

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan jumlah dokumen di Internet saat ini sangat pesat. Penanganan proses pencarian dokumen diinternet diperlukan teknik khusus untuk mengambil dokumen yang relevan dengan keinginan pengguna. *Information Retrieval* (IR) yang selanjutnya akan disebut sebagai IR adalah teknik yang dapat digunakan untuk mengambil dokumen yang relevan dengan permintaan pengguna , IR bertujuan untuk menghasilkan dokumen yang relevan dengan permintaan pengguna secara cepat dan tepat [1]. *Query* adalah suatu kemampuan untuk menampilkan data dari *database* untuk diolah lebih lanjut yang biasanya diambil dari *database* , informasi tertentu dari sebuah *database* yang tertulis dalam format tertentu. *Query* dapat didefinisikan sebagai perintah-perintah untuk mengakses data pada *database* [2].

Pada IR dokumen yang dikembalikan oleh sistem ke pengguna tidak selalu relevan terhadap keinginan pengguna. Kejadian ini dikarenakan *query* yang dimasukkan pengguna sering tidak jelas atau tidak spesifik, pengguna sering mengalami kesulitan untuk mendeskripsikan informasi yang sebenarnya dibutuhkan ke dalam bentuk kata. Tindakan yang dapat diambil untuk mengurangi hasil pengembalian dokumen yang tidak relevan yaitu dengan menerapkan teknik perbaikan *query* pada IR yaitu teknik *query expansion* [3].

Query yang singkat dan memiliki makna ambigu dapat bisa menyebabkan hasil pencarian tidak sesuai yang diinginkan, hal dapat diperbaiki dengan *Query Expansion*, yaitu merupakan salah satu teknik untuk menambah *term* tambahan baru pada *query* awal , sedangkan BM25 adalah pendekatan matematis yang bertujuan mengurutkan dokumen yang relevan dari *database* berdasarkan *query* [4] dimana ada bagian dari BM25 yang digunakan untuk mengurutkan *term* berdasarkan tingkat kepentingannya , Dengan kata lain, memformulasikan kembali *query* awal dengan tujuan untuk meningkatkan kinerja pengambilan dalam operasi pengambilan informasi. Motivasi utama *query expansion* adalah untuk menambahkan *term* yang bermakna yang membantu pengguna untuk menghilangkan *ambiguitas* bahasa alami dan mengekspresikan informasi dengan detail ke dalam *query* awal. Dengan menambahkan *term* terkait ke *query* awal juga dapat memperbanyak jumlah dokumen yang teridentifikasi, maka kemungkinan dokumen relevan yang dikembalikan juga meningkat [5]. Sehingga membuat *query* tersebut menjadi lebih akurat dan mendapatkan hasil yang lebih sesuai keinginan pengguna. Tetapi tidak semua *query* perlu di-*expansion*, beberapa faktor yang

mempengaruhinya. Misal subyek *query* yang singkat tetapi akurasiya tinggi, maka tidak perlu di-*expansion* lagi.

Perbaikan *query* dengan menambahkan *term* yang relevan disebut *query expansion*, dengan *query* awal dari user ditambah dengan *term* yang signifikan kemudian dilakukan pencarian ulang yang akan meningkatkan performa *query*, teknik *query expansion* sering digunakan untuk meningkatkan efisiensi proses pengambilan dokumen, serta untuk mengatasi *ambiguitas* suatu kata yang sama namun dengan konteks yang berbeda[4], perhitungan perangkikan dokumen relevan menggunakan metode BM25 dan BM25+ karena Metode BM25 merupakan metode yang terkenal untuk perangkikan dalam menemukan relevansi dokumen. Metode ini baik dalam mencocokkan *query* yang dimasukkan pengguna dengan dokumen yang terdapat di *database* [6], dan pembobotan menggunakan TF-IDF digunakan karena metode ini merupakan metode yang paling umum digunakan pada *information retrieval* untuk perhitungan bobot suatu *term*. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat [7]

Kata kunci yang tepat dibutuhkan oleh IR untuk dapat mengembalikan daftar dokumen yang berkualitas tinggi kepada pengguna. Banyak dokumen yang tidak relevan akan dikembalikan apabila kata kunci yang masukkan terlalu umum. Ini telah menjadi masalah bagi pengguna ketika mereka tidak yakin tentang konten yang mereka butuhkan atau kesulitan menggambarkan konteks informasi yang dibutuhkan hanya dalam beberapa kata kunci. [5], meningkatkan keakuratan hasil pencarian bisa dengan perbaikan *query* menggunakan *query expansion*

Penggunaan metode BM25 dan BM25+ dan *query expansion* diharapkan dapat mencapai nilai presisi yang tinggi. Pada penelitian ini BM25+ akan digunakan di bagian *query expansion* dan sebagai pembanding BM25 juga digunakan dibagian *query expansion* ,dikarenakan pengguna sulit menggambarkan informasi yang diinginkan ke dalam bentuk *query*. Dengan bantuan dari *query expansion* diharapkan *query* baru mampu mengembalikan dokumen sesuai dengan yang diinginkan pengguna.

1.2 Rumusan masalah

1.2.1 Permasalahan

Permasalahan yang umum terjadi adalah penggunaan *query* yang terlalu singkat dan ambigu. Sehingga jika sistem IR BM25 menggunakan *query* singkat dan ambigu tersebut sebagai input , sistem akan mengembalikan hasil pencarian yang tidak sesuai dengan yang diinginkan pengguna.

1.2.2 Metode Usulan

Query yang singkat dan ambigu dapat diatasi dengan melakukan *Query Expansion* menggunakan metode kemiripan BM25+ dan BM25 untuk meningkatkan hasil pencarian yang sesuai dengan yang diinginkan

1.2.3 Pertanyaan Penelitian

Seberapa efektif hasil pencarian dengan *Query Expansion* menggunakan metode kemiripan BM25+ dibandingkan BM25 ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Bertujuan untuk mengukur nilai presisi metode BM25 dan BM25+ dengan dataset berita olahraga jika diterapkan dalam *Query Expansion*

1.3.2 Manfaat

Diharapkan kita sebagai akademisi dapat menentukan metode yang sesuai dengan *Query Expansion* untuk dataset berita olahraga. Jika dengan penggunaan metode BM25+ dengan dataset berita olahraga memiliki nilai presisi yang tinggi, maka BM25+ layak di terapkan dalam *Query Expansion* di sistem IR yang mungkin sedang dibangun oleh rekan-rekan akademisi lain maupun untuk kebutuhan umum

1.4 Batasan-batasan

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Bahasa untuk *query* yang akan di perluas/ekspansi adalah bahasa Indonesia
2. Dokumen yang digunakan sebanyak 7000 artikel berita olahraga berbahasa indonesia yang di *crawl* dari situs <https://sport.detik.com/sepakbola/indeks> dengan rentang waktu 11 november 2019 – 12 november 2020
3. Penelitian skripsi ini hanya menghitung presisi hasil pencarian

1.5 Sistematika Laporan

Sistematika laporan penelitian ini adalah sebagai berikut :

1. BAB I PENDAHULUAN

Bab ini membahas mengenai Latar belakang, Rumusan masalah, Tujuan dan Manfaat, Batasan masalah, dan Sistematika laporan.

2. BAB II KAJIAN PUSTAKA

Bab ini membahas tentang Landasan teori, Penelitian sebelumnya dan Contoh penggunaan metode

3. BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan (algoritma), arsitektur (Input, Proses, Ouput), Dataset, tahapan penelitian, serta rencana pengujian.

4. BAB IV IMPLEMENTASI SISTEM

Bab ini menjelaskan hasil dari penelitian yang dilakukan. Hasil yang diperoleh berupa kekurangan, kelebihan, dan uji coba yang telah dilakukan.

5. BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan penelitian berikutnya