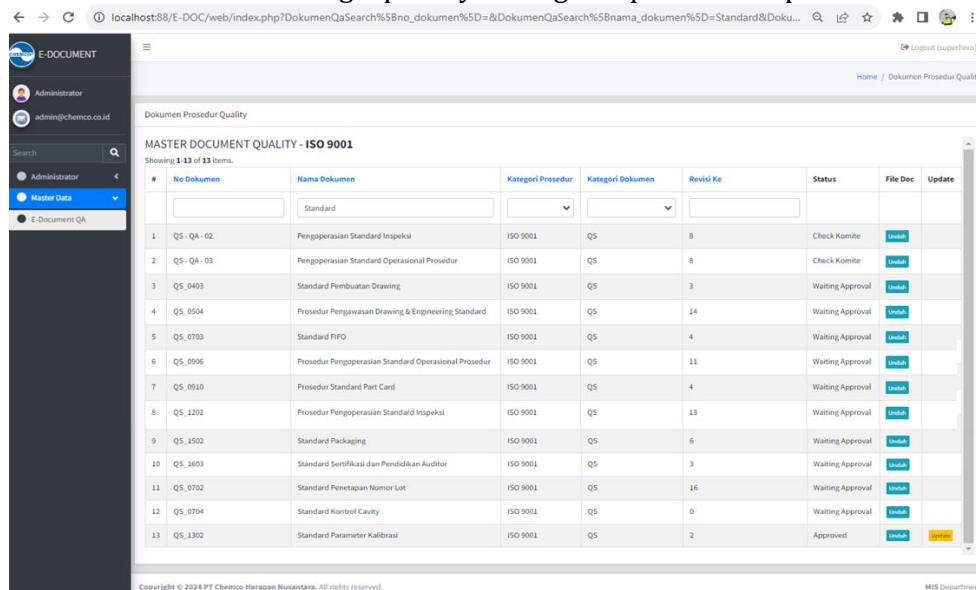
Interpretasi nilai F1-Score dapat dilakukan dengan memperhatikan beberapa skenario sebagai berikut :

1. Nilai F1-Score = 1 (Tinggi)
2. Nilai F1-Score Mendekati 1 (Tinggi)
3. Nilai F1-Score Mendekati 0 (Rendah)
4. Nilai F1-Score = 0 (Rendah)

HASIL DAN PEMBAHASAN

Mesin pencari dokumen di PT Chemco Harapan Nusantara menggunakan bahasa pemrograman JavaScript. Gambar 2 merupakan tampilan dan model algoritma mesin pencari dokumen di PT Chemco Harapan Nusantara setelah ditambahkan metode *Case Insensitive Search* dengan kueri menggunakan huruf Kapital seluruhnya.

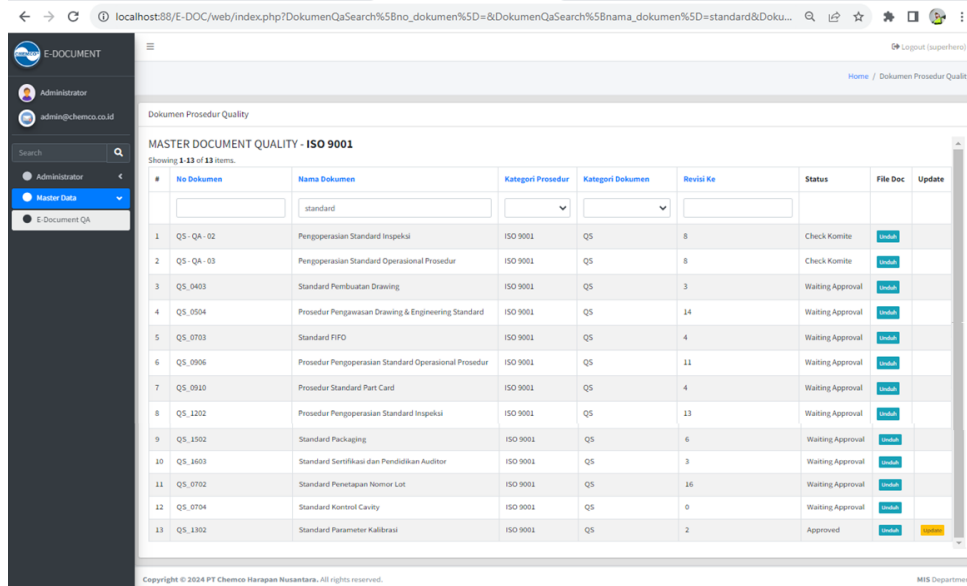
Gambar 2. Tampilan layar dengan input huruf kapital



#	No Dokumen	Nama Dokumen	Kategori Prosedur	Kategori Dokumen	Revisi Ke	Status	File Doc	Update
1	QS-QA-02	Pengoperasian Standard Inspeksi	ISO 9001	QS	8	Check Komite	Link	
2	QS-QA-03	Pengoperasian Standard Operasional Prosedur	ISO 9001	QS	8	Check Komite	Link	
3	QS_0403	Standard Pembuatan Drawing	ISO 9001	QS	3	Waiting Approval	Link	
4	QS_0504	Prosedur Pengawasan Drawing & Engineering Standard	ISO 9001	QS	14	Waiting Approval	Link	
5	QS_0703	Standard FIFO	ISO 9001	QS	4	Waiting Approval	Link	
6	QS_0906	Prosedur Pengoperasian Standard Operasional Prosedur	ISO 9001	QS	11	Waiting Approval	Link	
7	QS_0910	Prosedur Standard Part Card	ISO 9001	QS	4	Waiting Approval	Link	
8	QS_1202	Prosedur Pengoperasian Standard Inspeksi	ISO 9001	QS	13	Waiting Approval	Link	
9	QS_1502	Standard Packaging	ISO 9001	QS	6	Waiting Approval	Link	
10	QS_1603	Standard Sertifikasi dan Pendidikan Auditor	ISO 9001	QS	3	Waiting Approval	Link	
11	QS_0702	Standard Penetapan Nomor Lot	ISO 9001	QS	16	Waiting Approval	Link	
12	QS_0704	Standard Kontrol Cavity	ISO 9001	QS	0	Waiting Approval	Link	
13	QS_1302	Standard Parameter Kalibrasi	ISO 9001	QS	2	Approved	Link	Update

Adapun untuk Gambar 3 merupakan tampilan mesin pencari dokumen di PT Chemco Harapan Nusantara setelah ditambahkan metode *Case Insensitive Search* yang diinput kueri dengan menggunakan huruf kecil seluruhnya.

Gambar 3 . Tangkapan layar dengan input huruf kecil



Jika dilihat dari gambar 2 dan gambar 3, dengan input huruf besar dan kecil didapat hasil output yang sama, yaitu 13 dokumen. Algoritma pencarian yang tidak peka terhadap huruf besar atau kecil (*case insensitive*) bekerja dengan meminimalkan perbedaan antara huruf besar dan kecil selama proses pencarian. Tujuannya adalah untuk memastikan bahwa pencarian kata kunci dapat mengembalikan hasil yang relevan, tanpa memperdulikan apakah kata kunci tersebut ditulis dengan huruf besar, huruf kecil, atau kombinasi keduanya. Berikut adalah narasi yang menggambarkan cara kerja umum dari algoritma tersebut dalam konteks sistem pengambilan informasi. Pertama, proses dimulai dengan fase pra-pemrosesan baik pada dokumen yang akan diindeks maupun pada kueri pencarian yang dimasukkan oleh pengguna. Dalam fase ini, semua teks diubah menjadi format yang seragam, biasanya dengan mengkonversi seluruh karakter ke huruf kecil. Sebagai contoh, kata "Inspeksi", "INSPEKSI", dan "inspeksi" semua akan diubah menjadi "inspeksi". Konversi ini memastikan bahwa perbedaan kapitalisasi tidak mempengaruhi proses pencarian dan pengindeksan selanjutnya.

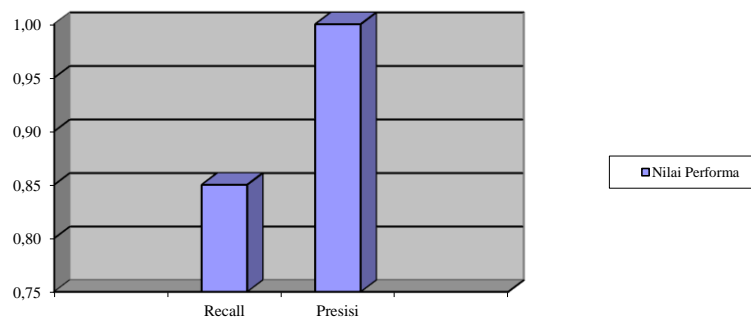
Selanjutnya, sistem melanjutkan ke fase pengindeksan, di mana dokumen yang telah dinormalisasi akan diproses. Setiap kata atau token dalam dokumen diidentifikasi dan diindeks dalam struktur data yang disebut indeks terbalik (*inverted index*), yang memungkinkan pencarian cepat, karena semua teks telah dinormalisasi ke huruf kecil, indeks hanya perlu menyimpan satu versi dari setiap kata, tanpa memperhatikan kapitalisasi aslinya. Ketika pengguna mengajukan kueri pencarian, kueri tersebut juga akan menjalani normalisasi yang serupa, di mana semua huruf dalam kueri akan diubah menjadi huruf kecil. Langkah ini memastikan bahwa kueri pencarian sesuai dengan format data yang telah diindeks, memungkinkan sistem untuk secara efisien mencari kata kunci dalam indeks terbalik tanpa perlu khawatir tentang perbedaan kapitalisasi. Setelah kueri dinormalisasi, sistem melakukan pencarian dalam indeks terbalik untuk menemukan semua dokumen yang mengandung kata kunci yang dicari. Karena proses normalisasi telah menghilangkan perbedaan kapitalisasi, pencarian ini dapat dengan mudah menemukan kata kunci tanpa memperhatikan bagaimana kata tersebut awalnya dikapitalisasi dalam dokumen.

Dengan menggunakan algoritma diatas mesin pencari dokumen *quality* sudah dapat secara fleksibel mengenali *query* yang diinput oleh pengguna dengan berbagai variasi penempatan huruf capital dan huruf kecil. Kemudian dilakukan pengamatan perubahan performa mesin pencari dokumen *quality* sebelum dan sesudah menggunakan metode *case insensitive search*. Hasil penelitian mendapatkan bahwa performa mesin pencari dokumen *quality* meningkat dalam hal respon waktu, tingkat keakuratan dokumen yang diinginkan pengguna serta meningkatnya alternatif informasi yang ditampilkan saat pengguna memasukkan *query* pada mesin pencari dokumen tersebut. Data hasil penelitian tersebut dapat dilihat pada Tabel dan Grafik berikut :

Tabel 1. Nilai Performa Mesin Pencari Dokumen

No.	Nama Aspek	Nilai	
		Sebelum	Sesudah
1.	Recall	0	0.85
2.	Presisi	0	1.0

Gambar 4. Peningkatan Performa Mesin Pencari Dokumen *Quality* PT Chemco Harapan Nusantara



Berdasarkan Gambar 4. Peningkatan Performa Mesin Pencari Dokumen *Quality* PT Chemco harapan Nusantara terlihat mesin pencari dokumen sudah dapat meningkatkan performanya dengan mencapai *efficient search* dengan ditemukannya hasil pencarian yang presisi dan relevan meskipun *query* yang diinput memiliki perbedaan letak huruf capital dan huruf kecil nya yang akhirnya berdampak pada efisiensi pencarian karena pengguna tidak lagi membuang waktu untuk memilih dan memilah informasi mana yang relevan atau mendekati relevan sehingga efisiensi tercapai. Selain itu dengan meningkatnya nilai *Recall* dan *Presisi* maka khasanah informasi yang diinginkan menjadi lebih kaya dan akurat sehingga efektivitas kerja dalam hal pencarian dokumen *quality* tercapai. Keseimbangan kinerja mesin pencari dokumen *quality* tersebut dapat dilihat performanya dengan melihat hasil *Recall*, *Presisi* & *F1 Score* . Dari hasil penelitian didapatkan nilai *Presisi* 1,0 dan nilai *Recall* 0,85 sehingga perolehan nilai *F1-Score* adalah 0.91. Hasil ini menunjukkan bahwa nilai *F1-Score* termasuk pada kategori mendekati 1.0 (tinggi) artinya dapat diinterpretasikan bahwa penerapan metode *case insensitive search* memiliki kinerja yang baik dalam mencapai keseimbangan antara *Presisi* dan *Recall* pada mesin pencari dokumen *quality* PT Chemco Harapan Nusantara. Meskipun mungkin ada beberapa kehilangan atau kelebihan dalam pengambilan dokumen relevan, keseimbangan antara *Presisi* dan *Recall* masih terjaga dengan baik. Penerapan metode *case insensitive search* dalam desain sistem pencarian meliputi:

1. Kemudahan Penggunaan : Metode *case insensitive search* dirancang untuk meningkatkan kemudahan penggunaan. Pengguna tidak perlu memperhatikan apakah mereka memasukkan huruf besar atau kecil saat melakukan pencarian, karena hasil pencarian akan mencakup semua kemungkinan variasi huruf.
2. Konsistensi Pencarian: Dengan mengabaikan perbedaan huruf besar dan kecil, sistem pencarian menciptakan konsistensi dalam perilaku pencarian. Hal ini membantu pengguna untuk memiliki pengalaman yang lebih seragam, mengurangi kebingungan, dan membuat proses pencarian lebih dapat diprediksi.
3. Meningkatkan Ruang Lingkup Pencarian: Pencarian yang tidak peka terhadap case memastikan bahwa hasil pencarian mencakup semua kemungkinan variasi penulisan. Seperti contoh tampilan layar mesin pencari dokumen diatas dengan input , pencarian "Prosedur standar Inspeksi" akan mencakup hasil yang mencantumkan "prosedur standar inpeksi," "prosedur Stándar inspeksi ," dan sebagainya.
4. Pemenuhan Diversitas Pengguna: Dengan metode *case insensitive search*, sistem pencarian dapat lebih baik mengakomodasi diversitas pengguna yang memiliki gaya penulisan yang berbeda. Penggunaan yang konsisten dari huruf besar atau kecil dapat bervariasi di antara individu, dan metode ini membantu mengatasi variasi ini.
5. Pengurangan Kesalahan Pencarian: Mengingat bahwa manusia dapat melakukan kesalahan ketika memasukkan huruf besar atau kecil saat melakukan pencarian, metode *case insensitive search* membantu mengurangi kemungkinan kesalahan manusiawi dan meningkatkan akurasi hasil pencarian.
6. Adaptasi Terhadap Gaya Penulisan Natural: Masyarakat cenderung mengadopsi gaya penulisan yang lebih santai dan kurang formal, terutama dalam lingkungan digital seperti media sosial. Metode *case insensitive search* membantu sistem pencarian untuk lebih baik beradaptasi dengan gaya penulisan natural ini.
7. Keterbukaan Terhadap Bahasa: Beberapa bahasa memiliki aturan penulisan yang lebih kompleks daripada yang lain. Metode *case insensitive search* dapat membantu meningkatkan keterbukaan terhadap variasi bahasa dan aturan penulisan yang berbeda di seluruh dunia.

KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan metode *case insensitive search* membawa sejumlah manfaat signifikan dalam meningkatkan efektivitas dan efisiensi kerja pada sistem *information retrieval*. Integrasi metode ini tidak hanya memperbaiki interaksi pengguna tetapi juga memperkuat kualitas hasil pencarian secara keseluruhan. Penerapan metode *case insensitive search* memberikan kemudahan pencarian, konsistensi pencarian, peningkatan ruang lingkup, adaptasi terhadap gaya penulisan natural, reduksi kesalahan pengguna, dan peningkatan ketersediaan informasi dengan mengabaikan perbedaan huruf besar dan kecil, sistem menciptakan konsistensi dalam perilaku pencarian. Hal ini membantu mengurangi kebingungan pengguna dan membuat pengalaman pencarian lebih dapat diprediksi.

REFERENSI

- Amrullah, F., dkk., (2018). Analisis Perbandingan Fitur Search Engine. *Informatics Journal*, 3(1),17-19.
- Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Informations Retrieval* (2nd ed.). USA : ACM Books Press.
- Camacho-Collados, J., & Pilehvar, M. T. (2018). On the role of text preprocessing in neural network architectures. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and*



- Interpreting Neural Networks for NLP* (pp. 40-46).
- Garcia, A., & Smith, J. (2023). Advancements in Document Retrieval Systems: A Case Insensitive Approach. *Journal of Information Science*, 35(2), 145-162.
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, 15(5), 1–22. <https://doi.org/10.1371/journal.pone.0232525>
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25(1), 114–146. <https://doi.org/10.1177/1094428120971683>
- Ibrihich, S., Oussous, A., Ibrihich, O., & Esghir, M. (2022). A Review on recent research in information retrieval. *Procedia Computer Science*, 201, 777-782.
- Işik, M., & Dağ, H. (2020). The impact of text preprocessing on the prediction of review ratings. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(3), 1405–1421. <https://doi.org/10.3906/elk-1907-46>
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismih'ok, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods*, 21(3), 733–765. <https://doi.org/10.1177/1094428117722619>
- Kim, S., & Lee, H. (2018). User-Friendly Document Searches: The Impact of Case Insensitive Queries. *International Journal of Human-Computer Interaction*, 30(4), 301-318.
- Mourão, E., Pimentel, J. F., Murta, L., Kalinowski, M., Mendes, E., & Wohlin, C. (2020). On the performance of hybrid search strategies for systematic literature reviews in software engineering. *Information and Software Technology*, 123(July 2019). <https://doi.org/10.1016/j.infsof.2020.106294>
- Rodriguez, J., & White, L. (2020). Advancements in Case Insensitive Search: Impact on Document Retrieval Efficiency. *Journal of Information Technology Research*, 15(2), 87-102.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1). <https://doi.org/10.1088/1757-899X/874/1/012017>
- Roshdi, A., & Roohparvaar, A. (2015). Review: Information Retrieval Techniques and Applications. *International Journal of Computer Networks and Communications Security*, 3(9), 373–377
- Sharma P, Agrawal A, Alai L, G. A. (2017). Challenges and techniques in preprocessing for twitter data. *International Journal of Engineering Science and Computing*, 7 (4): 6611-6613.
- Sivarajkumar, S., Mohammad, H. A., Oniani, D., Roberts, K., Hersh, W., Liu, H., He, D., Visweswaran, S., & Wang, Y. (2024). Clinical Information Retrieval: A Literature Review. *Journal of Healthcare Informatics Research*, 1–31. <https://doi.org/10.1007/s41666-024-00159-4>
- Tabassum, A., & Patil, R. R. (2020). A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing. *International Research Journal of Engineering and Technology*, June, 4864–4867. www.irjet.net
- Thompson, R., & Anderson, K. (2022). Enhancing User Experience in Document Retrieval Using Case Insensitive Search: A Comparative Study. *Journal of Computer Science and Technology*, 40(3), 321-336.