

Pembangunan Perangkat Lunak Peringkat Dokumen dari Banyak Sumber Menggunakan *Sentence Scoring* dengan Metode TF-IDF

Fabianus Hendy Evan
Magister Teknik Informatika
Universitas Atma Jaya
Yogyakarta, Indonesia
hendyevan@gmail.com

Y. Sigit Purnomo W.P.
Magister Teknik Informatika
Universitas Atma Jaya
Yogyakarta, Indonesia
sigit@staff.uajy.ac.id

Pranowo
Magister Teknik Informatika
Universitas Atma Jaya
Yogyakarta, Indonesia
pran@staff.uajy.ac.id

Abstract—Kebutuhan untuk meringkas dokumen sangat penting saat kita memasuki era di mana banyak sekali terdapat informasi yang disajikan. Makalah ini menyajikan sistem peringkat dokumen otomatis, yang menghasilkan ringkasan untuk dokumen dari berbagai sumber. Sistem ini berdasarkan pada identifikasi dan ekstraksi kalimat penting dari dokumen yang menjadi sumber informasi. Untuk menemukan kata-kata penting yang ada di dalam dokumen, makalah ini menggunakan sentence scoring dengan metode TF-IDF.

Metode TF-IDF ini memberikan nilai kepada masing-masing kata sesuai dengan tingkat penggunaannya di seluruh sumber dokumen yang ada. Sentence scoring dilakukan dengan menjumlah nilai dari masing-masing kata sehingga didapatkan nilai akhir dari masing-masing kalimat. Kalimat yang akan digunakan sebagai ringkasan di seleksi sesuai dengan peringkat di dalam masing-masing dokumen. Sentence scoring menggunakan metode TF-IDF dapat menghasilkan ringkasan dokumen yang bersifat ekstraktif yang memudahkan pengguna dalam memahami inti dari sebuah dokumen.

Keywords—TF-IDF, peringkat dokumen, sentence scoring

I. PENDAHULUAN

I.1 Latar Belakang

Kemajuan teknologi Internet yang semakin pesat telah memaksa manusia untuk berusaha mengikutinya. Hal tersebut juga di iringi dengan semakin banyaknya informasi yang tersebar secara luas di dunia maya. Jumlah halaman yang tersedia di Internet hampir dua kali lipat setiap tahunnya [8].

Mesin pencari memberikan banyak referensi media online dari sebuah kata kunci yang di berikan oleh user. Dari berbagai media online tersebut seringkali berita yang di sajikan memiliki topik yang sama, namun berbeda dalam penyajian beritanya.

Hal ini dapat membingungkan pengguna ketika ingin mendapatkan informasi yang relevan terkait suatu topik tertentu karena banyaknya sumber berita yang beredar.

Makalah ini memberikan pendekatan peringkat dokumen sebagai solusi untuk masalah ini. Ringkasan dapat didefinisikan sebagai teks yang dihasilkan dari satu atau lebih teks, yang menyampaikan informasi penting dalam teks asli dengan rasio kompresi yang cukup dan secara signifikan berisi teks lebih sedikit dari sumber [9].

Peringkat dokumen dari banyak sumber menyediakan ringkasan dari informasi yang terdapat di berbagai media online. Peringkat dokumen dari banyak sumber adalah proses penyaringan informasi penting dari satu set dokumen untuk menghasilkan informasi yang lebih relevan untuk pengguna. Hal ini dapat dilihat sebagai perpanjangan peringkat dokumen tunggal yang hanya menyediakan ringkasan dari satu dokumen saja.

Di dalam penelitian ini digunakan metode TF-IDF untuk membuat sebuah ringkasan dari banyak sumber dokumen. Metode TF-IDF digunakan untuk memberikan nilai pada kalimat yang dianggap sebagai inti dari dokumen. Kalimat tersebut nantinya akan di gabungkan sehingga terbentuk sebuah ringkasan dari dokumen-dokumen yang ada.

I.2 Tujuan Penelitian

Adapun tujuan yang ingin diperoleh dari penelitian ini adalah mengelompokkan kalimat dari beberapa dokumen menggunakan sentence scoring dengan metode TF-IDF untuk membuat sebuah ringkasan akhir.

1.3 Batasan Masalah

Persoalan penelitian ini masih terbuka luas dan melebar dilihat dari berbagai aspek dan sisi. Oleh karena itu, agar penelitian tetap menjadi fokus diberikan beberapa batasan dan asumsi sebagai berikut:

- Ringkasan akhir yang dihasilkan berupa ringkasan ekstraktif.
- Dokumen yang di ringkas adalah artikel dari media online.

II. TINJAUAN PUSTAKA

II.1 Summarization

Peringkasan dokumen adalah proses pengambilan teks dari sebuah dokumen dan membuat sebuah ringkasan yang mempunyai informasi yang lebih berguna bagi user [10]. Perbedaan antara *single-document summarization* dan *multi-document summarization* adalah, *multi-document summarization* lebih kompleks daripada *single-document summarization*. Faktor yang membuat *multi-document summarization* lebih kompleks antara lain:

- Setiap artikel yang di gunakan di tulis oleh penulis yang berbeda yang mempunyai perbedaan gaya menulis dan struktur dokumen.
- Beberapa artikel mungkin mempunyai pandangan yang berbeda tentang suatu topik yang sama.
- Fakta dan sudut pandang dapat berubah kapanpun setiap dokumen di buat sehingga dapat menimbulkan konflik dalam penyediaan informasi.

Secara umum terdapat dua model ringkasan dokumen yaitu model ekstraktif dan model abstraktif [3,4,5,6]. Model berbasis ekstraktif mengambil kalimat yang sudah ada di dalam dokumen dan kemudian menggabungkannya untuk membentuk sebuah ringkasan [13], sehingga memungkinkan ringkasan untuk meningkatkan informasi secara keseluruhan tanpa menambah panjang ringkasan [15]. Meskipun ringkasan yang dihasilkan dengan cara ini mungkin kurang koherensi, pendekatan ekstraktif masih populer saat ini karena mudah untuk diterapkan.

Dengan tujuan untuk menghasilkan ringkasan yang koheren dan gramatikal, ringkasan abstraktif menciptakan ringkasan dengan cara sintesis dan menulis ulang kalimat. Kalimat tersebut berdasarkan pemahaman kontekstual dan linguistik, yang sangat tergantung pada teknik bahasa dan analisis yang mendalam. Rather than hanya menggunakan teknik tunggal saja, beberapa peneliti tertarik regenerasi sebagai pasca-proses untuk menghasilkan ringkasan ekstraktif, yaitu, membuat pemangkasan atau revisi berdasarkan ringkasan ekstraktif [6,7].

Referensi [5], membuat sebuah framework di mana *multi-dokumen summarization* dapat di modelkan menggunakan teori informasi jarak. Ringkasan terbaik di definisikan sebagai yang mempunyai informasi jarak yang minimal (kondisi informasi jarak) ke seluruh dokumen. Referensi [9] menjelaskan sebuah sistem peringkasan dokumen dengan model ekstraktif yang

mengambil kalimat dari banyak dokumen sesuai dengan *cluster centroids* dari dokumen.

II.2 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode yang biasa di pakai dalam pembobotan sebuah kata di dalam sistem pencarian informasi [1]. Metode TF-IDF menghitung nilai dari masing-masing kata di dalam dokumen menggunakan frekuensi kata tersebut muncul. Kata dengan nilai TF-IDF yang tinggi berarti bahwa kata tersebut mempunyai hubungan yang kuat dengan dokumen di mana kata tersebut muncul, di asumsikan bahwa jika kata tersebut muncul di dalam query maka akan memiliki ketertarikan bagi pengguna [12]. TF-IDF juga dapat digunakan dalam pembobotan kata untuk mencari keputusan yang relevan [15].

Prosedur dalam implementasi TF-IDF terdapat perbedaan kecil di dalam semua aplikasinya, tetapi pendekatannya kurang lebih sama. TF (*Term Frequency*) merupakan frekuensi sebuah kata terdapat di dalam sebuah dokumen. Nilai dari TF didapat menggunakan persamaan:

$$TF(t) = f_{t,d} / \sum t, d \quad (1)$$

dimana $f_{t,d}$ merupakan frekuensi sebuah kata (t) muncul di dalam dokumen d , sedangkan $\sum t, d$ merupakan total keseluruhan kata yang terdapat di dalam dokumen d [1].

Kemudian untuk menghitung nilai IDF (*Inverse Document Frequency*) dari sebuah kata di dalam kumpulan dokumen menggunakan persamaan:

$$IDF(t) = \log(|D|/f_{t,D}) \quad (2)$$

$|D|$ merupakan jumlah dokumen yang ada di dalam koleksi, sedangkan $f_{t,D}$ merupakan jumlah dokumen di mana t muncul di dalam D [2,13].

Dalam koleksi dokumen D , sebuah kata t dan dokumen individu $d \in D$, dapat di hitung nilai TF-IDF menggunakan rumus:

$$TF - IDF(t) = TF(t) * IDF(t) \quad (3)$$

Diasumsikan bahwa $|D| \sim f_{t,D}$, ukuran dari kumpulan dokumen hampir sama dengan frekuensi t di dalam D . Jika $1 < \log(|D|/f_{t,D}) < c$ untuk sebuah konstanta c dengan nilai yang kecil, maka w_d akan lebih kecil daripada $f_{t,d}$ tetapi tetap bernilai positif. Hal berarti bahwa w mempunyai relasi yang biasa dengan seluruh dokumen tetapi tetap menyimpan beberapa informasi yang penting [12].

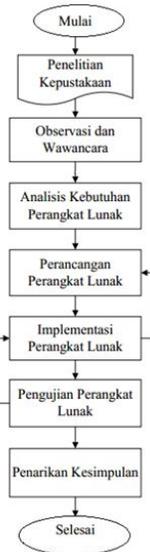
Sebagai contoh adalah ketika metode TF-IDF menemukan kata android di dalam kumpulan dokumen yang berisi artikel tentang android. Hal ini juga berlaku untuk beberapa kata seperti kata sambung yang tidak mempunyai arti yang relevan terhadap sebuah dokumen. Kata sambung tersebut di beri nilai yang rendah menggunakan TF-IDF.

Jika $f_{t,d}$ bernilai besar dan $f_{t,D}$ bernilai kecil, maka nilai dari $\log(|D|/f_{t,D})$ cenderung tinggi dan juga TF-IDF(t) akan bernilai tinggi. Kata dengan nilai TF-IDF yang tinggi berarti bahwa kata tersebut merupakan sebuah kata yang sangat penting di dalam dokumen d tetapi tidak di dalam kumpulan dokumen D .

III. METODOLOGI PENELITIAN

III.1 Proses Penelitian

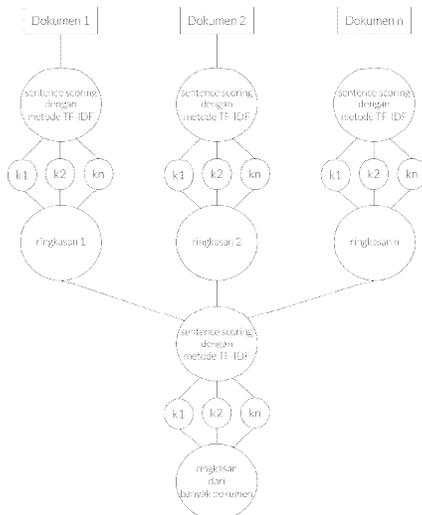
Tahapan proses metodologi penelitian yang dilakukan dalam penelitian ini terdapat pada gambar berikut:



Gambar 1. Diagram alir metodologi penelitian

III.2 Proses Peringkasan Dokumen

Tahapan proses peringkasan dokumen yang dilakukan dalam penelitian ini dapat di lihat seperti gambar berikut:



Gambar 2. Diagram alir pembentukan ringkasan dokumen dari banyak sumber

Proses tersebut dimulai dari pembuatan ringkasan dokumen tunggal. Setiap kalimat di dalam masing-masing dokumen diberi skor dengan menggunakan metode TF-IDF, kemudian dipilih beberapa kalimat dengan skor tertinggi. Kalimat tersebut yang nantinya akan menjadi ringkasan dari masing-masing dokumen. Proses selanjutnya adalah menggabungkan

semua ringkasan dokumen tunggal tersebut menjadi satu dan melakukan perhitungan skor kalimat menggunakan metode TF-IDF. Kalimat-kalimat dengan skor tertinggi akan menjadi ringkasan akhir dari banyak dokumen tersebut.

IV. HASIL DAN PEMBAHASAN

Uji coba yang dilakukan adalah dengan meringkas multi dokumen menggunakan TF-IDF. Makalah ini memberikan dokumen masukan sebanyak 2 yang mempunyai topik yang mirip. Dokumen ini di dapat dari media online yang banyak terdapat di dunia maya. Langkah-langkah peringkasan dokumen tersebut adalah:

1. Dokumen di pecah ke dalam kalimat-kalimat individu dan menghilangkan tanda penghubung yang terdapat di masing-masing kalimat seperti titik, koma, tanda seru, petik dan sebagainya. Kemudian diubah menjadi *lowercase*.
2. Masing-masing kalimat tersebut di pecah lagi menjadi kata-kata individu. Proses pemecahan kalimat menjadi kata diberikan filter yaitu kata sambung. Setelah itu hitung TF dari kata tersebut di dalam sebuah dokumen menggunakan persamaan (1).
3. Cari nilai IDF dari masing-masing kata tersebut dengan melihat pada seluruh dokumen menggunakan persamaan (2).
4. Masing-masing kata tersebut di hitung nilai TF-IDF nya menggunakan persamaan (3). Dari hasil perhitungan TF-IDF di dapatkan kata-kata yang mempunyai nilai yang berbeda, tergantung dari seberapa sering kata tersebut muncul di dokumen.
5. Lakukan *sentence scoring* dari sebuah dokumen menggunakan nilai TF dari masing-masing kata. Kalimat yang mempunyai skor tertinggi dipilih untuk kemudian di jadikan ringkasan dari sebuah dokumen.
6. Ringkasan-ringkasan dari masing-masing dokumen tersebut kemudian digabungkan menjadi satu.

Agar terbentuk sebuah ringkasan akhir, lakukan pemilihan kalimat ini menggunakan metode TF-IDF dan *sentence scorings* sama seperti ketika memilih kalimat yang akan digunakan sebagai ringkasan pada sebuah dokumen tunggal.

Makalah ini menguji perhitungan TF-IDF untuk *sentence scoring* pada dua dokumen untuk mendapatkan ringkasan dari ke dua dokumen tersebut. Hasil *sentence scoring* menggunakan metode TF-IDF pada dokumen tunggal:

Dokumen I:

"Pertarungan bisnis media sosial di Indonesia makin panas. Salah satu aplikasi jejaring sosial yang sedang naik daun di Indonesia, Path, baru saja menyatakan mendapat suntikan dana segar dari Bakrie Global Group sebesar US\$ 25 juta atau sekitar Rp 304 miliar.

"Kami sangat bahagia bisa bergabung dengan pertumbuhan Path," kata CEO Bakrie Anindya Bakrie seperti dikutip Recode.net, Sabtu, 11 Januari 2014. "Dengan tim manajemen yang solid dan rencana bisnis yang bagus, Path

akan terus menghubungkan banyak orang Indonesia secara personal, bermakna, dan produktif."

Kucuran modal segar ini membuat Path sudah mengantongi modal US\$ 65 juta. Sebelumnya, pada 2011, Path sudah mendapat investasi sebesar US\$ 10 juta.

CEO Path Dave Morin sangat berbahagia dengan kerja sama ini. Recode.net mengisahkan transaksi itu hingga akhirnya mencapai kata sepakat. "Setelah melalui jalan sulit selepas tahun 2013 yang begitu menantang," kata Dave Morin kepada Recode.net.

Tak disebutkan berapa besar saham Path yang dikuasai Bakrie. Yang pasti, sebelum Bakrie masuk, sudah ada investor-investor terdahulu seperti Greylock Partners, Kleiner Perkins, Index Ventures, Insight Venture Partners, Redpoint Venture Partners, dan First Round Capital.

Pendiri dan CEO Path, Dave Morin, dalam berbagai kesempatan menyatakan dia sudah lama ingin menggandeng investor dari Asia, khususnya Asia Tenggara. Di wilayah ini, jumlah pengguna Path tumbuh luar biasa. "Bisnis kami di Asia sangat kuat. Kami selalu berupaya mencari mitra untuk berekspansi," ujar Morin.

Jumlah pengguna Path di seluruh dunia saat ini mencapai 23 juta orang. Sekitar 25 persen pengguna, sekitar enam juta, datang dari Indonesia. Dari jumlah itu, pengguna yang aktif diperkirakan mencapai empat juta orang.

"Jumlah pengguna Path di sini (Indonesia) bahkan lebih besar dari Amerika Serikat," ujar Dave Morin saat bertandang ke Jakarta, November lalu. Itu membuat saya datang ke sini (Jakarta). Saya ingin mempelajari karakter orang-orang di sini," jelas Morin. Ia menegaskan suntikan modal baru itu akan dimanfaatkan untuk menguatkan pertumbuhan Path."

TABLE I. DOKUMEN I

No	Kalimat	Skor
1	Pertarungan bisnis media sosial di Indonesia makin panas.	0.0429
2	Salah satu aplikasi jejaring sosial yang sedang naik daun di Indonesia, Path, baru saja menyatakan mendapat suntikan dana segar dari Bakrie Global Group sebesar US\$ 25 juta atau sekitar Rp 304 miliar.	0.2046
3	"Kami sangat bahagia bisa bergabung dengan pertumbuhan Path," kata CEO Bakrie Anindya Bakrie seperti dikutip Recode.net, Sabtu, 11 Januari 2014.	0.1452
4	"Dengan tim manajemen yang solid dan rencana bisnis yang bagus, Path akan terus menghubungkan banyak orang Indonesia secara personal, bermakna, dan produktif."	0.1072
5	Kucuran modal segar ini membuat Path sudah mengantongi modal US\$ 65 juta.	0.1056
6	Sebelumnya, pada 2011, Path sudah mendapat investasi sebesar US\$ 10 juta.	0.0924
7	CEO Path Dave Morin sangat berbahagia dengan kerja sama ini.	0.0891
8	Recode.net mengisahkan transaksi itu hingga akhirnya mencapai kata sepakat.	0.0396
9	"Setelah melalui jalan sulit selepas tahun 2013 yang begitu menantang," kata Dave Morin kepada Recode.net.	0.0792
10	Tak disebutkan berapa besar saham Path yang dikuasai Bakrie.	0.0792
11	Yang pasti, sebelum Bakrie masuk, sudah ada investor-investor terdahulu seperti Greylock Partners, Kleiner Perkins, Index Ventures, Insight Venture Partners, Redpoint Venture Partners, dan First Round Capital.	0.1122
12	Pendiri dan CEO Path, Dave Morin, dalam berbagai kesempatan menyatakan dia sudah lama ingin menggandeng investor dari Asia, khususnya Asia Tenggara.	0.1419
13	Di wilayah ini, jumlah pengguna Path tumbuh luar biasa.	0.0792
14	"Bisnis kami di Asia sangat kuat.	0.0363
15	Kami selalu berupaya mencari mitra untuk berekspansi," ujar Morin.	0.0429
16	Jumlah pengguna Path di seluruh dunia saat ini mencapai 23 juta orang.	0.1254
17	Sekitar 25 persen pengguna, sekitar enam juta, datang dari Indonesia.	0.0924
18	Dari jumlah itu, pengguna yang aktif diperkirakan mencapai empat juta orang.	0.0792
19	"Jumlah pengguna Path di sini (Indonesia) bahkan lebih besar dari Amerika Serikat," ujar Dave Morin saat bertandang ke Jakarta, November lalu.	0.1617
20	Itu membuat saya datang ke sini (Jakarta).	0.0198
21	Saya ingin mempelajari karakter orang-orang di sini," jelas Morin.	0.0396
22	Ia menegaskan suntikan modal baru itu akan dimanfaatkan untuk menguatkan pertumbuhan Path.	0.0825

Pada tabel I dapat dilihat nilai masing-masing kalimat dari dokumen I yang di dapatkan menggunakan metode TF-IDF. Dari seluruh nilai yang ada di ambil beberapa kalimat dengan nilai terbesar. Dalam percobaan ini kalimat yang diambil berjumlah 4 dan nantinya akan menjadi ringkasan dari dokumen tersebut. Pada contoh dokumen I di dapat kalimat 2, 3, 12 dan 19.

Dokumen II:

"Path mendapat pendanaan putaran ketiga (Seri C) dari beberapa investor, yang totalnya mencapai 25 juta dollar AS. Dalam investasi kali ini, Path mendapat investor baru salah satunya dari Indonesia, yaitu Bakrie Global Group.

Pendiri dan CEO Path, Dave Morin, mengatakan, pihaknya memang mencari strategi dan fokus mengembangkan bisnis di Asia, terutama Asia Tenggara, yang menjadi tempat di mana layanan Path tumbuh.

Sebelumnya, Morin telah melakukan kunjungan bisnis ke Indonesia pada awal November 2013. Menurut sumber KompasTekno, salah satu tujuan kunjungan itu adalah terkait investasi dengan kelompok Bakrie.

Pada Juli 2013, Business Manager Path, Andreas Bezamat-Homer, juga berkunjung ke Indonesia. Ia menyatakan keseriusan Path untuk menggarap bisnis di Indonesia.

Selain Bakrie Global Group, investor lain yang berpartisipasi dalam pendanaan Seri C ini termasuk Greylock

Partners, Kleiner Perkins, Index Ventures, Insight Venture Partners, Redpoint Venture Partners, dan First Round Capital.

Jadi total dana 25 juta dollar AS atau sekitar Rp 300 miliar untuk Path "dikeroyok" oleh pihak-pihak di atas, bukan hanya dikeluarkan oleh pihak Bakrie Global Group saja."

TABLE II. DOKUMEN II

No	Kalimat	Skor
1	Path mendapat pendanaan putaran ketiga (Seri C) dari beberapa investor, yang totalnya mencapai 25 juta dollar AS.	0.18072
2	Dalam investasi kali ini, Path mendapat investor baru salah satunya dari Indonesia, yaitu Bakrie Global Group.	0.21084
3	Pendiri dan CEO Path, Dave Morin, mengatakan, pihaknya memang mencari strategi dan fokus mengembangkan bisnis di Asia, terutama Asia Tenggara, yang menjadi tempat di mana layanan Path tumbuh.	0.22886
4	Sebelumnya, Morin telah melakukan kunjungan bisnis ke Indonesia pada awal November 2013.	0.09036
5	Menurut sumber KompasTekno, salah satu tujuan kunjungan itu adalah terkait investasi dengan kelompok Bakrie.	0.10841
6	Pada Juli 2013, Business Manager Path, Andreas Bezamat-Homer, juga berkunjung ke Indonesia.	0.11444
7	Ia menyatakan keseriusan Path untuk menggarap bisnis di Indonesia.	0.10842
8	Selain Bakrie Global Group, investor lain yang berpartisipasi dalam pendanaan Seri C ini termasuk Greylock Partners, Kleiner Perkins, Index Ventures, Insight Venture Partners, Redpoint Venture Partners, dan First Round Capital.	0.28308
9	Jadi total dana 25 juta dollar AS atau sekitar Rp 300 miliar untuk Path "dikeroyok" oleh pihak-pihak di atas, bukan hanya dikeluarkan oleh pihak Bakrie Global Group saja.	0.24695

Pada tabel II dapat dilihat nilai masing-masing kalimat dari dokumen I yang di dapatkan menggunakan metode TF-IDF. Dari seluruh nilai yang ada di ambil beberapa kalimat dengan nilai terbesar. Dalam percobaan ini kalimat yang diambil berjumlah 4 dan nantinya akan menjadi ringkasan dari dokumen

tersebut. Pada contoh dokumen II di dapat kalimat 2, 3, 8 dan 9.

Kemudian hasil ringkasan dari kedua dokumen tersebut di gabungkan dan di seleksi lagi dengan memberikan nilai baru pada masing-masing kalimat menggunakan metode TF-IDF

TABLE III. HASIL SENTENCE SCORING RINGKASAN DOKUMEN I DAN DOKUMEN II

No	Kalimat	Skor
1	Salah satu aplikasi jejaring sosial yang sedang naik daun di Indonesia, Path, baru saja menyatakan mendapat suntikan dana segar dari Bakrie Global Group sebesar US\$ 25 juta atau sekitar Rp 304 miliar.	0.10855
2	"Kami sangat bahagia bisa bergabung dengan pertumbuhan Path," kata CEO Bakrie Anindya Bakrie seperti dikutip Recode.net, Sabtu, 11 Januari 2014.	0.1002
3	Pendiri dan CEO Path, Dave Morin, dalam berbagai kesempatan menyatakan dia sudah lama ingin menggandeng investor dari Asia, khususnya Asia Tenggara.	0.05845
4	"Jumlah pengguna Path di sini (Indonesia) bahkan lebih besar dari Amerika Serikat," ujar Dave Morin saat bertandang ke Jakarta, November lalu.	0.09185
5	Dalam investasi kali ini, Path mendapat investor baru salah satunya dari Indonesia, yaitu Bakrie Global Group.	0.02505
6	Pendiri dan CEO Path, Dave Morin, mengatakan, pihaknya memang mencari strategi dan fokus mengembangkan bisnis di Asia, terutama Asia Tenggara, yang menjadi tempat di mana layanan Path tumbuh.	0.1002
7	Selain Bakrie Global Group, investor lain yang berpartisipasi dalam pendanaan Seri C ini termasuk Greylock Partners, Kleiner Perkins, Index Ventures, Insight Venture Partners, Redpoint Venture Partners, dan First Round Capital.	0.24215
8	Jadi total dana 25 juta dollar AS atau sekitar Rp 300 miliar untuk Path "dikeroyok" oleh pihak-pihak di atas, bukan hanya dikeluarkan oleh pihak Bakrie Global Group saja.	0.1169

Dari tabel di atas dapat di lihat skor yang baru untuk masing-masing kalimat dari kedua ringkasan yang digabungkan. Dari hasil tersebut kemudian dipilih 3 kalimat dengan nilai terbesar untuk di jadikan ringkasan akhir dari ke dua dokumen tadi. Di dapat kalimat 1, 7 dan 8 yang di gunakan sebagai ringkasan akhir dari ke dua dokumen tersebut.

Hasil ringkasan akhir tersebut adalah:

"Salah satu aplikasi jejaring sosial yang sedang naik daun di Indonesia, Path, baru saja menyatakan mendapat suntikan dana segar dari Bakrie Global Group sebesar US\$ 25 juta atau sekitar Rp 304 miliar. Selain Bakrie Global Group, investor lain yang berpartisipasi dalam pendanaan Seri C ini termasuk Greylock Partners, Kleiner Perkins, Index Ventures, Insight Venture Partners, Redpoint Venture Partners, dan First

Round Capital. Jadi total dana 25 juta dollar AS atau sekitar Rp 300 miliar untuk Path "dikeroyok" oleh pihak-pihak di atas, bukan hanya dikeluarkan oleh pihak Bakrie Global Group saja."

Round Capital. Jadi total dana 25 juta dollar AS atau sekitar Rp 300 miliar untuk Path "dikeroyok" oleh pihak-pihak di atas, bukan hanya dikeluarkan oleh pihak Bakrie Global Group saja."

Dari hasil kedua dokumen tersebut dapat dilihat bahwa ringkasan yang dihasilkan adalah kalimat-kalimat yang mempunyai keterkaitan dengan topik dari dokumen, dalam hal ini topiknya adalah Path. Dengan menggunakan metode TF-IDF maka kata-kata yang mempunyai relevansi tinggi terhadap topik dari sebuah dokumen akan mendapat skor yang tinggi, sehingga kalimat yang memiliki kata tersebut juga akan mendapatkan skor tinggi.

V. KESIMPULAN

menggunakan metode TF-IDF ini nilai dari masing-masing kata dapat digunakan untuk memberikan skor pada tiap kalimat sehingga menghasilkan sebuah ringkasan dari banyak dokumen. Penggunaan metode ini juga dapat

menyeleksi kata mana yang jarang dan sering muncul dalam sebuah kumpulan dokumen, sehingga dalam proses meringkas dokumen dapat dicari topik yang sesuai dengan dokumen tersebut.

Untuk lebih mengoptimalkan pemilihan kalimat yang akandi gunakan sebagai ringkasan, dalam penelitian ke depan mungkin akan lebih baik jika di gabungkan dengan metode yang mencari kemiripan antara kata dan kalimat sehingga tidak terjadi adanya redundansi.

November 2007. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1284916.1285161>

REFERENCES

- [1] Aizawa, A. 2003. An information-theoretic perspective of TF-IDF measures. *Inf. Process. Manage.* 39, 1, 45–65.
- [2] Berger, A et al (2000). Bridging the Lexical Chasm: Statistical Approaches to Answer Finding. In *Proc. Int. Conf. Research and Development in Information Retrieval*, 192-199.
- [3] Jing H. 2000. Sentence reduction for automatic text summarization. In: *Proceedings of 6th conference on applied natural language processing (ANCL'00)*, pp 310–315.
- [4] Knight K, Marcu D. 2002. Summarization beyond sentence extraction: a probabilistic approach to sentence compression. *Artif Intell* 139(1):91–107.
- [5] Long, Chong, Huang, Min-Lie, Zhu, Xiao-Yan and Li, Ming. 2010. A New Approach for Multi-Documen Update Summarization. *JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY* 25(4): 739-749 July 2010. DOI 10.1007/s11390-010-1057-8.
- [6] Mani I. 2001. Automatic summarization. John Benjamin's Publishing Company, Amsterdam.
- [7] Otterbacher JC, Radev DR, Luo A. 2002. Revisions that improve cohesion in multi-document summaries: a preliminary study. In: *Proceedings of the ACL02 workshop on automatic summarization*, pp 27 – 36.
- [8] Radev, D.R. and Fan, W. 2000, "Automatic summarization of search engine hit lists". *Proceedings of the ACL-2000 workshop on recent advances in natural language processing and information retrieval*, Hong Kong, 2000, pp. 99-109.
- [9] Radev, D. R., Jing, H., & Budzikowska, M. (2000). Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *ANLP/NAACL Workshop on Summarization* Seattle, WA.
- [10] Radev, D. R., Hovy, E., and McKeown, K. 2002. "Introduction to the special issue on summarization. *Computational Linguistics*", 28(4):399-408.
- [11] Radev, Dragomir R., Hongyan Jing, Malgorzata Budzikowska, 2010. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies.
- [12] Ramos, Juan. 2000. Using TF-IDF to Determine Word Relevance in Document Queries. Department of Computer Science, Rutgers University, 23515 BPO Way, Piscataway, NJ, 08855.
- [13] Salton, G. & Buckley, C. (1988). Term-weighting approache sin automatic text retrieval. In *Information Processing & Management*, 24(5): 513-523.
- [14] Turner, J. and E. Charniak, "Supervised and unsupervised learning for sentence compression," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ser. ACL '05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 290– 297. [Online]. Available:<http://dx.doi.org/10.3115/1219840.1219876>.
- [15] Wu, H. C., Luk, R. W. P., Wong, K. F., and Kwok, K. L. 2008. Interpreting TF-IDF term weights as making relevance decisions. *ACM Trans. Inform. Syst.* 26, 3, Article 13 (June 2008), 37 pages. <http://doi.acm.org/10.1145/1361684.136168>
- [16] Zajic D., B. J. Dorr, J. Lin, and R. Schwartz, "Multi-candidate reduction: Sentence compression as a tool for document summarization tasks," *Inf. Process. Manage.*, vol. 43, pp. 1549–1570,