

## ***Query Expansion Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet***

Mahdarani Dwi Laxmi<sup>1</sup>, Indriati<sup>2</sup>, Mochammad Ali Fauzi<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya,  
Email: <sup>1</sup>mahdarani24@gmail.com, <sup>2</sup>indriati.tif@ub.ac.id, <sup>3</sup>moch.ali.fauzi@ub.ac.id

### **Abstrak**

*Query Expansion* secara umum merupakan suatu teknik untuk menambahkan query pada information retrieval dalam teknik relevance feedback. Query awal akan ditambahkan dengan beberapa term atau kata pada query untuk memberikan kemudahan dalam proses *information retrieval*. Temu Kembali Informasi dimulai dengan penyediaan beberapa koleksi dokumen yang akan digunakan. Dengan menggunakan operasi teks akan diproses menjadi suatu berkas inverted index. Untuk menemukannya, penelitian ini menggunakan metode pembobotan TF-IDF dan *algoritme cosine similarity* berbasis wordNet. Dengan menggunakan wordNet dilakukan penambahan query untuk memperbaiki sebuah teks tertentu sehingga sesuai dengan konsep kalimat tertentu. Pada penelitian ini akan digunakan synset berupa relasi kata hiponim untuk ditambahkan kedalam query. Berdasarkan hasil pengujian menggunakan *precision@20* dari 10 query didapatkan nilai presisi rata-rata sebesar 0,7. Artinya probabilitas sistem dapat menemukan kembali dokumen yang relevan tanpa menggunakan *query expansion* adalah sebesar 70%. Berdasarkan hasil pengujian menggunakan *precision@20* dari 10 query didapatkan nilai presisi rata-rata sebesar 0,52. Artinya probabilitas sistem dapat menemukan kembali dokumen yang relevan tanpa menggunakan *query expansion* adalah sebesar 52%.

**Kata kunci:** *query expansion, sistem temu kembali informasi, TF-IDF, algoritme cosine, wordnet.*

### **Abstract**

*Query Expansion is generally a technique for adding queries in information retrieval in relevance feedback techniques. The initial query will be added with several terms or words in the query to facilitate the process of information retrieval. Information Retrieval begins with the provision of several collections of documents to be used. Using text operations will be processed into an inverted index file. To find it, this research uses TF-IDF weighting method and wordNet based cosine similarity algorithm. By using wordNet, a query is added to correct a particular text so that it matches the concept of a particular sentence. In this research will be used synset in the form of a hyponym word relation to be added to the query. Based on the results of testing using precision @ 20 from 10 queries, the average precision value was 0.7. This means that the probability of the system can rediscover the relevant documents without using the query expansion is 70%. Based on the results of testing using precision @ 20 from 10 queries obtained an average precision value of 0.52. This means that the probability of the system can rediscover the relevant documents without using the query expansion is 52%.*

**Keywords:** *query expansion, information retrieval system, TF-IDF, cosine algorithm, wordnet.*

## **1. PENDAHULUAN**

Perkembangan teknologi informasi pada dunia maya saat ini merupakan salah satu masalah utama dalam mengakses dan mencari informasi yang lebih berguna dan lebih selektif. Menurut (Warsita, 2008) mengenai

teknologi informasi yaitu sarana dan prasarana untuk memperoleh, menafsirkan, menyimpan, dan menggunakan data secara bermakna. Namun masalah dalam pencarian informasi tersebut lebih besar daripada sekedar mengakses informasi. Hal ini disebabkan oleh meningkatnya kebutuhan pengguna digital akan

pencarian informasi yang lebih cepat dan lebih efisien. Banyaknya layanan sistem informasi yang menawarkan pencarian informasi lebih cepat sehingga menjadi salah satu cara yang ditawarkan bagi pengguna digital.

Banyaknya layanan sistem informasi pada dunia maya, salah satunya yaitu Twitter. Twitter merupakan salah satu media sosial yang membantu penggunanya mendapatkan berbagai informasi melalui postingan berbagai orang yang disebut kicauan (tweets) dengan jumlah karakter maksimal sebanyak 200 kata. Pengguna Twitter dari tahun ke tahun di berbagai negara mengalami peningkatan. Pengguna yang telah menggunakan Twitter maka akan memiliki hak akses untuk mendapatkan berbagai informasi yang ada di Twitter. Contohnya pada pengguna yang ingin mendapatkan informasi dari salah satu akun maka pengguna tersebut harus menjadi pengikut (*follower*) pada akun tersebut.

Sistem Temu Kembali Informasi atau Information Retrieval System yaitu bagian dari Ilmu Komputer yang berisi tentang pengambilan informasi dari beberapa dokumen yang berdasarkan pada isi dan konteks dokumen itu sendiri. Penelitian yang dilakukan Gerald J. Kowalski dalam bukunya "Information Storage and Retrieval System Theory and Implementation", sistem temu kembali informasi yaitu merupakan suatu sistem yang dapat melakukan penyimpanan, pencarian dan pemeliharaan sebuah informasi. Informasi didalamnya berisi sebuah teks (data numerik dan tanggal), audio, video, gambar dan objek multimedia lainnya.

Sistem temu kembali informasi memiliki tujuan untuk memberikan sebuah informasi yang relevan bagi pengguna. Dokumen-dokumen yang relevan dipengaruhi oleh query yang diberikan pengguna. Sebagian pengguna kesulitan untuk merepresentasikan informasi ke dalam query yang sesuai untuk melakukan pencarian sehingga user tidak mendapatkan dokumen yang sesuai dengan keinginan pengguna. Permasalahan tersebut diperlukan sebuah solusi untuk memperbaiki query yang di masukkan oleh pengguna yaitu query expansion. Query expansion tersebut akan menghasilkan query yang lebih maksimal dalam sebuah sistem pencarian dan dapat meningkatkan kepuasan pengguna dalam pencarian dokumen. Ide yang digunakan yaitu mengubah query yang lebih relevan dengan

cara menambah, membuang dan mengubah bobot tiap kata pada kueri.

Pada penelitian sebelumnya mengenai query expansion salah satunya bersumber pada penelitian yang dilakukan oleh Pamungkas, Zanwar Yoga, (2015) dalam penelitiannya Query Expansion Pada Sistem Temu Kembali Informasi Dokumen Berbahasa Indonesia Menggunakan Pseudo Relevance Feedback. Hasilnya bahwa query expansion dapat berpengaruh terhadap nilai precision. Dari beberapa percobaan yang dilakukan, precision akan mengalami kenaikan nilai. Namun pada nilai recall cenderung tetap. Kenaikan nilai yang terjadi pada precision dan ketetapan nilai pada recall menyebabkan nilai f-measure meningkat. Hal ini disebabkan karena query semakin spesifik dan jumlah dokumen cenderung tetap. Penambahan jumlah kata akan menaikkan rangking dari dokumen yang relevan. Pada penelitian query expansion menggunakan pseudo relevance feedback menghasilkan kenaikan akurasi sebesar 0,17 dan kenaikan recall sebesar 0,03 dan kenaikan rata-rata sebesar 0,18. Namun, untuk pengembangan lebih lanjut tentang penelitian ini yaitu perlu diperhatikan dan pencarian metode lain dalam pemilihan kata baru sehingga kata yang di tampilkan dan dipilih oleh user semakin baik dan memiliki kesamaan yang tinggi dengan query awal user.

Pada penelitian yang dilakukan oleh (Rahayuni, 2011) dalam penelitiannya Ekspansi Kueri Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Menggunakan Thesaurus. Hasilnya bahwa ekspansi kueri dengan thesaurus dapat meningkatkan hasil temu kembali informasi. Panjang ekspansi kueri mempengaruhi hasil temu kembali informasi dan semakin panjang ekspansi kueri yang diberikan maka dokumen yang di-retrieve semakin tidak relevan. Penambahan kata thesaurus pada ekspansi kueri mempengaruhi hasil temu kembali. Pada evaluasi dokumen menggunakan nilai recall dan precision menggunakan ekspansi kueri menghasilkan temuan kembali dokumen yang relevan sebesar 100% dengan nilai *average precision* sebesar 0,076. Namun untuk pengembangan lebih lanjut mengenai penelitian ini diperlukan uji coba dengan menggunakan dokumen yang lebih beragam dan jumlah dokumen diperbanyak.

Terdapat berbagai macam metode yang sudah dilakukan sebelumnya dalam

menentukan query expansion seperti beberapa uraian dan penelitian yang telah dilakukan sebelumnya, maka penelitian ini akan membangun sebuah sistem untuk menemukan informasi yang lebih akurat untuk user dengan memanfaatkan hiponim sebagai kata kunci berbasis WordNet untuk memperkaya query pengguna. WordNet merupakan kamus dimana didalamnya terdapat sinonim set (synset) (Li, et al., 016). Istilah hiponim dalam WordNet yang memiliki kedekatan dengan query diharapkan dapat untuk membangun query yang baru. Hasil dari query ini diharapkan dapat menghasilkan temu kembali yang lebih baik dalam pencarian dokumen dengan apa yang diinputkan oleh pengguna.

## 2. TINJAUAN PUSTAKA

### 2.1. Query Expansion

Definisi Query Expansion secara umum merupakan suatu teknik untuk menambahkan query pada information retrieval dalam teknik relevance feedback. Query awal akan ditambahkan dengan beberapa term atau kata pada query untuk memberikan kemudahan dalam proses *information* retrieval. Teknik ini digunakan untuk memodifikasi query agar memenuhi kebutuhan informasi yang dibutuhkan oleh user. Proses penambahan query dengan mencari sinonim dalam bentuk unstemmed-term dari query. Menurut Manning et.al (2009) penambahan input pada koleksi dokumen yang diberikan dengan memberi tanda pada dokumen yang dianggap relevan. Selanjutnya input ini digunakan untuk memperbaiki query yang telah diberikan user. Dalam query expansion user memberikan input tambahan pada query awal berupa kata maupun frase.

### 2.2. Sistem Temu Kembali Informasi

Temu Kembali Informasi dimulai dengan penyediaan beberapa koleksi dokumen yang akan digunakan. Dengan menggunakan operasi teks akan diproses menjadi suatu berkas inverted index. Menurut Rijsbergen (1979) dalam (Rusidi,2008) Information Retrieval System tidak menyatakan bahwa terdapat masalah. Sistem tersebut hanya memberitahukan keterangan yang ada keterkaitannya dengan dokumen permintaan pengguna. Dalam proses awal temu kembali informasi, pengguna diminta merepresentasikan

ulang kebutuhan dalam bentuk query. Selanjutnya query akan diproses dengan operasi serupa seperti pada koleksi dokumen. Query yang telah diproses menghasilkan dokumen hasil temu kembali.

### 2.3. Wordnet

*WordNet* adalah sebuah kamus yang merupakan pengembangan dari kamus kata Thesaurus, dikembangkan oleh Universitas Princeton dan berisi pemodelan leksikal Bahasa Inggris. Didalam WordNet terdapat synset (synonym set), didalamnya terdapat kata – kata yang memiliki keterkaitan dan relasi semantic dalam berbagai kategori kata seperti kata benda, kata kerja, kata sifat, dan kata keterangan Zhang, et al. (2009). Dalam synset satu kata dengan kata lainnya dapat saling menggantikan tanpa merusak atau mengubah makna dari kalimat dalam satu konteks tertentu.

Dengan menggunakan wordNet yang merupakan sebuah kamus kata yang telah di-expert maka dapat dilakukan penambahan query untuk memperbaiki sebuah teks tertentu sehingga sesuai dengan konsep kalimat tertentu Zhang, et al. (2009) . Pada penelitian ini akan digunakan synset berupa relasi kata hiponim untuk ditambahkan ke dalam query.

### 2.4. Processing

Dokumen-dokumen yang ada kebanyakan tidak memiliki struktur yang pasti sehingga informasi di dalamnya tidak bisa diekstrak secara langsung dan tidak semua kata mencerminkan makna/isi yang terkandung dalam sebuah dokumen. Preprocessing diperlukan untuk memilih kata yang akan digunakan sebagai indeks. Indeks ini adalah kata-kata yang mewakili dokumen yang nantinya digunakan untuk membuat pemodelan untuk Information Retrieval maupun aplikasi text mining lain. Singkatnya Preprocessing adalah merubah teks menjadi term index yang bertujuan untuk menghasilkan sebuah set term index yang bisa mewakili dokumen. Ada beberapa langkah dalam menerapkan preprocessing yakni :

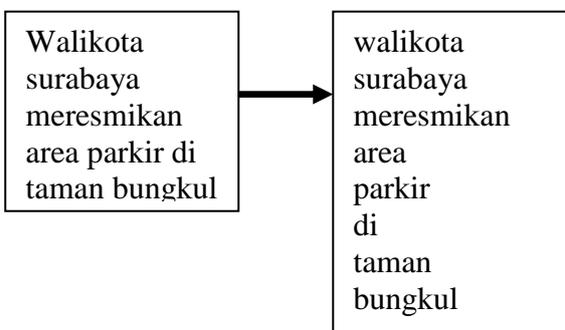
#### 1. Parsing

Proses awal disini dilakukan pemecahan struktur dokumen menjadi komponen yang terpisah, singkatnya adalah menentukan mana yang akan dijadikan satu dokumen. Contohnya buku dengan 500 halaman bisa dipecah menjadi 500 dokumen, dimana per

halaman dikelompokkan menjadi 1 dokumen.

2. Lexical Analysis atau Tokenisasi

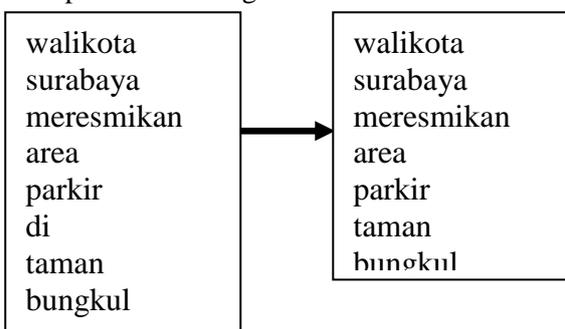
Pada proses ini dilakukan penghilangan angka, tanda baca dan karakter selain huruf alfabet, karena karakter-karakter tersebut dianggap sebagai pemisah kata (delimiter) dan tidak memiliki pengaruh terhadap pemrosesan teks. Serta perubahan semua kata menjadi kata dengan awalan huruf kecil. Proses tokenisasi akan digambarkan pada Gambar 1 di bawah ini.



Gambar 1. Contoh proses tokenisasi

3. Stopword Removal atau Filtering

Merupakan tahap pengambilan dari hasil token, yaitu kata-kata apa saja yang akan digunakan untuk merepresentasikan dokumen. Stoplist atau stopwords adalah kata-kata yang tidak deskriptif (tidak penting) yang dapat dibuang dengan pendekatan bag-of-words.



Gambar 2. Contoh proses filtering

4. Stemming

Proses pengubahan dari suatu kata menjadi kata dasar, dimana setiap kata yang berimbuhan akan berubah menjadi kata dasar. Proses ini nantinya tidak digunakan

karena berdasarkan penelitian sebelumnya. Fitur tanpa stemming memberikan akurasi lebih baik, ini menunjukkan bahwa pencari berita hoax ditentukan secara leksikal (Rasywir & Purwarianti, 2015). Leksikal adalah suatu makna yang nyata dalam kehidupan kita, jadi makna leksikal adalah arti sebenarnya yang dijelaskan oleh kata tersebut.

2.5. Pembobotan Term Frequency-inverse Document Frequency (TF-IDF)

Algoritme Term Frequency - Inverse Document Frequency (TF-IDF) merupakan algoritme yang digunakan untuk menghitung bobot suatu kata (term) masing-masing dokumen terhadap kata kunci. Term frequency (tf) adalah frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan, sedangkan Inverse document frequency (idf) merupakan suatu statistik yang mengkarakteristikan sebuah term dalam keseluruhan koleksi dokumen. Dan Document Frequency (DF) merupakan banyaknya dokumen dimana suatu kata tertentu muncul. TF merupakan pembobotan yang sederhana dimana penting tidaknya sebuah kata diasumsikan sebanding dengan jumlah kemunculan kata tersebut dalam dokumen, sementara IDF adalah pembobotan yang mengukur seberapa penting sebuah kata dalam dokumen. Nilai pembobotan TF x IDF akan tinggi jika nilai TF besar dan kata yang diamati tidak ditemukan di banyak dokumen. Maka untuk mencari nilai TF-IDF menggunakan Persamaan 2.1 :

- a. Term Frequency  
Term Frequency (TF)

$$W_{tf,t,d} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.1)$$

Keterangan :

$W_{tf,t,d}$  = Frekuensi kemunculan kata t dalam dokumen d

- b. Inverse Document Frequency (IDF)  
 $idf_t = \log_{10} N/df_t \quad (2.2)$

Keterangan:

$df_t$  = Banyak dokumen yang memuat

$N$  = jumlah total dokumen

c. TF.IDF Weighting

$$W_{t,d} = W_{tf,t,d} \times idf_t \quad (2.3)$$

d. Normalisasi

$$W_{t,d} = \frac{W_{t,d}}{\sqrt{\sum_{t=1}^n W_{t,d}^2}} \quad (2.4)$$

### 2.6. Ukuran Kemiripan

Pada pengukuran kemiripan dokumen yaitu menggunakan cosine similarity. Pengukuran ini digunakan untuk menghitung nilai sudut cosinus antara dua vektor yang merepresentasikan dokumen dan query. Nilai cosinus antara dokumen dan query didefinisikan sebagai berikut:

Tanpa normalisasi:

$$CosSim(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^t (W_{ij} \cdot W_{iq})}{\sqrt{\sum_{i=1}^t W_{ij}^2 \cdot \sum_{i=1}^t W_{iq}^2}} \quad (2.5)$$

Dengan normalisasi berdasarkan persamaan sebelumnya:

$$CosSim(d_j, q) = \vec{d}_j \cdot \vec{q} = \sum_{i=1}^t (W_{ij} \cdot W_{iq}) \quad (2.6)$$

Keterangan :

- $d_j$  = Data Latih
- $q$  = Data Uji
- $W_{ij}$  = Nilai pembobotan kata pada dokumen latih
- $W_{iq}$  = Nilai pembobotan kata pada dokumen uji

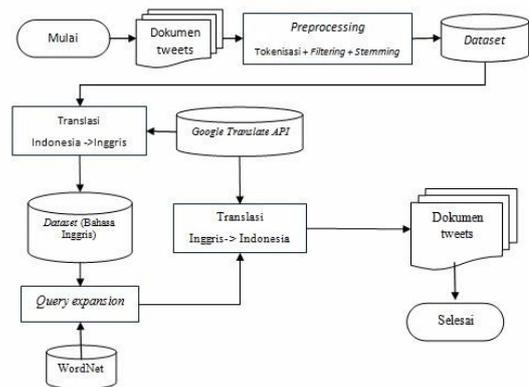
## 3. METODE PENELITIAN

### 3.1. Dataset

Pengumpulan data didapatkan dari akun media sosial Twitter. Data berupa tweets yang ada dalam situs tersebut. Nantinya tweet tersebut dilakukan proses query expansion dan dokumen relevan maupun tidak yang telah ditentukan oleh pakar. Pakar berasal dari LSO Display Filkom Universitas Brawijaya yang melakukan penilaian terhadap dokumen -dokumen yang ditampilkan dalam suatu query yang diinputkan.

### 3.2. Perancangan Algoritme

Pada tahap ini merupakan tahap perancangan algoritma yang akan digunakan dalam penelitian ini. Akan dijabarkan alur dan proses serta langkah – langkah dari algoritma tersebut. Selain itu pada tahap perancangan algoritma menggambarkan jalannya algoritma yang digunakan.



Gambar 3. Perancangan Algoritme

### 3.3. Teknik Pengujian dan Analisis

Pada tahap ini adalah menguji Algoritme yang telah dirancang untuk penelitian ini. Algoritme klasifikasi akan diterapkan ke dalam data uji yang telah ada sebelumnya. Tujuan dari pengujian adalah menemukan kesalahan yang mungkin terjadi dalam proses klasifikasi. Setelah didapatkan hasil pengujian kemudian dilanjutkan dengan analisis terhadap hasil yang didapat, bagaimana sebuah kondisi terjadi dan bagaimana cara mengatasi kesalahan yang muncul. Tujuan analisis juga melihat apakah penelitian sudah mencapai tujuan yang diinginkan.

Pengujian yang akan dilakukan pada penelitian ini ada dua jenis pengujian:

- a. Pengujian akurasi sistem Query Expansion pada sistem temu kembali informasi berbahasa indonesia dengan metode pembobotan tf-idf dan algoritma Cosine Similarity menggunakan thesaurus.
- b. Pengujian Treshold pada jumlah kata yang diekspansi terhadap akurasi sistem.

## 4. PENGUJIAN DAN PEMBAHASAN

Dalam penelitian ini akan digunakan 3 jenis pengujian, yakni pengujian dengan Precision at Kuntuk menentukan apakah

query yang dimasukkan relevan terhadap dokumen yang ditemukan kembali sebanyak *top K*, lalu dihitung tingkat akurasi. Pengujian menggunakan *query expansion* menghasilkan relevansi antara dokumen dan query yang telah di *query expansion* dilakukan secara manual terhadap 20 penilaian, masing-masing penilaian memasukkan query dan menilai dokumen yang relevan hingga urutan ke-20. Pengujian tanpa menggunakan *query expansion* menghasilkan relevansi antara dokumen dan query tanpa di *query expansion* dilakukan secara manual terhadap 20 penilaian.

**4.1. Evaluasi dengan Precision at K**

*Precision at K* digunakan dalam evaluasi untuk menghitung tingkat akurasi sistem. Pengujian ini dilakukan untuk menentukan apakah query yang dimasukkan relevan terhadap dokumen yang ditemukan kembali sebanyak *top K*, lalu dihitung tingkat akurasi. Jumlah dokumen yang akan dicoba sebanyak 376 dokumen dan jumlah *query*-nya sebanyak 10 *query*. Daftar *query* dan *query expansion* yang telah di *query expansion* akan diuji tertera pada Tabel 6.1 dan daftar dokumennya terdapat pada lampiran A. Top K yang akan diuji untuk tiap *query* adalah *top 20*.

**Tabel 1**  
**Daftar Query dan Query Expansion**

No	ID Query	Query	Query Expansion
1	Q-01	Reptil	Anaspid, Diaspid, Sina spid, Diaspida
2	Q-02	Kendaraan	bumper car, kerajinan militer vehicle, roket, skibob, kereta luncur, penggilingan, wheeled vehicle
3	Q-03	Burung	burungmalam, bird of passage, protoavis, Archaeopteryx, Sinomis, Ibero-

4	Q-04	Gunung	Black Hills, puncak gunung, ben, gunung bawah laut, gunung berapi
5	Q-05	Manusia	dunia, Homo erectus, Homo soloensis, Homo habilis, Homo sapiens, Neandertal man, Rhodesian man
6	Q-06	Memasak	koki, Kue kering, fry cook, preserver, roaster, bumbu, Petani
7	Q-07	Serangga	ephemeron, holometabola, defoliator, penyerbuk, gallfly, mecopteran, collembolan, proturan, kumbang, web spinner, kutu, bird louse, kutu, dipterous insect, leaf miner, hymenopterous insect, pekerja, rayap, orthopterous insect, phasmid, dictyopterous insect, bug, hemipterous insect, heteropterous insect, homopterous insect, psocopterous insect, ephemerid, stonefly, neuropteran, odonate, trichopterous insect, thysanuran insect, thysanopter, earwig, lepidopterous insect, kepompong, imago, ratu
8	Q-08	Kerja	mencuci, tindakan, pekerjaan, pekerjaan, operasi, layanan, bersinar, angkat berat, rumah tangga, pekerjaan rumah, menyetrika, pekerjaan yang sibuk, penebangan, loose end, kerja malam, dokumen, welfare work, tenaga kerja, subbing, penyelidikan, peduli, tugas, misi, pekerjaan persiapan, waktu kerja, usaha, kursus, mengikuti, kerajinan, besi, lacework, vemis, pekerjaan kulit, hasil karya, kerja logam, kerawang, polikrom, pekerjaan perak,

9	Q-09	Hewan	hama, makhluk, creepy-crawly, pelembar anak panah, pengintip, homeotherm, poikilotherm, range animal, anak nakal, pemulung, work animal, binatang dalam negeri, pengumpan, migrator, penganiaya, membelai, orang tabah, pengganti, marine animal, wanita, pria, dewasa, muda, berdarah mumi, raksasa, selamat, mutan, herbivora, insektivora, acrodont, pleurodont, zooplankton, embrio, chordate, invertebrata, metazoan, omnivora, predator, mangsa, permainan, hexapod, berkaki dua, larva, pembalap, fiksional animal, tawanan, pasangan
10	Q-10	Makanan	kuning telur, comfort food, panganan, tarif, bahan makanan, makanan bergizi, komisariat, makan, miraculous food, minuman, air, makanan jiwa, chyme, sisa, makanan segar, kenyamanan food, cokelat, baked goods, daging, Semacam spageti, makanan sehat, makanan cepat saji, makanan sarapan, menghasilkan, kelapa, dika bread, ikan, makanan laut, mentega, yogurt, keju, airtotor, makan

**4.2. Perbandingan Pengujian Tanpa Menggunakan Query Expansion Dan Menggunakan Query Expansion**

Pengujian menggunakan *query expansion* menghasilkan relevansi antara dokumen dan

query yang telah di query expansion dilakukan secara manual terhadap 20 penilaian, masing-masing penilaian memasukkan query dan menilai dokumen yang relevan hingga urutan ke-20. Sedangkan pengujian tanpa menggunakan *query expansion* juga menghasilkan relevansi antara dokumen dan query tanpa di query expansion dilakukan secara manual terhadap 20 penilaian, masing-masing penilaian memasukkan query dan menilai dokumen yang relevan hingga urutan ke-20. Hasil pengujiannya pada Tabel 2.

**Tabel 2.**  
**Hasil Pengujian dengan Precision @K**

Query	Dengan Query Expansion	Tanpa Query Expansion
Q-1	P@20=16/20=0.8	P@20=15/20=0.75
Q-2	P@20=13/20=0.6	P@20=15/20=0.75
Q-3	P@20=16/20=0.8	P@20=11/20=0.55
Q-4	P@20=10/20=0.5	P@20=11/20=0.55
Q-5	P@20=12/20=0.6	P@20=13/20=0.65
Q-6	P@20=14/20=0.7	P@20=8/20=0.4
Q-7	P@20=16/20=0.8	P@20=10/20=0.5
Q-8	P@20=15/20=0.75	P@20=5/20=0.25
Q-9	P@20=14/20=0.7	P@20=8/20=0.4
Q-10	P@20=14/20=0.7	P@20=8/20=0.4
Rata-rata	0.7 = 70%	0.52 52%

### 4.3. Skenario Pengujian.

Pada skenario ini dilakukan beberapa pengujian pada *Precision@20 Precision@15, Precision@10, Precision@5*. Dimana pengujian ini digunakan untuk mengukur *threshold* pada peringkat K. Query yang digunakan hanya query Q-01. Dimana Q-01 telah di query expansion. Dokumen yang dihitung ialah dokumen relevan teratas sejumlah K dan mengabaikan dokumen peringkat yang berada dibawah K. Berikut hasil pengujian *Precision@K* yang di tunjukkan pada tabel 3.

**Tabel 3.**  
**Skenario pengujian 2**

	Query	Query+1	Query+2	Query+3	Query+4
P@20	0.75	0.65	0.65	0.7	0.8
P@15	0.733	0.6	0.66	0.733	0.933
P@10	0.9	0.9	0.9	0.9	0.9
P@5	0.8	0.8	0.8	0.8	1

### 5. KESIMPULAN

Berdasarkan pada hasil penelitian menerapkan Query Expansion pada sistem temu kembali informasi berbahasa indonesia dengan metode pembobotan tf-idf dan algoritma Cosine Similarity berbasis WordNet serta hasil pengujian dan performa dari hasil Query Expansion pada sistem temu kembali informasi berbahasa indonesia dengan metode pembobotan tf-idf dan algoritma Cosine Similarity berbasis WordNet, maka dapat ditarik beberapa kesimpulan.

Berdasarkan hasil perbandingan pengujian menggunakan *precision@20* dari 10 query didapatkan nilai presisi rata-rata sebesar 0,7. Artinya probabilitas sistem dapat menemukan kembali dokumen yang relevan tanpa menggunakan query expansion adalah sebesar 70%. Sedangkan hasil pengujian menggunakan *precision@20* dari 10 query didapatkan nilai presisi rata-rata sebesar 0,52. Artinya probabilitas sistem dapat menemukan kembali dokumen yang relevan tanpa menggunakan query expansion adalah sebesar 52%.

Untuk penggunaan *WordNet* berbahasa inggris kurang efektif saat dataset berbahasa indonesia karena perlu melakukan proses *translat*, maka disarankan untuk menggunakan *WordNet* berbahasa indonesia. Untuk menghasilkan kata ekspansi yang lebih

beragam dan akurat maka disarankan untuk mencoba menggunakan relasi antar kata selain hiponim.

### 6. DAFTAR PUSTAKA

Manning, C. D., Raghavan, P. & Schutze, H., 2009. Introduction to Information Retrieval. Online Edition penyunt. Cambridge, England: Cambridge University Press.

Gerald, J. Kowalski, 2000. Information Storage and Retrieval Systems: Theory and Implementation, United States

Nutri Rahayuni 2011. Ekspansi Kueri pada Sistem Temu Kembali Informasi Berbahasa Indonesia Menggunakan Thesaurus, Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.

Pamungkas, Z.Yoga., Indriati., Ridok, A, 2015. Query Expansion pada Sistem Temu Kembali Informasi Dokumen Berbahasa Indonesia Menggunakan Pseudo Relevance Feedback, Informatika, Universitas Brawijaya.

Rasywir, Errissya dan Ayu Purwarianti. 2015. Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. Jurnal Cybermatika, (2)3, pp. 1-8.

Rusidi. 2008. Ekspansi Kueri dalam Sistem Temu Kembali Informasi Berbahasa Indonesia Menggunakan Peluang Bersyarat. [Skripsi]. Bogor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.

Warsita, Bambang. 2008. Teknologi Pembelajaran : Landasan & Aplikasinya, Jakarta: Rineka

Zhang, J., Deng, B., & Li, X. (2009). Concept Based Query Expansion Using WordNet. Beijing: Dept. Electronic Engineering, Tsinghua Univ. Beijing, 100084, China.