Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)
Vol. 1, No. 1, April 2014, hlm. 39-44

PERINGKASAN DOKUMEN BAHASA INDONESIA BERBASIS NON-NEGATIVE MATRIX FACTORIZATION (NMF)

Achmad Ridok

Program Studi Ilmu Komputer, Universitas Brawijya
Email : acridokb@ub.ac.id

(Naskah masuk: 10 Januari 2014, diterima untuk diterbitkan: 17 Februari 2014)

Abstrak

Peningkatan teknologi informasi telah memicu peningkatan dokumen teks digital secara massif termasuk dokumen berbahasa Indonesia. Penggalian informasi dari dokumen berupa ringkasan secara otomatis sangat dibutuhkan. Pada penelitian ini peringkasan otomatis menggunakan Nonnegatif Matrix Factorization (NMF) telah dikembangkan. Sistem dievaluasi dengan membandingkan ringkasan sistem dengan ringkasan dari 3 orang pakar terhadap 100 dokumen bahasa Indonesia . Hasil evaluasi menunjukkan ringkasan sistem mempunyai rata-rata presisi dan recall masing-masing 0.19724 dan 0.34085. Sedangkan evaluasi ringkasan antar pakar mempunyai rata-rata presisi dan recall masing-masing 0.68667 dan 0.70642..

Kata kunci: *peringkasan dokumen, NMF*

Abstract

Improvement of information technology has led to increased massively digital text documents, including documents of Indonesian language. Extracting information from documents such as automatic summary is needed. In this study peringkasan automatically using non-negative Matrix Factorization (NMF) has been developed. The system was evaluated by comparing summary of system with summary of of three experts on 100 Indonesian documents. The evaluation shows summary of the system has an average precision and recall respectively 0.19724 and 0.34085. While the summary of an expert evaluation had an average precision and recall respectively 0.68667 and 0.70642.

Keywords: *text summarization, NMF*

1. PENDAHULUAN

Peningkatan teknologi informasi telah memicu terjadinya peningkatan dokumen teks digital secara masif. Penggalian informasi yang terkandung dalam suatu dokumen teks tersebut diperlukan dengan cara untuk membangkitkan informasi yang mencakup keseluruhan dokumen secara ringkas yang disebut peringkasan. Dengan memberikan ringkasan suatu dokumen dapat disajikan inti dokumen secara singkat namun memenuhi keperluan pembaca untuk mengetahui secara cepat isi dokumen tanpa harus membaca seluruh dokumen. Peringkasan berbasis computer yang mampu membangkitkan ringksasan dokumen secara otomatis.

Penelitian ini telah dimulai sejak tahun 1958 oleh Luhn (Luhn, 1958, 159–165) dan terus berkembang sampai sekarang sejalan dengan meningkatnya data teks berbasis digital. Secara umum terdapat dua pendekatan yakni ekstraksi dan abstraksi. Pendekatan yang pertama berusaha membangkitkan ringkasan dokumen berdasarkan kata-kata atau kalimat-kalimat yang ada dalam teks asal berdasarkan tingkat kepentingannya. Pendekatan yang kedua melakukan peringkasan teks dengan cara menginterpretasikan teks asal dan menuliskan kembali ke dalam versi yang lebih singkat tetapi mempunyai semantik yang sama (Lin, 1999, pages 81–94). Berdasarkan konten atau isi, suatu peringkasan dibagi kedalam generik dan ringkasan berbasis query. Ringkasan berbasis generik mengasumsikan pengguna tidak mempunyai pemahaman awal tentang teks yang akan diringkas sehingga dapat digunakan oleh sembarang tipe pengguna (Karel Jezek, 2008, pp. 1-12). Sedangkan peringkasan berbasis query dibuat berdasarkan topik yang diinginkan oleh pengguna. Pengguna mempunyai informasi umum tentang teks yang akan diringkas dan mencari informasi spesifik dengan cara menjawab suatu pertanyaan (Karel Jezek, 2008, pp. 1-12).

Secara umum metode otomatisasi generik peringkasan dokumen terbagi dalam dua katagori : supervisi dan unsupervisi (Mani, 2001). Metode supervise didasarkan pada algoritma yang menggunakan sejumlah besar ringkasan buatan manusia sebagai data latih dan hasilnya adalah model peringkasan. Dengan demikian ringkasan hasil sistem sangat tergantung pada data latih dan jika pengguna ingin mengubah tujuan ringkasan maka ia harus merekontruksi kembali model dan data latihnya (Amini, 2002). Metode unsupervisi tidak memerlukan data latih ringkasan buatan

manusia untuk melatih sistem (Nomoto, 2001). Sampai saat ini terdapat dua metode generic unsupervisi yakni (*Latent Semantic Analysis*) LSA (Gong, 2001) dan (*Non-negative Matrix Factorization*) NMF (Ju-Hong Lee, 2009 20–34). Kedua metode ini mencoba membuat ringkasan dokumen dengan cara membangkitkan fitur semantik dari dokumen. Metode yang pertama mencoba membangkitkan ringksan dokumen berdasarkan keterkaitan semantic antar kata, sedang metode yang kedua melakukan peringkasan berdasarkan keterkaian semantic antar kalimat. Namun demikian metode yang terakhir memberikan kinerja yang lebih baik dibandingkan LSA (Ju-Hong Lee, 2009 20–34).

Pemanfaatan fitur semantik dalam membangkitkan ringkasan dokumen berbahasa Indonesia masih jarang dilakukan. Walaupun metode ekstraksi telah banyak digunkan diantaranya sebagaimana telah dilakukan oleh Ridok dan Cahyo (Ridok , 2013). Oleh karenanya pada penelitian ini dikembangkan sistem peringkasan dokumen bahasa Indonesia menggunakan kemampuan NMF.

Artikel ini disusun dengan sistematika sebagai berikut: bagian pertama berisi pendahuluan, selanjutnya akan diuraikan beberapa penelitian terkait pada bagian tinjaun pustaka. Setelah diuraikan metode penelitian dan data akan dibahas implementasi dan hasil sistem. Selanjutnya akan ditutup dengan kesimpulan dan saran pengembangan.

2. Tinjauan Pustaka

Peringkasan merupakan proses untuk mendapatkan intisari informasi terpenting dari suatu dokumen sehingga diperoleh versi yang lebih ringkas suatu dokumen. Secara umum terdapat dua tipe peringkasan yaitu ekstraktif dan abstraktif. Ekstraktif meringkas suatu dokumen dengan memilih sebagian dari kalimat yang ada dalam dokumen asli. Metode ini menggunakan statistik, linguistik, dan heuristik atau kombinasi dari semuanya dalam menetapkan kalimat ringkasannya. Sedangkan metode abstraktif melakukan peringkasan dengan cara menginterpretasi teks asal melalui proses transformasi suatu kalimat asli. Meskipun ringkasan yang dihasilkan oleh manusia bersifat tidak ekstraktif, akan tetapi kebanyakan penelitian mengenai peringkas ini adalah ekstraktif yang memberikan hasil yang lebih baik apabila dibandingkan dengan peringkasan abstraktif (Erkan, 2004).

Metode peringkasan tesks secara umum terdapat dua yakni peringkasan tersupervisi dan peringkasan non supervisi. Peringkasan tersupervisi diawali tahun 1995 oleh Kupiec dan kawan-kawan yang mengusulkan pemamfaatan metode statistik. Pada metode ini peringkasan dilakukan dengan menggunakan aturan Bayes yang dilatih pada hasil peringkasan manual (Kupiec, Pedersen, and Chen (1995). Chuang and Yang (2000) mengusulkan metode supervisi yang menggunakan teknik mesin pembelajaran (*machine learning*) untuk mengekstrak kalimat. Metode ini membagi kalimat-kalimat ke dalam segmen-segmen yang di representasikan oleh sekumpulan fitur yang telah didefinisikan sebelumnya. Peringkasan dilatih untuk mengestrak bagian-bagian kalimat penting berdasarkan pada kumpulan fitur ini. Amini and Gallinari (2002) mengusulkan algoritma semi-supervisi untuk melatih model pengklasifikasi peringkasan teks. Metode ini menggunakan CEM (*classification exptectation maximum*) sebagai suatu metode semi-supervisi dengan cara member label beberapa item data yang digunakan bersama-sama dengan sejumlah besar data yang belum berlabel untuk pelatihan. Yeh dan kawan-kawan (2005) mengusulkan suatu peringkasan baru yang dapat dilatih untuk peringkasan dokumen dengan menggunakan beberapa fitur dokumen. Posisi kalimat dirangking dan selanjutnya digunakan algoritma genetic untuk melatih fungsi skor.

2.1. Metode LSA

Metode LSA memanfaatkan kemampuan metode aljabar dalam mencoba menemukan kesamaan antar kalimat dan kesamaan antar kata menggunakan *Singular Value Decomposition* (SVD) yang dikenal dengan *Latent Semantix Analysis (LSA)* (Landauer et al., 1998). Metode ini mengektrak arti dari kata-kata dan kesamaan dari kalimat-kalimat memamfaatkan informasi penggunaan kata-kata tersebut dalam konteks disamping itu metode ini juga merepresentasikan arti dari kalimat-kalimat secara simultan. Suatu kalimat direpresentasikan sebagai kombinasi liner dari fitur semantik. Akan tetapi fitur semantik yang diporoleh dari LSA didekomposisi dari sejumlah besar bobot *term* positif dan negatif. Akibatnya arti dari fitur semantik tidak dapat ditangkap secara intuitif dan cakupan artinya menjadi tidak jelas. Demikian juga bobot dari fitur semantik suatu kalimat dapat mempunyai nilai positif dan negatif. Oleh karenya, suatu kalimat yang direpresentasikan sebagai kombinasi liner dari banyak fitur sematik yang tidak penting. Akibatnya metode peringkasan dengan metode ini gagal mengekstrak kalimat-kalimat yang mempunyai arti. (Lee & Seung, 1999; Zha,2002).

Sejalan dengan metode LSA yang memamfaatkan kemampuan metode aljabar adalah (*Non-Negative Matrix Factorization*) NMF. Metode ini selain termasuk metode unsupervisi sehingga tidak memerlukan proses pembelajaran juga mempunyai keuntungan fitur semantik yang diekstrak dapat diinterpretasi lebih intuitif dibandingkan dengan hasil yang diekstraksi menggunakan metode LSA. Hal ini terjadi karena metode ini hanya memamfaatkan nilai tidak negatif. Selanjutnya suatu kalimat dapat direpresentasikan sebagai kombinasi liner dari fitur semantik yang

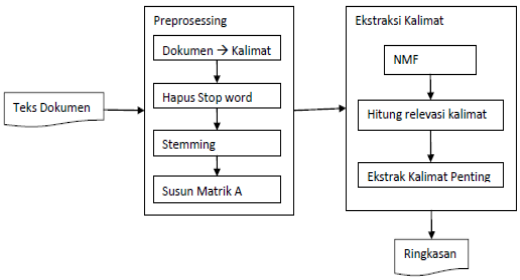
lebih intuitif sehingga cakupan arti dari fitur semantik menjadi sempit. Dengan demikian kemungkinan untuk mengekstraksi kalimat-kalimat penting lebih besar.

2.2. Metode NMF

Metode ini merepresentasikan objek-objek individu sebagai kombinasi liner tidak negative dari bagian informasi yang diekstrak dari sejumlah objek yang berukuran besar. Cara kerja metode ini adalah memecah dokumen teks ke dalam kalimat-kalimat dan menghitung frekuensi masing-masing term dalam kalimat yang direpresentasikan dengan matrik tidak negative A berukuran mxn, m jumlah term dan n jumlah kalimat dalam dokumen. Matrik A didekomposisi ke dalam suatu perkalian matrik fitur semantik berukuran mxr W dan matrik variabel semantik tidak negatif berukuran rxn H. Nilai r dipilih lebih kecil dari m atau n sehingga total ukuran W dan H lebih kecil dari matrik A.

3. Perancangan Sistem

Perancangan sistem peringkasan secara umum mengacu pada (Ju-Hong, dkk. 2009) terdiri dua tahap yakni preprosessing dan ekstraksi kalimat sebagaimana digambarkan pada gambar 1.



Gambar 1. Rancangan program peringkasan dengan NMF(Ju-Hong, dkk. 2009)

Pada tahap pra proses suatu dokumen teks dipecah ke dalam kalimat-kalimat tunggal dan semua *stopword* dibuang berdasarkan daftar kata stop hasil. Dalam hal ini daftar kata stop digunakan acuan hasil penelitian Tala tahun 2003. Langkah selanjutnya dilakukan proses steaming untuk mendapatkan akar kata masing-maing term menggunakan algoritma Porter yang telah diadaptasi ke dalam bahasa Indosenia (Tala, 2003). Selanjutnya bobot masing-masing term akan disimpan dalam matrik A menggunakan persamaan 1.

$$A_{ji} = Wgt(j,i) \quad (1)$$

Wgt(j,i) adalah fungsi pembobotan dari term j pada kalimat i. Dalam hal ini terdapat beberapa fungsi pembobotan yang akan digunakan dalam penelitian ini sebagaimana pada persamma 2 sampai dengan 6

$$A(i,j) = t(i,j) \quad (2)$$

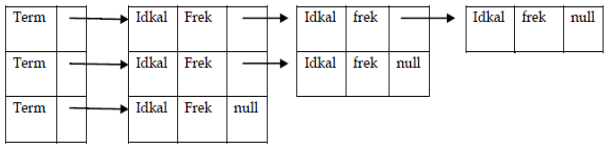
$$A(i,j) = \log(i + t(i,j)) \quad (3)$$

$$A(i,j) = \log(i + t(i,j))*\log(N/n(i)) \quad (4)$$

$$A(i,j) = 0.5 + 0.5*(t(i,j) / \max(t(i,j))) \quad (5)$$

$$A(i,j) = (0.5+0.5*(t(i,j)/ (\max(t(i,j)))*\log(N/n(i)) \quad (6)$$

Untuk memudahkan pengaksesan kembali terutama pada saat menghitung matrik A dan penentukan kalimat ringkasan, maka pelu disusun struktur data sebagaimana gambar 2. Setiap term unik akan dicatat berapa jumlah kemunculannya pada setiap kalimat. Struktur data ini akan dipanggil pada saat pembacaan suatu dokumen. Matrik A selanjutnya dapat disusun berdasarkan isi dari struktur data di atas



Gambar.2. Rancangan Struktur Data

Proses peringkasan dimulai dengan mengolah matrik A menggunakan algoritma MNF untuk mendapatkan matrik fitur semantik non-negatif W dan matrik variable semantik non negative H masing-masing menggunakan persamaan 7 dan 8. Selanjutnya berdasarkan hasil perhitungan algoritma MNF dihitung relevasi kalimatnya menggunakan persamaan 9. Nilai bobot relevansi kalimat ini akan digunakan untuk memilih kalimat yang akan dijadikan ringkasan.

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} \frac{\left(W^T A\right)_{\alpha\mu}}{\left(W^T WH\right)_{\alpha\mu}} \quad (7)$$

$$W_{i\alpha} \leftarrow W_{i\alpha} \frac{\left(AH^T\right)_{i\alpha}}{\left(WHH^T\right)_{i\alpha}} \quad (8)$$

$$A_{*J} = \sum_{l=1}^{r} H_{ij} W_{*l} \quad (9)$$

Generic Relevance kalimat ke j =

$$\sum_{i=1}^{r} \left(H_{ij} \cdot weight(H_{i*})\right) \quad (10)$$

$$weight(H_{i*}) = \frac{\sum_{q=1}^{n} H_{iq}}{\sum_{p=1}^{r}\sum_{q=1}^{n} H_{pq}} \quad (11)$$

3. 1. Algoritma NMF

Algoritma NMF dimulai dengan menginisialisai secara acak matrik W_{mxk} dan H_{kxn}, k

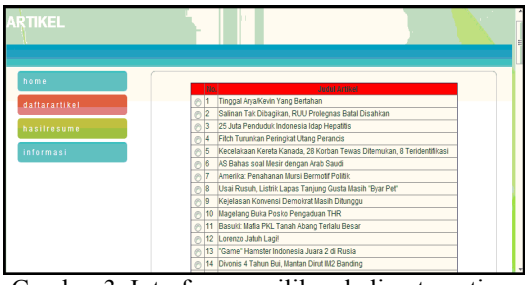
adalah jumlah kalimat yang akan diekstrak. Proses perhitungan dilakukan untuk menentukan matrik H dan W sampai jumlah maksimum ulangan yang diberikan. Terdapat beberapa algorimat untuk menentukan NMF, akan tetapi yang digunakan dalam penelitian ini adalah algoritma yang dikembangkan oleh Lee dan Seung (2000). Gambaran secara teknik algoritma ini adalah sebagai berikut :

Inisialisasi matrik W_{mxk} berukuran mxk secara acak
Inisialisai matrik H_{kxn} berukuran kxn secara acak
for i ← 1 to maksimum iterasi do
$H = H .* (W^T A)./(W^T WH + 10^{-9});$
$W=W. * (AH^T)./(WHH^T + 10^{-9});$
End for

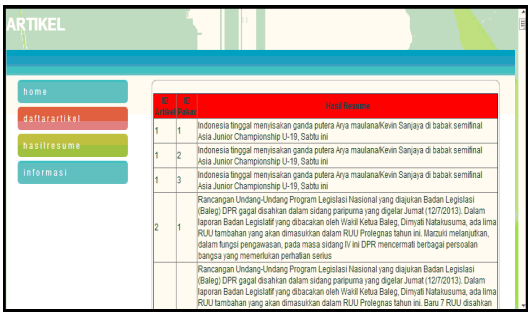
3.2. Data

Data yang digunakan untuk peringkasan dikumpulan 100 artikel yang di-*download* dari internet dengan alamat www.kompas.com pada tanggal 8 Juli sampai tanggal 15 Juli 2013. Selanjutnya masing-masing artikel dibuat ringkasan secara manual oleh 3 orang pakar bahasa Indonesia sebagai acuan.

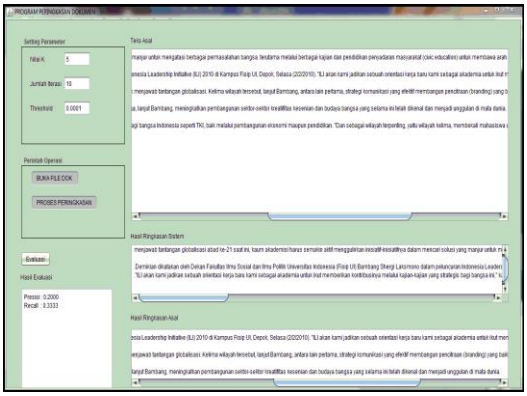
Proses penugujian dari para ahli dilakukan secara interaktif dengan memilih kalimat yang diaggap penting mewakili ringkasan dengan interface sebagaimana gambar 3 dan gambar 4. Hasil entri data tersebut akan disimpan dalam database dengan rancangan antar tabel sebagaimana gambar 5 dan dibandingkan dengan hasil dari sistem sebagaimana gambar 6.



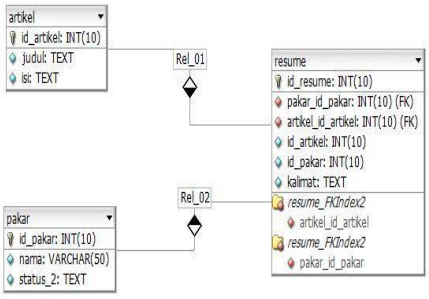
Gambar 3. Interface pemilihan kalimat penting



Gambar 4. Tampilan hasil ringkasan pakar



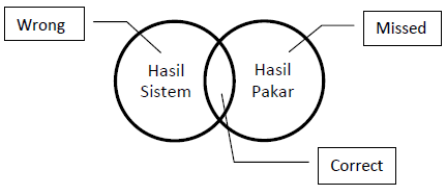
Gambar 5. Interface sistem peringkasan



Gambar 5. Relasi antar table

3.3. Metrik Evaluasi

Evaluasi yang digunakan pada pernelitian ini bersifat intrinsik, yaitu pengevaluasi dengan cara membuat ringkasan yang ideal kemudian hasilnya dibandingkan dengan ringkasan sistem. Evaluasi ini mengacu pada (Hovy, 2003) menggunakan persamaan 12 dan 13. Pengertian *correct* pada persamaan tersebut adalah jumlah kalimat yang berhasil di ekstrak sistem sesuai dengan kalimat yang diekstrak manusia. Pengertian *wrong* adalah jumlah kalimat yang diekstrak sistem tetapi tidak terdapat dalam kalimat yang diekstrak manusia dan *missed* adalah jumlah kalimat yang diekstrak manusia tetapi tidak diekstrak oleh sistem. Hubungan ketiga variabel ini dapat diilustrasikan pada gambar 6.



Gambar 6. Hubungan antara wrong, missed dan correct

$$precision = \frac{correct}{correct + wrong} \qquad (12)$$

$$recall = \frac{correct}{correct + missed} \qquad (13)$$

4. Hasil dan pembahasan

Berdasarkan hasil pengumpulan data hasil ringkasan para pakar diperoleh fakta bahwa ringkasan antar ketiga pakar hasilnya sangat bervariatif dengan rata-rata presisi dan recall diantara masing-masing adalah 0.68667 dan 0.70642. Selanjutnya hasil ringkasan dari masing-masing pakar digunakan untuk bahan evaluasi dari sistem.

Proses evaluasi sistem dilakukan dalam beberapa langkah. Langkah pertama adalah menentukan ringkasan dari seluruh teks artikel menggunakan lima scenario pembobotan sebagaimana disebutkan pada bab 3 dengan berbagai variari ulangan mulai dari 1 sampai dengan 100. Hal ini dilakukan untuk mengetahui pada ulangan keberapa perhitungan NMF mulai konvergen. Pada setiap ulangan sudah ditetapkan jumlah r adalah 5 dan untuk nilai awal W dan H ditetapkan 0.5. Langkah selanjutnya adalah mengevaluasi masing-masing hasil ringkasan dari sistem dengan hasil ringkasan para pakar menggunakan evaluator. Langkah terakhir menghitung rata-rata presisi dan recall untuk setiap perulangan.

Rata-rata presisi pada setiap pembobotan dapat dilihat pada tabel 2. Tanda w1 sampai dengan w5 masing-masing merepresentasikan fungsi pembobotan sebagaimana persamaan 2 sampai dengan 6. Dengan nilai parameter yang telah ditetapkan di atas evaluasi dilakukan mulai dari perulangan 2 sampai dengan 100. Namun demikian ternyata semua ulangan telah konvergen rata-rata pada ulangan 15. Nilai presisi diperoleh dengan menghitung nilai rata-rata presisi dari 100 buah artikel yang berasal dari 3 pakar. Selanjutnya nilai rata-rata dari ketiga pakar itulah yang akan dijadikan nilai rata-rata umum dan hasilnya ditampilkan sebagaimana pada tabel 1.

Tabel 1. Rata-rata nilai presisi untuk masing-masing pembobotan

	w1	w2	w3	w4	w5
2	0.19724	0.22111	0.18924	0.19324	0.18924
3	0.17952	0.19044	0.19324	0.19724	0.18924
4	0.16886	0.19044	0.18924	0.19724	0.18924
5	0.16486	0.19044	0.18524	0.19724	0.18924
6	0.16486	0.19044	0.18124	0.19724	0.18924
7	0.16486	0.19044	0.17724	0.19724	0.18924
8	0.16486	0.19044	0.18124	0.19724	0.18924
9	0.16086	0.19044	0.18124	0.19724	0.18924
10	0.16086	0.19044	0.17857	0.19724	0.18924
11	0.16086	0.19044	0.17857	0.19724	0.18924
12	0.16086	0.19044	0.17857	0.19724	0.18924
13	0.16086	0.19044	0.17857	0.19724	0.18924
14	0.16086	0.19044	0.17857	0.19724	0.18924
15	0.16086	0.19044	0.17857	0.19724	0.18924

Berdasarkan pada tabel 1 dapat diketahui bahwa yang mempunyai presisi paling baik adalah pembobotan persamaan 5. Namun demikian secara umum presisi hasil dari sistem masih di bawah 0.2, artinya jika hasil ringkasannya sebanyak 5 kalimat maka yang benar sesuai dengan ringkasan hasil pakar adalah 1 sedangkan yang 4 tidak sesuai.

Hasil perbandingan recall untuk masing-masing pembobotan adalah sebagaimana pada tabel 2. Proses perhitungan untuk mendapatkan recall sama dengan proses untuk mengitung presisi.

Tabel 2. Rata-rata nilai recall untuk masing-masing pembobotan

	w1	w2	w3	w4	w5
2	0.34085	0.23661	0.32585	0.33329	0.32585
3	0.31774	0.211	0.33335	0.34085	0.32585
4	0.30939	0.21239	0.32529	0.34085	0.32585
5	0.30183	0.21122	0.31779	0.34085	0.32585
6	0.30183	0.21122	0.30896	0.34085	0.32585
7	0.30183	0.21122	0.3014	0.34085	0.32585
8	0.30183	0.21122	0.30946	0.34085	0.32585
9	0.29317	0.21122	0.30946	0.34085	0.32585
10	0.29317	0.21122	0.30585	0.34085	0.32585
11	0.29317	0.21122	0.30585	0.34085	0.32585
12	0.29317	0.21122	0.30585	0.34085	0.32585
13	0.29317	0.21122	0.30585	0.34085	0.32585
14	0.29317	0.21122	0.30585	0.34085	0.32585
15	0.29317	0.21122	0.30585	0.34085	0.32585

Dari tabel 2 di atas dapat diketahui bahwa nilai presisi yang paling baik terjadi pada persmaan 5 dengan rata-rata secara keseluruhan adalah 0.34. Berdasarkan rumus recall berarti jika 1 kalimat hasil ringkasan dari sistem benar maka 2 kalimat hasil ringkasan pakar yang tidak cocok dengan ringkasana sistem. Jika ringkasan hasil pakar sebanyak 5 kalimat maka 2 diantaranya adalah benar sesuai dengan ringkasan sistem.

Secara umum peringkasan dokumen secara otomatis menggunakan NMF ini hasilnya kurang memuaskan. Hal ini ditandai dengan rendahnya nilai rata-rata presisi dan recall. Penyebab terjadinya hal ini adalah sumber data acuan sebagai *benchmark* yaitu hasil ringkasan pakar mempunyai nilai rata-rata presisi 0.68667 dan recall 0.70642. Rendahnya nilai presisi dan recall ini menandakan bahwa antara para pakar sendiripun masih terdapat bias antara satu dengan yang lain dalam membuat suatu ringkasan. Dengan nilai presisi dan recall tersebut dapat diartikan jika 2 orang pakar membuat ringkasan suatu teks dengan masing-masin6 6 kalimat dan 7 kalimat, maka 4 kalimat saja yang benar sedangkan 2 dan 3 kalimat tidak sama. Dengan data acuan seperti ini tentu hasil sistem akan lebih banyak bias yang ditimbulkan.

5. Kesimpulan dan Saran

Telah berhasil dikembangkan sistem peringkasan otomatis menggunakan NMF dengan rata-rata presisi dan recall dibawah 0,2. 2. Fungsi pembobotan yang relatif lebih baik adalah $A(i,j) = 0.5 + 0.5*(t(i,j) / \max(t(i,j))$. Rata-rata presisi dan recall hasil ringkasan antara pakar berkisar 0,6 dan

0.7. Dengan rendahnya presisi dan recal dari masing-masing pakar ini menyebabkan bias pada hasil sistem.

Pada penelitian ini belum dilakukan uji coba terhadap sensitifitas steaming terhadap sistem secara keseluruhan. Untuk itu pada penelitian berikutnya perlu diuji sejauh mana pengaruh steming terhadap sistem peringkasan menggunakan metode ini.

Sebagaimana dijelaskan pada bagain pembahasan, pada penelitian ini nilai H dan W sudah ditetapkan pada awal proses. Untuk itu masih perlu diuji sejauh mana pengaruh inisialisai dengan nilai acak pada marik H dan W terhadap kinerja sistem.

6. Daftar Pustaka

ACHMAD RIDOK ,TRI CAHYO ROMADHONA, 2013, Peringkas Dokument Otomaris Menggunakan Metode Fuzzy Model Sistem Inferensi Mamdani, Dalam Proceedings Seminar Nasional Teknologi Informasi dan Multimedia . - Yogyakarta : STIMIK AMIKOM, Vols. 1 07-19.

AMINI M. R., & GALLINARI, P., 2002, The use of unlabeled data to improve supervised learning for text summarization, In Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'02) . - Tampere, Finland. : [s.n.], Vols. (pp. 105–112).

BARZILAY R. and ELHADAD, M, 1997, Using Lexical Chains for Text Summarization. In Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization, pages 10-17.

ERCAN G. and CICEKLI I, 2008, Lexical Cohesion based Topic Modeling for Summarization, InProceedings of 9th Int. Conf. Intelligent Text Processing and Computational Linguistics (CICLing-2008), pages 582-592.

ERKAN G. and D.R RADEV, 2004, Lexrank : Graph-based centrality as salience in text summarization. JAIR

FIRMIN T. and M.J. CHRZANOWSKI, 1999, An Evaluation of Automatic Text Summarization Systems, The MIT Press : Cambridge.

GONG Y., & LIU, X. 2001, Generic text summarization using relevance measure and latent semantic analysis, In Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'01). - New Orleans, USA. Vols. (pp. 19–25).

HOVY E, 2003, Text Summarization, In Book The Oxford Handbook of Computational Linguistic, auth. Mitkov R. Oxford: Oxford University Press.

HOVY E. and LIN, C-Y, 1999, Automated Text Summarization in SUMMARIST, In book Advances in Automatic Text Summarization. Maybury I. Mani and M.T. : The MIT Press, pages 81-94.

JU-HONG LEE SUN PARK, CHAN-MIN AHN , DAEHO KIM, 2009, Automatic generic document summarization based on non-negative, In Information Processing and Management 45. Elsevier Ltd, 20–34.

KAREL JEZEK and JOSEF STEINBERGER, 2008, Automatic Text Summarization (the state of the art 2007 and new challenges), Znalosti . - 2008, pp. 1-12.

LIN EDUARD HOVY and CHIN YEW, 1999, Automated text summarization in SUMMARIST, MIT Press, 1999, pages 81–94.

LUHN H.P, 1958, The Automatic Creation of Literature Abstracts, IBM Journal of Research Development.

MANI I. and M.T. MAYBURY, 1999, Advance in Automatic Text Summarization. Cambridge : The MIT, Press.

MIHALCEA R. and TARAU, P, 2004, Text-rank – bringing order into texts, In Proceeding of the Conference on Empirical Methods in Natural Language Processing.

QAZVINIAN V. and RADEV, D.R, 2008, Scientific paper summarization using citation summary networks.

TALA, FADILLAH Z. 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. Master of Logic Project. Institute for Logic, Language and Computation. Universiteit van Amsterdam. The Netherlands.

ZHA H, 2002, Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering, In Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'02), Tampere, Finland. : (pp. 113–120).