

SIMPLE QUERY SUGGESTION* UNTUK PENCARIAN ARTIKEL MENGUNAKAN *JACCARD SIMILARITY

Komang Rinarta

Program Studi Sistem Komputer
Stikom Bali

Bali, Indonesia

Email: komangrinarta@gmail.com

ABSTRAK

Pencarian artikel merupakan sebuah proses yang digunakan untuk mencari artikel yang relevan sesuai dengan kata kunci yang dimasukkan oleh pengguna. Konsep sederhana dari pencarian artikel adalah pengguna memasukkan kata kunci, kemudian menekan tombol pencarian sehingga memberikan hasil pencarian. Proses pencarian dengan menggunakan konsep sederhana tersebut memerlukan waktu dan pencarian yang diberikan terkadang kurang sesuai dengan artikel yang dicari. Makalah ini merupakan penerapan dari *Jaccard Similarity* dalam contoh simple query suggestion. *Jaccard similarity* digunakan sebagai query suggestion karena sistem dapat membandingkan kata kunci yang diberikan dengan data yang terdapat didalam *database* yang disebut sebagai *Jaccard Index*. Sistem yang dibangun memberikan suggestion setiap kali user mengetikkan kata kunci dan hasil pencarian akan tampil secara akurat dengan waktu proses 7 mili detik dengan nilai *Jaccard Similarity* antara 0 hingga 1 pada 492 data artikel yang diambil dari Prosiding KNS&I STIKOM Bali.

Kata kunci: *simple query suggestion*, pencarian artikel, *jaccard similarity*

A. PENDAHULUAN

Perkembangan pencarian data di internet sangat pesat didukung dengan perkembangan perangkat keras. Google, Bing dan Yahoo secara konsisten dilaporkan sebagai website yang digunakan sebagai mesin pencari. Mereka menguasai pasar pencarian data sebanyak 96%. Google menguasai 66,9%, Bing 18,1% dan Yahoo 11,1%. Dalam beberapa tahun terakhir ketiganya mengenalkan tentang dynamic term/query suggestions untuk mendukung pengguna dalam query formulation, reformulation atau query expansion [1].

Query suggestion banyak digunakan pada website pencarian informasi untuk mempermudah mencari informasi yang relevan. Query suggestion berhubungan erat dengan form autocompletion pada input form sebuah website. Fitur autocomplete dikenal pertama kali pada aplikasi desktop dengan cara user mengetikkan sesuatu dan suggestion akan tampil secara otomatis. [2]

Jaccard similarity adalah sebuah algoritma yang digunakan untuk membandingkan dokumen yang satu dengan yang lainnya berdasarkan kata yang dimilikinya. *Jaccard similarity* biasanya digunakan untuk membandingkan dokumen dan menghitung nilai kemiripan dari dokumen tersebut. Dengan kemampuan inilah *jaccard similarity* digunakan dalam dynamic query suggestion untuk memberikan saran-saran pencarian dalam hal pencarian artikel.

$$jaccard\ sim(A,B) = P(A \cap B) | P(A \cup B)$$

Website merupakan salah satu teknologi yang digunakan untuk memberikan simulasi *form autocomplete*. Teknologi web yang banyak berkembang saat ini adalah teknologi *web responsive*. *Web responsive* adalah desain web yang dapat digunakan secara fleksibel dibuka dari komputer dengan berbagai resolusi layar. Selain dengan teknologi responsive, website juga dibangun dengan menggunakan teknologi AJAX dan bahasa pemrograman PHP serta database MySQL. Web konvensional pada umumnya tidak dapat dilihat dengan baik pada resolusi display yang berbeda-beda.

AJAX merupakan teknologi yang menggabungkan beberapa teknik pemrograman javascript, PHP dan MySQL yang ditampilkan dengan menggunakan HTML sebagai *desain user interface*. AJAX dapat merubah sebagian halaman website tanpa harus merubah keseluruhan halaman website seperti yang website konvensional lakukan sehingga permintaan terhadap halaman website



Gambar 1. Query suggestion

dapat dilakukan kapan saja tanpa menggunakan waktu tunggu untuk meminta halaman website secara keseluruhan.

B. TINJAUAN PUSTAKA

Autocomplete, atau yang lebih dikenal dengan *word completion*, adalah fitur yang disediakan oleh banyak web browser, surel, antarmuka mesin pencari, *source code editor*, *tools* pada query database, pengolah kata (*word processor*), dan *interpreter* pada *command line*. *Autocomplete* juga terdapat dan sudah terintegrasi dalam teks editor yang umum digunakan. Kegunaan dari fitur *autocomplete* ini adalah menampilkan perkiraan kata atau frase yang akan dimasukkan tanpa harus mengetikkan keseluruhan kata [2] [3].

Autocomplete melibatkan program yang dapat melakukan prediksi terhadap sebuah kata atau frasa yang pengguna ingin tulis tanpa harus menulis keseluruhan kata atau frasa secara lengkap. *Autocomplete* bekerja ketika pengguna menulis huruf pertama atau beberapa huruf/karakter dari sebuah kata, program yang melakukan prediksi akan mencari satu atau lebih kemungkinan kata sebagai pilihan. Jika kata yang dimaksud ada dalam pilihan kata prediksi maka kata yang dipilih tersebut akan disisipkan pada teks [2].

Dynamic Query Suggestion juga dikenal sebagai pre-submission query suggestion atau query autocompletion mengacu pada sebuah system atau pradigma dimana sebagian keyword yang diketik oleh pengguna mengarah pada kumpulan saran yang ditampilkan secara otomatis yang memungkinkan dipilih oleh user. Hasil dari dynamic query suggestion merupakan bagian lengkap dari keyword yang dimasukkan oleh pengguna dan ketika pengguna memilih salah satu dari saran suggestion tersebut, maka system akan memberikan hasil yang sesuai [4].

Jaccard Algorithm atau juga disebut dengan Jaccard Index dan juga dikenal dengan Jaccard Coefficient adalah sebuah nilai yang didapatkan dari membandingkan kemiripan ataupun perbedaan dari sebuah data set. Perhitungan koefisien Jaccard similarity antara dua kumpulan data adalah didapat dari hasil pembagian antara jumlah data yang sama dengan keseluruhan data yang ada. Jaccard index dapat dirumuskan sebagai berikut [5] [6] :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Jaccard Distance adalah perhitungan ketidaksamaan antara data set. Hal tersebut dapat disebut sebagai invers Jaccard Coefficient yang mengurangkan 1 dengan nilai Jaccard Similarity. Jaccard distance dapat dirumuskan sebagai berikut [5]:

$$J_s(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

Dalam sebuah artikel berjudul “AJAX: A New Approach to Web Applications”, Jesse James Garret menjelaskan dengan lengkap mengenai teknologi AJAX (*Asynchronous JavaScript And XML*) yang diyakini dapat menutup gap antara aplikasi berbasis desktop dan web. AJAX memiliki perbedaan dengan model web tradisional [7].

Istilah AJAX (*Asynchronous JavaScript And XML*) dalam pengembangan web menjadi populer. AJAX ini sendiri bukan merupakan bahasa pemrograman baru, AJAX hanya merupakan sebuah teknik pemanfaatan object XMLHttpRequest dengan javascript untuk berkomunikasi dengan server secara Asynchronous, dengan pemanfaatan object XMLHttpRequest ini kita dapat membuat proses berjalan secara background atau bekerja dibelakang layar sementara user dapat tetap berinteraksi dengan halaman web yang ada. Pemanfaatan tehnik Asynchronous ini jugalah yang telah mendorong pengembangan web menjadi lebih kaya atau banyak yang penulisnya dengan istilah pengembangan *Rich Internet Application* (RIA) atau WEB 2.0. [8]

C. METODE PENELITIAN

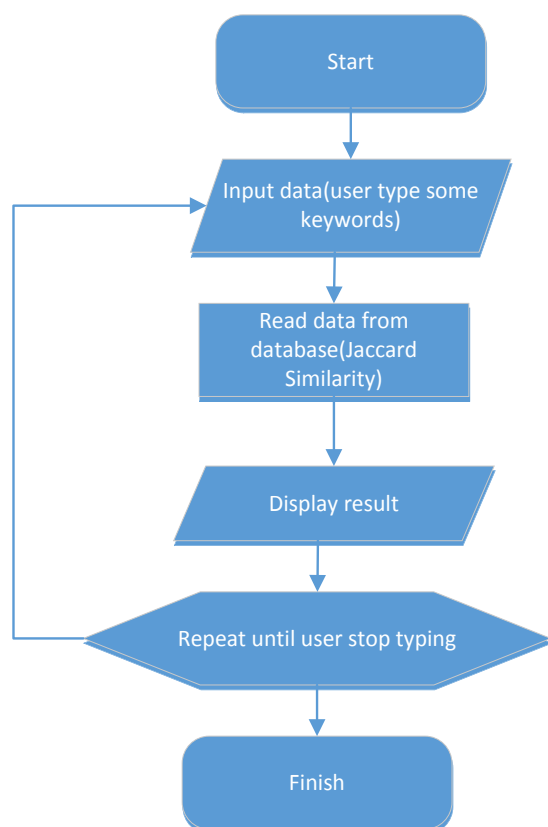
Jaccard similarity diimplementasikan sebagai sebuah algoritma untuk memberikan suggestion kepada pengguna sesuai dengan diagram pada gambar 2.

Simple query suggestion dengan menggunakan jaccard algorithm dimulai dari pengguna mengetikkan sesuatu pada bagian text field, kemudian system akan mencari artikel dari database dengan menghitung nilai jaccard similarity dan system akan menampilkan hasilnya. Hasil dari query suggestion mungkin satu atau lebih sesuai dengan data yang terdapat di database yang berhubungan dengan keyword yang dimasukkan oleh pengguna. System akan berhenti melakukan pencarian ketika pengguna berhenti mengetik atau ketika pengguna memilih salah satu query suggestion yang diberikan.

Bentuk matematis Jaccard similarity diubah menjadi bentuk SQL untuk membaca data dari database. Ketika pengguna memasukkan “signal indera pro”, system akan mencari dan menghitung nilai dari jaccard index dan jika hasil yang diharapkan adalah “SIGNAL PROCESSING OF RADAR INDERA” maka hasil dari jaccard index adalah 0,6 sesuai dengan rumusan berikut

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|signal indera pro|}{|SIGNAL PROCESSING OF RADAR INDERA|}$$

$$= \frac{|3|}{|5|} = 0.6$$



Gambar 2. Proses query suggestion

Hasil konversi bentuk matematis menjadi bentuk SQL adalah sebagai berikut :

```

SELECT *, (if (locate('signal',
judulMakalah)>0,1,0)+if (locate('indera',
judulMakalah)>0,1,0)+if
(locate('processing',
judulMakalah)>0,1,0)+if (locate('radar',
judulMakalah)>0,1,0)+if (locate('of',
judulMakalah)>0,1,0)+0) / (LENGTH(judulMakalah)
- LENGTH(REPLACE(judulMakalah, ' ', ''))
+ 1 +if (locate('signal',
judulMakalah)>0,0,1)+if (locate('indera',
judulMakalah)>0,0,1)+if
(locate('processing',
judulMakalah)>0,0,1)+if (locate('radar',
judulMakalah)>0,0,1)+if (locate('of',
judulMakalah)>0,0,1)+0) as score FROM
makalah order by score desc LIMIT 0, 10
    
```

D. HASIL DAN PEMBAHASAN

Aplikasi dibangun dengan menggunakan Bahasa pemrograman PHP dalam bentuk website dan MySQL sebagai databasenya dan juga menggunakan AJAX sebagai teknik dalam memberikan suggestion. Data yang digunakan didalam aplikasi ini adalah data prosiding KNS&I STIKOM Bali dengan 492 data artikel. Nilai Jaccard similarity didapatkan dari pembagian jumlah kata kunci dengan jumlah semua data yang digunakan. Ketika pengguna mengetikkan kata

“fuzzy” system akan mencari dan menghitung jaccard index dan data yang ada adalah “SIGNAL PROCESSING OF RADAR INDERA” sehingga menghasilkan nilai Jaccard index sama dengan 0 sesuai dengan perhitungan sebagai berikut :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{||}{|SIGNAL PROCESSING OF RADAR INDERA| \cup |0|} = \frac{0}{|5|} = 0$$

Ketika pengguna memasukkan “signal indera processing of radar”, system akan mencari dan menghitung jaccard index dan hasil yang diharapkan adalah “SIGNAL PROCESSING OF RADAR INDERA” maka nilai jaccard index adalah 1 sesuai dengan perhitungan berikut :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|signal indera processing of radar|}{|SIGNAL PROCESSING OF RADAR INDERA| \cup |5|} = \frac{5}{|5|} = 1$$



Gambar 3. Aplikasi query suggestion

Gambar menunjukkan hasil dari query suggestion ketika user memasukkan “signal indera pro” dan system akan menampilkan suggestion, waktu proses dan juga nilai dari jaccard index. System membutuhkan waktu sekitar 7 milidetik untuk mencari artikel yang sesuai dengan kata kunci dari pengguna dengan nilai jaccard index sebesar 0,6. Ketika salah satu suggestion dipilih oleh pengguna maka system akan menampilkan informasi detail dari artikel tersebut.



Gambar 4. Pengguna memilih salah satu suggestion

Tabel 1. Pengujian aplikasi

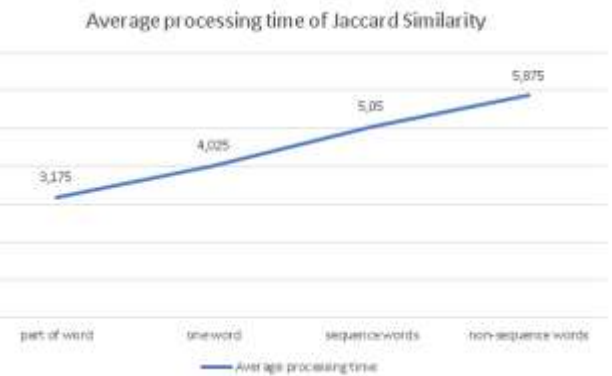
Process	Input	Result
Proses pencarian	Sebagian dari kata kunci	System menampilkan hasil yang sesuai
Proses pencarian	Satu kata kunci	System menampilkan hasil yang sesuai
Proses pencarian	Kata kunci yang berurutan	System menampilkan hasil yang sesuai
Proses pencarian	Beberapa kata kunci yang tidak berurutan	System menampilkan hasil yang sesuai

Tabel 2. Waktu proses pada 492 data

Input	Processing time
Sebagian dari kata kunci	2.0 milidetik
Satu kata kunci	6.5 milidetik
Kata kunci yang berurutan	7.4 milidetik
Beberapa kata kunci yang tidak berurutan	7.6 milidetik



Gambar 5. Grafik perbandingan jumlah data dengan waktu proses



Gambar 6. Waktu proses rata-rata

E. KESIMPULAN

Paper ini mengimplementasikan jaccard algorithm untuk membuat simple query suggestion dan mencari data yang sesuai berdasarkan kata kunci yang dimasukkan oleh pengguna. Query suggestion berjalan dengan baik dengan PHP, MySQL dan AJAX untuk membantu pengguna dalam mencari artikel yang dibutuhkan pada website. Berdasarkan hasil pengujian, aplikasi memberikan hasil yang akurat sebagai query suggestion sesuai dengan kata kunci yang dimasukkan pengguna baik sebagian kata kunci, satu buah kata kunci maupun beberapa kata kunci. Aplikasi pencarian membutuhkan waktu sekitar 7 mili detik untuk mengasilkan suggestion pada 492 data artikel.

REFERENSI

- [1] A. Shiri dan L. Zvyagintseva, "Dynamic Query Suggestion in Web Search Engines: A Comparative Examination," dalam *42nd Annual Conference of the Canadian Association of Information Science & Inaugural Librarians' Research Institute Symposium Brock University. St. Catharines, Ontario, 2014.*
- [2] Y. Primadani, "Simulasi Algoritma Levenshtein Distance Untuk Fitur Autocomplete Pada Aplikasi Katalog Perpustakaan," Universitas Sumatera Utara, Medan, 2014.
- [3] L. Y. Banowosari, A. Darmawan, K. Kurniawan dan M. Mitchell, "Analisis Pada Fitur Autocomplete Suggestion Dan Semantik Pada Pencarian Di Mesin Pencari Google," dalam *Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2014)*, Jakarta, 2014.
- [4] D. Hawking dan K. Griffiths, "An enterprise search paradigm based on extended query auto-completion. Do we still need search and navigation?," dalam *ADCS '13 Proceedings of the 18th Australasian Document Computing Symposium (pp. 18-25)*, New York, 2013.

- [5] S. Niwattanakul, J. Singthongchai, E. Naenudorn dan S. Wanapu, "Using Jaccard Coefficient For Keywords Similarity," dalam *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I*, Hong Kong, 2013.
- [6] I. E. Vorontsov, I. V. Kulakovskiy dan a. V. J. Makeev, "Jaccard index based similarity measure to compare transcription factor binding site models," dalam *Algorithms for Molecular Biology 2013*, 2013.
- [7] S. Handayaningsih dan W. Pujiyono, "istem Konsultasi Dan Laporan Pembimbingan Tugas Akhir," Yogyakarta, 2010.
- [8] H. Kurniawan, "Pengembangan Sistem Manajemen Reservasi Ruang Online Dengan Zk Ajax Framework," Universitas Islam Negeri Syarif Hidayatullah, Jakarta, 2011.
- [9] S. Plansangket dan J. Q. Gan, "A query suggestion method combining TF-IDF and Jaccard Coefficient for interactive web search," *Artificial Intelligence Research 4*, vol. 4, p. 119, 2015.
- [10] Sugiyanto, B. Surarso dan A. Sugiharto, "Analisa Performa Metode Cosine dan Jaccard Pada Pengujian Kesamaan Dokumen," *Jurnal Masyarakat Informatika*, vol. 5, pp. 1-8, 2014.
- [11] J.-M. Yang, R. Cai, F. Jing, S. Wang, L. Zhangy dan W.-Y. Ma, "Search-based Query Suggestion," dalam *Proceedings of the 17th ACM conference on Information and knowledge management. ACM 2008*, California, 2008.