



# **TUNING PARAMETER DALAM BM25 SIMILARITY PADA TEMU KEMBALI PENYAKIT AKASIA**

**AGITHA MINA SARI**



**DEPARTEMEN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2017**

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural



## PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul *Tuning Parameter* dalam BM25 *Similarity* pada Temu Kembali Penyakit Akasia adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, September 2017

*Agitha Mina Sari*  
NIM G64130035

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

## ABSTRAK

AGITHA MINA SARI. *Tuning Parameter* dalam BM25 *Similarity* pada Temu Kembali Penyakit Akasia. Dibimbing oleh JULIO ADISANTOSO dan YENI HERDIYENI.

Akasia merupakan tanaman hutan yang sering digunakan untuk pembuatan pulp. Penyakit tanaman seringkali menjadi kendala dalam memelihara hutan tersebut. Informasi yang tersedia saat ini masih berupa dokumen tertulis. Oleh karena itu, diperlukan suatu temu kembali informasi untuk pencarian dokumen penyakit tersebut. Temu kembali informasi dikembangkan dalam beragam model, salah satunya model peluang. Penelitian ini menggunakan model peluang dengan fungsi kemiripan BM25 yang memiliki suatu variabel yang dapat diubah nilainya, yang disebut dengan *tuning parameter*. Modifikasi nilai dari *tuning parameter* ini bertujuan untuk meningkatkan kinerja hasil pencarian dengan fungsi kemiripan BM25. Modifikasi nilai dari *tuning parameter* meningkatkan nilai rata-rata *normalized discounted cumulative gain* (NDCG) dari hasil pencarian 20 *query* pendek yang terdiri atas dua sampai empat term, yang pada awalnya sebesar 0.935236 menjadi 0.941762.

Kata Kunci: akasia, BM25, *information retrieval*, *tuning parameter*

## ABSTRACT

AGITHA MINA SARI. *Tuning Parameter* of BM25 *Similarity* in Information Retrieval *Acacia* Diseases. Supervised by JULIO ADISANTOSO and YENI HERDIYENI.

*Acacia* is a forest plants that is often used for wood pulp. Plant diseases are often be an obstacle in maintaining the forest. The information currently available is still a written document. Therefore, an information retrieval is needed to search for the disease documents. Information retrievals are developed in a variety of models, one of which is the opportunity model. In this research, we use opportunity model with BM25 similarity function which has a variable that can be changed its value, which is called tuning parameter. Modification of the value of this parameter tuning aims to improve performance with BM25 similarity function. The modification of the value of the parameter tuning increases the average value of normalized discounted cumulative gain (NDCG) from searching result of 20 short queries consisting of two to four terms from 0.935236 to 0.941762.

Keywords: *Acacia*, BM25, *information retrieval*, *tuning parameter*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

## **TUNING PARAMETER DALAM BM25 SIMILARITY PADA TEMU KEMBALI PENYAKIT AKASIA**

**AGITHA MINA SARI**

Skripsi  
sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer  
pada  
Departemen Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2017**



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural U

Pengujian Dr Imas Sukaesih Sitanggang, SSi MKom



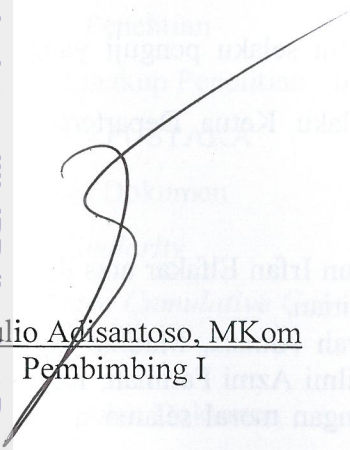
Judul Skripsi: *Tuning Parameter* dalam BM25 *Similarity* pada Temu Kembali Penyakit Akasia  
Nama : Agitha Mina Sari  
NIM : G64130035

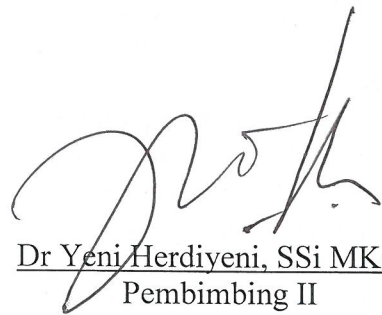
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.


© Hak cipta milik IPB (Institut Pertanian Bogor)

Disetujui oleh

  
Ir. Julio Adisantoso, MKom  
Pembimbing I

  
Dr. Yeni Herdiyeni, SSi MKom  
Pembimbing II

Diketahui oleh

  
Dr. Ir. Agus Buono, MSi MKom  
Ketua Departemen



Tanggal Lulus:

08 SEP 2017'

Bogor Agricultural



## PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Februari 2016 ini ialah temu kembali informasi, dengan judul *Tuning Parameter* dalam *BM25 Similarity* pada Temu Kembali Penyakit Akasia. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ibunda Siti Aminah, ayahanda Agus Dwianto, kakak, adik, dan seluruh keluarga atas segala doa dan kasih sayangnya.
2. Bapak Ir Julio Adisantoso, MKom dan Ibu Dr Yeni Herdiyeni, SSi MKom selaku pembimbing yang telah memberikan bimbingan, saran, dan arahan dalam melakukan penelitian ini.
3. Ibu Dr Imas Sukaesih Sitanggang, SSi MKom selaku penguji yang telah memberikan saran dalam penyusunan skripsi ini.
4. Bapak Dr Ir Agus Buono, MSi MKom selaku Ketua Departemen Ilmu Komputer.
5. Seluruh dosen Departemen Ilmu Komputer.
6. Seluruh Staf Departemen Ilmu Komputer.
7. Sella Monika, Muhammad Alvian Supriadi, dan Irfan Elfakar atas dukungan dan kerjasamanya sebagai rekan satu tim penelitian.
8. Nurwasilah, Dewi Asiah Shofiana, Siti Syarah Annisa, Maulita Agustina, Agrippina, Fitri Yanti, Gishella Erdyaning, Hilmi Azmi Fatimah, Keke Putri Utami, dan Shafira Fanni atas doa dan dukungan moral selama menempuh pendidikan di Departemen Ilmu Komputer.
9. Seluruh rekan-rekan Ilmu Komputer angkatan 50.

Semoga karya ilmiah ini dapat memberikan manfaat.

Bogor, September 2017

*Agitha Mina Sari*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.





- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## DAFTAR ISI

DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
DAFTAR LAMPIRAN	vi
PENDAHULUAN	1
Latar Belakang	1
Perumusan Masalah	1
Tujuan Penelitian	2
Manfaat Penelitian	2
Ruang Lingkup Penelitian	2
TINJAUAN PUSTAKA	2
Praproses Dokumen	3
BM25 <i>Similarity</i>	3
<i>Discounted Cumulative Gain (DCG)</i>	4
METODE	5
Pengumpulan Dokumen	5
Praproses Dokumen	6
<i>Similarity</i>	6
Evaluasi	6
HASIL DAN PEMBAHASAN	7
Praproses Dokumen	7
<i>Similarity</i>	10
Evaluasi	13
SIMPULAN DAN SARAN	20
Simpulan	20
Saran	20
DAFTAR PUSTAKA	20
LAMPIRAN	22
RIWAYAT HIDUP	36



## DAFTAR TABEL

1	Nilai <i>tuning parameter</i> yang diujikan	7
2	Kombinasi awalan akhiran yang tidak diizinkan dalam algoritme <i>stemming</i> Nadief dan Andriana (Agusta 2009)	9
3	Cara menentukan tipe awalan untuk kata yang diawali dengan “te-” dalam algoritme <i>stemming</i> Nadief dan Andriana (Agusta 2009)	10
4	Jenis awalan berdasarkan tipe awalannya dalam algoritme <i>stemming</i> Nadief dan Andriana (Agusta 2009)	10
5	Contoh hasil pembobotan dengan TF	10
6	Contoh urutan dokumen hasil pencarian pada Tuning	13
7	Hasil normalisasi nilai DCG pada sebelum dan setelah <i>tuning</i>	14
8	Nilai NDCG setiap <i>query</i> pada nilai parameter $b$ dan nilai $k_1 = 0.2$	15
9	Perbedaan hasil evaluasi pada BM25 <i>default</i> dan BM25 <i>tuning</i> pada <i>query</i> yang terdiri atas dua term "akar menebal"	17
10	Perbedaan hasil evaluasi pada BM25 <i>default</i> dan BM25 <i>tuning</i> pada <i>query</i> yang terdiri atas tiga term "batang berwarna putih"	18
11	Perbedaan hasil evaluasi pada BM25 <i>default</i> dan BM25 <i>tuning</i> pada <i>query</i> yang terdiri atas empat term "daun tertutup kerak hitam"	19

## DAFTAR GAMBAR

1	Tahapan penelitian	6
2	Struktur fail XML	8
3	Simulasi pengaruh nilai DF terhadap IDF pada $N = 14$	11
4	Simulasi pengaruh parameter $k_1$ terhadap nilai TF modifikasi ( $b = 0$ )	11
5	Simulasi pengaruh parameter $b$ terhadap nilai modifikasi TF ( $k_1=1.2$ )	12
6	Laju nilai rata-rata NDCG pada parameter $k_1$ ( $b=0.75$ )	14

## DAFTAR LAMPIRAN

1	Nilai relevansi dokumen pada tiap <i>query</i> (rel <sub>i</sub> )	22
2	Daftar <i>stopwords</i> Bahasa Indonesia	23
3	Term hasil proses <i>stemming</i>	28
4	Daftar panjang dokumen	30
5	Nilai NDCG pada setiap <i>tuning</i>	31

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## PENDAHULUAN

### Latar Belakang

Akasia merupakan salah satu tanaman hutan yang tumbuh di daerah Asia Pasifik. Tidak jarang tanaman ini dijadikan sebagai tanaman program pengembangan hutan oleh negara-negara di Asia Pasifik. Selain itu, tanaman ini juga merupakan salah satu jenis tanaman yang ditanam pada Hutan Tanaman Industri (HTI) untuk keperluan pembuatan pulp. Sekitar 1.3 juta hektar hutan tanaman akasia dibangun di Indonesia dengan tujuan produksi kayu pulp (Dephut 2003). Luas area hutan tanaman akasia di Indonesia mencapai 67% dari total luas area hutan tanaman akasia di dunia (FAO 2002). Salah satu tantangan dalam mengelola HTI tersebut adalah penyakit tanaman yang sering menyerang. Penyakit tanaman akasia dapat disebabkan oleh berbagai faktor, seperti hama, bakteri, dan serangga. Gejala dan area penyerangan penyakitnya pun berbeda-beda, penyakit tanaman akasia dapat menyerang pada bagian daun, batang, maupun akar (Irianto 2006). Saat ini informasi mengenai penyakit akasia masih dalam bentuk dokumen tertulis. Oleh sebab itu dibutuhkan suatu temu kembali informasi untuk melakukan pencarian dokumen tersebut.

Temu Kembali Informasi (TKI) merupakan proses yang berkaitan merepresentasikan, menyimpan, dan menemukan informasi dokumen yang sesuai dengan *query* (Sagayam *et al.* 2012). Saat melakukan pencarian dokumen digunakan kemiripan (*similarity*) antara kata pencarian yang diberikan oleh pengguna dengan setiap dokumen yang ada. Salah satu fungsi kemiripan yang digunakan dalam mesin pencarian adalah BM25. BM25 merupakan suatu fungsi yang menggabungkan atribut dokumen, seperti frekuensi global, frekuensi term, dan panjang dokumen (Jones *et al.* 2000). Pada penelitian Saputra (2013) mengenai *tuning parameter* dalam fungsi OKAPI BM25 dalam mesin pencari dokumen berbahasa Indonesia, dibandingkan kinerja fungsi kemiripan pada model peluang (BM25 *similarity*) dengan model ruang vektor (*cosine similarity*). BM25 memiliki nilai *Average Precision* (AVP) yang lebih tinggi, yaitu 0.5885 dibandingkan dengan model ruang vektor, yaitu 0.5327.

Fungsi kemiripan BM25 memiliki suatu variabel yang dapat diubah-ubah nilainya sesuai dengan kebutuhan dengan tujuan untuk mendapatkan hasil pencarian yang lebih baik, yaitu *tuning parameter*. Penelitian ini dilakukan untuk menguji apakah pencarian dengan BM25 dapat menghasilkan banyak dokumen yang relevan. Selain itu dibuktikan juga pengaruh dari perubahan *tuning parameter* yang akan dimodifikasi sedemikian rupa untuk mendapatkan hasil pencarian dengan nilai akurasi yang lebih baik.

### Perumusan Masalah

Rumusan permasalahan pada penelitian ini adalah:

1. Apakah fungsi kemiripan BM25 dapat mengukur relevansi secara akurat antara *query* masukan dengan dokumen yang dibutuhkan sehingga hasil pencarian dokumen penyakit sesuai dengan keinginan *user*?



2. Apakah modifikasi dari nilai *tuning parameter* dalam fungsi kemiripan BM25 dapat menghasilkan pencarian dokumen penyakit dengan hasil evaluasi yang lebih baik?

### Tujuan Penelitian

Penelitian ini bertujuan untuk menguji seberapa besar pengaruh modifikasi nilai dari *tuning parameter* yang ada dalam fungsi kemiripan BM25 terhadap evaluasi dari hasil pencarian dokumen penyakit akasia dalam bahasa Indonesia.

### Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk memudahkan pengembangan sistem temu kembali penyakit akasia dengan hasil modifikasi nilai *tuning parameter* BM25 *similarity* yang telah dilakukan pada penelitian ini.

### Ruang Lingkup Penelitian

Lingkup dari penelitian ini, yaitu:

1. Terdapat 14 penyakit akasia yang digunakan.
2. Dokumen yang digunakan merupakan dokumen berbahasa Indonesia.

## TINJAUAN PUSTAKA

### Sistem Temu Kembali Informasi (TKI)

Sistem temu kembali informasi merupakan suatu sistem yang menemukan informasi yang sesuai dengan kebutuhan pengguna dari kumpulan informasi secara otomatis. Prinsip kerja sistem temu kembali informasi adalah memberikan jawaban dari sekumpulan dokumen untuk menjawab pertanyaan (*query*) yang diberikan oleh pengguna. Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan (Salton 1989). Sistem Temu Kembali Informasi bertujuan untuk menjawab kebutuhan informasi user dengan sumber informasi yang tersedia dalam kondisi seperti sebagai berikut (Salton 1989):

1. Mempresentasikan sekumpulan ide dalam sebuah dokumen menggunakan sekumpulan konsep.
2. Terdapat beberapa pengguna yang memerlukan ide, tapi tidak dapat mengidentifikasikan dan menemukannya dengan baik.
3. Sistem temu kembali informasi bertujuan untuk mempertemukan ide yang dikemukakan oleh penulis dalam dokumen dengan kebutuhan informasi pengguna yang dinyatakan dalam bentuk *keyword query* atau istilah penelusuran.

## Praproses Dokumen

Praproses merupakan tahap awal dalam temu kembali informasi, salah satunya adalah *indexing*. *Indexing* merupakan proses persiapan yang dilakukan terhadap dokumen agar dokumen siap untuk diproses (Singh dan Saini 2014). Proses yang dilakukan dalam *indexing* ada empat tahap sebagai berikut.

- 1 Tokenisasi merupakan proses pemisahan suatu rangkaian karakter berdasarkan karakter spasi, dan mungkin pada waktu yang bersamaan dilakukan juga proses penghapusan karakter tertentu, seperti tanda baca. Token seringkali disebut sebagai istilah (*term*) atau kata, sebagai contoh sebuah token merupakan suatu urutan karakter dari dokumen tertentu yang dikelompokkan sebagai unit semantik yang berguna untuk diproses (Salton 1989).
- 2 Penghapusan *stopword* merupakan proses penghapusan term yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Penghapusan *stopwords* memiliki banyak keuntungan, yaitu akan mengurangi *space* pada tabel *term index* hingga 40% atau lebih (Baeza dan Neto 1999).
- 3 *Stemming* adalah proses pemotongan kata menjadi kata dasar. Proses *stemming* dilakukan dengan cara menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi bahasa Indonesia yang benar (Tala 2003).
- 4 Pembobotan merupakan proses pemberian bobot atau nilai pada term yang ada pada dokumen. Proses pemberian bobot didasarkan pada fungsi pembobotan yang digunakan, misalnya *term frequency* (TF). TF menghitung jumlah kemunculan setiap term  $t$  dalam sebuah dokumen  $d$ . Hasil dari pembobotan akan digunakan dalam perhitungan fungsi kemiripan.

### BM25 Similarity

Salah satu fungsi kemiripan yang digunakan dalam sistem temu kembali adalah BM25. BM25 merupakan hasil dari percobaan beberapa variasi fungsi *Best Match* pada model peluang. Pembentukan fungsi BM25 berawal dari masalah panjang dokumen yang beragam. Penggunaan fungsi Okapi pada sistem pencarian dasar dalam *Text Retrieval Conference* (TREC) telah mengalami banyak perubahan. TREC merupakan serangkaian *workshop* mengenai *Information retrieval* pada area penelitian yang berbeda-beda, salah satunya mengenai fungsi kemiripan Okapi BM25. Pada TREC-1 sampai dengan TREC-8 terdapat banyak perubahan pada fungsi Okapi. Pada TREC-1, fungsi Okapi dalam sistem pencarian tidak memberikan akurasi yang baik karena pembobotan dalam fungsi tersebut tidak memperhatikan aspek panjang dokumen di dalam nilai *term frequency* (TF). Oleh karena itu terbentuklah fungsi BM25 yang memiliki fungsi dengan 3 prinsip pembobotan yang baik, yaitu memiliki *inverse document frequency* (IDF), TF, dan memiliki fungsi normalisasi dari panjang dokumen (*document length normalization*) (Robertson dan Walker 1999).

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

TF merupakan jumlah kemunculan suatu term pada suatu dokumen. IDF dapat diartikan sebagai nilai kepentingan dari sebuah term. Semakin jarang sebuah term muncul dalam koleksi dokumen atau dengan kata lain jumlah dokumen (*document frequency*) yang mengandung term tersebut sedikit, maka term tersebut akan semakin penting, begitu pula sebaliknya. Nilai DF akan di-*inverse* untuk memberikan bobot yang besar pada term langka dan memberikan bobot yang lebih kecil untuk term yang sering muncul dalam koleksi yang disebut dengan bobot IDF. IDF dihitung dengan membagi jumlah koleksi dokumen dengan DF dari sebuah term. Jadi semakin kecil nilai DF dari suatu term, bobot IDF term tersebut akan semakin besar. Kemudian akan dihitung nilai logaritma (log) dari IDF untuk memperkecil efek dari IDF (Manning *et al.* 2009). Fungsi kemiripan BM25 adalah sebagai berikut:

$$BM25 = \sum_{t \in q} \left( \log \frac{N}{df_t} \right) \cdot \frac{(k_1 + 1)tf_{td}}{k_1((1-b) + b \times (L_d / L_{ave})) + tf_{td}} \cdot \frac{(k_3 + 1)tf_{tq}}{k_3 + tf_{tq}} \quad (1)$$

dengan  $q$  adalah *query*,  $N$  adalah jumlah dokumen dalam korpus,  $df_t$  adalah jumlah dokumen yang mengandung term  $t$ ,  $tf_{td}$  adalah jumlah term  $t$  yang muncul pada dokumen  $d$ ,  $tf_{tq}$  adalah jumlah term  $t$  yang muncul pada *query*  $q$ ,  $L_d$  adalah panjang dokumen  $d$ ,  $L_{ave}$  adalah rata-rata panjang dokumen, dan  $k_1$ ,  $k_3$ ,  $b$  adalah *tuning parameter* (Manning *et al.* 2009).

Keberhasilan dalam BM25 adalah modifikasi nilai TF dan parameter bebas dalam formulanya ( $k_1$  dan  $b$ ) (Lv dan Zhai 2011). Adapun nilai *tuning parameter* yang direkomendasikan oleh Robertson dan Walker (1999) yang telah terbukti efektif dan memberikan keakuratan yang baik yaitu:  $k_1 = 1.2$ ;  $k_3 = 1000$ ;  $b = 0.75$ . Sebelumnya nilai tersebut juga telah digunakan pada penelitian Saputra (2013) dan menghasilkan nilai akurasi yang cukup tinggi. Menurut Jones *et al.* (2000), semakin besar parameter  $k_1$  meningkatkan nilai modifikasi TF hampir secara linear. Selain itu, Jones *et al.* (2000) juga menyatakan bahwa parameter  $b$  sangat berkaitan dengan panjang dokumen yang cukup mempengaruhi *score* pada BM25 melalui modifikasi nilai TF.

### Discounted Cumulative Gain (DCG)

Dalam sistem temu kembali, *discounted cumulative gain* (DCG) digunakan untuk menghitung efektivitas dari algoritme temu kembali berdasarkan posisi dokumen di hasil pencarian (Jarvelin dan Kekalainen 2002). Nilai DCG semakin mendekati 1 jika hasil pengurutan dokumen semakin relevan terhadap *query* yang diberikan. Berikut merupakan formula dari akumulasi nilai relevansi urutan dari posisi  $i$  sampai  $p$ :

$$DCG_p = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2(i)} \quad (2)$$

dengan  $rel_i$  adalah nilai relevansi urutan posisi  $i$  dan  $p$  adalah jumlah dokumen hasil pencarian. Sebagai contoh sebuah pencarian menghasilkan enam dokumen,

kemudian dilakukan penilaian relevansi dari enam dokumen tersebut dengan nilai 0 sampai dengan 3. Nilai 3 berarti dokumen tersebut sangat relevan. Nilai 1 sampai 2 berarti "berada di tengah-tengah", dan nilai 0 berarti tidak relevan. Kemudian diperoleh nilai DCG sebesar 8.10 dengan perhitungan menggunakan Persamaan (2).

Nilai yang dihasilkan akan berbeda karena pada setiap pencarian bisa menghasilkan keluaran jumlah dokumen yang berbeda. Oleh karena itu, hasil dari DCG harus dinormalisasikan menggunakan perhitungan normalisasi DCG (NDCG) dengan formula sebagai berikut:

$$NDCG_p = \frac{DCG_p}{IDCG_p}$$

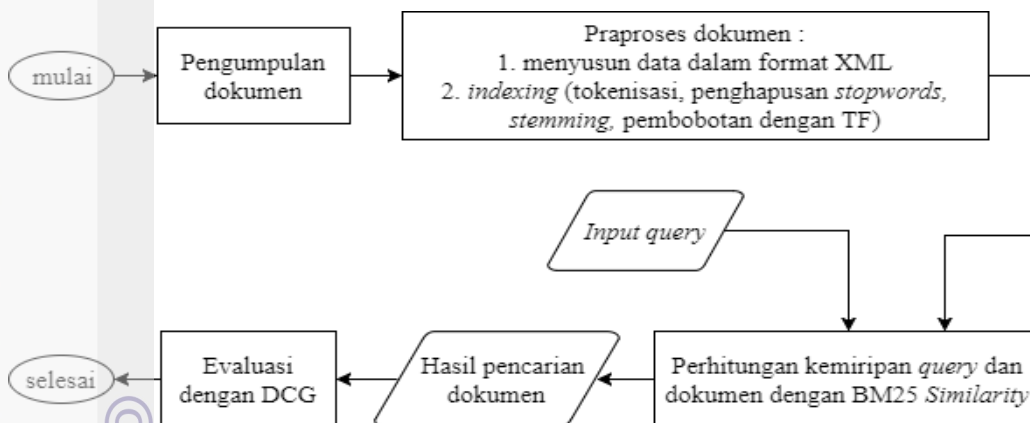
dengan IDCG adalah nilai peringkat ideal yang didapat dengan mengurutkan nilai relevansi dari DCG yang akan dinormalisasikan (Jarvelin dan Kekalainen 2002). Jika nilai relevansi pada contoh sebelumnya adalah 3, 2, 3, 0, 1, 2, maka nilai  $i$  diurutkan akan menjadi 3, 3, 2, 2, 1, 0 sehingga nilai ideal DCG(IDCG) adalah 8.69. Jadi nilai NDCG pencarian tersebut adalah  $8.10/8.69 = 0.932$ .

## METODE

Penelitian ini menggunakan 14 dokumen mengenai penyakit tanaman akasia yang diperoleh dari dokumentasi PT Arara Abadi oleh Indrayadi dan Mardai (2012a, 2012b). Dokumen tersebut berisi informasi mengenai penyakit busuk akar, busuk hati, keriting daun, bercak daun, embun hitam, layu fusarium, cacar daun, busuk batang, kanker batang, ceratocytis, hawar daun, rebah kecambah, penyakit tepung, dan bercak bergaris. Tahapan penelitian ini ditunjukkan pada Gambar 1. Tahapan tersebut melingkupi pengumpulan dokumen, praproses dokumen yang terdiri atas tiga tahap, yaitu menyusun data ke dalam format XML dan membuat indeks, selanjutnya dilakukan perhitungan dengan BM25 *similarity*, dan evaluasi hasil pencarian dokumen dengan DCG.

### Pengumpulan Dokumen

Dokumen mengenai penyakit tanaman akasia diperoleh dari dokumentasi PT Arara Abadi. Adapun informasi yang dikumpulkan dari dokumen tersebut adalah gejala, patogen, penyebaran, tanaman inang, pengendalian, tempat, dan area penyebaran penyakit akasia.



Gambar 1 Tahapan penelitian

### Praproses Dokumen

Tahapan setelah dokumen terkumpul adalah praproses. Tahap pertama yang dilakukan adalah menyusun data dalam dokumen ke dalam format XML. Tahap kedua adalah melakukan *indexing*, proses yang dilakukan dalam membuat indeks adalah tokenisasi, pembuangan *stopwords*, *stemming*, dan pembobotan dengan *TF* yaitu menghitung jumlah kemunculan setiap term  $t$  dalam sebuah dokumen  $d$ .

### Similarity

Pada tahap perhitungan *similarity*, nilai-nilai yang didapatkan dari proses pembobotan akan digunakan kembali pada perhitungan dari *similarity*. Nilai-nilai dari perhitungan *similarity* tersebut akan membentuk suatu sistem *ranking*, yang akan mengurutkan dokumen-dokumen berdasarkan tingkat kemiripan tertinggi ke tingkat kemiripan terendah dengan fungsi kemiripan BM25. Penelitian ini menggunakan fungsi kemiripan BM25 dengan melakukan modifikasi pada nilai *tuning parameter*. *Tuning parameter* adalah suatu variabel yang dapat diubah-ubah nilainya sesuai dengan kebutuhan dengan tujuan untuk mendapatkan hasil pencarian yang lebih baik.

Pada Persamaan (1) terdapat variabel yang disebut *tuning parameter*, yaitu  $k_1$ ,  $k_3$ , dan  $b$ . Nilai  $k_1 = 1.2$ ;  $k_3 = 1000$ ;  $b = 0.75$  akan diubah sesuai dengan kebutuhan sehingga dapat menghasilkan pencarian dengan skor kemiripan yang lebih baik. Menurut Robertson dan Walker (1999), nilai  $k_1$  dan  $b$  masing-masing di-set default 1.2 dan 0.75. Nilai  $b$  akan di-*tuning* mulai dari 0.75 sampai dengan yang paling kecil yaitu 0.15 dengan interval 0.15. Untuk nilai dari  $k_1$  yang awalnya bernilai 1.2, akan di-*tuning* dengan nilai antara 0.2 sampai dengan 4.2 dengan interval 1.0.

### Evaluasi

Hasil pencarian dokumen menggunakan fungsi kemiripan BM25 sebelum dan setelah dilakukan *tuning* dievaluasi dengan *discounted cumulative gain*

Hak Cipta Dilindungi Undang-Undang

© Hak cipta Institut Pertanian Bogor

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



(DCG). Setelah didapatkan urutan untuk setiap *query*, kemudian dilakukan perhitungan normalisasi DCG untuk setiap *query* pada setiap nilai  $k_1$  dan  $b$ . Nilai relevansi dokumen pada setiap *query* ( $rel_i$ ) terdapat pada Lampiran 1. Perhitungan DCG untuk setiap *query* pada setiap nilai  $k_1$  dan  $b$  akan diambil nilai rataannya dan dipetakan, kemudian akan didapatkan hasil evaluasi terbaik. Tabel 1 merupakan nilai *tuning parameter* yang diujikan. Perhitungan *score* BM25 dilakukan dengan bahasa pemrograman Python 2.7.13 kemudian dievaluasi dengan menghitung normalisasi DCG menggunakan Microsoft Excel.

Tabel 1 Nilai *tuning parameter* yang diujikan

Nama	$k_1$	$B$
Tuning1	0.2	0.75
Tuning2	0.2	0.60
Tuning3	0.2	0.45
Tuning4	0.2	0.30
Tuning5	0.2	0.15
Tuning6 ( <i>default</i> )	1.2	0.75
Tuning7	1.2	0.60
Tuning8	1.2	0.45
Tuning9	1.2	0.30
Tuning10	1.2	0.15
Tuning11	2.2	0.75
Tuning12	2.2	0.60
Tuning13	2.2	0.45
Tuning14	2.2	0.30
Tuning15	2.2	0.15
Tuning16	3.2	0.75
Tuning17	3.2	0.60
Tuning18	3.2	0.45
Tuning19	3.2	0.30
Tuning20	3.2	0.15
Tuning21	4.2	0.75
Tuning22	4.2	0.60
Tuning23	4.2	0.45
Tuning24	4.2	0.30
Tuning25	4.2	0.15

## HASIL DAN PEMBAHASAN

### Praproses Dokumen

Dilakukan dua tahapan dalam praproses dokumen, tahap pertama yang dilakukan adalah menyusun data dalam dokumen ke dalam format XML. Fail XML yang disusun berisi penyakit tanaman akasia. Struktur fail XML yang disusun seperti pada Gambar 2. Tahap kedua adalah melakukan *indexing*, yaitu proses persiapan yang dilakukan terhadap dokumen agar dokumen siap untuk

diproses. Proses yang dilakukan dalam *indexing* adalah tokenisasi, penghapusan *stopwords*, dan *stemming*. Data yang digunakan adalah data dalam *tag* <tanaman\_inang></tanaman\_inang>, <deskripsi></deskripsi>, <gejala></gejala>, <patogen></patogen>, dan <area\_penyakit></area\_penyakit>.

```
<?xml version="1.0" encoding="UTF-8"?>
<penyakit id="1">
  <nama>
    <lokal> nama penyakit (lokal) </lokal>
    <asing> nama penyakit (asing) </asing>
  </nama>
  <tanaman_inang>
    Acacia Crassicarpa atau Acacia Mangium
  </tanaman_inang>
  <deskripsi>
    deskripsi penyakit
  </deskripsi>
  <gejala>
    <ciri> gejala pertama </ciri>
    <ciri> gejala kedua, dan seterusnya </ciri>
  </gejala>
  <patogen>
    <jamur> jamur 1 </jamur>
    <jamur> jamur 2, dan seterusnya </jamur>
    <bakteri> bakteri 1 </bakteri>
    <bakteri> bakteri 2, dan seterusnya </bakteri>
  </patogen>
  <penyebaran>
    <melalui> media penyebaran 1 </melalui>
    <melalui> media penyebaran 2, dan seterusnya </melalui>
  </penyebaran>
  <pengendalian>
    <cara> cara pengendalian 1 </cara>
    <cara> cara pengendalian 2, dan seterusnya </cara>
  </pengendalian>
  <tempat> tempat pengendalian : nursery atau plantation
</tempat>
  <area_penyakit> area serangan, misalkan daun </area_penyakit>
</penyakit>
```

Gambar 2 Struktur fail XML

Proses tokenisasi dilakukan dengan membaca setiap karakter pada sebuah dokumen untuk memperoleh kata tunggal. Pada proses ini, kata didefinisikan sebagai unit terkecil sebuah dokumen yang dipisahkan oleh karakter selain huruf seperti spasi, titik, koma, dan tanda baca lainnya. Setelah hasil tokenisasi diperoleh, selanjutnya dilakukan penghilangan *stopwords* bahasa Indonesia pada Lampiran 2 yang telah digunakan pada penelitian Tala (2003). Proses tokenisasi dan penghapusan *stopwords* dilakukan menggunakan *tm package* pada Rstudio yang menghasilkan 284 term unik dari 14 dokumen.

Setelah itu dilakukan *stemming* dengan Sastrawati *library* pada bahasa pemrograman Python 2.7.13 yang dibuat oleh Bobby Nazief dan Mirna Adriani. Algoritme *stemming* dengan Sastrawi memiliki tahap-tahap sebagai berikut (Agusta 2009):

1. Cari kata yang akan di-*stem* dalam kamus. Jika ditemukan maka diasumsikan

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

- bahwa kata tersebut adalah *root word*. Maka algoritme berhenti.
- 2 *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.
  - 3 Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritme berhenti. Jika tidak maka ke langkah 3a.
    - a Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritme berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
    - b Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
  - 4 Hapus *Derivation Prefix*. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
    - a Periksa kombinasi awalan-akhiran yang tidak diizinkan pada Tabel 2. Jika ditemukan maka algoritme berhenti, jika tidak pergi ke langkah 4b.
    - b *For i = 1 to 3*, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5, jika sudah maka algoritme berhenti. Jika awalan kedua sama dengan awalan pertama algoritme berhenti.
  - 5 Melakukan *recoding*.
  - 6 Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

Tabel 2 Kombinasi awalan akhiran yang tidak diizinkan dalam algoritme *stemming* Nadief dan Andriana (Agusta 2009)

Awalan	Akhiran yang tidak diizinkan
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan

Tipe awalan ditentukan melalui langkah-langkah berikut:

- 1 Jika awalannya adalah: “di-”, “ke-”, atau “se-” maka tipe awalannya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.
- 2 Jika awalannya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.
- 3 Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.
- 4 Jika tipe awalan adalah “none” maka berhenti. Jika tipe awalan adalah bukan “none” maka awalan dapat dilihat pada Tabel 3. Hapus awalan jika ditemukan.

Hasil dari proses *stemming* diperoleh 239 term unik dari 14 dokumen yang dapat dilihat pada Lampiran 3. Hasil praproses terbentuk dalam matriks yang telah diboboti dengan TF. Contoh hasil pembobotan TF dapat dilihat pada Tabel 5.

Tabel 3 Cara menentukan tipe awalan untuk kata yang diawali dengan “te-” dalam algoritme *stemming* Nadief dan Andriana (Agusta 2009)

Set 1	Following characters			Tipe awalan
	Set 2	Set 3	Set 4	
“-r-“	“-r-“	-	-	none
“-r-“	Vowel	-	-	ter-luluh
“-r-“	not (vowel or “-r-“)	“-er-“	Vowel	ter
“-r-“	not (vowel or “-r-“)	“-er-“	not vowel	ter-
“-r-“	not (vowel or “-r-“)	not “-er-“	-	ter
not (vowel or “-r-“)	“-er-“	Vowel	-	none
not (vowel or “-r-“)	“-er-“	not vowel	-	te

Tabel 4 Jenis awalan berdasarkan tipe awalannya dalam algoritme *stemming* Nadief dan Andriana (Agusta 2009)

Tipe awalan	Awalan yang harus dihapus
di-	di-
ke-	ke-
se-	se-
te-	te-
ter-	ter-
ter-luluh	Ter

Tabel 5 Contoh hasil pembobotan dengan TF

ID dokumen	Term							
	abu	acacia	acaciae	air	akar	alami	ambil	ambrosia
1	0	2	0	1	1	0	0	0
2	0	2	0	0	0	0	0	0
3	0	2	0	0	0	0	0	0
4	0	2	0	0	0	0	0	0
5	0	1	0	0	0	0	0	0
6	0	2	1	0	0	0	0	0
7	0	2	0	0	2	0	0	0
8	0	1	0	0	0	0	1	0
9	0	3	0	0	0	0	0	0
10	0	2	0	0	8	0	0	0
11	0	2	0	0	3	1	0	0

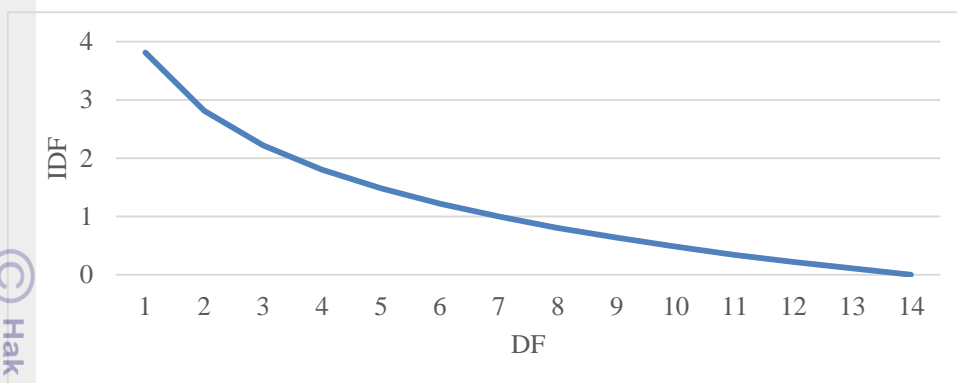
### Similarity

Terdapat beberapa faktor yang mempengaruhi hasil perhitungan fungsi kemiripan BM25 pada Persamaan (1), yaitu nilai IDF, modifikasi TF, dan normalisasi panjang dokumen. Gambar 3, Gambar 4, dan Gambar 5 merupakan simulasi data palsu (*dummy*) untuk mengetahui pengaruh nilai IDF,  $k_1$ , dan  $b$  dalam Persamaan (1). Nilai IDF menyatakan pendistribusian term secara luas dalam koleksi dokumen. Berdasarkan Gambar 3 dapat dilihat bahwa semakin banyak jumlah dokumen yang mengandung term yang diinginkan, maka semakin

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

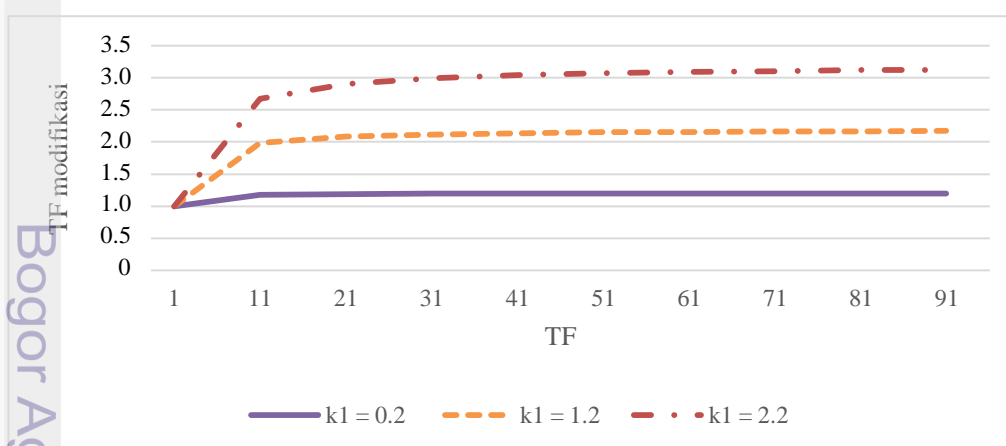
kecil nilai IDF yang dihasilkan. Hal itu terjadi karena term tersebut dianggap kurang penting, sebaliknya jika jumlah dokumen yang mengandung term yang diinginkan sedikit, term tersebut memiliki nilai kepentingan yang lebih tinggi.



Gambar 3 Simulasi pengaruh nilai DF terhadap IDF pada N = 14

Parameter  $k_1$  mempengaruhi hasil perhitungan fungsi kemiripan BM25 dengan memodifikasi nilai TF. Jika parameter  $b$  pada Persamaan (1) diberi nilai nol, maka nilai dari normalisasi panjang dokumen tidak akan mempengaruhi hasil modifikasi TF. Gambar 4 berisi hasil modifikasi TF dengan nilai  $k_1 = 0.2$ ,  $k_1 = 1.2$ , dan  $k_1 = 2.2$  tanpa melibatkan panjang dokumen. Jika nilai parameter  $k_1$  semakin tinggi maka perbedaan hasil modifikasi TF yang dihasilkan semakin jauh antara dokumen yang memiliki TF yang tinggi dengan dokumen yang memiliki TF yang rendah. Parameter  $k_1$  memberikan batas nilai tertinggi sebesar  $k_1 + 1$ . Jika nilai  $k_1 = 1.2$  maka nilai TF akan dimodifikasi ke dalam selang nilai 1 sampai hampir mendekati  $k_1 + 1 = 2.2$ .

Parameter  $b$  pada Persamaan (1) sangat berkaitan dengan normalisasi panjang dokumen yang mempengaruhi *score* pada BM25 melalui modifikasi nilai TF. Untuk mengetahui pengaruh dari parameter  $b$  pada modifikasi TF, nilai  $k_1$  diatur *default* menjadi 1.2 dengan nilai  $L_d$  yang lebih besar, lebih kecil, dan sama dengan nilai  $L_{ave}$ . Misalkan  $L_{ave} = 2$ , diberikan  $L_d = 4$  untuk panjang dokumen yang lebih besar dari rata-rata panjang dokumen,  $L_d = 1$  untuk panjang dokumen yang lebih kecil dari rata-rata panjang dokumen, dan  $L_d = 2$  untuk panjang dokumen yang sama dengan rata-rata panjang dokumen.



Gambar 4 Simulasi pengaruh parameter  $k_1$  terhadap nilai TF modifikasi ( $b = 0$ )

Hak Cipta Dilindungi Undang-Undang

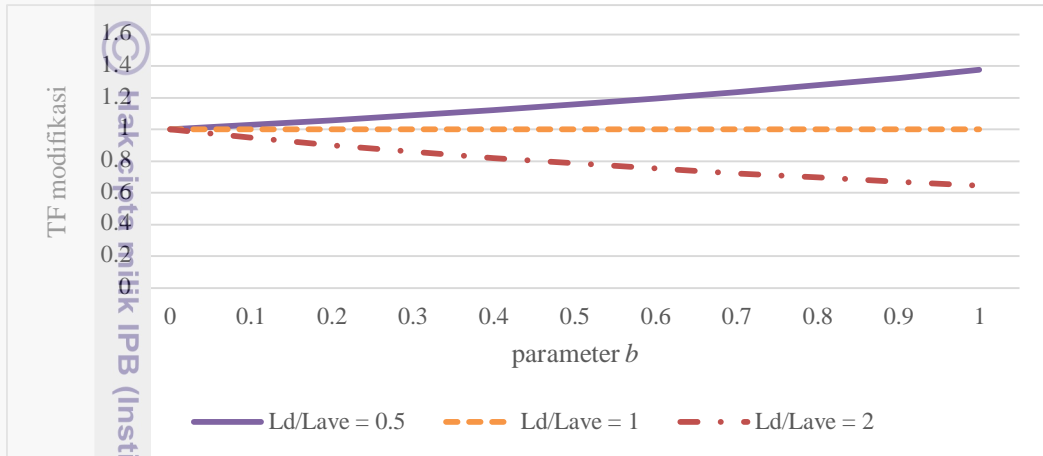
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Pada Gambar 5 parameter  $b$  memodifikasi nilai  $TF = 1$  dengan meningkatkan, menurunkan, atau sama sekali tidak mengubah nilai  $TF$ . Jika suatu dokumen memiliki panjang yang lebih kecil dari rata-rata panjang dokumen maka kenaikan parameter  $b$  akan meningkatkan nilai  $TF$ . Jika suatu dokumen memiliki panjang yang lebih besar dari rata-rata panjang dokumen maka kenaikan parameter  $b$  akan menurunkan nilai  $TF$ . Tetapi jika suatu dokumen memiliki panjang yang sama dengan rata-rata panjang dokumen maka kenaikan parameter  $b$  tidak akan mengubah nilai  $TF$ .



Gambar 5 Simulasi pengaruh parameter  $b$  terhadap nilai modifikasi TF ( $k_l=1.2$ )

Persamaan (1) juga mengandung parameter  $k_3$  yang dapat diubah-ubah nilainya. Penelitian ini menggunakan nilai  $k_3 = 1000$  sesuai dengan rekomendasi Robertson dan Walker (1999). Parameter  $k_3$  mempengaruhi nilai  $tf_{tq}$  (jumlah term  $t$  yang muncul pada  $query$   $q$ ). Namun, dalam penelitian ini hanya digunakan 20  $query$  pendek yang terdiri dari dua sampai dengan empat term dan dalam sebuah  $query$  tidak mengandung term yang sama, maka tidak dilakukan *tuning* pada parameter  $k_3$ .

Perhitungan kemiripan antara  $query$  dengan dokumen dilakukan pada tahap *Similarity*. Terdapat 20  $query$  yang digunakan berupa gejala penyakit yang terdiri dari dua sampai empat term.  $Query$  tersebut dihitung nilai kemiripannya terhadap dokumen dengan fungsi kemiripan BM25, kemudian nilai BM25 dari setiap  $query$  terhadap dokumen yang relevan diurutkan dari yang terbesar. Perhitungan dilakukan menggunakan bahasa pemrograman Python 2.7.13 dengan memodifikasi kode program milik Hirakawa (2014). Contoh hasil pengurutan nilai BM25 untuk setiap  $query$  pada nilai  $k_l = 1.2$  dan  $b = 0.75$  ditunjukkan pada Tabel 6. Dengan keterangan id dokumen 1: rebah kecambah, 2: bercak daun, 3: hawar daun, 4: busuk batang, 5: penyakit tepung, 6: bercak bergaris, 7: layu fusarium, 8: cacar daun, 9: keriting daun, 10: busuk akar, 11: busuk hati, 12: embun hitam, 13: kanker batang, dan 14: ceratocystis.

Tabel 6 Contoh urutan dokumen hasil pencarian pada Tuning6

Query	Urutan id dokumen hasil pencarian dokumen dengan BM25
akar mengkerut	7, 10, 14, 11, 1
patah batang	11, 4, 3, 2, 14, 7, 13,1
pembusukan batang	11, 14, 7, 4, 3, 10, 13, 1, 2
akar kehitaman	14, 11, 7, 10, 1, 12, 13, 8, 2, 4, 3
tanaman roboh	1, 11, 9, 7, 2, 6, 12, 4, 3, 13, 10, 5, 14
daun menguning	9, 8, 2, 10, 11, 5, 12, 4, 3, 14, 7, 6
akar kemerahan	10, 14, 11, 7, 2, 4, 3, 1
akar menebal	10, 11, 8, 14, 9, 7, 1
bercak hitam	2, 4, 3, 8, 7, 6, 5, 12, 13, 9, 14, 11
bercak memanjang	6, 1, 2, 5, 4, 3, 9, 8, 7
batang membengkak	13, 11, 8, 14, 7, 4, 3, 1, 2
daun rontok	7, 5, 8, 12, 2, 4, 3, 14, 9, 10, 6
tepung putih	7, 10, 5, 14, 11
batang berbercak hitam	7, 4, 3, 2, 13, 14, 11, 8, 6, 5, 12, 9, 1
daun berbercak coklat	2, 8, 7, 14, 5, 4, 3, 6, 9, 11, 12, 10
batang berwarna putih	14, 11, 10, 7, 12, 5, 13, 4, 3, 1, 2
daun berbercak kemerahan	2, 4, 3, 10, 5, 6, 9, 8, 7, 12, 14
daun berbercak kekuningan	2, 9, 8, 5, 10, 4, 3, 6, 11, 7, 12, 14
penebalan jaringan daun	8, 10, 13, 9, 1, 11, 5, 12, 2, 4, 3, 14, 7, 6
daun tertutup kerak hitam	12, 5, 10, 8, 2, 7, 4, 3, 14, 13, 11, 9, 6

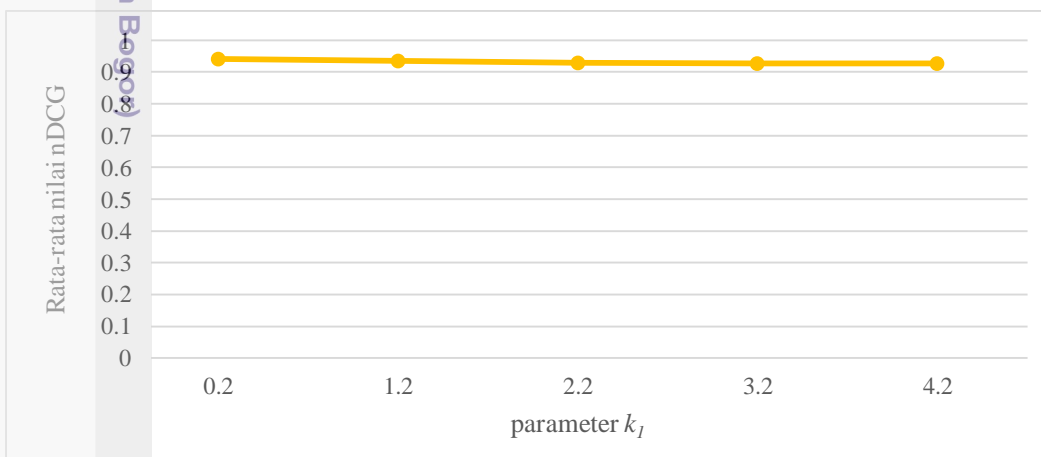
### Evaluasi

Setelah didapatkan urutan untuk setiap *query*, dilakukan perhitungan DCG untuk setiap *query* pada setiap nilai  $k_1$  dan  $b$ . Tabel 7 merupakan nilai NDCG untuk setiap *query* sebelum dilakukan *tuning* (nilai  $k_1 = 1.2$  dan  $b = 0.75$ ) dan setelah dilakukan *tuning* (nilai  $k_1 = 0.2$  dan  $b = 0.30$ ). Gambar 6 merupakan laju nilai rata-rata NDCG pada nilai parameter  $k_1$  pada nilai  $b$  yang di-set default 0.75, dan Tabel 8 merupakan nilai NDCG setiap *query* pada nilai parameter  $b$  dan nilai  $k_1 = 0.2$ .

Berdasarkan Gambar 6, semakin kecil nilai  $k_1$  yang digunakan, maka semakin besar nilai NDCG yang dihasilkan. Berdasarkan hal tersebut, dapat dikatakan bahwa nilai  $k_1$  yang rendah bertujuan agar perbedaan hasil modifikasi TF antara dokumen yang memiliki nilai TF yang tinggi dengan dokumen yang memiliki nilai TF yang rendah tidak akan jauh berbeda. Hal tersebut disebabkan pencarian term pada *query* uji cenderung lebih sering menghasilkan nilai TF yang kecil. Oleh karena itu, dapat dikatakan bahwa selang nilai TF pada evaluasi menggunakan *query* uji sangat kecil sehingga parameter  $k_1$  akan lebih sesuai jika nilainya kecil.

Tabel 7 Hasil normalisasi nilai DCG pada sebelum dan setelah *tuning*

ID	Query	NDCG sebelum <i>tuning</i>	NDCG setelah <i>tuning</i>
1	akar mengkerut	0.971727	0.971727
2	patah batang	0.991409	0.991409
3	pembusukan batang	0.894949	0.894949
4	akar kehitaman	0.938133	0.893421
5	tanaman roboh	0.894661	0.894661
6	daun menguning	0.985277	0.985277
7	akar kemerahan	0.891771	0.987480
8	akar menebal	0.986489	1.000000
9	bercak hitam	0.988527	0.988527
10	bercak memanjang	0.844916	0.830101
11	batang membengkak	1.000000	1.000000
12	daun rontok	0.707096	0.762779
13	tepung putih	0.952451	0.952451
14	batang berbercak hitam	0.998275	0.997618
15	daun berbercak coklat	1.000000	0.986473
16	batang berwarna putih	0.965356	0.995033
17	daun berbercak kemerahan	1.000000	1.000000
18	daun berbercak kekuningan	0.992183	0.987901
19	penebalan jaringan daun	0.874189	0.907134
20	daun tertutup kerak hitam	0.827315	0.808293
Rata-rata		0.935236	0.941762



Gambar 6 Laju nilai rata-rata NDCG pada parameter  $k_1$  ( $b=0.75$ )

Evaluasi terhadap parameter  $b$  pada Tabel 8 menggunakan nilai  $k_1 = 0.2$  karena pada saat  $k_1 = 0.2$  didapatkan rata-rata NDCG yang paling tinggi seperti yang terdapat pada Gambar 6. Berdasarkan Tabel 8 dapat dilihat bahwa rata-rata NDCG tertinggi didapatkan pada nilai  $b = 0.3$  sebesar 0.941762. Tetapi, secara keseluruhan nilai evaluasi yang diperoleh tidak menghasilkan perubahan yang

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



signifikan, semua nilai pada parameter  $b$  saat nilai  $k_1 = 0.2$  menghasilkan nilai NDCG 0.94 jika dilakukan pembulatan nilai.

Tabel 8 Nilai NDCG setiap *query* pada nilai parameter  $b$  dan nilai  $k_1 = 0.2$

<i>Query ID</i>	$b = 0.75$	$b = 0.60$	$b = 0.45$	$b = 0.30$	$b = 0.15$
1	0.971727	0.971727	0.971727	0.971727	0.971727
2	0.991409	0.991409	0.991409	0.991409	0.991409
3	0.894949	0.894949	0.894949	0.894949	0.894949
4	0.893421	0.893421	0.893421	0.893421	0.893421
5	0.894661	0.894661	0.894661	0.894661	0.910642
6	0.985277	0.985277	0.985277	0.985277	0.943775
7	0.987481	0.987481	0.987481	0.987480	0.987481
8	1.000000	1.000000	1.000000	1.000000	1.000000
9	0.988527	0.988527	0.988527	0.988527	0.988527
10	0.844916	0.830102	0.830102	0.830101	0.830102
11	1.000000	1.000000	1.000000	1.000000	1.000000
12	0.707096	0.707096	0.726372	0.762779	0.762779
13	0.952451	0.952451	0.952451	0.952451	0.952451
14	0.997618	0.997618	0.997618	0.997618	0.997618
15	0.986473	0.986473	0.986473	0.986473	0.986473
16	0.995033	0.995033	0.995033	0.995033	0.995033
17	1.000000	1.000000	1.000000	1.000000	1.000000
18	0.987901	0.987901	0.987901	0.987901	0.987901
19	0.907134	0.907134	0.907134	0.907134	0.907134
20	0.827315	0.827315	0.819947	0.808293	0.808293
Rata-rata	0.940669	0.939929	0.940524	0.941762	0.940486

Nilai  $b$  yang semakin besar akan memberikan perubahan peningkatan atau penurunan nilai modifikasi TF yang semakin besar pula pada panjang dokumen yang lebih kecil atau lebih besar dari panjang rata-rata dokumen pada koleksi dokumen. Nilai panjang dokumen untuk setiap dokumen terdapat pada Lampiran 4. Keberhasilan parameter  $b$  pada penelitian ini ditentukan oleh kesesuaian hasil perhitungan BM25 menggunakan nilai  $b$  tertentu dengan nilai  $rel_i$  pada setiap *query*. Terdapat *query* yang lebih sesuai jika perubahan (peningkatan dan penurunan) modifikasi TF terhadap panjang dokumen lebih sedikit, seperti pada *query* 5 yang memperoleh hasil evaluasi lebih besar jika menggunakan nilai  $b$  lebih kecil dari 0.30 dan *query* 12 yang memperoleh hasil evaluasi lebih besar jika menggunakan nilai  $b$  lebih kecil atau sama dengan 0.30.

Terdapat juga *query* yang lebih sesuai jika perubahan (peningkatan dan penurunan) nilai modifikasi TF terhadap panjang dokumen lebih tinggi, seperti pada *query* 10 dan 20 yang memperoleh hasil evaluasi lebih besar jika menggunakan nilai  $b$  lebih besar dari 0.30 dan *query* 6 yang memperoleh hasil evaluasi lebih besar jika menggunakan nilai  $b$  lebih besar atau sama dengan 0.30. *Tuning* pada parameter  $b$  tidak memberikan satu nilai yang optimal disebabkan oleh panjang dokumen yang relatif sama, sehingga terkadang nilai  $b$  yang kecil dapat mengasilkan akurasi yang tinggi contohnya pada *query* 5 dan 12, sebaliknya

nilai  $b$  yang besar juga dapat mengasilkan akurasi yang tinggi contohnya pada *query* 6, 10 dan 20

*Tuning* terbaik diperoleh pada nilai  $k_1 = 0.2$  dan  $b = 0.30$ . Keseluruhan hasil NDCG untuk setiap *tuning* terdapat pada Lampiran 5. Tabel 7 berisi nilai NDCG pada sebelum dan setelah *tuning*. Terdapat beberapa *query* yang memiliki nilai NDCG yang berbeda. Hal ini disebabkan oleh kesesuaian nilai relevansi ( $rel_i$ ) terhadap nilai  $k_1$  dan  $b$ . Beberapa *query* lebih sesuai menggunakan nilai  $k_1 = 0.2$  dan  $b = 0.30$  karena perbedaan nilai modifikasi TF yang lebih sedikit antara dokumen yang memiliki nilai TF yang tinggi dengan dokumen yang memiliki nilai TF yang rendah. Selain itu, hal tersebut juga disebabkan oleh perubahan yang lebih sedikit pada nilai modifikasi TF, yaitu mengalami peningkatan pada dokumen yang memiliki panjang dokumen yang lebih kecil dari rata-rata panjang dokumen dan mengalami penurunan pada dokumen yang memiliki panjang dokumen yang lebih besar dari rata-rata panjang dokumen. Sebaliknya, jika *query* lebih sesuai dengan perbedaan nilai modifikasi TF yang lebih jauh dan perubahan nilai modifikasi TF yang lebih tinggi, *query* tersebut menghasilkan nilai evaluasi yang lebih tinggi jika menggunakan nilai  $k_1 = 1.2$  dan  $b = 0.75$ .

Sebagai contoh *query* yang terdiri atas dua term, yaitu “akar menebal” memiliki nilai evaluasi yang lebih tinggi dengan nilai  $k_1 = 0.2$  dan  $b = 0.30$  (BM25 *tuning*) dibandingkan dengan nilai  $k_1 = 1.2$  dan  $b = 0.75$  (BM25 *default*). Berdasarkan Tabel 9, dapat dilihat bahwa urutan hasil pencarian dengan Tuning4 dan Tuning6 berbeda pada urutan ke-4 dan urutan ke-5. Pada Tuning4, dokumen 9 berada pada urutan ke-4 dan dokumen 14 berada pada urutan ke-5, sebaliknya pada Tuning6 dokumen 14 berada pada urutan ke-4 dan dokumen 9 berada pada urutan ke-5. Nilai relevansi *query* “akar menebal” pada dokumen 9 adalah 1 dan 0 untuk dokumen 14 (Lampiran 1), maka hasil NDCG pada Tuning4 lebih tinggi karena urutan dokumen 9 berada di atas dokumen 14.

Pada Tabel 9, hasil modifikasi TF “akar” dan “tebal” menggunakan  $k_1 = 0.2$  memiliki nilai yang lebih kecil jika dibandingkan dengan  $k_1 = 1.2$  karena dimodifikasi ke dalam selang nilai 1 sampai mendekati 1.2, sedangkan TF menggunakan  $k_1 = 1.2$  dimodifikasi ke dalam selang nilai 1 sampai hampir mendekati 2.2. Nilai TF “tebal” pada dokumen 9 yang bernilai 1 dimodifikasi menjadi 1 pada  $k_1 = 0.2$  maupun  $k_1 = 1.2$ . Nilai TF “akar” pada dokumen 14 yang semula bernilai 3 dimodifikasi menjadi 1.125 pada  $k_1 = 0.2$  dan 1.571428 pada  $k_1 = 1.2$ .

Panjang dokumen 14 lebih besar jika dibandingkan dengan rata-rata panjang dokumen, maka dengan mengaktifkan parameter  $b$  nilai modifikasi TF “akar” akan mengalami penurunan. Nilai modifikasi TF “tebal” akan mengalami peningkatan karena panjang dokumen 9 lebih kecil dibanding rata-rata panjang dokumen. Peningkatan dan penurunan nilai modifikasi TF pada penggunaan  $b = 0.30$  lebih sedikit daripada  $b = 0.75$ . Oleh sebab itu, *query* “akar menebal” tidak sesuai jika menggunakan nilai  $k_1 = 1.2$  dan  $b = 0.75$  karena terdapat perbedaan yang jauh pada modifikasi nilai TF “akar” yang bernilai 3 dan modifikasi nilai TF “tebal” yang bernilai 1. Selain itu, peningkatan nilai modifikasi TF “akar” akibat penggunaan nilai  $b = 0.75$  lebih tinggi jika dibandingkan dengan nilai  $b = 0.30$  membuat nilai modifikasi TF “akar” pada dokumen 14 semakin tinggi dan akan menghasilkan skor BM25 yang semakin tinggi setelah dikalikan dengan nilai IDF

“akar”, sehingga dokumen 14 berada pada urutan yang lebih tinggi daripada dokumen 9.

Tabel 9 Perbedaan hasil evaluasi pada BM25 *default* dan BM25 *tuning* pada *query* yang terdiri atas dua term "akar menebal"

	Dokumen 9	Dokumen 14
Panjang dokumen	40	59
TF "akar"	-	3.000000
TF "tebal"	1.000000	-
Modifikasi TF "akar" pada $k_1 = 1.2$ , $b = 0$	-	1.571428
Modifikasi TF "akar" pada $k_1 = 0.2$ , $b = 0$	-	1.125000
Modifikasi TF "tebal" pada $k_1 = 1.2$ , $b = 0$	1.000000	-
Modifikasi TF "tebal" pada $k_1 = 0.2$ , $b = 0$	1.000000	-
Modifikasi TF "akar" BM25 <i>default</i>	-	1.568969
Modifikasi TF "tebal" BM25 <i>default</i>	1.149045	-
Modifikasi TF "akar" BM25 <i>tuning</i>	-	1.124846
Modifikasi TF "tebal" BM25 <i>tuning</i>	1.016109	-
IDF "akar"	1.485426827	
IDF "tebal"	1.807354922	
Urutan BM25 <i>tuning</i>	10,11,8,9,14,7,1	
Urutan BM25 <i>default</i>	10,11,8,14,9,7,1	

Contoh *query* yang terdiri atas tiga term adalah “batang berwarna putih” memiliki nilai evaluasi yang lebih tinggi dengan nilai  $k_1 = 0.2$  dan  $b = 0.30$  (BM25 *tuning*) dibandingkan dengan nilai  $k_1 = 1.2$  dan  $b = 0.75$  (BM25 *default*). Berdasarkan Tabel 10, dapat dilihat bahwa urutan hasil pencarian dengan Tuning4 dan Tuning6 berbeda pada urutan ke-3 dan urutan ke-4. Pada Tuning4 dokumen 7 berada pada urutan ke-3 dan dokumen 10 berada pada urutan ke-4, sebaliknya pada Tuning6 dokumen 10 berada pada urutan ke-3 dan dokumen 7 berada pada urutan ke-4. Nilai relevansi *query* “batang berwarna putih” pada dokumen 7 adalah 3 dan 1 untuk dokumen 10 (Lampiran 1), maka hasil NDCG pada Tuning4 lebih tinggi karena urutan dokumen 7 berada di atas dokumen 10.

Pada Tabel 10, hasil modifikasi TF “batang”, “warna” dan “putih” menggunakan  $k_1 = 0.2$  memiliki nilai yang lebih kecil jika dibandingkan dengan  $k_1 = 1.2$  karena dimodifikasi ke dalam selang nilai 1 sampai hampir mendekati 1.2, sedangkan TF menggunakan  $k_1 = 1.2$  dimodifikasi ke dalam selang nilai 1 sampai hampir mendekati 2.2. Nilai TF “warna” dan “putih” pada dokumen 7 yang bernilai 1 dimodifikasi menjadi 1 pada  $k_1 = 0.2$  maupun  $k_1 = 1.2$ . Nilai TF “warna” dan “putih” pada dokumen 10 yang semula bernilai 2 dan 3 dimodifikasi menjadi 1.090909 dan 1.125 pada  $k_1 = 0.2$ , sedangkan pada  $k_1 = 1.2$  menjadi 1.375 dan 1.571428.

Panjang dokumen 7 dan 10 lebih besar jika dibandingkan dengan rata-rata panjang dokumen, maka dengan mengaktifkan parameter  $b$ , nilai modifikasi TF “batang”, “warna”, dan “putih” akan mengalami penurunan. Penurunan nilai modifikasi TF pada penggunaan  $b = 0.30$  lebih sedikit daripada  $b = 0.75$ . Oleh sebab itu, *query* “batang berwarna putih” tidak sesuai jika menggunakan nilai  $k_1 =$

1.2 dan  $b = 0.75$  karena terdapat penurunan yang lebih besar pada modifikasi nilai TF “warna” dan “putih” akibat penggunaan nilai  $b = 0.75$  jika dibandingkan dengan nilai  $b = 0.30$ , sehingga membuat nilai modifikasi TF “warna” dan “putih” pada dokumen 7 semakin rendah, jadi dokumen 7 berada pada urutan yang lebih rendah daripada dokumen 10.

Tabel 10 Perbedaan hasil evaluasi pada BM25 *default* dan BM25 *tuning* pada *query* yang terdiri atas tiga term "batang berwarna putih"

	Dokumen 7	Dokumen 10
Panjang dokumen	79	75
TF "batang"	6.000000	-
TF "warna"	1.000000	2.000000
TF "putih"	1.000000	3.000000
Modifikasi TF "batang" pada $k_1 = 1.2, b = 0$	1.833333	-
Modifikasi TF "batang" pada $k_1 = 0.2, b = 0$	1.161290	-
Modifikasi TF "warna" pada $k_1 = 1.2, b = 0$	1.000000	1.375000
Modifikasi TF "warna" pada $k_1 = 0.2, b = 0$	1.000000	1.090909
Modifikasi TF "putih" pada $k_1 = 1.2, b = 0$	1.000000	1.571428
Modifikasi TF "putih" pada $k_1 = 0.2, b = 0$	1.000000	1.125000
Modifikasi TF "batang" BM25 <i>default</i>	1.756744	-
Modifikasi TF "warna" BM25 <i>default</i>	0.875133	1.274461
Modifikasi TF "putih" BM25 <i>default</i>	0.875133	1.482334
Modifikasi TF "batang" BM25 <i>tuning</i>	1.157384	-
Modifikasi TF "warna" BM25 <i>tuning</i>	0.982860	1.082627
Modifikasi TF "putih" BM25 <i>tuning</i>	0.982860	1.119114
IDF “batang”	0.807354922	
IDF “warna”	1.485426827	
IDF “putih”	1.485426827	
Urutan BM25 <i>tuning</i>	14,11,7,10,12,5,13,4,3,1,2	
Urutan BM25 <i>default</i>	14,11,10,7,12,5,13,4,3,1,2	

Contoh *query* yang terdiri atas empat term yaitu “daun tertutup kerak hitam” memiliki nilai evaluasi yang lebih rendah dengan nilai  $k_1 = 0.2$  dan  $b = 0.30$  (BM25 *tuning*) dibandingkan dengan nilai  $k_1 = 1.2$  dan  $b = 0.75$  (BM25 *default*). Berdasarkan Tabel 11, dapat dilihat bahwa urutan hasil pencarian dengan Tuning4 dan Tuning6 berbeda di urutan ke-4, ke-5 dan urutan ke-6. Pada Tuning4 dokumen 7 berada pada urutan ke-4, dokumen 8 pada urutan ke-5, dan dokumen 2 berada pada urutan ke-6, sebaliknya pada Tuning6 dokumen 7 berada pada urutan ke-6 setelah dokumen 8 dan dokumen 2. Nilai relevansi *query* “daun tertutup kerak hitam” pada dokumen 2 dan dokumen 8 adalah 1, sedangkan dokumen 7 memiliki nilai relevansi 0 (Lampiran 1), maka hasil NDCG pada Tuning4 lebih rendah karena urutan dokumen 7 berada di atas dokumen 8 dan dokumen 2.

Pada Tabel 11, hasil modifikasi TF “daun” dan “hitam” menggunakan  $k_1 = 0.2$  memiliki nilai yang lebih kecil jika dibandingkan dengan  $k_1 = 1.2$  karena dimodifikasi ke dalam selang nilai 1 sampai hampir mendekati 1.2, sedangkan TF menggunakan  $k_1 = 1.2$  dimodifikasi ke dalam selang nilai 1 sampai hampir

mendekati 2.2. Pada  $k_1 = 0.2$  nilai TF = 2 dimodifikasi menjadi 1.090909, nilai TF = 3 dimodifikasi menjadi 1.125, nilai TF = 6 dimodifikasi menjadi 1.161290, dan nilai TF = 7 dimodifikasi menjadi 1.666667. Pada  $k_1 = 1.2$  nilai TF = 2 dimodifikasi menjadi 1.375, nilai TF = 3 dimodifikasi menjadi 1.571428, nilai TF = 6 dimodifikasi menjadi 1.833333, dan nilai TF = 7 dimodifikasi menjadi 1.878048.

Tabel 11 Perbedaan hasil evaluasi pada BM25 *default* dan BM25 *tuning* pada *query* yang terdiri atas empat term "daun tertutup kerak hitam"

	Dokumen 2	Dokumen 7	Dokumen 8
Panjang dokumen	65	79	61
TF "daun"	6.000000	3.000000	7.000000
TF "tutup"	-	-	-
TF "kerak"	-	-	-
TF "hitam"	2.000000	3.000000	2.000000
Modifikasi TF "daun" pada $k_1 = 1.2, b = 0$	1.833333	1.571428	1.878048
Modifikasi TF "daun" pada $k_1 = 0.2, b = 0$	1.161290	1.125000	1.166667
Modifikasi TF "hitam" pada $k_1 = 1.2, b = 0$	1.375000	1.571428	1.375000
Modifikasi TF "hitam" pada $k_1 = 0.2, b = 0$	1.090909	1.125000	1.090909
Modifikasi TF "daun" BM25 <i>default</i>	1.808521	1.462149	1.869541
Modifikasi TF "hitam" BM25 <i>default</i>	1.333826	1.462149	1.359150
Modifikasi TF "daun" BM25 <i>tuning</i>	1.160058	1.117691	1.662637
Modifikasi TF "hitam" BM25 <i>tuning</i>	1.087653	1.117691	1.089677
IDF "daun"	0.34792330342		
IDF "hitam"	0.63742992062		
Urutan BM25 <i>tuning</i>	12,5,10,7,8,2,4,3,14,13,11,9,6		
Urutan BM25 <i>default</i>	12,5,10,8,2,7,4,3,14,13,11,9,6		

Panjang dokumen 2, 7 dan 8 lebih besar jika dibandingkan dengan rata-rata panjang dokumen, maka dengan mengaktifkan parameter  $b$  ( $b = 0.30$  pada Tuning4 dan  $b = 0.75$  pada Tuning6), nilai modifikasi TF "daun" dan "hitam" akan mengalami penurunan. Penurunan nilai modifikasi TF pada penggunaan  $b = 0.30$  lebih sedikit daripada  $b = 0.75$ . Oleh sebab itu, *query* "daun tertutup kerak hitam" tidak sesuai jika menggunakan nilai  $k_1 = 0.2$  dan  $b = 0.30$  karena terdapat penurunan yang lebih sedikit pada modifikasi nilai TF "daun" dan "hitam" akibat penggunaan nilai  $b = 0.30$  jika dibandingkan dengan nilai  $b = 0.75$ , sehingga membuat nilai modifikasi TF "daun" dan "hitam" pada dokumen 7 tidak berbeda jauh dengan dokumen 2 dan dokumen 8. Skor pada dokumen 7 tersebut menjadi semakin tinggi setelah dikalikan dengan nilai IDF "daun" dan "hitam", sehingga dokumen 7 berada pada urutan yang lebih tinggi daripada dokumen 2 dan dokumen 8.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## SIMPULAN DAN SARAN

### Simpulan

Berdasarkan hasil penelitian ini, telah dilakukan proses modifikasi nilai dari *tuning parameter* yang ada pada fungsi kemiripan BM25. Modifikasi nilai dari *tuning parameter* tersebut dapat meningkatkan kinerja hasil pencarian dokumen. Hal tersebut disebabkan oleh *tuning parameter*  $k_1$  yang mempengaruhi nilai modifikasi TF dan parameter  $b$  yang mempengaruhi normalisasi panjang dokumen. Nilai rata-rata NDCG dengan menggunakan BM25 sebelum dilakukan *tuning* (nilai  $k_1 = 1.2$  dan  $b = 0.75$ ) yaitu 0.935236, sedangkan setelah dilakukan *tuning* nilai rata-rata NDCG yang terbesar terdapat pada nilai  $k_1 = 0.2$  dan  $b = 0.30$  yaitu 0.941762. Keduanya memiliki nilai NDCG yang cukup besar, sehingga dinyatakan fungsi kemiripan BM25 dapat mengukur relevansi cukup akurat antara *query* masukan dengan dokumen yang dibutuhkan untuk dokumen dan *query* yang digunakan pada penelitian ini.

### Saran

Penelitian ini hanya menggunakan 14 dokumen penyakit akasia. dan hanya mengujikan 20 *query* dan terdiri atas dua sampai dengan empat kata. Untuk penelitian selanjutnya diharapkan dapat menggunakan dokumen yang lebih beragam dan melakukan pengujian dengan *query* yang berbeda, lebih beragam, dan terdiri dari banyak kata. Diharapkan penggunaan nilai  $k_1$ ,  $k_3$ , dan  $b$  yang lebih beragam dalam melakukan modifikasi dari *tuning parameter* sehingga memungkinkan didapatkannya nilai NDCG yang lebih baik. Selain itu, diharapkan juga dalam evaluasi untuk penelitian selanjutnya, penilaian relevansi dokumen dilakukan oleh pakar.

## DAFTAR PUSTAKA

- Agusta L. 2009. Perbandingan algoritme stemming porter dengan algoritme nazief dan andriani untuk stemming dokumen teks bahasa indonesia. *Konferensi Nasional Sistem dan Informatika*. 9(36):196-201.
- Baeza RY, Neto R. 1999. *Modern Information Retrieval*. Boston (US): Addison Wesley-Pearson international edition.
- [Dephut] Departemen Kehutanan. 2003. Pengembangan Hutan Tanaman Industri (HTI)-Pulp 2002. Jakarta (ID): Departemen Kehutanan. Di dalam: Krisnawati H, Kallio M, Kanninen M, editor. *Acacia mangium Willd. Ekologi, Silvikultur dan Produktivitas* [internet]. [Waktu dan tempat pertemuan tidak diketahui]. Bogor (ID): CIFOR. Hlm 1-14; [diunduh 2017 Jan 9]. Tersedia pada: [http://www.cifor.org/publications/pdf\\_files/Books/Bk\\_risnawati1101.pdf](http://www.cifor.org/publications/pdf_files/Books/Bk_risnawati1101.pdf)
- [FAO] Food and Agriculture Organization of the United Nations. 2002. Tropical forest plantation areas 1995 data set by D. Pandey. Di dalam: Varmola M,

- Del Lungo A, editor. *Forest Plantations Working Paper 18*; 2002 Mei; Roma, Italia. Roma (IT): Forest Resources Division, FAO. hlm 1-13.
- Hirakawa N. 2014. BM25 [internet]. [diunduh 2017 Mei 2]. Tersedia pada: <https://github.com/nhirakawa/BM25>
- Indrayadi H, Mardai. 2012a. *Pedoman Pengenalan Pengendalian Hama Penyakit Acacia dan Eucalyptus di Plantation*. Riau (ID): Divisi Penelitian dan Pengembangan Kehutanan Sinarmas Forestry.
- Indrayadi H, Mardai. 2012b. *Pedoman Pengendalian Hama Penyakit Acacia dan Eucalyptus di Nursery*. Riau (ID): Divisi Penelitian dan Pengembangan Kehutanan Sinarmas Forestry.
- Irianto RSB, Barry KM, Mohammed CL. 2006. Heart Rot and Root Rot in Acacia Mangium: Identification and Assessment. Di dalam: Potter K, Rimbawanto A, Beadle C, editor. *Heart Rot and Root Rot in Tropical Acacia Plantations*; 2006 Feb 7-9; Yogyakarta, Indonesia. Canberra: ACIAR Proceedings No. 124. hlm 26-33.
- Jarvelin K, Kekalainen J. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422-446.
- Jones KS, Walker S, Robertson S. 2000. A probabilistic model of information retrieval: development and comparative experiments. *Information Processing and Management*. 36:809-840.
- Lv Y, Zhai CX. 2011. Adaptive term frequency normalization for BM25. Di dalam: *CIKM '11 Proceedings of the 20th ACM international conference on Information and knowledge management*; 2011 Okt 24-28; Glasgow, Scotland (UK): ACM New York. hlm 1985-1988.
- Manning C, Raghavan P, Schütze H. 2009. *Introduction to Information Retrieval*. Cambridge (UK): Cambridge University Press.
- Robertson SE, Walker S. 1999. Okapi/Keenbow at TREC-8. Di dalam: *Proceedings of TREC-8*; 1999 Nov 16-19; Maryland, United States of America. Maryland (US): NIST. hlm 151-162.
- Sagayam R, Srinivasan S, Roshni S. 2012. A survey of text mining: retrieval, extraction and indexing techniques. *IJCER*. 2(5):1443-1444.
- Salton G. 1989. *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Boston (US): Addison – Wesley Publishing Company, Inc.
- Saputra T. 2013. Tuning Parameter dalam Fungsi OKAPI BM25 pada Mesin Pencari Teks Bahasa Indonesia [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Singh V, Saini B. 2014. An effective pre-processing algorithm for information retrieval systems. *International Journal of Database Management Systems (IJDMS)*. 6(6):13-24.
- Tala FZ. 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia [tesis]. Amsterdam (NL): University of Amsterdam.

Lampiran 1 Nilai relevansi dokumen pada tiap *query* (reli)

Query*	Nilai relevansi dokumen													
	ID dokumen**													
Hak Cipta Dilindungi Undang-Undang	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	0	0	0	3	0	0	1	1	0	0	0
2	2	3	3	3	0	0	1	0	0	0	3	0	1	1
3	0	2	2	2	0	0	3	0	0	1	3	0	1	1
4	0	1	1	1	0	0	2	1	0	1	3	1	1	1
5	3	0	0	0	0	0	1	0	0	1	3	0	0	2
6	0	2	0	0	0	0	0	3	3	3	0	0	0	0
7	0	1	1	1	0	0	1	0	0	3	1	0	0	0
8	0	0	0	0	0	0	0	1	1	3	1	0	0	0
9	0	3	3	3	1	1	3	1	1	0	1	1	1	1
10	0	1	1	1	2	3	1	1	1	0	0	0	0	0
11	0	0	0	0	0	0	0	1	0	0	3	0	3	0
12	0	2	0	0	0	1	3	0	0	2	0	0	0	1
13	0	0	0	0	3	0	3	0	0	2	0	0	0	1
14	0	3	3	3	1	1	3	2	1	0	1	1	2	2
15	0	3	1	1	1	1	2	3	1	0	0	0	0	1
16	0	0	0	0	1	0	3	0	0	1	3	0	0	3
17	0	3	3	3	1	1	1	1	1	1	0	0	0	0
18	0	3	1	1	1	1	1	3	3	2	0	0	0	0
19	1	0	0	0	0	0	0	3	3	1	1	0	1	0
20	0	1	1	1	0	0	0	1	0	0	0	3	1	0

\*dengan keterangan *query* 1: akar mengkerut, 2: patah batang, 3: pembusukan batang, 4: akar kehitaman, 5: tanaman roboh, 6: daun menguning, 7: akar kemerahan, 8: akar menebal, 9: bercak hitam, 10: bercak memanjang, 11: batang membengkak, 12: daun rontok, 13: tepung putih, 14: batang berbercak hitam, 15: daun berbercak coklat, 16: batang berwarna putih, 17: daun berbercak kemerahan, 18: daun berbercak kekuningan, 19: penebalan jaringan daun, 20: daun tertutup kerak hitam.

\*\*dengan keterangan ID dokumen 1: rebah kecambah, 2: bercak daun, 3: hawar daun, 4: busuk batang, 5: penyakit tepung, 6: bercak bergaris, 7: layu fusarium, 8: cacar daun, 9: keriting daun, 10: busuk akar, 11: busuk hati, 12: embun hitam, 13: kanker batang, dan 14: ceratocystis.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural



Lampiran 2 Daftar *stopwords* Bahasa Indonesia

Term	Term	Term	Term
ada	bahkan	berikutnya	bung
adalah	bahwa	berjumlah	cara
adanya	bahwasanya	berkali-kali	caranya
adapun	baik	berkata	cenderung
agak	bakal	berkehendak	cukup
agakny	bakalan	berkeinginan	cukupkah
agar	balik	berkenaan	cukuplah
akan	banyak	berlainan	cuma
akankah	bapak	berlalu	dahulu
akhir	baru	berlangsung	dalam
akhiri	bawah	berlebihan	dan
akhirnya	beberapa	bermacam	dapat
akibat	begini	bermacam-macam	dari
aku	beginian	bermaksud	daripada
akulah	beginikah	bermula	datang
amat	beginilah	bersama	dekat
amatlah	begitu	bersama-sama	demi
anda	begitukah	bersiap	demikian
andalah	begitulah	bersiap-siap	demikianlah
antar	begitupun	bertanya	dengan
antara	bekerja	bertanya-tanya	depan
antaranya	belakang	berturut	di
apa	belakangan	berturut-turut	dia
apaan	belum	bertutur	diakhiri
apabila	belumah	berujar	diakhirinya
apakah	benar	berupa	dialah
apalagi	benarkah	besar	diantara
apatah	benarlah	betul	diantaranya
artinya	berada	betulkah	diberi
asal	berakhir	biasa	diberikan
asalkan	berakhirlah	biasanya	diberikannya
atas	berakhirnya	bila	dibuat
atau	berapa	bilakah	dibuatnya
ataukah	berapakah	bisa	didapat
ataupun	berapalah	bisakah	didatangkan
awal	berapapun	boleh	digunakan
awalnya	berarti	bolehkah	diibaratkan
bagai	berasal	bolehlah	diibaratkannya
bagaikan	berawal	buat	diingat
bagaimana	berbagai	bukan	diingatkan
bagaimanakah	berdatangan	bukankah	diinginkan
bagaimanapun	beri	bukanlah	dijawab
bagi	berikan	bukannya	dijelaskan
bagian	berikut	bulan	dijelaskannya

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 2 Lanjutan

Term	Term	Term	Term
dikarenakan	ditanya	ingat	kapanpun
dikatakan	ditanyai	ingat-ingat	karena
dikatakannya	ditanyakan	ingin	karenanya
dikerjakan	ditegaskan	inginkah	kasus
diketahui	ditujukan	inginkan	kata
diketuainya	ditunjuk	ini	katakan
dikira	ditunjuki	inikah	katakanlah
dilakukan	ditunjukkan	inilah	katanya
dilalui	ditunjukkannya	itu	ke
dilihat	ditunjuknya	itukah	keadaan
dimaksud	dituturkan	itulah	kebetulan
dimaksudkan	dituturkannya	jadi	kecil
dimaksudkannya	diucapkan	jadilah	kedua
dimaksudnya	diucapkannya	jadinya	keduanya
diminta	diungkapkan	jangan	keinginan
dimintai	dong	jangankan	kelamaan
dimisalkan	dua	janganlah	kelihatan
dimulai	dulu	jauh	kelihatannya
dimulailah	empat	jawab	kelima
dimulainya	enggak	jawaban	keluar
dimungkinkan	enggaknya	jawabnya	kembali
dini	entah	jelas	kemudian
dipastikan	entahlah	jelaskan	kemungkinan
diperbuat	guna	jelaslah	kemungkinannya
diperbuatnya	gunakan	jelasnya	kenapa
dipergunakan	hal	jika	kepada
diperkirakan	hampir	jikalau	kepadanya
diperlihatkan	hanya	juga	kesampaian
diperlukan	hanyalah	jumlah	keseluruhan
diperlukannya	hari	jumlahnya	keseluruhannya
dipersoalkan	harus	justru	keterlaluhan
dipertanyakan	haruslah	kadangkala	ketika
dipunyai	harusnya	kala	khususnya
diri	hendak	kalau	kini
dirinya	hendaklah	kalaulah	kinilah
disampaikan	hendaknya	kalaupun	kira
disebabkan	hingga	kalian	kira-kira
disebut	ia	kami	kiranya
disebutkan	ialah	kamilah	kita
disebutkannya	ibarat	kamu	kitalah
disini	ibaratkan	kamulah	kok
disinilah	ibaratnya	kan	kurang
ditambahkan	ibu	kapan	lagi
ditandaskan	ikut	kapankah	lagian

Hak Cipta Dilindungi Undang-Undang

© Hasil Cipta Milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 2 Lanjutan

Term	Term	Term	Term
lah	memungkinkan	mengucapkannya	oleh
lain	meminta	mengungkapkan	olehnya
lainnya	memintakan	menjadi	pada
lalu	memisalkan	menjawab	padahal
lama	memperbuat	menjelaskan	padanya
lamanya	mempergunakan	menuju	pak
lanjut	memperkirakan	menunjuk	paling
lanjutnya	memperlihatkan	menunjuki	panjang
lebih	mempersiapkan	menunjukkan	pantas
lewat	mempersoalkan	menunjuknya	para
lima	mempertanyakan	menurut	pasti
luar	mempunyai	menuturkan	pastilah
macam	memulai	menyampaikan	penting
maka	menaiki	menyangkut	pentingnya
makanya	menambahkan	menyatakan	per
makin	menandaskan	menyebabkan	percuma
malah	menanti	menyebutkan	perlu
malahan	menanti-nanti	menyeluruh	perlukah
manpu	menantikan	menyiapkan	perlunya
manpuhkah	menanya	merasa	pernah
mana	menanyai	mereka	persoalan
manakala	menanyakan	merekalah	pertama
manalagi	mendapat	merupakan	pertama-tama
masa	mendapatkan	meski	pertanyaan
masalah	mendatang	meskipun	pertanyakan
masalahnya	mendatangi	meyakini	pihak
masih	mendatangkan	meyakinkan	pihaknya
masihkah	menegaskan	minta	pukul
masing	mengakhiri	mirip	pula
masing-masing	mengakibatkan	misal	pun
mau	mengapa	misalkan	punya
maupun	mengatakan	misalnya	rasa
melainkan	mengatakannya	mula	rasanya
melakukan	mengenai	mulai	rata
melalui	mengerjakan	mulailah	rupanya
melihat	mengetahui	mulanya	saat
melihatnya	menggunakan	mungkin	saatnya
memang	menghendaki	mungkinkah	saja
memastikan	mengibaratkan	nah	sajalah
memberi	mengibaratkannya	naik	saling
memberikan	mengingat	namun	sama
membuat	mengingatkan	nanti	sama-sama
memerlukan	menginginkan	nantinya	sambil
memihak	mengira	nyaris	sampai
memiliki	mengucapkan	nyatanya	sampai-sampai

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 2 Lanjutan

Term	Term	Term	Term
sampaikan	seingat	semuanya	sini
sana	sejak	semula	sinilah
sangat	sejauh	sendiri	soal
sangatlah	sejenak	sendirian	soalnya
satu	sejumlah	sendirinya	suatu
saya	sekadar	seolah	sudah
sayalah	sekadarnya	seolah-olah	sudahkah
se	sekali	seorang	sudahlah
sebab	sekali-kali	sepanjang	supaya
sebabnya	sekalian	sepantasnya	tadi
sebagai	sekaligus	sepantasnyalah	tadinya
sebagaimana	sekalipun	seperlunya	tahu
sebagainya	sekarang	seperti	tahun
sebagian	sekecil	sepertinya	tak
sebaik	seketika	sepihak	tambah
sebaik-baiknya	sekiranya	sering	tambahnya
sebaiknya	sekitar	seringnya	tampak
sebaliknya	sekitarnya	serta	tampaknya
sebanyak	sekurang-	serupa	tandas
sebegini	kurangnya	sesaat	tandasnya
sebegitu	sekurangnya	sesama	tanpa
sebelum	sela	sesampai	tanya
sebelumnya	selain	sesegera	tanyakan
sebenarnya	selaku	sesekali	tanyanya
seberapa	selalu	seseorang	tapi
sebesar	selama	sesuatu	tegas
sebetulnya	selama-lamanya	sesuatunya	tegasnya
sebisanya	selamanya	sesudah	telah
sebuah	selanjutnya	sesudahnya	tempat
sebut	seluruh	setelah	tengah
sebutlah	seluruhnya	setempat	tentang
sebutnya	semacam	setengah	tentu
secara	semakin	seterusnya	tentulah
secukupnya	semampu	setiap	tentunya
sedang	semampunya	setiba	tepat
sedangkan	semasa	setibanya	terakhir
sedemikian	semasih	setidak-tidaknya	terasa
sedikit	semata	setidaknya	terbanyak
sedikitnya	semata-mata	setinggi	terdahulu
seenaknya	semaunya	seusai	terdapat
segala	sementara	sewaktu	terdiri
segalanya	semisal	siap	terhadap
segera	semisalnya	siapa	terhadapnya
seharusnya	sempat	siapakah	teringat
sehingga	semua	siapapun	teringat-ingat

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

### Lampiran 2 Lanjutan

Term	Term	Term	Term
terjadi	terus	turut	waduh
terjadilah	terutama	tutur	wah
terjadinya	tetap	tuturnya	wahai
terkira	tetapi	ucap	waktu
terlalu	tiap	ucapnya	waktunya
terlebih	tiba	ujar	walau
terlihat	tiba-tiba	ujarnya	walaupun
termasuk	tidak	umum	wong
ternyata	tidakkah	umumnya	yaitu
tersampaikan	tidaklah	ungkap	yakin
tersebut	tiga	ungkapnya	yakni
tersebutlah	tinggi	untuk	yang
tertentu	toh	usah	
tertuju	tunjuk	usai	

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 3 Term hasil proses *stemming*

Term	Term	Term	Term
abu	cma	inokium	mat
acacia	coklat	inokulum	mati
acaciae	collusom	jalin	meliola
air	comb	jamur	merah
akar	crassicarpa	jaring	mere
alami	cutting	jelek	minggu
ambil	cylindrocladium	jenis	miselium
ambrosia	damping	kaku	muda
area	daun	kanker	muka
ateloceuda	dibudidayakan	kapas	muncul
australe	die	kayu	musim
awat	digitata	kecambah	mycelia
back	dukung	kelas	nampak
bahaya	ering	kelayu	natural
bakteri	erysiphe	kelembaban	naung
bark	eucalyptus	kemarau	nekrotik
basah	faktor	kembang	nerotik
basidiomycetes	fase	kerak	normal
batang	fresh	keras	noxious
batas	fusarium	kering	nursery
bekas	gambut	kerut	nutrisi
bengkak	ganoderma	kirramyces	obligat
benjol	gantung	kondisi	oidium
bentuk	gar	kotor	open
berat	garis	kualitas	outbreak
bercak	gejala	kulit	pangkal
biak	gelap	kuning	pasir
bibit	generasi	kusam	passalora
biji	genetik	laboratorium	patah
botryodiplodia	genotipe	lahan	pathogen
botrytis	gerak	lapang	pengaruh
break	getah	lapis	perflexa
buah	gugur	layu	paeophleopspora
buka	halo	lebar	paeotrichoconis
bunga	hambat	lekuk	phellinus
buruk	hati	lignosus	philippii
busuk	hawar	limpah	phytophthora
cabang	helopelthis	lingkar	plant
cacar	hidup	lubang	plantation
campestris	hifa	luka	posisi
cendawan	hitam	lunak	pruning
cepat	honey	madu	pucat
ceratocystis	hujan	mangium	pucuk
clone	infeksi	mastoporum	pupuk

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 3 Lanjutan

Term	Term	Term	Term
puru	rontok	stool	tipis
putih	sakit	subresinosum	titik
pythium	sarang	subur	tua
ralstonia	sebar	sumber	tubuh
ranting	seedlot	tabur	tulang
rapat	segar	tahap	tumbuh
recover	sehat	tanah	tunas
rendam	serang	tanam	tutup
rentan	shoot	tangkai	ubah
resistensi	sifat	tawon	umur
rhizoctonia	singling	tebal	ungu
rhizomorfo	siram	tempel	utama
rigidoporus	solanacearum	temu	warna
ringan	spacing	tepung	waspada
robok	spaerotheca	terang	xanthomonas
rongg a	spora	tingkat	

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 4 Daftar panjang dokumen

ID dokumen	Nama dokumen	Panjang dokumen
1	Rebah kecambah	64
2	Bercak daun	65
3	Hawar daun	74
4	Busuk batang	74
5	Penyakit tepung	37
6	Bercak bergaris	50
7	Layu fusarium	79
8	Cacar daun	61
9	Keriting daun	40
10	Busuk akar	75
11	Busuk hati	72
12	Embun hitam	27
13	Kanker batang	43
14	Ceratocystis	59
Rata-rata panjang dokumen		58.57143

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Lampiran 5 Nilai NDCG pada setiap *tuning*

<i>Query</i> *	Tuning1	Tuning2	Tuning3	Tuning4	Tuning5
1	0.971727	0.971727	0.971727	0.971727	0.971727
2	0.991409	0.991409	0.991409	0.991409	0.991409
3	0.894949	0.894949	0.894949	0.894949	0.894949
4	0.893421	0.893421	0.893421	0.893421	0.893421
5	0.894661	0.894661	0.894661	0.894661	0.910642
6	0.985277	0.985277	0.985277	0.985277	0.943775
7	0.987481	0.987481	0.987481	0.987481	0.987481
8	1.000000	1.000000	1.000000	1.000000	1.000000
9	0.988527	0.988527	0.988527	0.988527	0.988527
10	0.844916	0.830102	0.830102	0.830102	0.830102
11	1.000000	1.000000	1.000000	1.000000	1.000000
12	0.707096	0.707096	0.726372	0.762779	0.762779
13	0.952451	0.952451	0.952451	0.952451	0.952451
14	0.997618	0.997618	0.997618	0.997618	0.997618
15	0.986473	0.986473	0.986473	0.986473	0.986473
16	0.995033	0.995033	0.995033	0.995033	0.995033
17	1.000000	1.000000	1.000000	1.000000	1.000000
18	0.987901	0.987901	0.987901	0.987901	0.987901
19	0.907134	0.907134	0.907134	0.907134	0.907134
20	0.827315	0.827315	0.819947	0.808293	0.808293
rata-rata	0.940669	0.939929	0.940524	0.941762	0.940486

\*dengan keterangan *query* 1: akar mengkerut, 2: patah batang, 3: pembusukan batang, 4: akar kehitaman, 5: tanaman roboh, 6: daun menguning, 7: akar kemerahan, 8: akar menebal, 9: bercak hitam, 10: bercak memanjang, 11: batang membengkak, 12: daun rontok, 13: tepung putih, 14: batang berbercak hitam, 15: daun berbercak coklat, 16: batang berwarna putih, 17: daun berbercak kemerahan, 18: daun berbercak kekuningan, 19: penebalan jaringan daun, 20: daun tertutup kerak hitam.

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 5 Lanjutan

<i>Query*</i>	Tuning6	Tuning7	Tuning8	Tuning9	Tuning10
1	0.971727	0.971727	0.971727	0.971727	0.971727
2	0.991409	0.991409	0.991409	0.991409	0.991409
3	0.894949	0.894949	0.894949	0.894949	0.894949
4	0.938133	0.893421	0.893421	0.893421	0.893421
5	0.894661	0.894661	0.894661	0.894661	0.910642
6	0.985277	0.985277	0.985277	0.943775	0.943775
7	0.891771	0.891771	0.891771	0.891771	0.891771
8	0.986489	0.986489	0.986489	0.977948	0.977948
9	0.988527	1.000000	1.000000	1.000000	1.000000
10	0.844916	0.830102	0.830102	0.830102	0.830102
11	1.000000	1.000000	1.000000	1.000000	1.000000
12	0.707096	0.707096	0.726372	0.762779	0.762779
13	0.952451	0.952451	0.952451	0.952451	0.952451
14	0.998275	0.998275	0.998275	0.998275	0.998275
15	1.000000	1.000000	0.998494	0.998494	0.998494
16	0.965356	0.965356	0.965356	0.995033	0.995033
17	1.000000	1.000000	1.000000	1.000000	1.000000
18	0.992183	0.992183	0.998653	0.998653	0.998653
19	0.874189	0.874189	0.874189	0.874189	0.874189
20	0.827315	0.827315	0.827315	0.819947	0.819947
rata-rata	0.935236	0.932834	0.934046	0.934479	0.935278

\*dengan keterangan *query* 1: akar mengkerut, 2: patah batang, 3: pembusukan batang, 4: akar kehitaman, 5: tanaman roboh, 6: daun menguning, 7: akar kemerahan, 8: akar menebal, 9: bercak hitam, 10: bercak memanjang, 11: batang membengkak, 12: daun rontok, 13: tepung putih, 14: batang berbercak hitam, 15: daun berbercak coklat, 16: batang berwarna putih, 17: daun berbercak kemerahan, 18: daun berbercak kekuningan, 19: penebalan jaringan daun, 20: daun tertutup kerak hitam.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 5 Lanjutan

<i>Query*</i>	Tuning11	Tuning12	Tuning13	Tuning14	Tuning15
1	0.971727	0.971727	0.971727	0.971727	0.971727
2	0.991408	0.991408	0.991408	0.991408	0.991408
3	0.894949	0.894949	0.894949	0.894949	0.894949
4	0.832848	0.832848	0.832848	0.832848	0.816986
5	0.894660	0.894660	0.894660	0.894660	0.910642
6	0.985276	0.985276	0.943774	0.943774	0.943774
7	0.891771	0.891771	0.891771	0.891771	0.891771
8	0.960971	0.960971	0.952430	0.952430	0.977948
9	0.988526	1.000000	1.000000	1.000000	1.000000
10	0.844915	0.830101	0.830101	0.830101	0.830101
11	1.000000	1.000000	1.000000	1.000000	1.000000
12	0.707095	0.707095	0.726372	0.762779	0.762779
13	0.952450	0.952450	0.952450	0.952450	0.952450
14	0.998274	0.998274	0.998274	0.998274	0.998274
15	1.000000	1.000000	1.000000	1.000000	0.998493
16	0.965355	0.965355	0.965355	0.965355	0.965355
17	1.000000	1.000000	1.000000	1.000000	1.000000
18	0.983098	0.983098	0.983098	0.985233	0.985233
19	0.874189	0.874189	0.874189	0.874189	0.874189
20	0.827314	0.827314	0.827314	0.827314	0.819947
rata-rata	0.928241	0.928074	0.926536	0.928463	0.929301

\*dengan keterangan *query* 1: akar mengkerut, 2: patah batang, 3: pembusukan batang, 4: akar kehitaman, 5: tanaman roboh, 6: daun menguning, 7: akar kemerahan, 8: akar menebal, 9: bercak hitam, 10: bercak memanjang, 11: batang membengkak, 12: daun rontok, 13: tepung putih, 14: batang berbercak hitam, 15: daun berbercak coklat, 16: batang berwarna putih, 17: daun berbercak kemerahan, 18: daun berbercak kekuningan, 19: penebalan jaringan daun, 20: daun tertutup kerak hitam.

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 5 Lanjutan

<i>Query*</i>	Tuning16	Tuning17	Tuning18	Tuning19	Tuning20
1	0.971727	0.971727	0.971727	0.971727	0.971727
2	0.991409	0.991409	0.991409	0.991409	0.991409
3	0.883856	0.883856	0.894949	0.894949	0.894949
4	0.832848	0.832848	0.832848	0.832848	0.832848
5	0.894661	0.894661	0.894661	0.894661	0.910642
6	0.985277	0.985277	0.943775	0.943775	0.943775
7	0.891771	0.891771	0.891771	0.891771	0.891771
8	0.960971	0.95243	0.95243	0.95243	0.95243
9	1.000000	1.000000	1.000000	1.000000	1.000000
10	0.844916	0.830102	0.830102	0.830102	0.830102
11	0.980255	0.980255	0.980255	0.980255	1.000000
12	0.707096	0.707096	0.726372	0.762779	0.762779
13	0.952451	0.952451	0.952451	0.952451	0.952451
14	0.998275	0.998275	0.998275	0.998275	0.998275
15	1.000000	1.000000	1.000000	1.000000	0.998494
16	0.965356	0.961882	0.961882	0.961882	0.961882
17	1.000000	1.000000	1.000000	1.000000	1.000000
18	0.983099	0.983099	0.983099	0.984446	0.984446
19	0.874189	0.874189	0.874189	0.874189	0.874189
20	0.827315	0.827315	0.827315	0.827315	0.827315
rata-rata	0.928241	0.928074	0.926536	0.928463	0.929301

\*dengan keterangan *query* 1: akar mengkerut, 2: patah batang, 3: pembusukan batang, 4: akar kehitaman, 5: tanaman roboh, 6: daun menguning, 7: akar kemerahan, 8: akar menebal, 9: bercak hitam, 10: bercak memanjang, 11: batang membengkok, 12: daun rontok, 13: tepung putih, 14: batang berbercak hitam, 15: daun berbercak coklat, 16: batang berwarna putih, 17: daun berbercak kemerahan, 18: daun berbercak kekuningan, 19: penebalan jaringan daun, 20: daun tertutup kerak hitam.

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## Lampiran 5 Lanjutan

<i>Query</i> *	Tuning21	Tuning22	Tuning23	Tuning24	Tuning25
1	0.971727	0.971727	0.971727	0.971727	0.971727
2	0.979172	0.991409	0.991409	0.991409	0.991409
3	0.883856	0.883856	0.883856	0.883856	0.883856
4	0.838157	0.832848	0.832848	0.832848	0.832848
5	0.894661	0.894661	0.894661	0.894661	0.910642
6	0.985277	0.943775	0.943775	0.943775	0.943775
7	0.891771	0.891771	0.891771	0.891771	0.891771
8	0.960971	0.95243	0.95243	0.95243	0.95243
9	0.988527	1.000000	1.000000	1.000000	1.000000
10	0.844916	0.830102	0.830102	0.830102	0.830102
11	0.9698	0.9698	0.9698	0.9698	0.9698
12	0.707096	0.707096	0.726372	0.762779	0.762779
13	0.952451	0.952451	0.952451	0.952451	0.952451
14	0.996927	0.996927	0.998275	0.998275	0.998275
15	1.000000	1.000000	1.000000	1.000000	0.998494
16	0.961882	0.961882	0.961882	0.961882	0.961882
17	1.000000	1.000000	1.000000	1.000000	1.000000
18	0.984446	0.984446	0.984446	0.984446	0.984446
19	0.874189	0.874189	0.874189	0.874189	0.874189
20	0.827315	0.827315	0.827315	0.827315	0.827315
rata-rata	0.925657	0.923334	0.924365	0.926186	0.92691

\*dengan keterangan *query* 1: akar mengkerut, 2: patah batang, 3: pembusukan batang, 4: akar kehitaman, 5: tanaman roboh, 6: daun menguning, 7: akar kemerahan, 8: akar menebal, 9: bercak hitam, 10: bercak memanjang, 11: batang membengkak, 12: daun rontok, 13: tepung putih, 14: batang berbercak hitam, 15: daun berbercak coklat, 16: batang berwarna putih, 17: daun berbercak kemerahan, 18: daun berbercak kekuningan, 19: penebalan jaringan daun, 20: daun tertutup kerak hitam.

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

## RIWAYAT HIDUP

Penulis dilahirkan di Depok pada tanggal 22 November 1994, anak kedua dari pasangan Bapak Agus Dwianto dan Ibu Siti Aminah. Pada tahun 2013 penulis lulus dari SMA Negeri 4 Depok dan pada tahun yang sama lulus seleksi masuk S1 IPB Departemen Ilmu Komputer melalui jalur SNMPTN. Pada September 2015 sampai dengan Januari 2016, penulis menjadi Asisten Praktikum Mata Kuliah Rangkaian Digital dan pada Februari sampai dengan Juni 2017 menjadi Asisten Praktikum Mata Kuliah Sistem Cerdas. Penulis juga pernah menjalani Praktik Kerja Lapang (PKL) pada Juli sampai dengan Agustus 2016 di PT Media Indonusa sebagai *System Developer*.

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural



## © Hak cipta milik IPB (Institut Pertanian Bogor)

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penerjemahan, atau untuk keperluan lain.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa





## © Hak cipta milik IPB (Institut Pertanian Bogor)

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penerjemahan, atau untuk keperluan lain.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa