

Metoda Text Mining untuk Pencarian Kata Di Internet

Nani K. Tachjar^{#1}

[#]*Institut Perbanas Jakarta*

Jl. Perbanas, Setiabudi, Kuningan, Jakarta 12940

Tel. (021) 525 2533, Fax. (021) 522 8460

¹*nanitachjar@gmail.com*

Abstrak - Pendidikan Tinggi bidang IT telah menghasilkan banyak lulusan. Namun, jumlah lulusan masih belum memenuhi kebutuhan tenaga kerja industri TI di dunia Indonesia, khususnya Bank Dunia. Lulusan TI dari perguruan tinggi yang memiliki kemampuan, alih-alih bekerja di luar negeri. Sementara yang memiliki kemampuan di bawah standar sebagai operator dan programmer, banyak dihasilkan oleh perguruan tinggi. Sinergi antara pendidikan dan industri masih belum ideal. Ini terjadi karena kurangnya *link and match* antara kurikulum universitas antara pendidikan dan industri, menyebabkan kelangkaan SDM IT. Berikut aplikasi pemetaan tenaga kerja yang menggunakan teknik text mining bidang ICT untuk memberikan informasi yang diperlukan oleh Perguruan Tinggi, terutama untuk kurikulum yang tinggi perbaikan maju dalam rangka untuk menyeimbangkan kebutuhan pasar.

Kata kunci : *text mining*, basis data, *filtering*, *stemming*

Abstract — Higher Education IT field has produced many graduates. However, the number of graduates is still not meet the labor needs of IT industry in the world Indonesia, especially the World Bank. IT graduates from universities that have the capability, instead of working abroad. While that has the ability under the standard as operators and programmers, many printed by the College. Synergy between education and industry is still not ideal. Which, this happens due to the lack of link and match between the curriculum of universities between education and industry, causing shortages of IT human resources. Here workforce mapping application that uses the ICT field text mining techniques in order to provide the information required by the College, especially for curriculum improvement College forward in order to balance the needs of the market.

Keywords — *text mining*, basis data, *filtering*, *stemming*

I. PENDAHULUAN

Sumber Daya Manusia (SDM) di bidang Teknologi Informasi & Komunikasi (TIK) masih dibutuhkan pasar, umumnya di Indonesia. Berdasarkan artikel di harian Kompas (25 November 2013)[1], dipaparkan oleh Kepala Badan Penelitian dan Pengembangan Sumber Daya Manusia (SDM) Kementerian Komunikasi dan Informatika Aizirman Djuan bahwa Indonesia masih kekurangan tenaga ahli di bidang teknologi informasi. Banyak sarjana TI yang hanya memiliki kemampuan dasar sebagai operator dan programmer, dan di sisi lain, yang ahli di bidang TI justru direkrut menjadi tenaga kerja di perusahaan asing di luar negeri.

Selain itu, terdapat artikel berjudul “Lulusan IT banyak yang mengecewakan” yang diungkap oleh Ahmad Bagus Santoso, Human Resource Departement PT. Indocyber Global Teknologi pada 16 Januari 2010[2] menyoroti mutu lulusan Perguruan Tinggi yang cenderung instan. Dimana kurikulum perguruan tinggi tidak selaras dengan kebutuhan industri IT.

Sehingga pelamar kerja tidak seperti yang diharapkan dan pemberi kerja sering merasa kecewa.

Kesimpulan yang dapat ditarik dari kedua artikel tersebut, bahwa tenaga kerja bidang TIK di Indonesia masih terjadi kekosongan dan ketidaksesuaian dengan pemenuhan kebutuhan. Pun, berdasarkan info lowongan kerja yang ditampilkan dalam situs www.jobsdb.com dengan kategori Information Technology, lowongan kerja yang masuk per hari dapat mencapai 200. Para pemberi lowongan kerja rata-rata mensyaratkan pengalaman 2 tahun[3]. Lowongan pekerjaan yang terdapat di situs tersebut masih sangat umum, dan tidak menyebutkan secara spesifik kebutuhan pengetahuan khusus mengenai ilmu perbankan. Setelah ditelusuri lebih dalam dengan 4 menggunakan kata kunci Lowongan kerja IT perbankan di situs tersebut terdapat 19 lapangan pekerjaan dari mulai tanggal 30 Desember 2013 hingga 28 Januari 2014. Situs pemberi kerja lainnya, seperti www.karir.com menayangkan lowongan kerja bidang IT sebanyak 160 selama kurun waktu 20 Desember 2013 sampai dengan 28 Januari 2014. Lain halnya dengan situs www.jobstreet.com, situs ini menampilkan 14 lowongan pekerjaan semenjak 30 Desember 2013 sampai dengan 26 Januari 2014 dengan memasukkan kata kunci

“Information technology bank” agar hasil yang ditampilkan hanya dengan kategori sesuai kata kunci[4].

Jika ditarik benang merah dari pernyataan-pernyataan di atas, dapat terlihat bahwa lowongan pekerjaan untuk tenaga IT masih terbuka. Hanya saja, bila dicermati pernyataan dari Aizirman Djusan masih terdapat kekosongan SDM IT di Indonesia yang menggaris bawahi adanya ketidaksesuaian antara pengajaran di perguruan tinggi dengan kebutuhan pasar. Oleh sebab itu untuk memahami karakteristik lulusan IT yang sesuai dengan kebutuhan industri, diperlukan pemetaan kebutuhan sumber daya manusia bidang TIK khususnya dalam dunia perbankan yang bertujuan agar perguruan tinggi sebagai pemasok SDM dapat mensinergikan kebutuhan pasar dengan kurikulum yang diajarkan[5].

II. METODE PENELITIAN

Penelitian ini dilakukan dengan menggunakan metoda waterfall yang terdiri dari beberapa tahapan, dimulai dari identifikasi permasalahan yang ada, pengumpulan literatur-literatur yang terkait, serta pengumpulan data dan kemudian merancang suatu sistem atau mengembangkan sistem yang sudah ada kemudian melakukan pengujian dan melakukan analisa hasil keluaran atau output sistem.

2.1 Identifikasi Masalah

Proses identifikasi masalah dengan melihat kebutuhan tenaga kerja bidang TIK dengan melihat penawaran kerja yang dikeluarkan oleh perusahaan pemberi kerja melalui Internet.

2.2 Pengumpulan data

Pengumpulan data dilakukan dengan mengumpulkan data kebutuhan tenaga kerja bidang TIK yang dikeluarkan pemberi kerja melalui website pencari atau pemberi kerja seperti jobsdb.com dan lain-lain. Data untuk ketersediaan tenaga kerja diperoleh dari website forlap.dikti.go.id.

2.3 Studi literatur

Studi *literature* dilakukan dengan kajian pustaka dari penelitian sebelumnya yang dilihat pada jurnal ilmiah yang berasal dari dalam maupun luar negeri juga karya ilmiah yang sesuai serta berita-berita di web yang mendukung penelitian ini. Jurnal ilmiah diperoleh dari artikel IEEE *Transaction*, sciencedirect.com mengenai web mining.

2.4 Text preprocess

Text Processing dapat diperoleh dengan menggunakan tiga cara, yaitu; analisa leksikal, penghilangan *stopwords*, *filtering* dan *stemming*.

III. HASIL DAN PEMBAHASAN

3.1 Deskripsi Objek Penelitian

Data yang tersebar di internet sebenarnya memiliki makna bila dapat diolah dengan tepat. Data yang dijadikan sebagai bahan penelitian merupakan kalimat dengan kata kunci

seperti lowongan kerja programmer, lowongan kerja IT analyst, lowongan kerja NET Programmer, vacancy Information Technology, lowongan kerja teknologi informasi, lowongan kerja teknik informatika, lowongan kerja system informasi.

Berdasarkan kata kunci ini kemudian diambil persyaratan yang terdapat di dalamnya untuk melamar kerja. Misalkan di dalam persyaratan untuk mengisi posisi lowongan kerja Junior .NET Developer dibutuhkan ijazah S-1 dengan IPK 3,00, mengerti dan memahami pemrograman berbasis objek, dan dapat menulis program dengan menggunakan SQL.



Gbr. 1 Contoh lowongan pekerjaan

Setelah itu disimpulkan lebih lanjut dengan mencari kata kunci lain yaitu lowongan pekerjaan Junior .NET Programmer dan menyarikan persyaratan yang dibutuhkan. Algoritma untuk mencari dan merangkum persyaratan tersebut yang akan dibangun oleh peneliti dan nantinya akan menghasilkan kesimpulan untuk pemetaan posisi pekerjaan akan dibutuhkan keahlian masing-masing. Pada akhirnya pemetaan ini akan diturunkan ke dalam kurikulum yang seharusnya bersinergi dengan perusahaan sehingga perusahaan tidak perlu lagi bersusah payah untuk mencari SDM yang berkualitas.

3.2 Hasil Analisis

Tahapan awal yang dilakukan adalah mengumpulkan kata-kata dari kalimat yang berbebaran di internet lalu menyimpannya dalam basis data. Basis data ini akan menentukan keakuratan kata kunci yang dilaksanakan pada proses akhir.

3.3 Pembuatan basis data

Lowongan kerja sebagai sumber informasi yang diperoleh akan dilakukan proses *tokenizing* dan *word counting* yang berguna untuk memecah kata-kata sekaligus untuk menghitung banyaknya kata yang muncul secara terus menerus (frekuensi). Proses ini akan dilakukan untuk setiap lowongan kerja yang disimpan.

Setelah pemecahan kalimat menjadi kata-kata dibuat penghitungan frekuensi kemunculan tiap kata dalam lowongan kerja yang akan berbentuk token untuk setiap kata. Langkah berikutnya adalah proses *filtering* yaitu menghilangkan semua kata yang tidak mengandung makna seperti nama hari, nama bulan, kata penghubung, kata sambung dan kata depan. Pengurangan kata-kata ini ditujukan agar tidak berpengaruh

pada penyimpanan sistem basis data dengan menyimpan kata-kata yang tidak dibutuhkan.

Tahap berikut yang dilakukan setelah *filtering* adalah proses *stemming*, yaitu mengubah kata menjadi kata dasar, dengan cara menghilangkan semua imbuhan yang terdapat dalam basis data kata (imbuhan; meng, kan, an, me, pe, i, di, dan lain-lain). Kata yang telah diproses ini selanjutnya disimpan ke dalam basis data dengan menggunakan cara konvensional.

Kata dalam bentuk token ini akan dipergunakan untuk dibandingkan dengan *Learning Outcomes* (LO) yang sudah ditetapkan oleh perguruan tinggi. Sehingga didapat apakah kebutuhan dunia industri sesuai dengan LO. Jika LO masih belum sesuai dengan hasil yang dengan basis data maka akan menjadi masukan bagi perguruan tinggi untuk memperbaiki kurikulumnya agar sesuai dengan kebutuhan pasar.

Gambar 2. Pembuatan basis data

Gbr. 2 Tahapan Pengerjaan

TABEL I.
PENYIMPANAN TOKEN DI BASIS DATA

No	Token	Frekuensi
1	Sarjana	3
2	MS	2
3	SQL	2
4	Server	2
5	Networking	3
6	Visual	4
7	Basic	2
8	C	2

IV. KESIMPULAN

Hasil dari penelitian ini dapat ditarik kesimpulan bahwa penggunaan text mining dapat membantu perguruan tinggi untuk menentukan *Learning Outcomes* (LO) dan mengembangkan kurikulum agar mahasiswa yang lulus dapat langsung terserap ke dunia industri.

SARAN

Penelitian ini dapat dilanjutkan ke tahapan berikutnya yaitu membangun aplikasi *text mining* dengan membuat aplikasi *text crawling* sebelumnya untuk menarik kata kunci dari seluruh lowongan yang ada di internet.

REFERENSI

- [1] Larose, D.T, 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken: Wiley-Interscience, John Wiley & Sons, Inc.
- [2] Sanchez, D., M.J. Martin-Bautista, I. Blanco, 2008, *Text Knowledge Mining: An Alternative to Text Data Mining*, IEEE International Conference on Data Mining Workshops, 978-0-7695-3503-6/08.
- [3] <http://bisniskeuangan.kompas.com/read/2013/11/25/1509248/Indonesia.Kurang.Tenaga.Ahli.Teknologi.Informasi>, diakses terakhir: 2 April 2014
- [4] <http://www.web-datamining.net/>, diakses terakhir: 16 Desember 2013
- [5] Fauzi, Andirian. <http://inet.detik.com/read/2010/01/16/180830/1279897/398/lulusan-ti-banyak-yang-mengecewa-kan>, diakses terakhir : 1 April 2014