

**OPTIMASI MESIN PENCARI DENGAN RE-RANKING HASIL  
PENCARIAN PRODUK BERDASARKAN KLIK USER  
STUDI KASUS: BLIBLI.COM**

Skripsi



oleh

**JUAN CHRISTIAN TJANDRA**

**71140006**

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA

2018

**OPTIMASI MESIN PENCARI DENGAN RE-RANKING HASIL  
PENCARIAN PRODUK BERDASARKAN KLIK USER  
STUDI KASUS: BLIBLI.COM**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai Salah Satu Syarat dalam Memperoleh Gelar  
Sarjana Komputer

Disusun oleh  
**JUAN CHRISTIAN TJANDRA**  
**71140006**

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA

2018

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

**OPTIMASI MESIN PENCARI DENGAN RE-RANKING HASIL  
PENCARIAN PRODUK BERDASARKAN KLIK USER  
STUDI KASUS : BLIBLI.COM**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 3 Agustus 2018



JUAN CHRISTIAN TJANDRA

71140006

## HALAMAN PERSETUJUAN

Judul Skripsi : OPTIMASI MESIN PENCARI DENGAN RE-RANKING HASIL PENCARIAN PRODUK BERDASARKAN KLIK USER  
STUDI KASUS : BLIBLI.COM

Nama Mahasiswa : JUAN CHRISTIAN TJANDRA

N I M : 71140006

Matakuliah : Skripsi (Tugas Akhir)

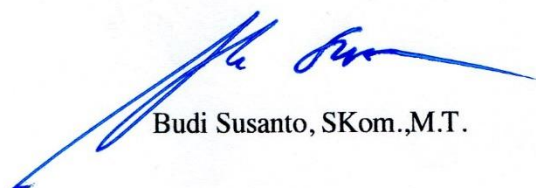
Kode : TIW276

Semester : Genap

Tahun Akademik : 2017/2018

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 3 Agustus 2018

Dosen Pembimbing I



Budi Susanto, SKom., M.T.

Dosen Pembimbing II



Gloria Virginia, S.Kom., MAI, Ph.D.

## HALAMAN PENGESAHAN

### OPTIMASI MESIN PENCARI DENGAN RE-RANKING HASIL PENCARIAN PRODUK BERDASARKAN KLIK USER STUDI KASUS : BLIBLI.COM

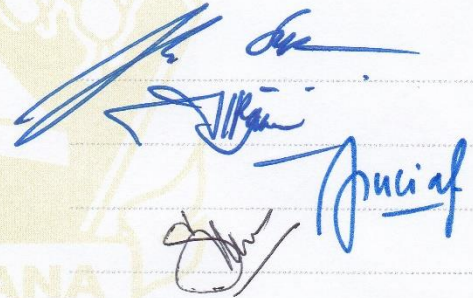
Oleh: JUAN CHRISTIAN TJANDRA / 71140006

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 26 Juli 2018

Yogyakarta, 3 Agustus 2018  
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom., M.T.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. Lucia Dwi Krisnawati, Dr. Phil.
4. Hendro Setiadi, M.Eng



Dekan



(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi



(Gloria Virginia, Ph.D.)

## UCAPAN TERIMAKASIH

Pertama-tama Penulis mengucapkan syukur kepada Tuhan karena atas berkat-Nya, Penulis dapat menyelesaikan tugas akhir dengan judul “Optimasi Mesin Pencari dengan *Re-Ranking* Hasil Pencarian Produk Berdasarkan Klik User Studi Kasus: Blibli.com”.

Banyak pihak yang terlibat dalam mendukung dan membantu Penulis dalam menyelesaikan tugas akhir ini. Maka dari itu, Penulis ingin mengucapkan banyak terima kasih kepada:

1. Drs. Yani S. Tjandra dan Nelfazia Devi selaku orang tua Penulis yang selalu mendoakan saya agar Penulis dapat menyelesaikan tugas akhir dan studinya.
2. Bapak Budi Susanto, S. Kom., M.T. selaku dekan Fakultas Teknologi Informasi sekaligus dosen pembimbing Penulis yang membimbing Penulis dalam tugas akhir ini dengan sabar.
3. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku Ketua Program Studi sekaligus dosen pembimbing kedua Penulis yang tidak lelah dalam memberikan bimbingan dan ide pada proses pengerjaan tugas akhir Penulis.
4. Bapak Ifnu Bima Fatkhan selaku *Senior Software Development Engineer* PT. Global Digital Niaga atau Blibli.com yang memberikan ide atau topik pada penelitian ini.
5. Teman-teman Blibli Future Program yang menjadi motivasi Penulis dalam menyelesaikan Tugas Akhir ini.
6. Eunike Indriani selaku HRD PT. Global Digital Niaga atau Blibli.com yang selalu memberi semangat dan mendengarkan keluh kesah Penulis.
7. Teman-teman Mockingjay di Solo yang selalu memberikan semangat pada Penulis.
8. Teman-teman kuliah Penulis yang juga memberikan semangat pada Penulis.

Akhir kata, Penulis sekali lagi mengucapkan banyak terimakasih kepada pihak-pihak yang disebutkan di atas dan juga pihak-pihak yang mungkin tidak bisa Penulis bisa sebutkan satu per satu. Semoga hasil dari penelitian ini dapat berguna bagi pihak yang terkait dan juga bagi pembaca.

©UKDW

## INTISARI

### OPTIMASI MESIN PENCARI DENGAN *RE-RANKING* HASIL PENCARIAN PRODUK BERDASARKAN KLIK *USER*

Studi Kasus: Blibli.com

Mesin pencari adalah hal yang sangat dibutuhkan untuk sebuah *e-commerce* dalam pemasaran produk yang ingin dijual. Blibli.com adalah salah satu *e-commerce* terbesar di Indonesia dengan model bisnis B2B2C (*Business to Business to Customer*). Untuk memberikan pelayanan yang terbaik, salah satu usaha yang bisa dilakukan adalah dengan memberi kemudahan calon pembeli dalam pencarian barang. Dalam pencarian produk, Blibli.com menggunakan mesin pencari solr dengan fungsi *scoring similarity* BM25. Data klik *user* bisa digunakan untuk *re-ranking* ulang hasil pencarian agar urutan atau *ranking* hasil pencarian yang dikeluarkan menjadi lebih baik.

Algoritma yang melibatkan klik *user* dalam melakukan *ranking* ulang hasil pencarian dapat diterapkan pada mesin pencari Solr. Implementasi algoritma *clickthrough data* dapat dibentuk menjadi sebuah *Plugin* yang dapat dipakai pada mesin pencari *solr*. Penelitian ini akan berpusat pada penerapan algoritma yang melibatkan klik *user* pada mesin pencari *solr*.

Penelitian berhasil memberikan kenaikan nilai *Mean Average Precision* sebesar 0.2122213703 dari 0.5886318369 sebelum penerapan algoritma *clickthrough data* menjadi 0.8008532072 setelah penerapan algoritma *clickthrough data*. Dapat disimpulkan bahwa algoritma *clickthrough data* dapat menaikkan kualitas *ranking* hasil pencarian produk Blibli.com pada mesin pencari *solr*. Kekurangan dari sistem ini adalah masih banyak barang yang tidak relevan tetap dikembalikan pada mesin pencari *solr* walaupun *ranking* atau posisinya cenderung berada di bawah dokumen yang relevan.

**Kata Kunci:** Solr, Solr Plugin, Re-ranking, Information Retrieval



## DAFTAR ISI

<b>PERNYATAAN KEASLIAN SKRIPSI</b>	<b>iii</b>
<b>HALAMAN PERSETUJUAN</b>	<b>iv</b>
<b>HALAMAN PENGESAHAN</b>	<b>v</b>
<b>UCAPAN TERIMAKASIH</b>	<b>vi</b>
<b>INTISARI</b>	<b>viii</b>
<b>DAFTAR ISI</b>	<b>ix</b>
<b>DAFTAR GAMBAR</b>	<b>xii</b>
<b>DAFTAR TABEL</b>	<b>xiii</b>
<b>BAB I</b>	
<b>PENDAHULUAN</b>	<b>1</b>
Latar Belakang	1
Rumusan Masalah	2
Batasan Masalah	2
Tujuan Penelitian	2
Manfaat Penelitian	2
Metodologi Penelitian	3
Metode Pengumpulan Data	3
Metode Pengembangan Sistem	3
Metode Evaluasi	4
Sistematika Penulisan	4
<b>BAB II</b>	
<b>TINJAUAN PUSTAKA</b>	<b>6</b>
Tinjauan Pustaka	6
Landasan Teori	7
Mesin Pencari Solr	7
Booster	7
Query Parser	7
Re-rank Query Parser	8
Information Retrieval di Mesin Pencari	8
Proses Indexing	8
Query Processing	9
TF/IDF	9

BM25	11
Re-Ranking	13
Algoritma Clickthrough Data	15
Teknik Penentuan Jumlah Sampel dan Pengambilan Sampel	16
Rumus Slovin	17
Bootstrap Sampling	17
Teknik Pengukuran Kappa	18
Metode Evaluasi Precision dan Recall	19
Metode Evaluasi Mean Average Precision (MAP)	20
Precision Pada n Dokumen Teratas (P@n)	20
Mean Average Precision (MAP)	21
<b>BAB III</b>	
<b>PERANCANGAN SISTEM</b>	<b>22</b>
Spesifikasi Perangkat	22
Spesifikasi Perangkat Keras	22
Spesifikasi Perangkat Lunak	22
Spesifikasi Server Amazon Web Service (AWS)	23
Perancangan Proses Pengumpulan Data	23
Perancangan Proses Pengumpulan Data Sampel Produk	23
Perancangan Proses Pengumpulan Data Label Produk	24
Perancangan Proses Pengumpulan Data Klik User	24
Perancangan Sistem	25
ClickthroughData Query Parser Plugin	25
Clickthroughdata Web Service	26
Search Platform	26
Perancangan Desain Antarmuka Sistem	27
Perancangan Pengujian Sistem	28
<b>BAB IV</b>	
<b>IMPLEMENTASI DAN ANALISIS SISTEM</b>	<b>29</b>
Implementasi Sistem	29
Implementasi Pengumpulan Data	29
Data Produk	29
Label	33
Klik User	35
Implementasi Algoritma Clickthrough Data	35
ClickthroughDataQParserPlugin	36
Clickthrough Data Web Service	38
Search Platform	39

Server Solr	40
Pembuatan Plugin dan Pengaturan pada Solr	41
Implementasi Antarmuka Sistem	43
Analisis dan Hasil Pengujian Sistem	44
<b>BAB V</b>	
<b>KESIMPULAN DAN SARAN</b>	<b>53</b>
Kesimpulan	53
Saran	53
<b>DAFTAR PUSTAKA</b>	<b>55</b>
<b>LAMPIRAN</b>	<b>58</b>

©UKDW

## DAFTAR GAMBAR

Gambar 2.1. Algoritma Pembobotan pada Mesin Pencari di Etsy	15
Gambar 2.2. Algoritma clickthrough data oleh Paulsson (2017)	16
Gambar 3.1. Diagram Sistem	25
Gambar 3.2. Use case diagram clickthroughdata Web Service	26
Gambar 3.3. Desain antarmuka aplikasi search platform	27
Gambar 4.1. Task pada google form untuk mengumpulkan klik user	36
Gambar 4.2. Alur kerja clickthroughdataQParserPlugin	37
Gambar 4.3. Struktur tabel basis data klik user	38
Gambar 4.4. Spesifikasi API mengambil klik berdasarkan query	39
Gambar 4.5. Spesifikasi API untuk mengumpulkan klik	39
Gambar 4.6. Flowchart pengumpulan data klik user	41
Gambar 4.7. Pengaturan build pom.xml pada maven project	47
Gambar 4.8. Aplikasi Search Platform Bagian Atas	43
Gambar 4.9. Aplikasi Search Platform Bagian Bawah	43
Gambar 4.10. Kurva hasil presisi pada query 1	44
Gambar 4.11. Kurva hasil presisi pada query 2	45
Gambar 4.12. Kurva hasil presisi pada query 3	45
Gambar 4.13. Kurva hasil presisi pada query 4	46
Gambar 4.14. Kurva hasil presisi pada query 5	46

Gambar 4.15. Kurva hasil presisi pada query 6	47
Gambar 4.16. Kurva hasil presisi pada query 7	47
Gambar 4.17. Kurva hasil presisi pada query 8	48

©UKDW

## DAFTAR TABEL

Tabel 2.1. Tabel Jumlah Label Terhadap Dokumen	19
Tabel 2.2. Tabel precision pada n dokumen teratas	21
Tabel 4.1. Jumlah Data Produk pada Setiap Kategori	29
Tabel 4.2. Jumlah relevan dokumen terhadap query	34
Tabel 4.3. Hasil Kesepakatan Kappa dari 5 Orang Juri	35
Tabel 4.4. Tabel Hasil Average Precision dan MAP	49
Tabel 4.5. Precision dan Recall pada Halaman Pertama dan Kedua	50
Tabel 4.6. Precision dan Recall pada Semua Dokumen yang Dikembalikan	51

## INTISARI

### OPTIMASI MESIN PENCARI DENGAN *RE-RANKING* HASIL PENCARIAN PRODUK BERDASARKAN KLIK *USER*

Studi Kasus: Blibli.com

Mesin pencari adalah hal yang sangat dibutuhkan untuk sebuah *e-commerce* dalam pemasaran produk yang ingin dijual. Blibli.com adalah salah satu *e-commerce* terbesar di Indonesia dengan model bisnis B2B2C (*Business to Business to Customer*). Untuk memberikan pelayanan yang terbaik, salah satu usaha yang bisa dilakukan adalah dengan memberi kemudahan calon pembeli dalam pencarian barang. Dalam pencarian produk, Blibli.com menggunakan mesin pencari solr dengan fungsi *scoring similarity* BM25. Data klik *user* bisa digunakan untuk *re-ranking* ulang hasil pencarian agar urutan atau *ranking* hasil pencarian yang dikeluarkan menjadi lebih baik.

Algoritma yang melibatkan klik *user* dalam melakukan *ranking* ulang hasil pencarian dapat diterapkan pada mesin pencari Solr. Implementasi algoritma *clickthrough data* dapat dibentuk menjadi sebuah *Plugin* yang dapat dipakai pada mesin pencari *solr*. Penelitian ini akan berpusat pada penerapan algoritma yang melibatkan klik *user* pada mesin pencari *solr*.

Penelitian berhasil memberikan kenaikan nilai *Mean Average Precision* sebesar 0.2122213703 dari 0.5886318369 sebelum penerapan algoritma *clickthrough data* menjadi 0.8008532072 setelah penerapan algoritma *clickthrough data*. Dapat disimpulkan bahwa algoritma *clickthrough data* dapat menaikkan kualitas *ranking* hasil pencarian produk Blibli.com pada mesin pencari *solr*. Kekurangan dari sistem ini adalah masih banyak barang yang tidak relevan tetap dikembalikan pada mesin pencari *solr* walaupun *ranking* atau posisinya cenderung berada di bawah dokumen yang relevan.

**Kata Kunci:** Solr, Solr Plugin, Re-ranking, Information Retrieval

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Mesin pencari (*search engine*) adalah hal yang sangat dibutuhkan untuk sebuah *e-commerce* dalam pemasaran produk yang ingin dijual. *Search engine* pada situs *e-commerce* yang baik tentu adalah *search engine* yang bisa memberikan hasil pencarian yang diinginkan oleh *customer*-nya.

Blibli.com adalah salah satu *e-commerce* terbesar di Indonesia dengan model bisnis B2B2C (*Business to Business to Customer*). Berdiri sebagai perusahaan di bidang *e-commerce*, Blibli.com mempunyai *core values* yang bernama RESPECT. Di dalam *core values* RESPECT tersebut, ada dua values yaitu *customer focus* dan *servicing*. Pelayanan terhadap *customer* Blibli.com menjadi hal yang sangat diperhatikan oleh perusahaan.

Untuk memberikan pelayanan yang terbaik, salah satu usaha yang bisa dilakukan adalah dengan memberi kemudahan calon pembeli dalam pencarian barang. Dalam memberikan kemudahan pencarian barang, tentunya diperlukan *search engine* yang bisa memberikan hasil yang diinginkan oleh pembeli.

Dalam pencarian produk, Blibli.com menggunakan mesin pencari solr dengan fungsi *scoring similarity* BM25. Pada mesin solr yang dipakai Blibli.com, faktor klik dari *user* belum dipakai sebagai pertimbangan dalam *ranking* dokumen. Solr mendukung *re-ranking* hasil pencarian dengan sistem pembobotan berdasarkan beberapa faktor termasuk data klik *user*.

Data klik *user* bisa digunakan untuk *re-ranking* ulang hasil pencarian agar urutan atau *ranking* hasil pencarian yang dikeluarkan menjadi lebih baik (Joachims, 2002). Algoritma *re-ranking* dibuat berdasarkan data klik yang pernah dilakukan oleh *user* agar hasil pencarian barang yang diberikan bisa lebih sesuai dengan harapan *user*. Ketika algoritma ini digunakan, semakin mesin pencari ini sering dipakai, maka hasil pencarian pun semakin relevan (Paulsson, 2017).



## 1.2. Rumusan Masalah

Latar belakang penelitian tersebut menunjukkan bahwa penting untuk dilakukan *re-ranking* terhadap fungsi pembobotan yang sudah ada. Masalahnya, apakah dengan penerapan algoritma *re-ranking* berdasarkan klik *user* pada mesin pencari *solr* dapat dihasilkan *ranking* hasil pencarian produk yang lebih baik? Metode evaluasi *Mean Average Precision* dan *Precision* pada  $n$  dokumen teratas dapat digunakan untuk mengukur kualitas *ranking* yang dihasilkan sistem.

## 1.3. Batasan Masalah

Batasan masalah yang akan dibuat dalam penelitian ini adalah sebagai berikut:

- Data yang dipakai uji coba adalah data sampel produk Blibli.com yang diambil dari 10000 data produk Blibli.com pada tahun 2017
- *Clickthrough data* adalah data klik yang dikumpulkan dari percobaan responden dalam menggunakan sistem selama seminggu
- Proses pemberian label dilakukan oleh 5 orang juri

## 1.4. Tujuan Penelitian

Tujuan penelitian ini adalah membuat *plugin* pada *solr* untuk membobot ulang hasil pencarian yang melibatkan data klik *user*.

## 1.5. Manfaat Penelitian

Manfaat dari penelitian yang dibuat adalah sebagai berikut:

1. Hasil search yang diberikan lebih sesuai dengan harapan *user*
2. Peningkatan pembelian barang

## 1.6. Metodologi Penelitian

Dalam penyelesaian penelitian ini, terdapat beberapa metodologi penelitian, yaitu metode pengumpulan data, metode pengembangan sistem, dan metode evaluasi

### 1.6.1. Metode Pengumpulan Data

Data produk yang diperoleh merupakan data yang diperoleh dari Blibli.com dan berupa kumpulan produk yang berbentuk JSON. Setiap produk berisi keterangan nama produk, nama *brand*, kategori, detail produk, *url*, deskripsi, harga, keterangan diskon, harga setelah diskon, *merchant id*. Data riwayat dokumen yang pernah di klik *user* dan *query* yang pernah dilakukan *user* diperoleh dengan cara mencari responden untuk menggunakan sistem pencarian selama 1 minggu dengan *query* atau *search keyword* yang sudah ditentukan sebelumnya.

Jumlah produk yang dipakai adalah 1386 produk. Jumlah tersebut didapat dari teknik penentuan jumlah sampel dengan rumus Slovin dari 10000 data produk Blibli.com yang diberikan pada tahun 2017. Pengambilan sampel dari 10000 menjadi 1386 produk menggunakan teknik *bootstrap sampling*.

Pemberian label pada 1386 produk yang dipakai untuk uji coba dilakukan secara manual oleh 5 orang. Hasil pelabelan 5 orang tersebut untuk setiap produknya akan dipilih label yang paling banyak. Tingkat kesepakatan kelima orang juri tersebut akan diukur juga dengan menggunakan Kappa.

### 1.6.2. Metode Pengembangan Sistem

Sistem akan dikembangkan sesuai dengan langkah-langkah berikut:

1. Pengembangan data klik dan produk yang memuat proses pengumpulan data, pembuatan korpus, dan pembuatan basis data klik *user*

2. Pembuatan *plugin* yang memuat algoritma *clickthrough data* (Paulsson, 2017)
3. Pelatihan sistem dengan pengumpulan data klik *user*
4. Evaluasi menggunakan metode  $P@n$  (*Precision* pada  $n$  dokumen teratas) dan *MAP* (*Mean Average Precision*)

### 1.6.3. Metode Evaluasi

Evaluasi dilakukan dengan cara pengujian terhadap sistem, apakah dokumen atau produk yang diberikan mesin pencari memiliki kualitas *ranking* yang baik dengan menggunakan metode *Precision* pada  $n$  dokumen teratas ( $P@n$ ) dan *MAP* (*Mean average Precision*). Menurut Manning, Raghavan, dan Schütze pada bukunya yang berjudul *Introduction to Information Retrieval* (2009), kedua *metrics* tersebut cocok digunakan untuk mengukur kualitas *ranking* hasil pencarian pada *search engine* yang mempunyai dokumen banyak.

Pada penelitian ini juga dihitung *precision* dan *recall* secara keseluruhan untuk mengetahui tingkat relevansi dari hasil pencarian mesin pencari solr dengan tambahan *plugin* yang dibangun.

### 1.7. Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini dibagi menjadi beberapa bab sebagai berikut :

Bab 1 merupakan PENDAHULUAN, berisi latar belakang masalah yang akan diteliti dan rencana penelitian yang akan dilakukan.

Bab 2 merupakan TINJAUAN PUSTAKA yang memuat uraian dari konsep-konsep atau teori-teori yang digunakan.

Bab 3 merupakan PERANCANGAN SISTEM yang memuat bagaimana sistem akan dibangun.

Bab 4 merupakan IMPLEMENTASI DAN ANALISIS SISTEM dimana bab ini menjelaskan bagaimana implementasi yang dibuat dan analisis dari sistem yang dibuat.

Dan Bab 5 adalah KESIMPULAN DAN SARAN yang berisi kesimpulan dari penelitian ini dan saran untuk pengembangan sistem dan penelitian kedepannya.

©UKDW

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil analisis yang telah ditulis pada Subbab 4.2, dapat ditarik kesimpulan bahwa *plugin* dengan penerapan algoritma *clickthrough data* dapat memberikan hasil *ranking* yang lebih baik. Nilai *Mean Average Precision* sebelum penambahan *plugin* sebesar 0,5886. Setelah penambahan *plugin*, nilai *MAP* naik senilai 0,2123 menjadi 0,8009.

Kekurangan pada penelitian ini adalah sebagai berikut:

1. Pembentukan korpus yang kurang baik. Jumlah dokumen terlabel relevan antar *query* tidak seimbang.
2. Jumlah *query* untuk evaluasi kurang banyak
3. Meskipun sistem ini dapat mengeluarkan seluruh dokumen yang relevan terhadap suatu *query* dengan nilai *recall* 1 tetapi nilai *precision* nya terbilang buruk. Ini menunjukkan bahwa sistem ini juga mengeluarkan banyak barang yang diberi label tidak relevan.

Karena fokus penelitian ini adalah peningkatan ranking yang sudah dijawab pada naiknya *MAP*, masalah dan kekurangan pada penelitian ini bisa dipakai untuk dijadikan penelitian selanjutnya.

#### 5.2. Saran

Sistem yang dibuat masih bisa dikembangkan menjadi lebih baik. Banyak sisi yang dapat dikembangkan dari sistem yang dibangun. Oleh karena hal tersebut dan juga adanya kekurangan pada penelitian ini, penulis memberikan beberapa saran untuk pengembangan sistem ini. Saran tersebut adalah sebagai berikut:

1. Algoritma *clickthrough data* yang dipakai dapat diganti dengan penerapan *Learning to Rank* karena algoritma *Learning to Rank* tidak hanya sekedar menambahkan skor ke dalam hasil *BM25*, tapi algoritma ini membuat

model baru dengan belajar dari *user behaviour* atau kebiasaan *user*. Solr sudah mendukung penerapan daripada algoritma *learning to rank* ini.

2. Algoritma *clickthrough data* yang diimplementasikan belum memperhatikan keterkaitan sebuah *term* dari suatu *query* yang diberikan terhadap klik *user*. Jika *user* melakukan klik terhadap dokumen yang tidak relevan dengan *query* yang diberikan, dokumen dan *query* tersebut tetap dimasukkan ke dalam basis data *click*. Kemudian saat algoritma *clickthrough data* dijalankan, penambahan nilai klik *user* akan tetap dilakukan terhadap dokumen yang mungkin tidak relevan terhadap *query* atau *term* tersebut. Untuk mengatasi hal itu, maka algoritma *clickthrough data* ini dapat dikembangkan lagi dengan cara memberikan analisis hubungan sebuah *term* dari *query* terhadap klik *user* yang dilakukan.
3. *Booster* yang digunakan pada penelitian ini adalah *booster* yang dipakai oleh Blibli.com sekarang ini. Penyesuaian seperti apa *booster* yang baik terhadap studi kasus produk Blibli.com ini bisa menjadi topik penelitian yang baru.
4. Pada sistem *clickthrough data web service*, basis data *user click* dapat dipindahkan ke dalam struktur penyimpanan yang lebih ringan dibanding *relational database mysql*. Struktur penyimpanan yang dimaksud bisa berupa file log pada mesin pencari *solr*. Dengan menyimpan data *user* langsung pada *solr*, mungkin dapat meningkatkan performa dan mengurangi ketergantungan (*dependent*) antar sistem.
5. Pembuatan korpus produk Blibli.com yang baik dapat dijadikan penelitian selanjutnya supaya ketika ingin meneliti algoritma pembobotan baru, hasil evaluasinya bisa lebih baik lagi.

## DAFTAR PUSTAKA

- Abdillah, A. A., Murfi, H., & Satria, Y. (2015). Uji Kinerja Learning to Rank dengan Metode Support Vector Regression. *IndoMS Journal on Industrial and Applied Mathematics*, 2(1), 15-25.
- Belkin, N. J. (1993). Interaction with texts: Information retrieval as information seeking behavior. *Information retrieval*, 93, 55-66.
- Bendersky, M., Wang, X., Metzler, D., & Najork, M. (2017, February). Learning from user interactions in personal search via attribute parameterization. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 791-799). ACM.
- Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2), 249-254.
- Condon, F. (2015). *Nice Docs Finish First: Designing Search Ranking for Fairness at Etsy*.
- Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines*. Boston, MA: AddisonWesley Professional.
- Joachims, T. (2002, July). Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 133-142). ACM.
- Kanhabua, N., & Nørnvåg, K. (2010, September). *Determining Time of Queries for Re-ranking Search Results*. In *ECDL* (Vol. 10, pp. 261-272).

- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *Introduction to information retrieval*. Cambridge: Cambridge University Press.
- Paulsson, A. (2017). *Using clickthrough data to optimize search result ranking: An evaluation of clickthrough data in terms of relevancy and efficiency*.
- Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133-142).
- Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4), 333-389.
- Shahi, D. (2016). *Apache Solr*. Apress.
- Setiawan, N. (2007). *Penentuan ukuran sampel memakai rumus slovin dan tabel krejcie-morgan: telaah konsep dan aplikasinya*. Abstrak.
- Smiley, D., Pugh, E., Parisa, K., & Mitchell, M. (2015). *Apache Solr enterprise search server*. Packt Publishing Ltd.
- Stephanie. (2017, October 17). Bootstrap Sample: Definition, Example. Retrieved from <http://www.statisticshowto.com/bootstrap-sample/>
- William, H. (2010, November 10). Measuring Search Relevance [Blog Post]. Retrieved from <https://www.ebayinc.com/stories/blogs/tech/measuring-search-relevance/>.
- Taylor, C. (2018, June 27). What Is Bootstrapping in Statistics? Retrieved from <https://www.thoughtco.com/what-is-bootstrapping-in-statistics-3126172>



Zhuang, Z., & Cucerzan, S. (2006, November). Re-ranking search results using query logs. In *Proceedings of the 15th ACM international conference on Information and knowledge management* (pp. 860-861). ACM.

©UKDW