

ANALISIS PENGUKURAN EFEKTIVITAS SEARCH ENGINE

Rafika¹, Yanuar Firdaus A.w.², Retno Novi Dayawati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Setiap search engine memiliki algoritma dan caranya sendiri untuk menghasilkan query result. Permasalahannya adalah sejauh mana query result yang dihasilkan oleh search engine yang document collectionnya berupa freetext document relevan dengan informasi yang diinginkan oleh user berdasarkan query inputan berupa Common Query Language. Dalam tugas akhir ini, Common Query Language yang digunakan adalah simple query dan boolean query. Karena keterbatasan untuk mengetahui cara dan bagaimana search engine menghasilkan query result, untuk menguji kehandalan dari query result tersebut dapat dilakukan dengan cara menganalisis efektivitas dari masing-masing search engine. Efektivitas dapat diukur dari beberapa parameter pengukuran diantaranya adalah precision. Precision adalah jumlah kelompok dokumen relevan dari jumlah dokumen yang ditemukan oleh system. Selain itu diperlukan sebuah penilaian subjektif yang dilakukan oleh user yang disebut dengan satisfaction. Satisfaction ini merupakan questioner yang akan menganalisis secara langsung interaksi antara user dengan search engine.

Dari hasil analisis hubungan antara nilai precision yang diperoleh dari sistem dan hasil satisfaction user dapat disimpulkan bahwa, baik simple maupun boolean query secara umum, Google dapat dikatakan memiliki efektivitas yang lebih baik daripada Yahoo.

Kata Kunci : Search engine. Free-Text Document, Common Query Language, Precision, Satisfaction.

Abstract

Each search engine has its own algorithm and how to produce a query result. The problem is the extent to which a query result generated by search engines which document collectionnya free-form text document with relevant information desired by the user's query input form berdasarkan Common Query Language. In this final task, the Common Query Language used are simple queries and boolean queries.

Because of the limitations and how to find out how the search engine query result, to test the reliability of the query result can be carried out by analyzing the effectiveness of each search engine. Can measure the effectiveness of several parameters such as Precision Measurement. Precision is the number of groups of relevant documents from a document found by the system. In addition, required a subjective assessment made by a user called with satisfaction. Satisfaction is a questioner who will analyze the direct interaction between the user with a search engine. From the results of the analysis of the relationship between values obtained from the Precision system, and the results satisfaction user can conclude that, whether simple or boolean queries in general, Google can be said to have the effectiveness that is better than Yahoo.

Keywords : Search engine. Free-Text Document, Common Query Language, Precision, Satisfaction.

1. Pendahuluan

1.1 Latar Belakang

Setiap orang dapat mencari informasi yang diinginkannya dengan menginputkan *keyword* atau *query* didalam *search engine*. *Search engine* akan mengembalikan *query result* berupa dokumen yang relevan dengan *query* yang diinputkan oleh *user*. Dari dokumen yang dihasilkan oleh *search engine*, *user* diharapkan puas terhadap dokumen yang dihasilkan. Kepuasan *user* terhadap dokumen yang dihasilkan inilah yang menjadi sebuah tantangan untuk memberikan suatu pengukuran terhadap dokumen yang dihasilkan dari inputan *user*. Sehingga diperlukan sebuah pengukuran dimana sebelum *user* menginputkan jenis *query* tertentu, *user* akan mengetahui jenis *query* seperti apa yang sesuai dengan *search engine* yang digunakan. Dalam hal ini yang menjadi *search engine* adalah Google dan Yahoo, dan jenis *query* yang digunakan berupa *common query language* yang *query*nya berupa simple dan boolean.

Permasalahan yang akan timbul adalah bagaimana cara mengukur kepuasan *user* terhadap dokumen yang dihasilkan. Kepuasan yang dimaksud disini adalah dokumen yang dihasilkan akan relevan dengan keinginan *user*. Namun karena keterbatasan untuk mengetahui algoritma dan cara bagaimana *search engine* menghasilkan *query result*, maka untuk menguji kehandalan dari *query result* tersebut dengan cara menganalisis efektivitas dari masing-masing *search engine*.

Efektivitas *search engine* diperlukan untuk mengukur sejauh mana hasil *query* yang disebut dengan dokumen relevan sesuai dengan yang diinginkan *user*. Efektivitas dapat di ukur dari beberapa parameter pengukuran diantaranya adalah *precision*. *Precision* adalah jumlah kelompok dokumen relevan dari jumlah dokumen yang ditemukan oleh *system*. *Precision* merupakan penilaian biner yang melihat efektivitas *search engine* dari hasil *query* relevan atau tidak relevan, ditemukan atau tidak ditemukan informasi yang diinginkan [7]. Selain itu *precision* mudah untuk diterapkan dalam mengukur efektivitas *search engine* [1].

Selain mengukur efektivitas dengan menggunakan *precision*, diperlukan sebuah penilaian subjektif yang dilakukan oleh *user* yang disebut dengan *satisfaction*. *Satisfaction* ini merupakan *questioner* yang akan menganalisis secara langsung interaksi antara *user* dengan *search engine*. Hal ini diperlukan untuk mengetahui sejauh mana penilaian *user* terhadap *query result* yang dihasilkan *search engine* secara subjektif.

1.2 Perumusan Masalah

Permasalahan yang dijadikan objek penelitian dan pengembangan tugas akhir ini adalah :

1. Bagaimana cara menentukan *search engine* yang akan dianalisis.
2. Bagaimana cara menentukan *Common Query Language* yang akan digunakan sebagai data testing untuk *search engine*.
3. Bagaimana cara menganalisis *precision search engine* berdasarkan *common query language*.

4. Bagaimana cara melakukan dan menganalisis *test satisfaction* pada *user*.

Dalam penelitian tugas akhir ini, objek penelitian dibatasi dengan ruang lingkup sebagai berikut :

1. *Search engine* yang dianalisis mengacu pada www.hitwise.com yaitu ada dua *search engine*; Google dan Yahoo. *Search engine* tersebut memiliki *document collection* berupa *free text document*.
2. *Query* yang digunakan berdasarkan *Common Query Language* berupa; *simple* dan *boolean queries*.
3. *Test Environment* dibatasi pada 20 hasil temuan pertama dalam *search engine*[17]. *Document collection* sebanyak 189 dokumen yang didapat dari 7 *sample query*.
4. Pengujian dilakukan oleh *system* dan *user testing*, *user testing* akan melakukan *satisfaction* terhadap Google dan Yahoo berupa mengisi *questioner* yang berhubungan dengan *test relevansi search engine*.
5. *Document collection* Google dan Yahoo didapat pada tanggal 22 Februari 2009, pukul 11.00-13.30.
6. *Query* yang digunakan sebagai *inputan* adalah bahasa inggris.
7. Dokumen yang dicari didalam *search engine* memiliki filetype:htm.
8. Menggunakan *stopword* bahasa inggris.

1.3 Tujuan

Menganalisis kesesuaian *Common Query Language* untuk *search engine* Google dan Yahoo, berdasarkan hasil *precision* dari *system* yang dibuat dan *satisfaction* dari *user*, yang diperoleh dari setiap *query input* pada *search engine* yang *document collectionnya* berupa *free-text document*.

1.4 Metodologi penyelesaian masalah

Metodologi pembahasan yang digunakan dalam penelitian Tugas Akhir ini adalah:

1. Studi literatur
Mengumpulkan literatur-literatur yang berhubungan dengan TA baik cetak maupun elektronik
2. Seleksi *search engine* dan *query*
Seleksi *search engine* dilakukan dengan mengacu kepada rilis yang dibuat oleh www.hitwise.com pada tanggal 26 April 2008. Ada dua *search engine* yang paling banyak digunakan yaitu; Google dan Yahoo. *Query inputan* mengacu pada *Common Query Language* berupa; *simple queries* dan *boolean*.
3. *Test environment* dan *test user*
Test environment dilakukan pada 20 hasil temuan pertama dalam *search engine*. *Query result* yang dihasilkan akan diukur *precisionnya*. *User testing* akan melakukan *satisfaction* terhadap ke dua *search engine* yaitu; Google dan Yahoo berupa mengisi *questioner* yang berhubungan dengan *test relevansi search engine*.
4. Analisis *search engine*
Analisis dilakukan dari hasil *system* dan *satisfaction* dari *user*. Dimana pengujian akan menilai *precision* dari setiap *Common Query Language* dari

- masing-masing *search engine* dengan skenario penentuan *term* sebagai query input dan hasil *satisfaction* dari *user* berupa *questioner*.
5. Penyusunan laporan dan penarikan kesimpulan



5. Kesimpulan dan Saran

5.1 Kesimpulan

5.1.1 Kesimpulan dari Sistem

Berdasarkan hasil analisis nilai *precision* yang ditemukan oleh sistem dari *document collection* menunjukkan bahwa:

- Untuk *simple query* Google lebih baik berdasarkan nilai *precision*.
- Untuk *single query* Google lebih baik berdasarkan nilai *precision*.
- Untuk *quote query* Yahoo lebih baik berdasarkan nilai *precision*.
- Untuk *boolean query* Google lebih baik berdasarkan nilai *precision*.
- Untuk *operator and query* Google lebih baik berdasarkan nilai *precision*.
- Untuk *operator or query* Google lebih baik berdasarkan nilai *precision*.

5.1.2 Kesimpulan dari Questioner

Berdasarkan 30 *user* yang berpartisipasi dalam mengisi *questioner*, dapat disimpulkan bahwa *user* merasa puas dengan *query result* yang dihasilkan oleh Google. Baik itu berupa *simple query* maupun *boolean query*.

Selain itu, *user* juga lebih familiar dan lebih mudah menggunakan Google untuk mencari informasi, dikarenakan Google dianggap lebih mudah digunakan dari pada Yahoo. Google lebih memfokuskan untuk menerima pencarian dari *user*.

5.1.3 Kesimpulan Umum

Hubungan antara nilai *precision* yang diperoleh dari sistem dan hasil *satisfaction user* dapat disimpulkan bahwa, baik *simple* maupun *boolean query* secara umum, Google dapat dikatakan memiliki efektivitas yang lebih baik daripada Yahoo.

5.1.4 Kelemahan dan Manfaat System

- Kelemahan *system* yang telah dibangun adalah tidak melakukan *threshold* terhadap penentuan *query* uji.
- Tidak memproses pengambilan *document collection* pada tanggal yang berbeda.
- Manfaat dari *system* ini adalah mempermudah *user* untuk menggunakan *common query language* sebagai referensi jenis *query* dalam menggunakan *search engine*. Namun *cost* yang dikeluarkan untuk menghasilkan *precision* untuk setiap *common query language* dibutuhkan *software* dan *hardware* yang mendukung *system* yang dibuat

5.2 Saran

Untuk penentuan *query* uji dapat menggunakan skenario yang berbeda, dengan menambahkan *threshold* pada *query* uji.

Daftar Pustaka

[1]	Adisantoso, Ir. Julio dan Rhida, Ahmad. 2004. <i>Corpus Dokumen Teks Bahasa Indonesia Untuk Pengujian Efektivitas Temu Kembali Informasi</i> . Departemen Ilmu Kompute, Fak. Matematika dan Ilmu Pengetahuan Alam, IPB.
[2]	Boston.lti.cs.cmu.edu/classes/15-482/Lectures/06-Evaluation.pdf. <i>Introduction to IR, NLP, MT, and Speech Evaluation</i> .
[3]	Chandra Astuti, Novika. 2007. <i>Tugas Akhir: Penanganan Gangguan pada Call Center 147 Berdasarkan Pengukuran Kepuasan Pelanggan Speedy Segmen Residensial di PT Telkom Divre III</i> . STT Telkom Bandung.
[4]	Febrian, jack. 2003. <i>Menggunakan Internet</i> . Penerbit: Informatika Bandung
[5]	Goutte, Cyril dan Gaussier, Eric. <i>A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation</i> . Xerox Research Centre Europe 6, chemin de Maupertuis, F-38240 Meylan, France
[6]	Hardi, Wisnu. 2006. <i>Precision and 6 search engine untuk ilmu perpustakaan dan informasi</i> . Peneliti pada Veritas Document Management, Jakarta.
[7]	http://members.home.nl/mackelenbergh/How to test information retrieval.htm
[8]	Kumar P.K. Suri, Rakesh dan Chauhan, R.K. <i>Search Engines Evaluation</i> . DESIDOC Bulletin of Information Technology , Vol. 25, No. 2, March 2005, pp. 3-10 © 2005, DESIDOC
[9]	Lewandowski, Dirk. 2008. <i>The Retrieval Effectiveness of Web Search Engines: Considering Results Descriptions</i> . Hamburg University of Applied Sciences, Faculty Design, Media and Information Department Information, Berliner Tor 5, D – 20249 Hamburg, Germany.
[10]	Mandala, Rila dan Hendra Setiawan. <i>Peningkatan Performansi Sistem Temu-Kembali Informasi dengan Perluasan Query Secara Otomatis</i> . Bandung: Institut Teknologi Bandung.
[11]	Muresan, Gheorghe. 2005. <i>Measuring Search Effectiveness: Lessons from Interactive TREC</i> . School of Communication, Information and Library Studies. Rutgers University.
[12]	YFA. <i>Slide Matakuliah IR: String Processing</i> . IT Telkom Bandung.
[13]	YFA. <i>Slide Matakuliah IR: Usability</i> . IT Telkom Bandung.
[14]	YFA. <i>Slide Matakuliah IR: Vector Method</i> . IT Telkom Bandung.
[15]	www.hitwise.com . <i>Percentage of U.S. Searches among Leading Search Engine Providers</i> . 12 juni 2008.
[16]	www.en.wikipedia.org/wiki/Search_engine . <i>Web Search Engine</i>
[17]	www.74.125.77.132/search?q=cache:UaIsL_mW_YMJ:endino.files.wordpress.com/2008/10/meningkatkan-posisi-website-pada-search-engine.pdf+mengapa+user+hanya+melihat+dari+dua+halaman+pertama+dalam+search+engine&hl=id&ct=clnk&cd=1&gl=id&client=firefox-a