



PENERAPAN *PSEUDO RELEVANCE FEEDBACK* PADA MESIN PENCARI *E-COMMERCE*

MUHAMMAD SYARIF



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2016**

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



@Hak cipta milik IPB University

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

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul Penerapan *Pseudo Relevance feedback* pada Mesin Pencari *E-Commerce* 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 disertasi ini.

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

Bogor, Agustus 2016

Muhammad Syarif
NIM G64120044

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



ABSTRAK

MUHAMMAD SYARIF. Penerapan *Pseudo Relevance Feedback* pada Mesin Pencari *E-Commerce*. Dibimbing oleh JULIO ADISANTOSO dan YENI HERDIYENI.

Electronic commerce menyediakan suatu fitur pencarian (mesin pencari) untuk memudahkan pengguna menemukan produk yang diinginkan. Mesin pencari menemukan dokumen berdasarkan *query* yang diberikan oleh pengguna. Tidak semua pengguna dapat memberikan kata kunci yang tepat sebagai *query*. Sehingga hasil pencarian yang dikembalikan tidak sesuai dengan yang diinginkan. *Pseudo relevance feedback* (PRF) merupakan salah satu metode temu kembali informasi yang digunakan untuk meningkatkan relevansi hasil pencarian. Penelitian ini menggunakan metode PRF untuk melakukan formulasi ulang *query* pada mesin pencari *e-commerce*. Formulasi ulang *query* dilakukan dengan menambahkan *term* pada *query* awal (perluasan *query*). Data yang digunakan merupakan dokumen produk yang dimiliki oleh PT. Global Digital Niaga. Penerapan PRF untuk perluasan *query* berdasarkan atribut *categories* memberikan hasil MAP yang lebih tinggi yaitu 0.6836 dan nilai MAP tanpa PRF yaitu 0.5990. Penerapan PRF berdasarkan atribut *name* dan *description* menghasilkan nilai MAP paling rendah yaitu 0.2940.

Kata kunci: *e-commerce*, perluasan *query*, *pseudo relevance feedback*, relevansi, mesin pencari.

ABSTRACT

MUHAMMAD SYARIF. Implementation of *Pseudo Relevance Feedback* in *E-Commerce Search Engine*. Supervised by JULIO ADISANTOSO and YENI HERDIYENI.

Electronic commerce provides a product search feature (search engine) to facilitate users to find a specific product. The search engine finds documents according to user's *query*. Not all user can formulate a good keyword as the *query*. It makes the search results irrelevant with the user's need. *Pseudo relevance feedback* (PRF) is one of information retrieval methods that is used to increase results relevance. This research proposed *pseudo relevance feedback* to reformulate *query* that implemented in *e-commerce search engine*. Data for this research is products document from PT. Global Digital Niaga. Implementation of PRF to expand *query* by product categories attribute give a MAP value of 0.6836 higher than MAP value without PRF (0.5990). The implementation of PRF to expand *query* by product name and description attribute give a lowest MAP value of 0.2940.

Keywords: *e-commerce*, *pseudo relevance feedback*, *query expansion*, *relevance*, *search engine*.



PENERAPAN *PSEUDO RELEVANCE FEEDBACK* PADA MESIN PENCARI *E-COMMERCE*

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
2016
MUHAMMAD SYARIF**

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



@Hak cipta milik IPB University

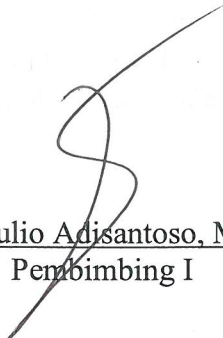
Hak Cipta Dilindungi Undang-undang

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


Judul Skripsi: Penerapan *Pseudo Relevance Feedback* pada Mesin Pencari *E-Commerce*

Nama : Muhammad Syarif
NIM : G64120044

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:

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



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah SWT atas nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan karya ilmiah ini. Penelitian mengambil tema temu kembali informasi dengan judul Penerapan *Pseudo Relevance Feedback* pada Mesin Pencari *E-Commerce* dan dilaksanakan sejak bulan November 2015.

Terima kasih penulis ucapkan kepada kepada ayah, ibu, serta seluruh keluarga, atas segala doa dan kasih sayangnya. Ungkapan terima kasih juga penulis sampaikan kepada Bapak Ir Julio Adisantoso, MKom dan Ibu Dr Yeni Herdiyeni, SSi Mkom selaku pembimbing, serta Bapak Muhammad Abrar Istiadi, SKomp MKom selaku penguji. Disamping itu, ucapan terima kasih penulis sampaikan kepada PT. Global Digital Niaga, Muhammad Ammar Imron, Rijen J P Sianturi, dan Rizki Adi Utomo atas kerjasamanya dalam penelitian yang telah dilaksanakan ini. Penulis juga mengucapkan terima kasih kepada Bapak Ifnu Bima dan Bapak Ronald Prasetya yang membimbing pengembangan sistem. Penulis juga mengucapkan terima kasih kepada Vector Ilmu Komputer IPB angkatan 49 dan Forum Komunikasi Mahasiswa Boyolali (FKMB) serta Keluarga Mahasiswa Nahdlatul Ulama IPB (KMNU IPB) khususnya angkatan 49 atas doa dan dukungan yang diberikan kepada penulis.

Semoga karya ilmiah ini bermanfaat.

Bogor, Agustus 2016

Muhammad Syarif

DAFTAR ISI

DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
DAFTAR LAMPIRAN	vi
PENDAHULUAN	1
Latar Belakang	1
Perumusan Masalah	2
Tujuan Penelitian	2
Manfaat Penelitian	2
Ruang Lingkup Penelitian	2
METODE	3
Pembuatan Indeks	4
<i>Scoring</i>	5
Inisialisasi <i>Query</i>	7
Formulasi Ulang <i>Query</i>	8
Evaluasi	9
HASIL DAN PEMBAHASAN	10
<i>Scoring</i>	10
Inisialisasi <i>Query</i>	11
Formulasi Ulang <i>Query</i>	11
Evaluasi <i>Pseudo Relevance Feedback</i>	12
SIMPULAN DAN SARAN	15
Simpulan	15
Saran	16
DAFTAR PUSTAKA	16
LAMPIRAN	18
RIWAYAT HIDUP	28

DAFTAR TABEL

1	Tabel <i>contingency</i>	9
2	MAP PRF berdasarkan atribut <i>name</i> dan <i>description</i>	12
3	MAP PRF berdasarkan atribut <i>categories</i>	13
4	Perbandingan MAP tanpa PRF, PRF berdasarkan <i>name</i> dan <i>description</i> , dan PRF berdasarkan <i>categories</i>	14

DAFTAR GAMBAR

1	Tahapan penelitian	3
2	Pengambilan <i>top-k</i> dokumen pada temu kembali awal	7
3	Pencarian dokumen dengan <i>query</i> monitor	11
4	Grafik 11 titik <i>recall</i> PRF berdasarkan atribut <i>name</i> dan <i>description</i>	13
5	Grafik 11 titik <i>recall</i> PRF berdasarkan atribut <i>categories</i>	14
6	Grafik 11 titik <i>recall</i> tanpa PRF, PRF berdasarkan atribut <i>name</i> dan <i>description</i> , dan PRF berdasarkan atribut <i>categories</i>	15

DAFTAR LAMPIRAN

1	Daftar <i>stopwords</i>	18
2	Perbandingan nilai <i>precision</i> yang diinterpolasikan ke dalam 11 titik <i>recall</i> ketika menggunakan atribut <i>name</i> dan <i>description</i> sebagai PRF	20
3	Kategori dokumen produk peringkat 24 teratas pada temu kembali awal	20
4	Perbandingan nilai <i>precision</i> yang diinterpolasikan ke dalam 11 titik <i>recall</i> ketika menggunakan atribut <i>categories</i> sebagai PRF	22
5	Formulasi ulang <i>query</i> berdasarkan atribut <i>categories</i> dengan <i>top-k</i> 5 dokumen	23
6	Formulasi ulang <i>query</i> berdasarkan atribut <i>name</i> dan <i>description</i> dengan <i>top-k</i> 10 dokumen	24
7	Dokumen hasil temu kembali dengan menggunakan <i>query</i> monitor	25
8	Perbandingan nilai <i>precision</i> yang diinterpolasikan ke dalam 11 titik <i>recall</i> ketika tanpa menggunakan PRF, PRF menggunakan atribut <i>categories</i> , dan PRF menggunakan atribut <i>name</i> dan <i>description</i>	27