

ANALISIS PADA FITUR *AUTOCOMPLETE SUGGESTION* DAN SEMANTIK PADA PENCARIAN DI MESIN PENCARI *GOOGLE*

*Lintang Y. Banowosari*¹
*Andi Darmawan*²
*Kega Kurniawan*³
*Marvin Mitchell*⁴

Universitas Gunadarma
*lintang@staff.gunadarma.ac.id*¹, *andi.lazvegas@gmail.com*²,
*keygen.gundar@gmail.com*³, *marvinmarpol@gmail.com*⁴

Abstrak

Perkembangan teknologi internet mendorong munculnya fitur dan inovasi terbaru untuk meningkatkan pengalaman dan kemudahan pengguna dalam menjelajahi dunia maya. Salah satu fitur tersebut adalah autocomplete suggestion. Fitur autocomplete suggestion merupakan fitur yang diimplementasikan pada web browser dan mesin pencari yang memungkinkan web browser atau mesin pencari untuk memberikan saran pencarian ketika baru beberapa kata diketikan dalam kolom pencarian atau address bar. Makalah ini membahas analisis tentang fitur autocomplete suggestion dan semantik dari sistem pencarian pada salah satu mesin pencari terpopuler di dunia, yaitu Google.

Kata Kunci: *Semantik, Autocomplete Suggestion, Mesin Pencari, Google*

PENDAHULUAN

Ketika menggunakan *web browser* untuk menjelajah di dunia maya maka pada kolom *address bar* saat mengetikan beberapa huruf/kata, *web browser* tersebut memberikan saran dari apa yang diketik, atau sama halnya ketika menggunakan mesin pencari untuk mencari suatu informasi, ketika baru mengetikan beberapa huruf/kata pada kolom pencarian, mesin pencari tersebut memberikan saran pencarian yang terkait dari apa yang diketikan. Ketika memilih saran tersebut, *web browser* atau mesin pencari akan langsung melengkapi kata kunci pencarian sesuai dengan saran yang dipilih.

Fitur tersebut dinamakan *autocomplete suggestion*. Fitur *autocomplete suggestion* memberikan pengalaman baru dan kemudahan bagi pengguna ketika berjelajah di internet. Salah satu mesin pencari yang mengimplementasikan fitur ini adalah *Google*. *Google* merupakan salah satu mesin pencari terpopuler di dunia. Oleh karena itu, makalah ini berisi tentang analisis cara kerja *autocomplete suggestion*, faktor yang mempengaruhi keberagaman pada *autocomplete suggestion* dan analisis semantik pada hasil pencarian mesin pencari *Google*. Metode yang digunakan dalam analisis ini adalah *Knowledge graph*.

Knowledge graph merupakan *knowledge base* yang digunakan *Google* untuk meningkatkan hasil pencarian dengan

semantic search dari informasi yang dikumpulkan dari berbagai sumber. Tujuannya adalah agar pengguna dapat menggunakan informasi ini untuk menyelesaikan permintaan mereka tanpa harus menavigasi ke situs lain dan menyusun informasi itu sendiri.

1. MESIN PENCARI

Mesin pencari adalah istilah yang digunakan untuk *website* tempat orang mencari (mesin pencari) seperti *Google.com*.^[1] Di Mesin Pencari seperti *Google*, ada ribuan bahkan orang mencari informasi dengan mengetikkan kata atau beberapa kata yang ingin mereka temukan informasi lebih detailnya.

2.1 Cara Kerja Mesin pencari

Cara kerja mesin pencari yaitu menyimpan informasi tentang banyak halaman web, yang diambil langsung dari WWW. Halaman-halaman ini diambil dengan *web crawle* (Frieur, 2013), peramban web otomatis mengikuti setiap pranala atau *link* yang dilihatnya. Isi setiap halaman lalu dianalisis untuk menentukan cara indeks-nya (misalnya, kata-kata diambil dari judul, subjudul, atau field khusus yang disebut meta tag). Data tentang halaman web disimpan dalam sebuah database indeks untuk digunakan dalam pencarian selanjutnya. Sebagian mesin pencari, seperti *Google*, menyimpan seluruh atau sebagian halaman sumber (yang disebut *cache*) maupun informasi tentang halaman web itu sendiri.

Selain halaman web, mesin pencari juga menyimpan dan memberikan informasi hasil pencarian berupa pranala yang merujuk pada file, seperti file audio, file video, gambar, foto dan sebagainya, serta informasi tentang seseorang, suatu produk, layanan, dan informasi beragam lainnya yang semakin terus berkembang sesuai dengan perkembangan teknologi informasi.

2.2 Semantic Search

Semantic search merupakan teknik pencarian data dalam *search query* yang bertujuan tidak hanya untuk mencari kata

kunci, tetapi juga untuk menentukan makna kontekstual dari kata kunci yang digunakan dalam pencarian (Graham, 2005).

Semantic search memberikan hasil yang lebih bermakna dengan mengevaluasi dan memahami frase pencarian dan menemukan hasil yang paling relevan dalam situs web, database, atau tempat penyimpanan data lainnya.

Semantic search bekerja pada prinsip-prinsip bahasa semantik. Tidak seperti algoritma *typical search*, *Semantic search* didasarkan pada konteks, substansi, maksud dan konsep frase yang dicari. *Semantic search* juga mencakup lokasi, sinonim dari istilah, tren saat ini, variasi kata dan unsur-unsur bahasa alami lainnya sebagai bagian dari pencarian. Konsep *Semantic search* berasal dari berbagai algoritma pencarian dan metodologi, termasuk kata *keyword-to-concept mapping*, *graph patterns* dan logika *fuzzy*.

Mesin pencari web besar seperti *Google* dan *Bing* menggabungkan beberapa unsur dari pencarian semantik. Jika *Google* menggunakan algoritma pemeringkatan seperti *PageRank* untuk memprediksi relevansi, *semantic search* menggunakan semantik, atau ilmu yang mempelajari arti/makna bahasa, untuk memperoleh hasil pencarian yang sangat relevan. Namun *Google* juga telah mengumumkan proyek *semantic search* nya sendiri.

2.3 Autocomplete Suggestion

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^[4]

Tujuan awal dari program *autocomplete* atau dalam hal ini kita menyebutnya dengan *word prediction* adalah untuk membantu meningkatkan kecepatan mengetik bagi orang-orang dengan kekurangan fisik, serta untuk membantu mereka mengurangi jumlah penekanan tombol keyboard yang dibutuhkan untuk menyelesaikan sebuah kata atau kalimat.

Pada mesin pencari, antarmuka fitur *autocomplete* menyediakan sugesti pencarian yang sesuai dengan apa yang pengguna minta atau menyediakan sugesti berdasarkan kata atau huruf yang belum pengguna input secara sempurna didalam search box. Hal ini biasa disebut dengan autosuggest atau incremental search.

Google bukanlah mesin pencari pertama yang mengimplementasikan *Autocomplete Suggestion*. *Google* mengimplemenasikan *Autocomplete suggestion* pada tahun 2008.

Google Autocomplete suggestion ditentukan oleh beberapa faktor, seperti popularitas dari kata kunci yang diketik pada kolom pencarian. Dan juga pada *Google Autocomplete suggestion* terdapat *suggestion* yang dihilangkan, seperti

pencarian dengan kata kunci yang berbau pornografi, kekerasan, kebencian, barang ilegal dan berbahaya, dan konten yang berisi pelanggaran hak cipta. [5]

Ada beberapa kemungkinan apabila penyaranan tidak muncul ketika mengetikan kata kunci pada pencarian, antara lain:

1. kata kunci yang dicari tidak populer.
2. kata kunci yang dicari masih baru. Membutuhkan waktu beberapa hari atau minggu bagi kata kunci populer yang baru untuk muncul sebagai penyaranan.
3. kata kunci yang dicari disalah artikan sebagai pelanggaran kebijakan.

2.4 Google Distance

Google Distance merupakan ukuran kemiripan semantik yang dihasilkan dari jumlah *hits* yang dilakukan mesin pencari *Google* untuk sekumpulan kata kunci. Kata kunci dengan arti yang sama atau serupa dalam *natural language*, kemiripan semantik-nya cenderung lebih tinggi dari pada kata-kata yang memiliki arti yang berbeda (Lukas, 2009). Secara spesifik, Normalisasi *Google Distance* (NGL) antara dua istilah x dan y adalah:

$$NGD(x, y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log M - \min\{\log f(x), \log f(y)\}}$$

M merupakan jumlah halaman web yang dicari oleh *Google*; $f(x)$ dan $f(y)$ merupakan jumlah *hits* untuk masing-masing istilah pencarian x dan y , dan $f(x,y)$ merupakan jumlah halaman web yang mengandung istilah x dan y .

Jika dua istilah pencarian x dan y tidak terdapat pada halaman web yang sama, tetapi terjadi secara terpisah, normalisasi *Google distance* antara mereka adalah tak terbatas. Jika kedua istilah terdapat pada web yang sama, *NGD* mereka adalah nol,

atau setara dengan koefisien antara x kuadrat dan y kuadrat.

2. PENGAMATAN DAN ANALISIS

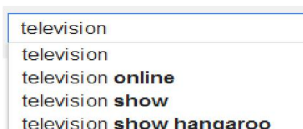
Berikut merupakan hasil analisis terhadap *Autocomplete suggestion* dan *Semantic Search* pada mesin pencari *Google*:

3.1 Fitur *Google Autocomplete suggestion*

Terdapat beberapa jenis penyaranan yang dapat terjadi pada fitur *Google autocomplete suggestion*, antara lain:

3.1.1 Saran Berdasarkan Pencarian Nyata

Saran yang diberikan *Google* berdasarkan dari pencarian yang sering dilakukan oleh orang lain.



Gambar 1 Suggestion Based on Real Searches

Sebagai contoh, ketika mengetikkan "television", maka *Google* akan memberikan *suggestion*:

television

television online

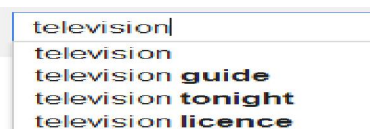
television show

television show hangaroo

Saran tersebut pencarian nyata yang pernah dilakukan orang lain. Salah satu faktor yang memungkinkan hal ini adalah popularitas *Google*. Jika banyak orang yang mengetikkan kata "television" kemudian disertai kata "show", hal itu dapat memunculkan "television show" sebagai saran.

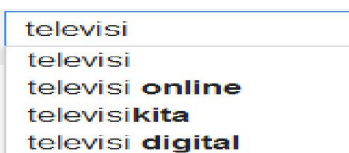
3.1.2 Saran Beragam Berdasarkan Lokasi dan Bahasa

Tidak semua orang melihat saran yang sama. Saran yang diberikan dibedakan berdasarkan Lokasi dan bahasa yang digunakan.



Gambar 2 Saran yang didasarkan pada wilayah

Misalnya dengan menggunakan *Google.co.id* (gambar 1) saran yang muncul akan berbeda apabila menggunakan *Google.co.uk* (gambar 2) walaupun kata kunci "television" yang diketikkan sama.



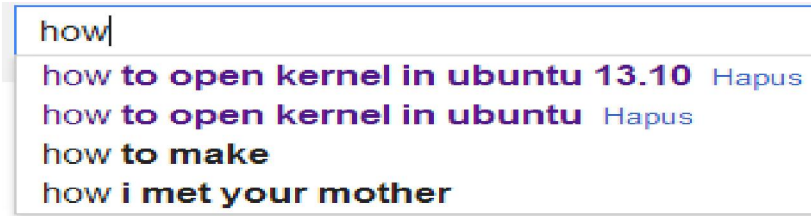
Gambar 3 Saran berdasarkan bahasa pencarian

Begitu pula dengan bahasa, walaupun pencarian menggunakan kata yang memiliki arti yang sama, apabila menggunakan bahasa yang berbeda (gambar 1 menggunakan Bahasa Inggris, gambar 3 menggunakan Bahasa Indonesia) maka *Google* akan memberikan saran yang berbeda.

Singkatnya, lokasi dan bahasa memiliki pengaruh terhadap saran yang diberikan. Perbedaan lokasi atau bahasa dapat memberikan sarann yang berbeda pula.

3.1.3 Saran Berdasarkan Pencarian Sebelumnya

Saran yang diberikan *Google* juga dapat berisi pencarian yang pernah dilakukan sebelumnya.



Gambar 4 Saran berdasarkan pencarian sebelumnya

Misalnya ketika kami melakukan pencarian dengan kata kunci "how", maka saran yang akan tampil adalah:

how to open kernel in ubuntu 13.10

how to open kernel in ubuntu

how to make

how i met your mother

saran pertama dan kedua merupakan pencarian yang pernah dilakukan sebelumnya, oleh karena itu warna huruf berbeda dengan saran yang lain, dan disamping saran terdapat label "hapus".

3.2 Spelling Correction pada Autocomplete suggestion

Dalam fitur *Autocomplete suggestion* juga terdapat *spelling correction* yang berfungsi untuk memperbaiki ejaan yang salah yang diketikkan pada kolom pencarian. Ada beberapa jenis kesalahan ejaan yang dapat diperbaiki secara otomatis dengan *spelling correction*, antara lain:

1. *Transposition*: merupakan kesalahan ejaan dimana adanya kesalahan posisi dari huruf dalam sebuah kata kunci.



Gambar 5 Kesalahan Transposisi ejaan

2. *Wrong Letter*: merupakan kesalahan ejaan dimana adanya kesalahan huruf dalam sebuah kata kunci.



Gambar 6 Kesalahan ejaan *Wrong Letter*

3. *Extra Letter*: merupakan kesalahan ejaan karena adanya kelebihan jumlah huruf dalam sebuah kata kunci.



Gambar 7 Kesalahan ejaan *Extra Letter*

4. *Missing Letter*: merupakan kesalahan ejaan dimana adanya kekurangan jumlah huruf dalam sebuah kata kunci.



Gambar 8 Kesalahan Ejaan *Missing Letter*

3.3 Semantic Search pada Mesin pencari Google

Sistem pendukung *Google* untuk *semantic search* dinamakan *Knowledge graph* yang mampu menghilangkan kebutuhan untuk *re-query* dan memberikan disambiguasi. *Google* menggunakan dua faktor dasar untuk menilai seberapa penting dan relevansinya setiap halaman web sebelum memberikan peringkat kepada halaman web tersebut. Faktor-faktor tersebut adalah:

1. Peringkat halaman (untuk mengukur popularitas dengan menganalisis *backlink*).
2. Relevansi (dengan menganalisis penggunaan kata kunci atau *search query* yang digunakan dalam halaman web tersebut).

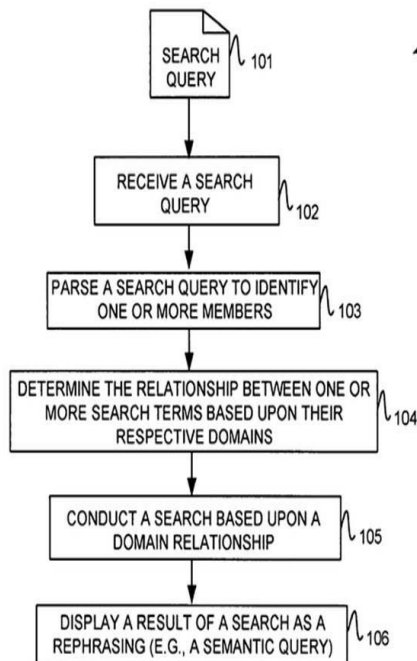
Bentuk perankingan berdasarkan 2 hal diatas tidak membantu dalam menemukan halaman-halaman web yang relevan dengan maksud yang diinginkan oleh para pencari, karena secara tidak langsung faktor popularitas dapat mengurangi peringkat semantik relevansi dari halaman web tersebut. Ini adalah alasan bahwa *Google* menggunakan semantik untuk mengidentifikasi dan memprioritaskan peringkat halaman web yang memiliki konten yang relevan secara semantik daripada hanya menghitung kata kunci dalam kata kunci dan *backlink* untuk menganalisis halaman-halaman web.

Query Processing Dalam Semantic Environment

Search query yang diterima oleh *Google* diurai (menggunakan *parser*) untuk mengidentifikasi satu atau lebih kata kunci. Dalam proses ini, sinonim atau istilah pengganti lainnya akan diidentifikasi. Sinonim ini dikenal sebagai calon sinonim dan calon sinonim tersebut akan dipecah dan diproses sebagai sinonim yang berkualitas (*Qualified Synonym*). Kemudian, *relationship engine* digunakan untuk mengidentifikasi hubungan antara anggota berdasarkan domain masing-masing. Yang dimaksud dengan domain disini adalah pemusatan kategori dari kalimat yang serupa.

Kata kunci dari *Search Query* akan diidentifikasi oleh domain yang merupakan kategori semantik yang memiliki koleksi pra-entitas yang telah ditetapkan. Hal ini membantu *Google* untuk menghubungkan istilah untuk pencocokan identitas yang terdekat (Salah satu point penting yang perlu dicatat di sini adalah bahwa *Google* hanya akan menemukan kata-kata yang berhubungan dalam *search query* dengan yang sudah ada dalam database-nya yang

merupakan *Knowledge graph*, oleh karena itu, beberapa *Search Query* meskipun sudah sesuai semantik mungkin tidak muncul).



Gambar 9 Pemrosesan *Query* pada *Google*

Sebuah pencarian yang terpisah akan dilakukan oleh mesin *query* menggunakan *Domain Matching Relationship* dan hasil akhir akan ditampilkan setelah semantik *Search Query* diidentifikasi (mesin *query* dapat mempluralkan atau mengulang kata-kata pada *Search Query* jika diperlukan). Oleh karena itu, dapat disimpulkan bahwa, sebuah *Search Query* yang kompleks yang diinput oleh pengguna dipecah-pecah, kemudian disederhanakan dengan melibatkan suatu proses yaitu dengan fitur *Google Distance* untuk mengukur kemiripan semantik dari kata kunci yang dicari, Setelah itu, halaman web yang relevan diidentifikasi dan ditampilkan sebagai hasil akhir.

KESIMPULAN

Kesimpulan dari hasil analisis yang diperoleh antara lain:

1. Fitur-fitur pada *autocomplete Google* menampilkan beberapa sugesti pencarian berdasarkan pada: letak geografis dan bahasa yang digunakan dalam pencarian, kata kunci populer yang sering dicari, dan riwayat pencarian dari pengguna yang bersangkutan. *Autocomplete Google* juga dapat membetulkan ejaan yang salah saat pengguna memasukkan kata kunci dengan adanya fitur *spelling correction*.
2. Sistem pencarian semantik pada *Google* berdasarkan pada dua hal yaitu: Peringkat halaman dan Relevansi, dimana menurut kami dua hal ini sangat bertentangan satu sama lain dan tidak saling mendukung dimana jika Peringkat halaman suatu halaman web tinggi maka akan semakin tinggi kemungkinan *Google* akan menampilkan hasil pencarian dengan urutan peringkat halaman dari mulai yang tertinggi hingga yang terendah tanpa memperhatikan relevansi isi dari halaman web yang bersangkutan. Atau dapat dijelaskan dengan sederhana semakin tinggi peringkat halaman maka nilai relevansi akan semakin diabaikan.
3. Sedangkan untuk proses bagaimana *Google* menerjemahkan *Search Query* atau kata kunci yang pengguna input dapat kami simpulkan *Google* menggunakan beberapa piranti seperti *parser*, *relationship engine*, *knowledge graph* dan *Domain Matching Relationship*.

Proses pencarian pada *Google* sudah cukup baik, malah terbilang sangat baik untuk jajaran mesin pencari yang ada di jagat internet ini. Namun hal itu belum menjamin bahwa *Google* akan menampilkan informasi atau halaman web yang diinginkan oleh pengguna. Untuk memperbaiki ini disarankan untuk lebih menekankan relevansi informasi pada sebuah halaman web daripada popularitasnya (*Page Rank*), dan mungkin *Google* dapat menjadikan pola kalimat di setiap lokasi pengguna sebagai

landasan pencarian dan menyarankan halaman web asing dengan cara menterjemahkan masukkan yang telah diberikan oleh pengguna, sehingga kemungkinan pengguna mendapatkan informasi yang sesuai dengan keinginannya akan lebih besar tercapai.

DAFTAR PUSTAKA

- Edy Victor Haryanto, *Rancang Bangun Prototype Mesin Pencari String Menggunakan Metode FuzzyString Matching*, Proseding Konferensi Nasional Sistem dan Informatika, STIKOM, Bali, 2011
- Frieo R Sudarja, *Perbandingan Algoritma String Matching dan Algoritma Auto Complete pada Aplikasi Kamus Kedokteran Dorland Berbasis Android*, Eprints, STMIK GI MDP, Palembang, 2013
- Google Support, *Autocomplete*, <http://support.google.com/websearch/answer/106230>. 14 April 2014
- Graham D., *Google's Search for Meaning*, New Scientist Online Magazine, UK, 2005
- Joydeep. B, *Detailed Analysis of Semantic Search and its role in Hummingbird Algorithm*, <http://www.searchenginepeople.com/blog/detailed-analysis-of-semantic-search-and-its-role-in-hummingbird-algorithm.html>, 2013
- Krush, A., *How to thrill Google Hummingbird: The SEO's guide [INFOGRAPHIC]*, Search Engine Journal, November, 2013
- Lukas Lukmana, *SEO Search Engine Optimization, Cara Cepat mendapatkan Rating Tinggi di Search Engine*. Penerbit C.V Andi Offset, Yogyakarta, 2009

Martins.B, Silva. Mario J, *Spelling Correction for Search Engine Queries*, Proceedings of EsTAL-04, SpringerVerlag, Spain, 2004

and Semidefinite Programming, Knowledge Media Technologies, First International Core-to-Core Workshop, Germany, 2006

Poland. J, Zeugmann. T, *Clustering the Google Distance with Eigenvectors*

Taufan Riyadi, *Jenis-jenis Search Engine*, IlmuKomputer.Com, 2009