



UNIVERSITAS INDONESIA

**EKSTRAKSI TOPIK UTAMA HARIAN PORTAL BERITA INDONESIA
ONLINE MENGGUNAKAN *NONNEGATIVE MATRIX FACTORIZATION***

SKRIPSI

**HANIF FATRIAL
0706261700**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM SARJANA MATEMATIKA
DEPOK
JUNI 2012**



UNIVERSITAS INDONESIA

**EKSTRAKSI TOPIK UTAMA HARIAN PORTAL BERITA INDONESIA
ONLINE MENGGUNAKAN *NONNEGATIVE MATRIX FACTORIZATION***

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana sains

**HANIF FATRIAL
0706261700**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM SARJANA MATEMATIKA
DEPOK
JUNI 2012**

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Hanif Fatrial
NPM : 0706261700
Tanda Tangan : 
Tanggal : 15 Juni 2012





HALAMAN PENGESAHAN

Skripsi ini diajukan oleh

Nama : Hanif Fatrial
NPM : 0706261700
Program Studi : Sarjana Matematika
Judul Skripsi : Ekstraksi Topik Utama Harian Portal Berita Online
Menggunakan Nonnegative Matrix Factorization.

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi S1 Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia

DEWAN PENGUJI

Pembimbing : Dr. rer. nat. Hendri Murfi, M.Kom ()
Penguji : Dr. Alhaji Akbar B., M.Sc ()
Penguji : Dra. Yahma Wisnani, M.Kom ()
Penguji : Dr. Yudi Satria, M.T ()

Ditetapkan di : Depok

Tanggal : 15 Juni 2012

KATA PENGANTAR

Alhamdulillah, Puji syukur penulis panjatkan kepada Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Sains Jurusan Matematika pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada

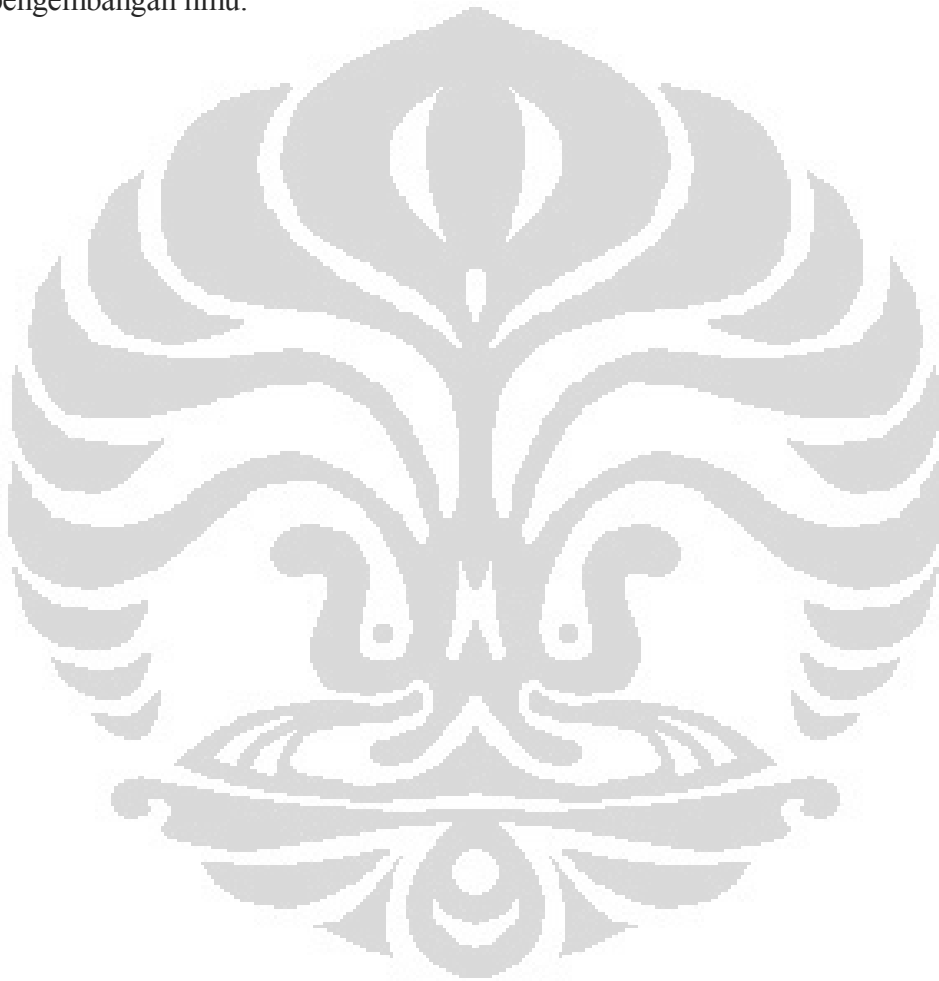
- (1) Bapak Dr. rer.nat. Hendri Murfi, M.Kom selaku pembimbing skripsi yang dengan sabar menuntun dan memberikan ilmu dalam penulisan skripsi ini.
- (2) Ibu Dra. Yahma Wisnani selaku pembimbing akademik yang selalu memberikan arahan dalam menjalani masa perkuliahan penulis.
- (3) Ayah dan Ibu yang terus mendukung penulis dengan do'a dan nasihat-nasihat beliau.
- (4) Irfayanto dan Sri Fatmayenti yang telah memberikan motivasi yang lebih kepada penulis.
- (5) Ibu Dr. Kiki Ariyanti Sugeng, Bpk Dr. Alhaji Akbar B. M.Sc, Bpk Prof Djati Kerami, Bpk Dr. Yudi Satria, M.T., Bpk Drs.Suryadi MT, M.T., yang telah hadir pada SIG 1 dan SIG 2, terima kasih telah memberikan saran yang membangun.
- (6) Seluruh dosen Departemen Matematika UI yang telah memberikan penulis ilmu yang bermanfaat untuk masa depan penulis.
- (7) Adit, Andi, Dhanar, Manda, Hikma, Shafira, Nedi, Widya yang telah memberikan kenangan-kenangan yang tak terlupakan.
- (8) Seluruh teman-teman angkatan 2007 yang telah menjalani masa perkuliahan bersama.
- (9) Angkatan 2005, 2006, 2008, 2009, 2010, 2011 yang menjadi teman didalam maupun diluar kelas perkuliahan.
- (10) Eka Mustikawati, S.Hum atas pembuatan topik utama secara manual

- (11) Anak-anak pondokan anugerah yang menjadi teman sepermainan diluar jam perkuliahan .

Penulis juga ingin mengucapkan terima kasih kepada seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu dalam penyusunan skripsi ini. Akhir kata, penulis mohon maaf jika terdapat kesalahan atau kekurangan dalam skripsi ini. Penulis berharap semoga skripsi ini bermanfaat bagi pengembangan ilmu.

Penulis

2012



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Hanif Fatrial
NPM : 0706261700
Program Studi : S1 Matematika
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

Ekstraksi Topik Utama Harian Portal Berita *Online* Menggunakan *Nonnegative Matrix Factorization*.

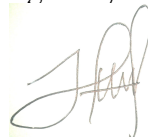
beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 15 Juni 2012

Yang menyatakan



(Hanif Fatrial)

ABSTRAK

Nama : Hanif Fatrial
Program Studi : Matematika
Judul : Ekstraksi Topik Utama Harian Portal Berita Indonesia *Online*
Menggunakan *Nonegative Matrix Factorization*.

Penelitian tugas akhir sarjana di Departemen Matematika Universitas Indonesia dapat berupa aplikasi matematika dalam kehidupan sehari-hari. Salah satu aplikasinya dapat diterapkan pada pencarian topik pada portal berita Indonesia *online* secara otomatis dengan bantuan mesin. Dengan transformasi kata-kata pada artikel portal berita Indonesia *online* kedalam bentuk matriks, dapat dilakukan proses *Latent Semantic Analysis* (LSA) dengan menggunakan metode *Nonegative Matrix Factorization* dalam mengekstraksi kata-kata pendukung topik dari sekumpulan dokumen. Pada skripsi ini akan dibahas mengenai implementasi *Latent Semantic Analysis* dengan menggunakan *Nonegative Matrix Factorization* (NMF) dalam ekstraksi kata-kata pendukung topik sehingga kata-kata tersebut dapat menginterpretasikan topik utama harian dari portal berita Indonesia *online*.

Kata Kunci : *latent semantic analysis, nonegative matrix factorization*, topik utama, portal berita indonesia, ekstraksi topik.

xiii+31 halaman : 8 tabel, 12 gambar
Daftar Pustaka : 8 (1997-2008)

ABSTRACT

Name : Hanif Fatrial
Program Study : Mathematics
Title : Daily Topic Extraction Indonesian Online News Portal Using
Non-negative Matrix Factorization.

There are a wide range of fields in mathematics that can be used as a final research in the Department of Mathematics, University of Indonesia, including the applications of mathematics for daily life. One of the applications can be applied to searching topic in Indonesia online news portal automatically with machines. Latent Semantic Analysis (LSA) using Matrix Factorization Nonnegative method can extract the words from a collection of documents which supporting the topic. This *skripsi* will be discussed on the implementation of Latent Semantic Analysis using Nonnegative Matrix Factorization in extraction for the words wich support topics that words can be interpretation of a topic Indonesian daily online news portal.

Keywords : latent semantic analysis, nonegative matrix factorization, news topic , Indonesian news portal, topic extraction.

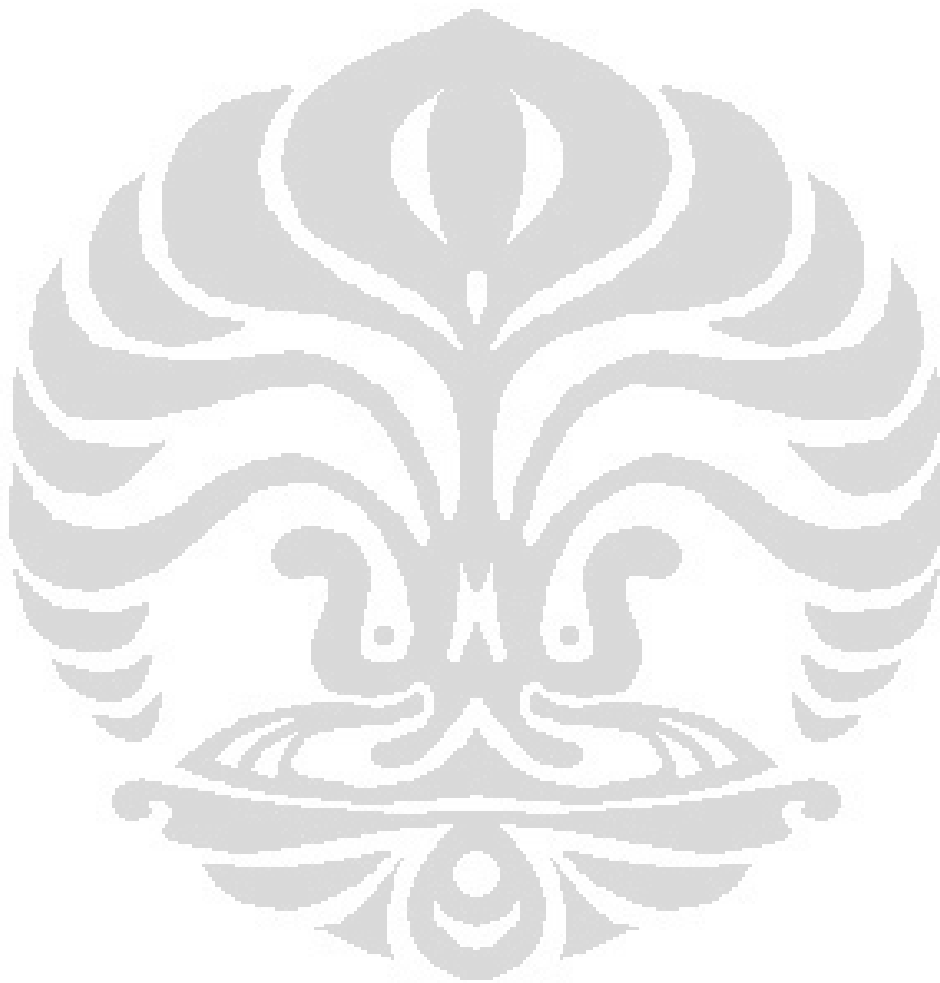
xiii+31 pages : 8 tables, 12 pictures

Bibliography : 8 (1997-2008)

DAFTAR ISI

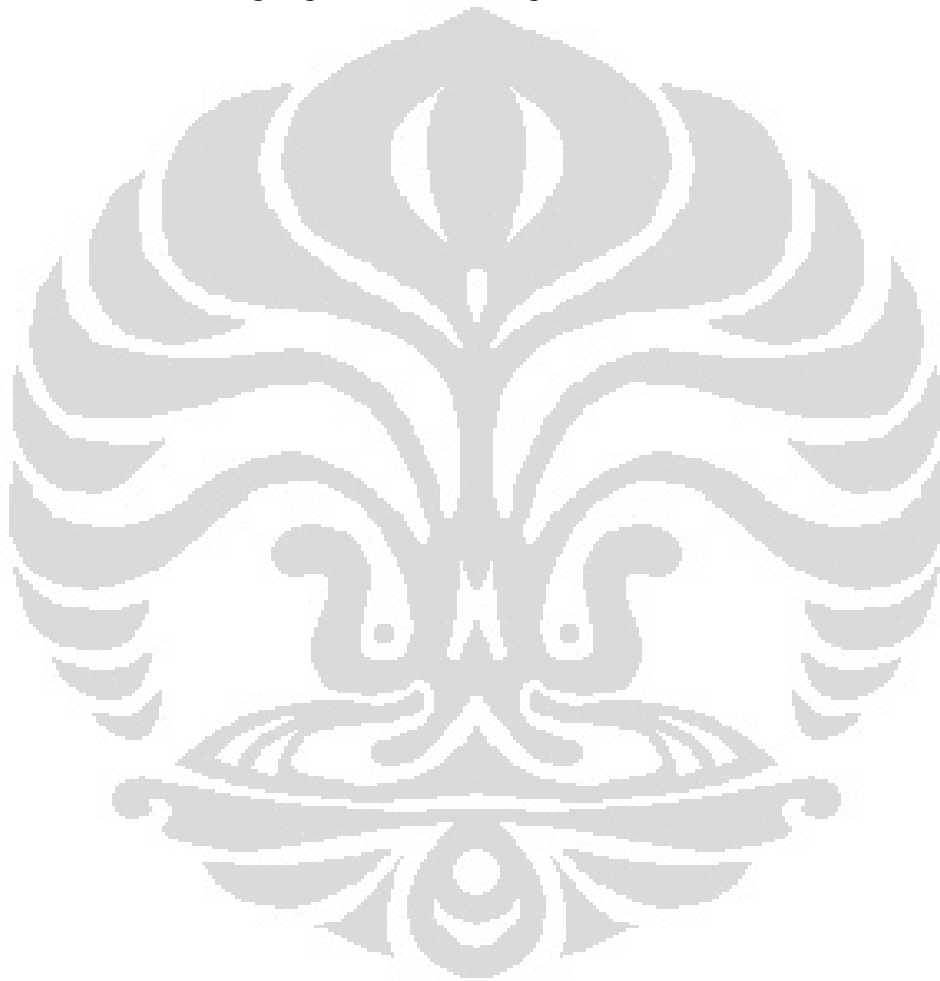
HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Metodologi Penelitian	3
1.4.1 Perumusan masalah dan studi literatur	3
1.4.2 Pengumpulan data	3
1.4.3 Implementasi algoritma ekstraksi topik	3
1.4.4 Simulasi	4
BAB 2 DASAR TEORI	5
2.1 Pembelajaran Mesin	5
2.1.1 <i>Supervised Learning</i>	5
2.1.2 <i>Unsupervised Learning</i>	5
2.2 Faktorisasi Matriks	6
2.2.1 Metode Langsung	6
2.2.2 Metode Aproksimasi	6
2.3 Norm	7
BAB 3 <i>NONNEGATIVE MATRIX FACTORIZATION</i>	8
3.1 <i>Latent Semantic Analysis</i>	8
3.2 Formulasi <i>Nonnegative Matrix Factorization</i>	8
3.2.1 <i>Multiplicative Update Rule (MUR)</i>	10
3.2.2 <i>Multiplicative Update Algorithm (MUA)</i>	13
BAB 4 SIMULASI	15
4.1 Perangkat Lunak dan Perangkat Keras	15
4.2 Preparasi Data	16
4.3 Ekstraksi Kamus Kata	17
4.3.1 Pemisahan Kata	18
4.3.2 Penyaringan Kamus Kata	19
4.4 Pembentukan Matriks Kata dokumen	20
4.5 Ekstraksi Topik Utama	21
4.6 Ekstraksi Topik dengan <i>Nonnegative Matrix Factorization</i>	22
4.7 Hasil Simulasi	22
4.7.1 Interpretasi Topik	23
4.7.2 Tren dari Kata-Kata Pendukung Topik	25
4.8 Perbandingan Hasil dengan Pencarian Manual	28
BAB 5 KESIMPULAN DAN SARAN	31

5.1 Kesimpulan.....	31
5.2 Saran.....	31
DAFTAR PUSTAKA.....	33
LAMPIRAN.....	34



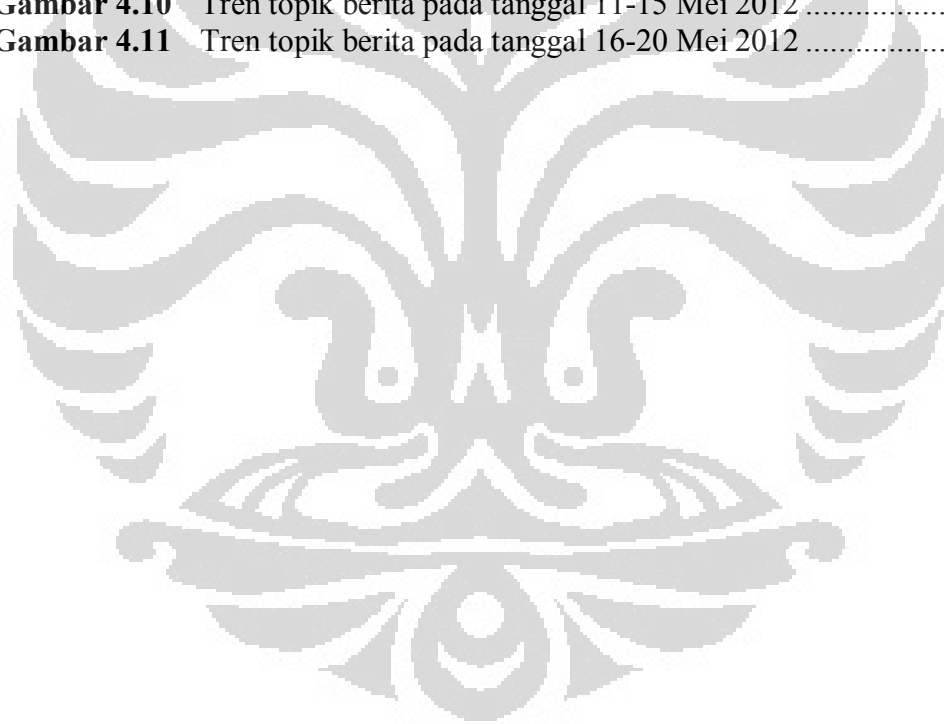
DAFTAR TABEL

Tabel 4.1 Hasil Simulasi untuk data pada tanggal 1 Mei 2012	23
Tabel 4.2 Hasil Simulasi untuk data pada tanggal 2 Mei 2012	24
Tabel 4.3 Hasil Simulasi untuk data pada tanggal 3 Mei 2012	24
Tabel 4.4 Hasil Simulasi untuk data pada tanggal 4 Mei 2012	25
Tabel 4.5 Perbandingan pada 1 Mei 2012 pukul 14.00 WIB	28
Tabel 4.6 Perbandingan pada 2 Mei 2012 pukul 14.00 WIB	29
Tabel 4.7 Perbandingan pada 3 Mei 2012 pukul 14.00 WIB	29
Tabel 4.8 Perbandingan pada 4 Mei 2012 pukul 14.00 WIB	30



DAFTAR GAMBAR

Gambar 1.1	Pertumbuhan pengguna <i>internet</i> di Indonesia	1
Gambar 3.1	Simulasi uji konvergensi MUR.....	14
Gambar 4.1	Ilustrasi Proses Simulasi.....	15
Gambar 4.2	Rata-rata kapasitas dari feed RSS dari portal berita Indonesia <i>online</i>	17
Gambar 4.3	Grafik jumlah artikel yang akan diproses per hari selama satu bulan.....	17
Gambar 4.4	Contoh tampilan sebuah artikel dalam <i>file</i> RSS	18
Gambar 4.5	Grafik jumlah kata berbeda yang dilibatkan per hari.....	19
Gambar 4.6	Proses pembentukan matriks kata dokumen V	20
Gambar 4.7	Proses interpretasi topik dari matriks bobot W.....	21
Gambar 4.8	Tren topik berita pada tanggal 1-5 Mei 2012	25
Gambar 4.9	Tren topik berita pada tanggal 6-10 Mei 2012	26
Gambar 4.10	Tren topik berita pada tanggal 11-15 Mei 2012	27
Gambar 4.11	Tren topik berita pada tanggal 16-20 Mei 2012	27

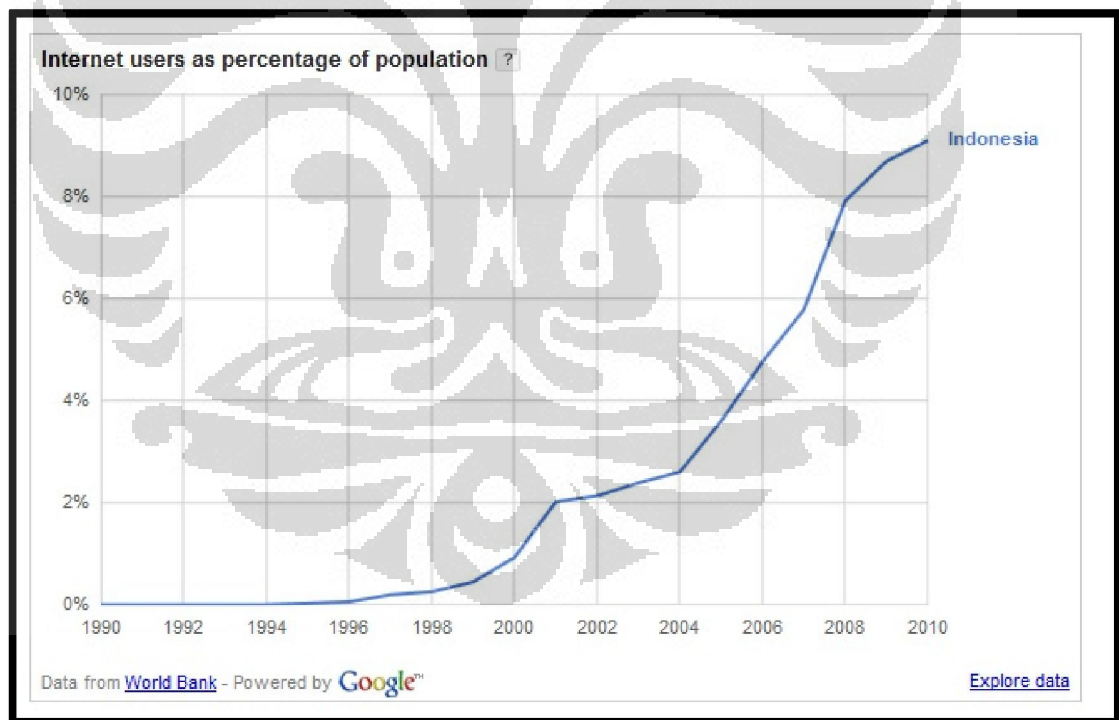


BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada saat sekarang *internet* mempunyai pengaruh yang besar terhadap arus informasi di dunia. Sehingga dengan perkembangan teknologi informasi yang semakin cepat telah menyebabkan penggunaan *internet* naik secara drastis. Cakupan jarak dan batas negara tidak lagi menjadi kendala bagi pengguna untuk mendapatkan informasi, sehingga *internet* menjadi salah satu kebutuhan dan gaya hidup baru yang tak terpisahkan dalam kehidupan manusia. Data dari bank dunia menyebutkan persentase populasi pengguna *internet* di Indonesia mencapai 9,1 % pada tahun 2010. Dikutip dari *teknoup.com*¹ pada tahun 2012 terdapat sekitar 30 juta penduduk Indonesia adalah pengguna *internet* yang merupakan pengguna terbesar se-Asia Tenggara.



[sumber : <http://www.google.com/publicdata/explore> , waktu akses : 13.00 WIB ,20 Maret 2012]

Gambar 1.1 Pertumbuhan pengguna *internet* di Indonesia

¹ <http://www.teknoup.com/news/15025/peningkatan-penggunaan-internet-jadikan-indonesia-sebagai-pasar-potensial/> , waktu akses : 13.00, 20 Maret 2012

Dalam hal penggunaan *internet* untuk mengakses berita, hasil penelitian sebuah lembaga *survey* yang pernah dikutip *detik.com*² menyebutkan adanya kecenderungan semakin banyak orang yang mencari berita melalui portal *online*. Hal ini disebabkan portal berita *online* ini dapat diakses kapan saja dan dimana saja. Meningkatnya pembaca portal berita *online* juga memicu munculnya portal-portal penyedia berita *online* yang baru dengan ciri khas masing-masing. Portal berita ini berlomba-lomba untuk menyediakan berita yang saat ini sedang terjadi sehingga memungkinkan arus berita di portal ini sangat cepat dengan jumlah artikel berita yang banyak. Akan tetapi, hal ini menjadi kendala bagi mereka yang hanya menginginkan topik utama dari berita pada saat itu, baik itu karena keterbatasan kesempatan atau waktu untuk mengakses berita dari berbagai portal maupun berita yang menjadi topik utama telah dianggap kadaluarsa karena keterbatasan portal tersebut dalam menampilkan berita.

Latent Semantic Analysis (LSA) adalah sebuah teori dan metode untuk ekstraksi dan representasi topik dari sekumpulan dokumen (Landauer dan Dumais, 1997). LSA menggunakan beberapa metode untuk proses faktorisasi matriks diantaranya *Singular Value Decomposition* dan *Non-negative Matrix Factorization* (Lee dan Seung, 1999). Salah satu matriks yang dihasilkan oleh metode SVD adalah matriks yang merepresentasikan topik dalam suatu kalimat. Akan tetapi matriks ini cenderung bersifat padat dan berisi bilangan negatif dan non-negatif pada elemen-elemennya, sehingga representasi bilangan-bilangan negatif terhadap topik sulit untuk diinterpretasi. Sedangkan matriks representasi yang dihasilkan oleh NMF memiliki entri non-negatif sehingga mudah untuk interpretasi matriks.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah di atas, masalah yang dibahas adalah bagaimana mengekstraksi topik utama harian dari portal berita *online* berbahasa Indonesia dengan menggunakan metode *Nonnegative matrix factorization* (NMF).

² <http://www.detik.com/>, waktu akses : 13.00 WIB, 20 Maret 2012

1.3 Tujuan Penelitian

Sesuai dengan permasalahan yang diangkat, maka tujuan pada penulisan ini adalah menerapkan metode *Nonnegative matrix factorization* untuk melakukan ekstraksi topik dari portal berita *online* berbahasa Indonesia dimana topik yang telah diekstrak dapat digunakan sebagai bahan analisa lanjutan.

1.4 Metodologi Penelitian

Penelitian ini dilakukan dengan langkah-langkah sebagai berikut:

1.4.1 Perumusan masalah dan studi literatur

Pada tahap ini akan dilakukan analisa terhadap masalah, pengumpulan bahan-bahan dan referensi untuk dijadikan bahan acuan dalam melakukan studi awal pemahaman konsep dan perumusan model sistem yang akan dibuat. Literatur-literatur ini diperoleh melalui penelusuran jurnal, makalah, buku, dan informasi lain yang terkait dengan penelitian ini.

1.4.2 Pengumpulan data

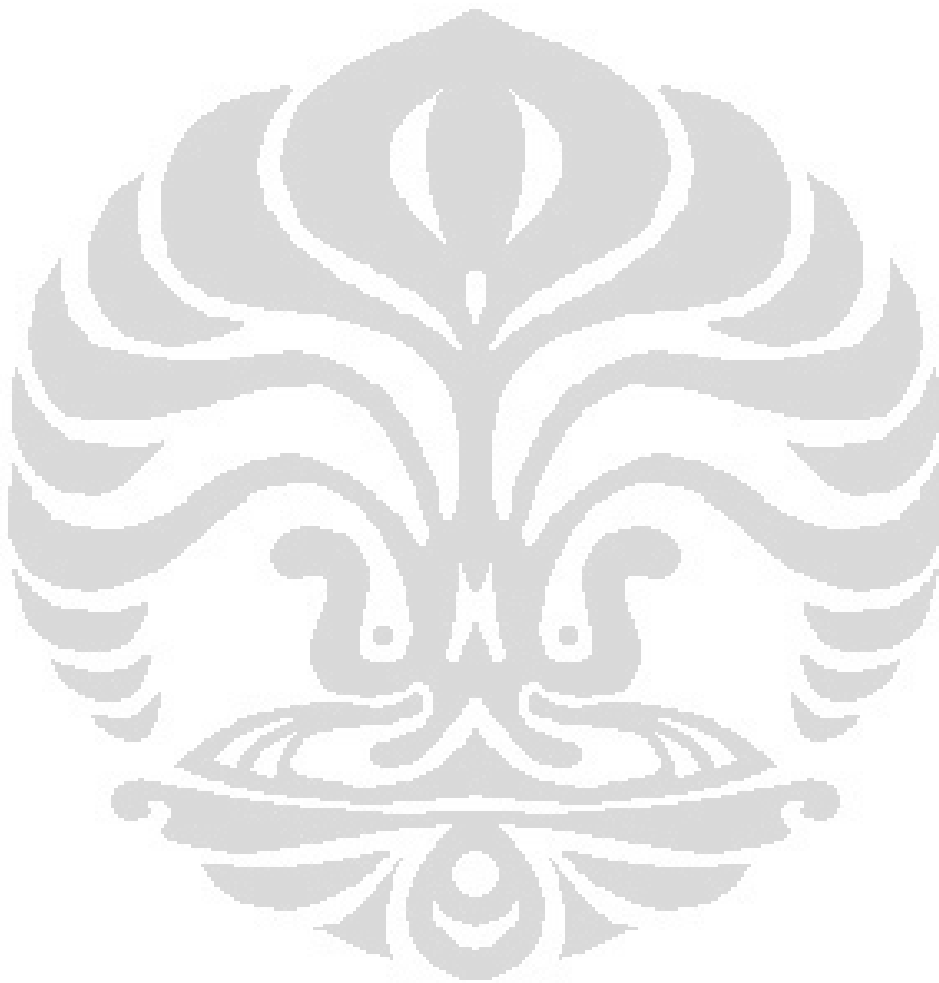
Dalam menyebarkan beritanya, portal-portal berita Indonesia *online* memiliki RSS sebagai ringkasan yang merupakan file *XML* sederhana yang memuat beberapa artikel beserta gambar maupun suara. Dengan menggunakan RSS ini, memungkinkan kita untuk mengunduh ringkasan dari berita-berita (artikel) yang di perbaharui oleh portal berita secara *online*. Dikarenakan arus berita pada portal penyedia berita pada saat sekarang begitu cepat, untuk penelitian ini akan diambil data-data dari RSS ini setiap hari dalam sebulan.

1.4.3 Implementasi algoritma ekstraksi topik

Tahap ini merupakan implementasi metode *Non-negative matrix factorization* untuk ekstraksi topik dalam bentuk algoritma. Pada tahap ini dirancang algoritma untuk mengubah teks pada file RSS ke dalam bentuk matriks kata dokumen dan algoritma untuk proses dekomposisi matriks kata dokumen dengan menggunakan metode NMF sehingga didapatkan ekstraksi topik yang diharapkan.

1.4.4 Simulasi

Pada tahap ini dilakukan penerjemahan algoritma yang dirancang kedalam bahasa pemrograman *python* sehingga dapat dimengerti oleh mesin (komputer) dan dilakukan eksekusi yang akan menampilkan hasil yaitu topik utama harian dari portal berita Indonesia *online*



BAB 2 DASAR TEORI

2.1 Pembelajaran Mesin

Pembelajaran Mesin (*Machine Learning*) merupakan bidang ilmu yang fokus pada bagaimana cara untuk mengkonstruksi sebuah program komputer yang secara otomatis dan dapat menjadi lebih baik dengan ‘*experience*’ (Mitchell, 1997). Sebuah kutipan yang terkenal dari Tom M. Mitchell mengenai definisi pembelajaran mesin adalah sebuah program komputer dikatakan belajar dari pengalaman E yang bergantung pada target T dan ukuran kinerja program P , jika kinerja dari program yang diukur dengan P terdapat di target T maka program komputer tersebut dikatakan belajar dari pengalaman E . Pembelajaran telah menerapkan beberapa konsep dan hasil dari berbagai bidang diantaranya statistik, kecerdasan buatan, filosofi, teori informasi, biologi, dll.

Berdasarkan tipe data pembelajarannya, pembelajaran mesin dapat dibagi menjadi :

2.1.1 *Supervised Learning*

Data pembelajaran disertai dengan nilai target pembelajaran pada masing-masing data (*labeled*). Data pembelajaran berbentuk $\{x_n, t_n\}$ untuk $n=1 \dots N$, dimana x adalah vektor input dan t adalah target. Pembelajaran *supervised* ini bertujuan membangun model yang dapat memenuhi target pembelajaran. Contoh dari pembelajaran supervised adalah *classification*, *regression*, *ordinal regression*, *ranking*, dll.

2.1.2 *Unsupervised Learning*

Data pembelajaran tidak disertai nilai target pembelajaran. Data pembelajaran berbentuk $\{x_n\}$ untuk $n=1 \dots N$, dimana x adalah vektor input. Pembelajaran *unsupervised* ini bertujuan untuk membangun model yang dapat menemukan variable atau komponen tersembunyi pada data pembelajaran. Pembelajaran ini dapat digunakan untuk beberapa kebutuhan seperti: *concept extraction*, *density estimation*, *clustering*, *dimensionality reduction*, *recommendation*, dll.

2.2 Faktorisasi Matriks

Metode faktorisasi matriks adalah proses pemecahan atau penguraian suatu matriks menjadi beberapa matriks. Tujuan dari faktorisasi ini adalah matriks-matriks hasil faktorisasi memiliki struktur tertentu sehingga membuat beberapa operasi akan menjadi lebih sederhana (efisien dari segi waktu komputasi) atau jumlah komponen yang lebih sedikit (efisien dalam segi alokasi memori).

Secara umum, metode faktorisasi dibagi menjadi dua kelompok, yaitu metode langsung dan metode aproksimasi.

2.2.1 Metode Langsung

Beberapa faktorisasi matriks dengan metode langsung yang banyak digunakan adalah:

1. Faktorisasi LU

Merupakan faktorisasi yang menguraikan suatu matriks $A_{m \times n}$ menjadi dua buah matriks, yaitu segitiga bawah ($L_{m \times n}$) dan matriks segitiga atas ($U_{m \times n}$).

2. Faktorisasi Cholesky

Merupakan faktorisasi yang menguraikan suatu matriks definit positif $A_{m \times m}$ menjadi dua buah matriks, yaitu matriks segitiga bawah ($G_{m \times m}$) dan matriks segitiga atas ($G_{m \times m}^T$).

3. Faktorisasi QR

Merupakan faktorisasi yang menguraikan suatu matriks $A_{m \times n}$ menjadi dua buah matriks, yaitu matriks orthogonal ($Q_{m \times m}$) dan matriks segitiga atas ($R_{m \times n}$).

2.2.2 Metode Aproksimasi

Beberapa metode aproksimasi yang sering digunakan adalah:

1. Singular Value Decomposition (SVD)

Merupakan faktorisasi yang menguraikan suatu matriks $A_{m \times n}$ menjadi tiga buah matriks, yaitu matriks ortogonal ($U_{m \times m}$), matriks diagonal ($U_{m \times m}$) matriks ortogonal ($V_{n \times n}$).

2. Nonnegative Matrix Factorization

Merupakan faktorisasi yang menguraikan suatu matriks $A_{m \times n}$ menjadi dua buah matriks, yaitu matriks $W_{m \times k}$ dan matriks $H_{k \times n}$ dimana $k = \text{rank}(A) = \min(m, n)$ dan semua elemen W dan H adalah *nonnegative*.

2.3 Norm

Definisi 2.1 Norm adalah suatu fungsi $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$, yaitu fungsi yang memetakan suatu vektor ke panjangnya yang bernilai riil, yang memenuhi sifat berikut, misal x dan y adalah vektor dan α adalah scalar $\in \mathbb{R}$, maka

1. $\|x\| \geq 0$ dan $\|x\| = 0$ jika dan hanya jika $x = 0$
2. $\|x + y\| \leq \|x\| + \|y\|$
3. $\|\alpha x\| \leq |\alpha| \|x\|$

Definisi 2.2 Norm dari suatu matriks merupakan norm vektor yang diterapkan dalam ruang vektor berdimensi mn untuk matriks berukuran $m \times n$, misal A dan B adalah matriks, maka :

1. $\|A\| \geq 0$ dan $\|A\| = 0$ jika dan hanya jika $A = 0$
2. $\|A + B\| \leq \|A\| + \|B\|$
3. $\|\alpha A\| \leq |\alpha| \|A\|$

Salah satu norm yang akan digunakan dalam skripsi ini adalah **Frobenius Norm** atau Hilbert-Schmidt Norm yang dinyatakan dalam bentuk berikut :

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} \quad (2.1)$$

BAB 3

NONNEGATIVE MATRIX FACTORIZATION

3.1 Latent Semantic Analysis

Latent semantic analysis (LSA) adalah metode yang dapat digunakan untuk mencari semantik tersembunyi dari sekumpulan dokumen dengan menggunakan *singular value decomposition* (SVD) untuk mereduksi dimensi dari matriks kata dokumen (Landauer dan Dumais, 2009). Semantik tersembunyi dapat berupa kata-kata yang bisa diinterpretasikan sebagai sebuah topik dari kumpulan dokumen.

Misalkan V adalah matriks kata dokumen berukuran $m \times n$ yang kolomnya adalah vektor dokumen. SVD dapat mendekomposisi V menjadi tiga matriks :

$$V = XYZ^T \quad (3.1)$$

dimana X adalah matriks *orthogonal* berukuran $m \times m$ yang kolomnya menyatakan vektor *singular* kiri dari V , Z adalah matriks *orthogonal* berukuran $n \times n$ yang kolomnya menyatakan vektor *singular* kiri dari V , dan Y adalah matriks diagonal berukuran $m \times n$ yang mengandung nilai *singular* $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min m,n}$ dari matriks V .

3.2 Formulasi Nonnegative Matrix Factorization

Latent semantic analysis dengan menggunakan *singular value decomposition* untuk mereduksi dimensi dari matriks kata dokumen, memiliki kelemahan dalam menggali hubungan antara kata dengan topik dari kumpulan dokumen. Pada faktorisasi SVD terdapat kecenderungan nilai matriks yang merepresentasikan hubungan antara kata dan topik bernilai negatif, sehingga akan menyulitkan dalam hal intrerpretasi hubungan tersebut. Untuk menutupi kelemahan ini digunakan *nonnegative matrix factorization* (NMF) pada LSA. Metode ini menghasilkan matriks representasi kata dengan topik yang bernilai non-negatif sehingga matriks ini lebih mudah untuk diinterpretasikan.

NMF merupakan metode faktorisasi matriks V yang berukuran $m \times n$ menjadi matriks $W_{m \times k}$ dan $H_{k \times n}$ yang bernilai non-negatif. Metode dekomposisi dengan NMF secara umum dapat dinyatakan dalam bentuk persamaan berikut.

$$V \approx WH \quad (3.1)$$

Matriks V merupakan matriks kata dokumen yang merepresentasikan teks dokumen dimana setiap entri dari vektor barisnya adalah banyaknya kata dalam setiap dokumen. Setiap entri dari vektor kolomnya merepresentasikan banyaknya kata di sebuah dokumen. W adalah **matriks bobot** yang setiap vektor barisnya merepresentasikan vektor dari masing-masing kata terhadap topik dan H adalah **matriks fitur** yang setiap vektor kolomnya merepresentasikan vektor dari masing-masing dokumen terhadap topik seperti yang diilustrasikan pada gambar 3.2 berikut :

$$\begin{array}{c}
 \begin{matrix} & d_1 & d_2 & \dots & d_n \\
 k_1 & \left(\begin{array}{cccc} e_{11} & e_{12} & \dots & e_{1n} \end{array} \right) \\
 k_2 & \left(\begin{array}{cccc} e_{21} & e_{22} & \dots & e_{2n} \end{array} \right) \\
 \dots & \left(\begin{array}{cccc} \dots & \dots & \dots & \dots \end{array} \right) \\
 k_m & \left(\begin{array}{cccc} e_{m1} & e_{m2} & \dots & e_{mn} \end{array} \right)
 \end{matrix} & \approx & \begin{matrix} & t_1 & t_2 & \dots & t_k \\
 k_1 & \left(\begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1k} \end{array} \right) \\
 k_2 & \left(\begin{array}{cccc} x_{21} & x_{22} & \dots & x_{2k} \end{array} \right) \\
 \dots & \left(\begin{array}{cccc} \dots & \dots & \dots & \dots \end{array} \right) \\
 k_m & \left(\begin{array}{cccc} x_{m1} & x_{m2} & \dots & x_{mk} \end{array} \right)
 \end{matrix} \times \begin{matrix} & d_1 & d_2 & \dots & d_n \\
 t_1 & \left(\begin{array}{cccc} y_{11} & y_{12} & \dots & y_{1n} \end{array} \right) \\
 t_2 & \left(\begin{array}{cccc} y_{21} & y_{22} & \dots & y_{2n} \end{array} \right) \\
 \dots & \left(\begin{array}{cccc} \dots & \dots & \dots & \dots \end{array} \right) \\
 t_k & \left(\begin{array}{cccc} y_{k1} & y_{k2} & \dots & y_{kn} \end{array} \right)
 \end{matrix}
 \end{array}$$

V
W
H

Keterangan : k_i = kata ke-i, d_i = dokumen ke-i, t_i = topik ke-i,
 e_{ij} = jumlah kata ke-i dalam dokumen ke-j,
 x_{ij} = nilai hubungan antara kata ke-i dengan topik ke-j,
 y_{ij} = nilai hubungan antara topik ke-i dengan dokumen ke-j.

Gambar 3.2 Ilustrasi *Nonnegative Matrix Factorization*

aproksimasi sehingga dicapai kondisi dimana perkalian matriks ini sama dengan matriks V . Untuk mencapai kondisi ini dibutuhkan sebuah kriteria untuk mengetahui perkalian matriks W dan H yang didapat mendekati nilai V yang disebut sebagai *Cost Function*. Model *Cost Function* dibangun dengan pengukuran jarak antara dua matriks nonnegatif A dan B , seperti yang dijelaskan dalam persamaan berikut.

$$\|A - B\|^2 = \sum_{ij} (a_{ij} - b_{ij})^2 \quad (3.2)$$

Dalam penelitian ini akan digunakan aturan *Frobenius Norm* pada persamaan (2.1) untuk menuju kondisi pada persamaan (3.1) yang merupakan pengembangan aturan *Cost Function* yang dijelaskan pada persamaan (3.2). Sehingga didapatkan formulasi untuk NMF sebagai berikut.

$$\begin{aligned} \min_{W, H} \quad & f(W, H) \equiv \frac{1}{2} \|V - WH\|_F^2 \\ \text{s.t} \quad & W_{mk} \geq 0, H_{kn} \geq 0, \forall m, n, k \end{aligned} \quad (3.3)$$

3.2.1 *Multiplicative Update Rule* (MUR)

Untuk menyelesaikan masalah NMF ada beberapa algoritma yang sering digunakan, diantaranya *additive update algorithm*, *multiplicative update algorithm*, *alternating least square algorithm*, dll. *Multiplicative Update Rule* adalah aturan yang populer dipakai untuk menyelesaikan permasalahan NMF sebagaimana yang disebutkan pada persamaan (3.3). Aturan ini diklaim oleh Lee dan Seung (2001) bahwa nilai dari *cost function* yang didapat dari pembaharuan matriks W dan H adalah *nonincreasing* dan *limit point* dari barisan W, H adalah titik *stationary* yang merupakan syarat penting untuk *local minimum*.

Teorema 3.1 Jarak *euclidean* $\|V - WH\|$ merupakan fungsi tidak naik (*nonincreasing*) dengan *update rule* :

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T V)_{a\mu}}{(W^T W H)_{a\mu}} \quad (3.4.a)$$

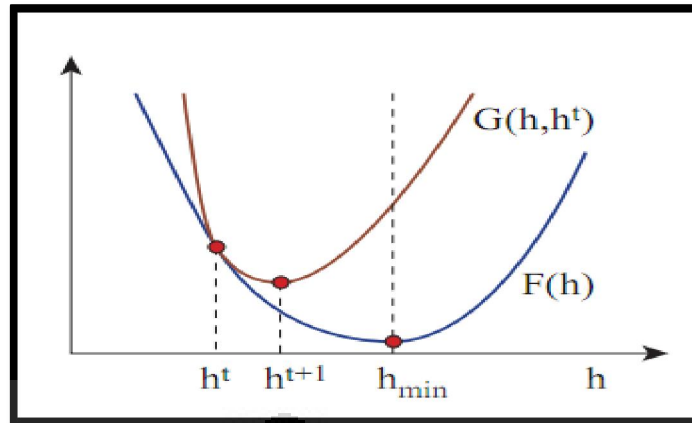
$$W_{ia} \leftarrow W_{ia} \frac{(V H^T)_{ia}}{(W H H^T)_{ia}} \quad (3.4.b)$$

Bukti. (Lee dan Seung, 2009) Untuk membuktikan teorema diatas akan diberikan definisi *auxiliary function*, *lemma 3.1* dan *lemma 3.2* sebagai berikut :

Definisi 3.1 $G(h, h')$ adalah *auxiliary function* untuk fungsi F jika memenuhi kondisi berikut :

$$G(h, h') \geq F(h), \quad G(h, h) = F(h) \quad (3.5)$$

Auxiliary function digunakan pada *lemma 3.1* dan *lemma 3.2* dibawah ini yang diilustrasikan oleh gambar berikut :



Gambar 3.3 Ilustrasi dari *auxiliary function*

Lemma 3.1 jika G adalah *auxiliary function* maka F nonincreasing terhadap *update* :

$$h^{t+1} = \arg \min_h G(h, h^t) \quad (3.6)$$

Bukti. $F(h^{t+1}) \leq G(h^{t+1}, h^t) \leq G(h^t, h^t) = F(h^t)$.

Catatan bahwa $F(h^{t+1}) = F(h^t)$ jika h^t adalah *local minimum* dari $G(h, h^t)$. Jika turunan dari F ada dan kontinu pada *neighborhood* dari h^t , hal ini juga mengimplikasikan bahwa turunan dari $\nabla F(h^t) = 0$. Maka, dengan iterasi *update* pada persamaan (3.6) didapatkan barisan estimasi yang konvergen ke *local minimum* $h_{min} = \arg \min_h F(h)$ dari fungsi objektif :

$$F(h_{min}) \leq \dots F(h^{t+1}) \leq F(h^t) \dots \leq F(h_2) \leq F(h_1) \leq F(h_0). \quad (3.7)$$

Selanjutnya akan ditunjukkan dengan pendefinisian $G(h, h^t)$ adalah *auxiliary function* untuk $\|V - WH\|$, akan mudah membuktikan teorema 3.1 dengan persamaan (3.6).

Lemma 3.2 jika $K(h^t)$ adalah matriks diagonal

$$K_{ab}(h^t) = \delta_{ab}(W^t W h^t)_a / h_a^t \quad (3.8)$$

maka

$$G(h, h^t) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2} (h - h^t)^T K(h^t) (h - h^t) \quad (3.9)$$

adalah *auxiliary function* untuk

$$F(h) = \frac{1}{2} \sum_i (v_i - \sum_a W_{ia} h_a)^2 \quad (3.10)$$

Bukti. Karena $G(h, h) = F(h)$ obvious, maka akan ditunjukkan $G(h, h^t) \geq F(h)$. Dengan membandingkan

$$F(h) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2} (h - h^t)^T (W^T W) (h - h^t) \quad (3.11)$$

Dengan persamaan (3.9) sehingga didapatkan $G(h, h^t) \geq F(h)$ ekuivalen dengan

$$0 \leq (h - h^t)^T [K(h^t) - W^T W] (h - h^t) \quad (3.12)$$

Untuk membuktikan persamaan (3.12) definit positif, anggap matriks :

$$M_{ab}(h^t) = h_a^t (K(h^t) - W^T W)_{ab} h_b^t \quad (3.13)$$

yang mengubah ukuran dari $K - W^T W$, kemudian $K - W^T W$ definit positif jika dan hanya jika M adalah :

$$\begin{aligned} v^T M v &= \sum_{ab} v_a M_{ab} v_b \\ &= \sum_{ab} h_a^t (W^T W)_{ab} h_b^t v_a^2 - v_a h_a^t (W^T W)_{ab} h_b^t v_b \\ &= \sum_{ab} (W^T W)_{ab} h_a^t h_b^t \left[\frac{1}{2} v_a^2 + \frac{1}{2} v_b^2 - v_a v_b \right] \\ &= \frac{1}{2} \sum_{ab} (W^T W)_{ab} h_a^t h_b^t (v_a - v_b)^2 \\ &\geq 0. \end{aligned}$$

Selanjutnya dapat dibuktikan teorema 3.1 dengan mengganti $G(h, h^t)$ pada persamaan (3.6) dengan persamaan (3.9) sehingga didapatkan *update rule*

$$h^{t+1} = h^t - K(h^t)^{-1} \nabla F(h^t) \quad (3.14)$$

karena persamaan (3.9) adalah *auxiliary function* maka F *nonincreasing* terhadap update rule diatas, merujuk pada lemma 1. Dengan menulis komponen dari persamaan (3.14) secara eksplisit didapat

$$h_a^{t+1} = h_a^t \frac{(W^T x)_a}{(W^T W h^t)_a} \quad (3.15)$$

Dengan aturan W dengan H pada lemma 3.1 dan lemma 3.2, maka akan didapat F juga *nonincreasing* terhadap *update rule* untuk W.

3.2.2 *Multiplicative Update Algorithm (MUA)*

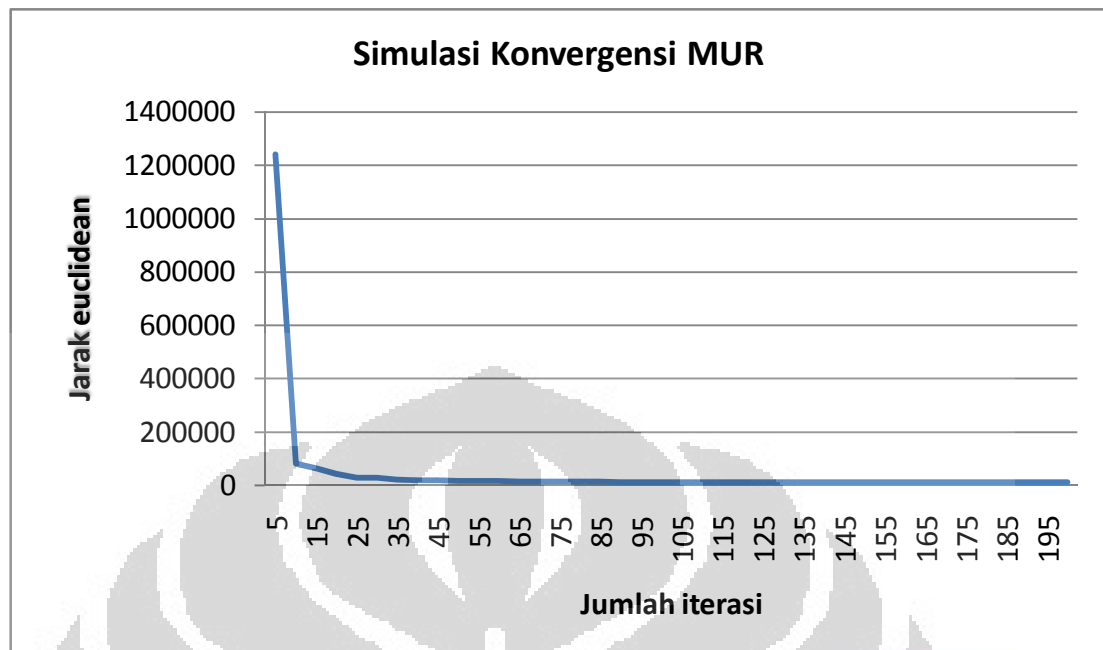
Dari *multiplicative update rule* diatas dibentuk sebuah algoritma untuk menyelesaikan masalah NMF (Berry, dkk, 2007) sebagai berikut :

Table 3.1 Multiplicative Update Algorithm

1:	$W = rand(m,k)$, where $w_{ij} \geq 0, \forall i,j$
2:	$H = rand(k,n)$, where $h_{ij} \geq 0, \forall i,j$
3:	for $i = 1$ to <i>maxiter</i> do
4:	$H = H \cdot (V^T V) ./ (W^T W H + 10^{-9})$
5:	$W = W \cdot (V H^T) ./ (W H H^T + 10^{-9})$
6:	end

Penambahan 10^{-9} pada algoritma diatas untuk menghindari pembagian dengan nol. Dengan algoritma ini, penulis mencoba memperkuat klaim dari Lee dan Seung (2009) dengan melakukan simulasi pada permasalahan yang diangkat dalam penulisan ini. Data yang digunakan untuk simulasi uji konvergensi *multiplicative update rule* adalah matriks kata dokumen dari sekumpulan dokumen berita Indonesia *online* pada tanggal 1 Mei 2011. Hasil dari simulasi

yang dilakukan dapat dilihat pada gambar berikut :

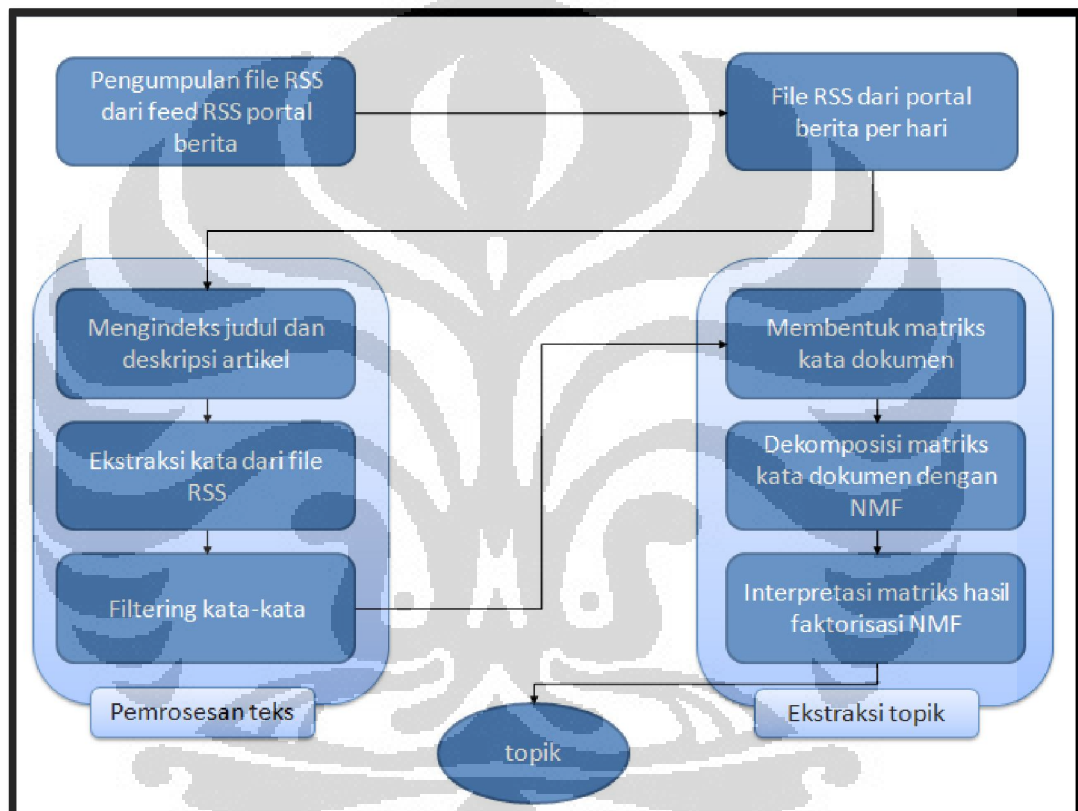


Gambar 3.1 Simulasi uji konvergensi MUR

Pada gambar 3.1 dapat dilihat bahwa pada iterasi ke-5 sampai dengan iterasi ke-100, penurunan jarak Euclidian antara matriks kata dokumen V dengan matriks W , H hasil aproksimasi dengan MUA pada tabel 3.1 menurun secara signifikan. Hal ini memperkuat klaim dari Lee dan Seung (2001).

BAB 4 SIMULASI

Bab ini akan menjelaskan hasil simulasi ekstraksi topik utama harian portal berita berbahasa Indonesia online dengan menggunakan metode NMF. Secara umum proses ekstraksi atau simulasi yang dilakukan dapat dijelaskan dalam gambar berikut:



Gambar 4.1 Ilustrasi Proses Simulasi

4.1 Perangkat Lunak dan Perangkat Keras

Simulasi dilakukan dengan menggunakan bantuan dari perangkat lunak dan perangkat keras sebagai berikut :

- Perangkat Lunak

Sistem Operasi : Windows 7 32bit

Bahasa Pemrograman : *Python 2.7.2*³

Modul tambahan : *Numpy 1.6.1*⁴ dan *Feedparser 5.1.2*.⁵

- Perangkat Keras

Processor : AMD Athlon X2 7750 *Dual-Core Processor*, ~2.7Ghz.

Memori : 2048 MB RAM

4.2 Preparasi Data

Proses pertama dari simulasi adalah persiapan data yang akan digunakan. Persiapan data ini terdiri dari penentuan sumber data. Data berasal dari *file* RSS beberapa portal berita nasional. *Feed* RSS dari portal berita *online* yang digunakan dalam penelitian ini berasal dari portal berita:

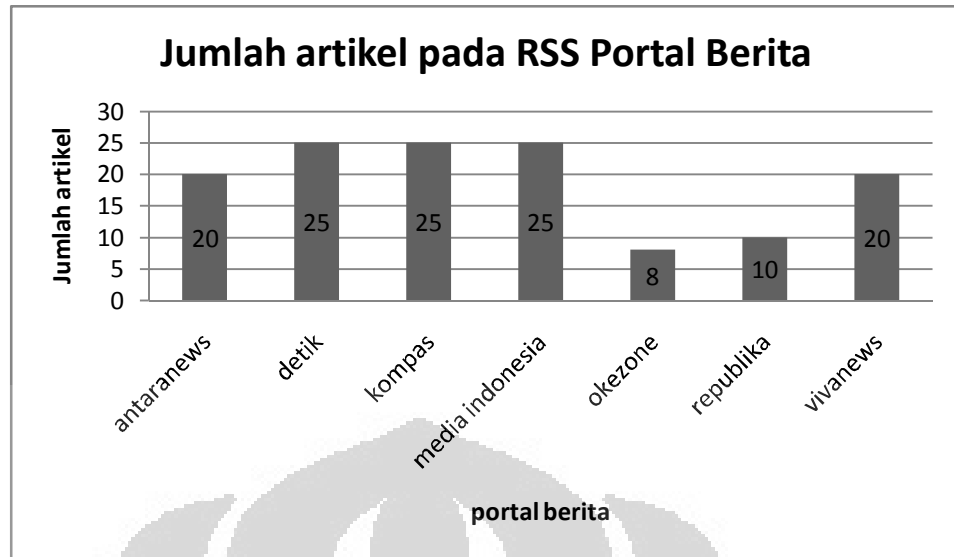
1. Kompas.com
2. Okezone.com
3. Detik.com
4. Vivanews.com
5. Antaranews
6. Republika *online*
7. Media Indonesia

Untuk sekali akses ke *feed* RSS dari portal berita diatas artikel yang dapat diunduh terbatas sesuai dengan kapasitas dari *feed* RSS tersebut menampilkan artikel berita. Berikut statistik jumlah artikel yang dapat ditampilkan oleh masing-masing *feed* RSS dari beberapa portal berita Indonesia *online*.

³ <http://www.python.org/>

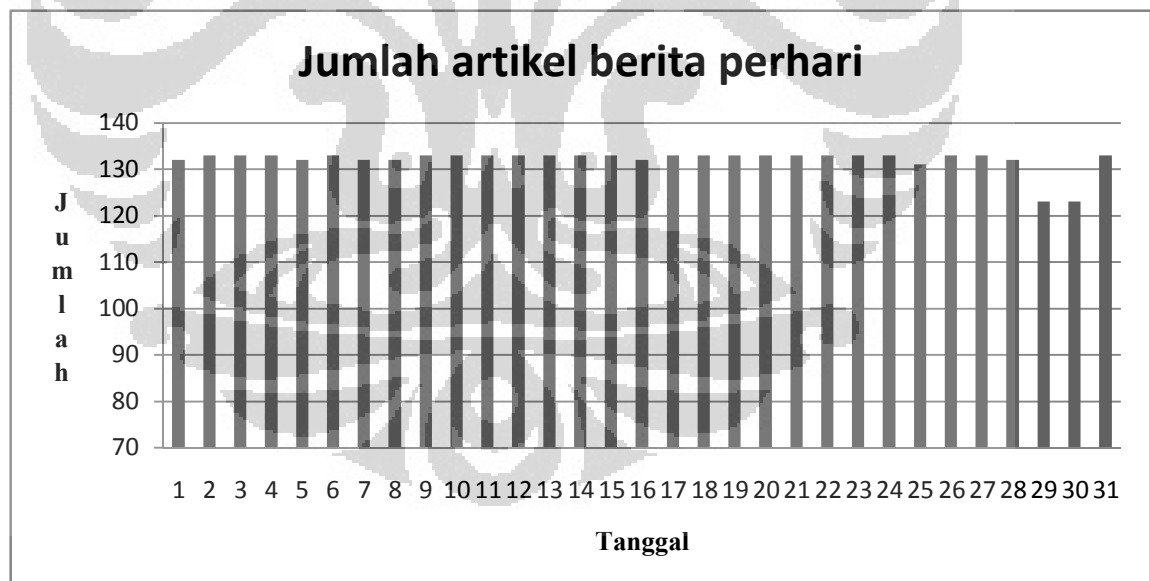
⁴ <http://numpy.scipy.org/>

⁵ <http://code.google.com/p/feedparser/>



Gambar 4.2 Rata-rata kapasitas dari feed RSS dari portal berita Indonesia *online*

Kemudian *file-file* RSS dari portal berita ini dikumpulkan selama satu bulan yaitu pada bulan Mei 2012 sehingga didapatkan statistik jumlah artikel dari *feed* RSS portal berita Indonesia *online* sebagai berikut:



Gambar 4.3 Grafik jumlah artikel yang akan diproses per hari selama satu bulan

4.3 Ekstraksi Kamus Kata

Setelah preparasi data selesai, akan dilakukan ekstraksi kamus kata yang dilakukan dengan cara sebagai berikut : pemisahan kata per kata yang disimpan

pada kamus kata kandidat topik dan penyaringan kamus kata dari kata-kata yang dianggap tidak relevan dengan topik.

4.3.1 Pemisahan Kata

Pada tahap ini, dengan menggunakan modul *feedparser* pada *python*, berikut contoh sebuah artikel berita dalam format RSS.

```
<?xml version="1.0" encoding="ISO-8859-1"?> <?xml-stylesheet href="feed.css" type="text/css"?> <rss
version="2.0"> <channel> <title>ANTARA - Berita Terkini</title> <description>Berita Terkini dari ANTARA
News</description> <link>http://www.antaranews.com</link> <lastBuildDate>Tue, 01 May 2012 13:00:02
+0700</lastBuildDate> <item> <title>Cahaya TV network siarkan berita ANTARA TV</title>
<link>http://www.antaranews.com/berita/1335853260/cahaya-tv-network-siarkan-berita-antara-tv</link>
<description> Cahaya TV Network menjalin kerja sama dengan Perum LKBN ANTARA untuk menyiarkan
rogram-program berita buatan ANTARA yang akan berlangsung hingga 2013.</description> <pubDate>Tue,
01 May 2012 13:21:00 +0700</pubDate> </item>
```

Gambar 4.4 Contoh tampilan sebuah artikel dalam *file* RSS

Pada gambar 4.3, bagian dengan warna merah merupakan bagian judul dan deskripsi dari sebuah artikel berita pada file RSS. Dengan bantuan modul *feedparser* file RSS tersebut dapat diindeks berdasarkan tautan, judul deskripsi, dll. Sehingga, dengan mudah kita mengambil bagian judul dan deskripsi dari artikel tersebut. Akan tetapi masih terdapat markup dan gambar yang terbawa oleh judul dan artikel yang diindeks oleh *feedparser*. Beberapa langkah yang dilakukan pada pemisahan kata adalah sebagai berikut :

Langkah 1. Penghapusan Markup dan Gambar

Pada gambar 4.3, modul *feedpaser* akan mengindeks judul dari artikel berita tersebut sebagai berikut :

```
'<title>Cahaya TV network siarkan ANTARA TV</title>'
```

Markup yang berupa `<title>` dari judul tersebut akan dihilangkan sehingga akan didapatkan kata-kata yang hanya terdapat pada judul artikel.

Langkah 2 Pemisahan kata

Selanjutnya akan dipisahkan kata-kata yang terdapat pada judul dan deskripsi artikel berita. Kata-kata pada judul dan deskripsi artikel, oleh program dibaca sebagai kumpulan string yang dipisahkan oleh spasi. Jadi

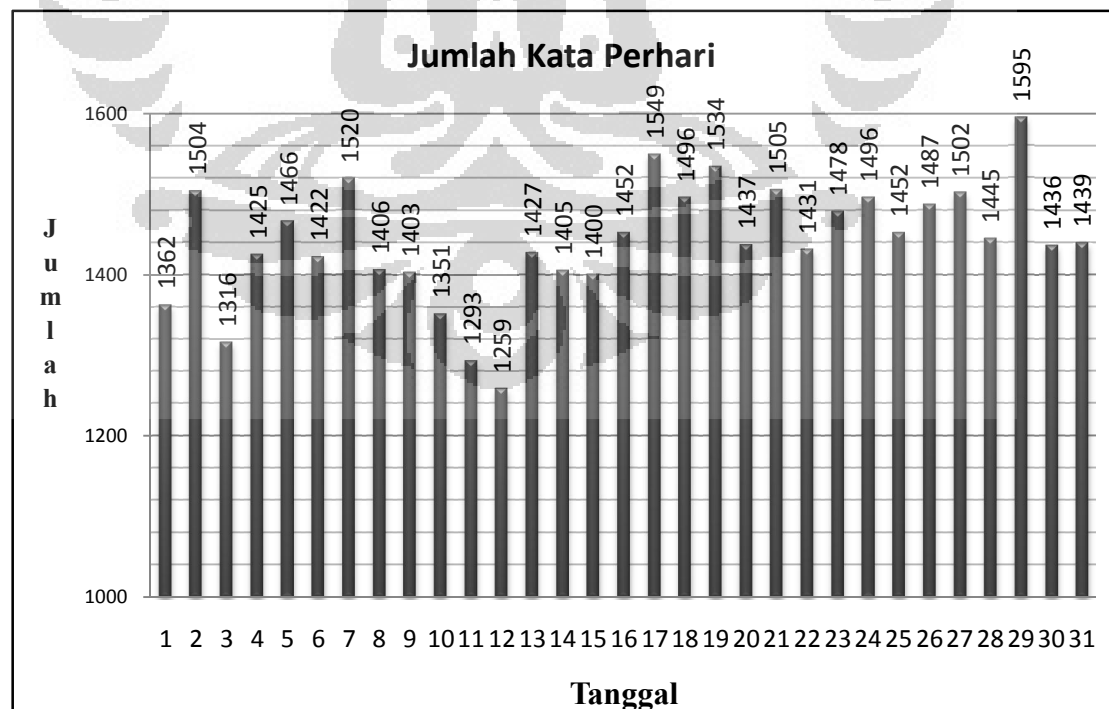
ketika program membaca spasi, dengan otomatis string yang telah dibaca akan disimpan sebagai sebuah kata pada kamus kata. Hal ini dilakukan terhadap semua artikel pada setiap file RSS pada satu hari, sehingga didapatkan kamus pada hari tersebut.

4.3.2 Penyaringan Kamus Kata

Pada tahap ini dilakukan penyaringan kata pada kamus kata, yaitu membuang kata-kata yang dianggap tidak terlalu memberi makna pada topik seperti kata penghubung, kata ganti orang, identitas portal, dll. Diantara kata-kata yang tidak member makna pada topik tersebut, kata-kata yang dibuang pada simulasi ini adalah sebagai berikut :

'saya', 'lebih', 'menjadi', 'di', 'dan', 'tidak', 'ke', 'sudah', 'ini', 'itu', 'tak', 'bisa', 'saat', 'masih', 'belum', 'yang', 'akan', 'dari', 'dengan', 'untuk', 'dalam', 'micom', 'co', 'id'.

Setelah dilakukan penyaringan dengan menghilangkan kata-kata diatas, maka akan didapatkan kamus kata kadidat topik yang akan diekstraksi. Pada gambar 4.5 dapat dilihat bahwa jumlah kata hasil ekstraksi dari file-file RSS yang dikumpulkan setiap harinya berbeda.

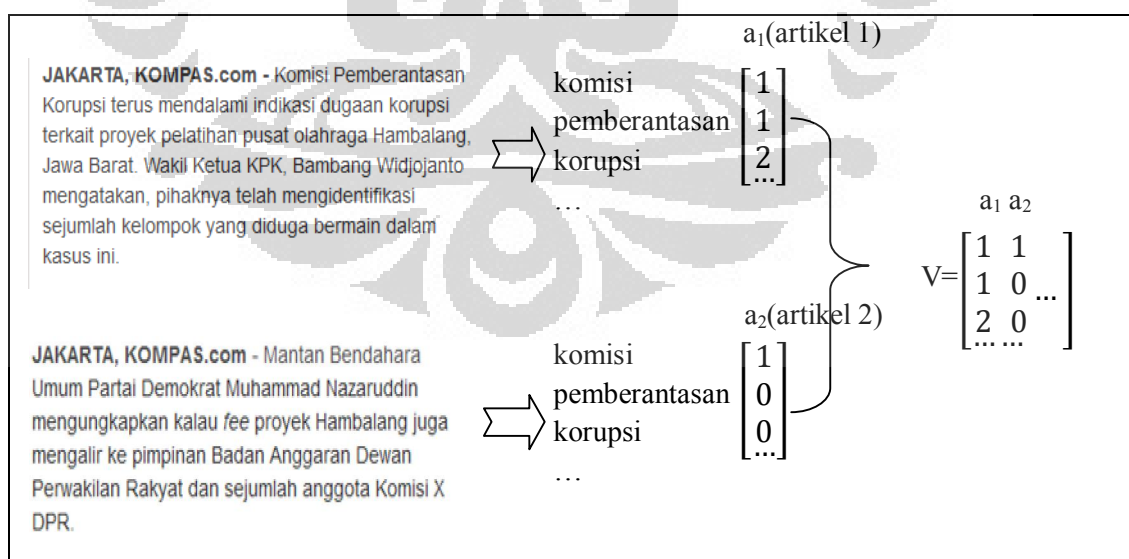


Gambar 4.5 Grafik jumlah kata berbeda yang dilibatkan per hari

4.4 Pembentukan Matriks Kata dokumen

Setelah mendapatkan kamus kata hasil ekstraksi kata, selanjutnya akan dibentuk sebuah matriks kata dokumen V yang merepresentasikan jumlah kata yang tersimpan pada kamus kata pada setiap artikel berita. Kolom-kolom dari matriks V merepresentasikan artikel sedangkan baris-baris dari matriks V merepresentasikan kata.

Pada proses pembentukan matriks kata dokumen, pertama dimulai dengan menghitung kemunculan semua kata-kata pada kamus kata pada satu artikel sehingga terbentuk sebuah vektor. Pada gambar 4.5, didalam artikel pertama kata-kata komisi, pemberantasan, dan korupsi pada kamus kata muncul sebanyak satu kali untuk kata komisi, satu kali untuk kata pemberantasan, dan dua kali untuk kata korupsi. Sehingga, dengan menghitung frekuensi kata-kata dalam kamus kata pada artikel pertama didapatkan vektor a_1 . Begitu juga pada artikel kedua, frekuensi kemunculan kata-kata dalam kamus kata akan membentuk vektor a_2 . Hal ini dilakukan untuk semua artikel pada setiap file RSS sehingga akan didapatkan vektor-vektor frekuensi kemunculan kata pada artikel. Selanjutnya akan didapatkan matriks kata dokumen V dengan vektor kolomnya adalah vektor-vektor frekuensi kata pada artikel.



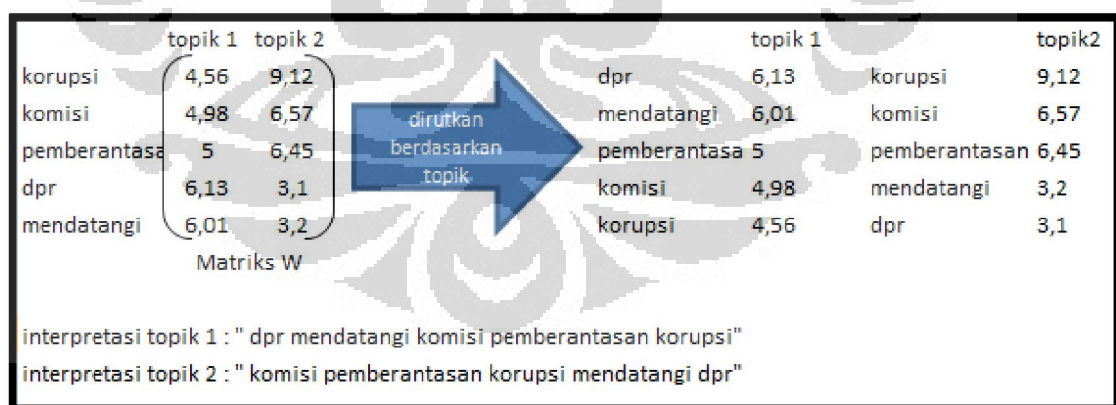
Gambar 4.6 Proses pembentukan matriks kata dokumen V

4.5 Ekstraksi Topik Utama

Topik adalah sebuah kata benda, yang memiliki makna pokok pembicaraan dalam diskusi, ceramah, karangan, dan sebagainya; bahan diskusi hal yang menarik perhatian umum pada waktu akhir-akhir ini (Kamus Besar Bahasa Indonesia, 2008). Pada skripsi ini topik utama dari portal berita *online* berbahasa Indonesia merupakan sebuah kalimat yang menggambarkan pokok pembicaraan dari artikel-artikel berita pada portal berita tersebut.

Secara manual, topik utama dari portal berita *online* berbahasa Indonesia ini dapat diekstraksi dengan membaca semua artikel berita. Sehingga secara intuisi diperoleh sebuah kalimat yang dinyatakan sebagai topik utama pada kumpulan artikel tersebut.

Dengan menggunakan NMF, topik utama dari portal berita *online* berbahasa Indonesia dapat diekstraksi dengan interpretasi dari beberapa kata yang memiliki bobot tertinggi pada matriks bobot W . Vektor kolom dari matriks W hasil faktorisasi dengan menggunakan NMF merepresentasikan hubungan topik dengan setiap kata pada kamus kata. Dengan mengambil sepuluh kata dengan bobot tertinggi pada masing-masing topik, dapat diinterpretasikan menjadi sebuah kalimat yang diklaim sebagai topik utama hasil ekstraksi dengan menggunakan NMF.



Gambar 4.7 Proses interpretasi topik dari matriks bobot W

Pada gambar 4.7 diberikan sebuah contoh matriks bobot W hasil faktorisasi dengan NMF yang akan diinterpretasikan menjadi topik utama. Pada vektor kolom pertama pada matriks W yang merupakan representasi topik 1 dengan kata-

kata yang mendukung topik tersebut. Pada topik 1 kata-kata dengan bobot tertinggi secara berurutan adalah ‘dpr’, ‘mendatangi’, ‘pemberantasan’, ‘komisi’, dan ‘korupsi’. Sehingga dari kata-kata pendukung topik 1 ini akan diinterpretasikan menjadi sebuah kalimat topik yaitu: “dpr mendatangi komisi pemberantasan korupsi”. Begitu juga dengan topik 2, dimana kata-kata pendukung topik dengan bobot tertinggi secara berurutan adalah ‘korupsi’, ‘komisi’, ‘pemberantasan’, ‘mendatangi’, dan ‘dpr’. Sehingga dari kata-kata tersebut diinterpretasikan menjadi sebuah kalimat topik yaitu: “komisi pemberantasan korupsi mendatangi dpr”.

4.6 Ekstraksi Topik dengan *Nonnegative Matrix Factorization*

Pada tahap ini, dilakukan faktorisasi matriks dengan metode NMF terhadap matriks kata dokumen V yang dibentuk dari file-file RSS yang telah dijelaskan pada subbab 4.2. Dengan menerapkan algoritma pada tabel 3.1 pada bahasa pemrograman python (lampiran 3) akan dilakukan proses faktorisasi sehingga menghasilkan matriks W yang dibutuhkan untuk interpretasi topik.

Pada simulasi ini pemilihan jumlah iterasi berdasarkan hasil simulasi uji konvergensi multiplicative update algorithm (gambar 3.4) dimana pada simulasi tersebut didapatkan penurunan nilai *cost function* secara signifikan terjadi sampai dengan itersi ke-100. Untuk jumlah topik yang akan diekstrak (nilai k), akan dipilih 5 topik yang akan diekstrak setiap harinya. Sehingga, jika kita memiliki matriks kata dokumen V berukuran $n \times m$ maka akan terbentuk matriks bobot W berukuran $n \times 5$.

4.7 Hasil Simulasi

Setelah mendapatkan matriks bobot W hasil faktorisasi matriks kata dokumen V dengan menggunakan NMF, akan diurutkan bobot dari masing-masing kata pada setiap topik dari yang terbesar ke terkecil. Sepuluh kata pertama yang memiliki bobot terbesar menjadi kata-kata pendukung topik yang akan diinterpretasikan menjadi sebuah kalimat topik.

4.7.1 Interpretasi Topik

Tabel 4.1 sampai tabel 4.6 menunjukkan hasil dari simulasi ekstraksi topik utama harian portal berita berbahasa Indonesia menggunakan *nonnegative matrix factorization* untuk kumpulan artikel berita pada tanggal 1-4 Mei 2012. Kolom kedua pada tabel tersebut menunjukkan kata-kata pendukung topik yang dihasilkan dari simulasi ini. Kolom ketiga adalah interpretasi kalimat topik utama yang dilakukan dengan intuisi penulis dengan menggunakan kata-kata pendukung topik yang bersesuaian.

Tabel 4.1 Hasil Simulasi untuk data pada tanggal 1 Mei 2012

Topik ke	Kata-kata pendukung topik	Interpretasi topik
1	'buruh', 'hari', 'indonesia', 'selasa', 'istana', 'depan', 'kpk', 'aksi', 'nasional', 'polri'	aksi pada hari buruh nasional didepan istana dan gedung kpk
2	'buruh', 'kpk', 'dpr', 'bus', 'ketua', 'partai', 'jalan', 'presiden', '2012', 'istana'	
3	'buruh', 'dpr', 'jalan', 'pendemo', 'istana', 'bus', 'massa', 'hari', 'ketua', 'merdeka'	Massa pendemo di jalan merdeka pada hari buruh.
4	'buruh', 'hari', 'indonesia', 'kpk', 'istana', 'jalan', 'tewas', 'presiden', 'polri', 'depan'	Hari buruh di Indonesia
5	'buruh', 'hari', 'selasa', 'may', 'istana', 'indonesia', 'pekerja', 'day', 'serikat', 'aksi',	Aksi serikat pekerja pada ahari buruh atau may day

Pada tabel 4.1 dapat dilihat bahwa kata 'buruh' dan 'hari' mendominasi kata-kata pendukung topik. Pada topik pertama, delapan kata pertama dapat dengan mudah untuk diinterpretasikan sedangkan untuk topik kedua lima kata pertama sulit untuk diinterpretasikan sehingga. Pada tabel 4.2, 4.3, dan 4.4 terlihat bahwa tidak ada kata yang sangat dominan. Jumlah bobot yang merata mengakibatkan kata-kata yang mendukung topik tidak berkaitan seperti pada tabel 4.2 , topik ke-2, sehingga sulit untuk diinterpretasikan. Hal ini juga mengakibatkan topik hasil ekstraksi memiliki dua kalimat kalimat hasil interpretasi, seperti pada tabel 4.3 topik ke-5 dimana topik ke-5 ini

diinterpretasikan sebagai “Jenazah endang rahayu sedyaningsih” dan “kasus kpk yang melibatkan anis matta dan sby”.

Tabel 4.2 Hasil Simulasi untuk data pada tanggal 2 Mei 2012

Topik ke	LSA (Kata-kata pendukung topik)	Interpretasi topik
1	'indonesia', 'kpk', 'menkes', 'kesehatan', 'kasus', 'endang', 'buruh', 'ktp', 'sedyaningsih', 'menteri'	Kesehatan menkes Endang Sedyaningsih
2	'kpk', 'tni', 'anggota', 'dpr', 'ad', 'indonesia', 'neneng', 'kasus', 'palmerah', 'restoran'	
3	'anas', 'kasus', 'indonesia', 'kpk', 'ketua', 'partai', 'buruh', 'kesehatan', 'demokrat', 'korupsi'	Kasus Anas ketua partai democrat di KPK indonesia
4	'kpk', 'menkes', 'kasus', 'anas', 'endang', 'nazaruddin', 'buruh', 'korupsi', 'angie', 'surat'	Kasus anas
5	'kpk', 'kasus', 'korupsi', 'anggota', 'barat', 'ketua', 'dpr', 'partai', 'tni', 'indonesia', 'komisi', 'anas'	Kasus korupsi anggota kpk

Tabel 4.3 Hasil Simulasi untuk data pada tanggal 3 Mei 2012

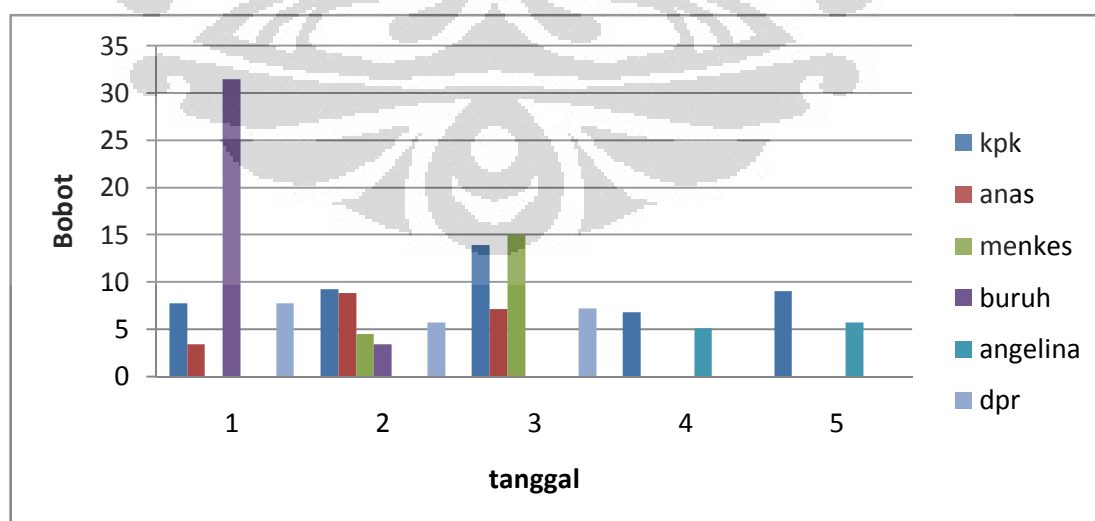
Topik ke	LSA (Kata-kata pendukung topik)	Interpretasi topik
1	'kpk', 'ketua', 'sby', 'api', 'emas', 'kasus', 'iswahyudi', 'sebagai', 'pemotor', 'patek'	
2	'partai', 'kpk', 'ketua', 'nasdem', 'endang', 'umum', 'anas', 'demokrat', 'kasus', 'rahayu'	Kasus kpk yang melibatkan ketua umum partai nasdem dan demokrat anas
3	'kasus', 'korupsi', 'endang', 'tersangka', 'dpr', 'ketua', 'anas', 'partai', 'dugaan', 'suap'	Dugaan suap tersangka kasus korupsi dpr dan anas ketua partai
4	'endang', 'kpk', 'rahayu', 'sedyaningsih', 'jenazah', 'partai', 'kesehatan', 'terakhir', 'kamis', 'hills'	Jenazah Endang Rahayu Sedyaningsih
5	'endang', 'kpk', 'rahayu', 'kasus', 'anis', 'sedyaningsih', 'matta', 'ketua', 'sby', 'jenazah'	Jenazah endang rahayu sedyaningsih, kasus kpk yang melibatkan anis matta dan sby

Tabel 4.4 Hasil Simulasi untuk data pada tanggal 4 Mei 2012

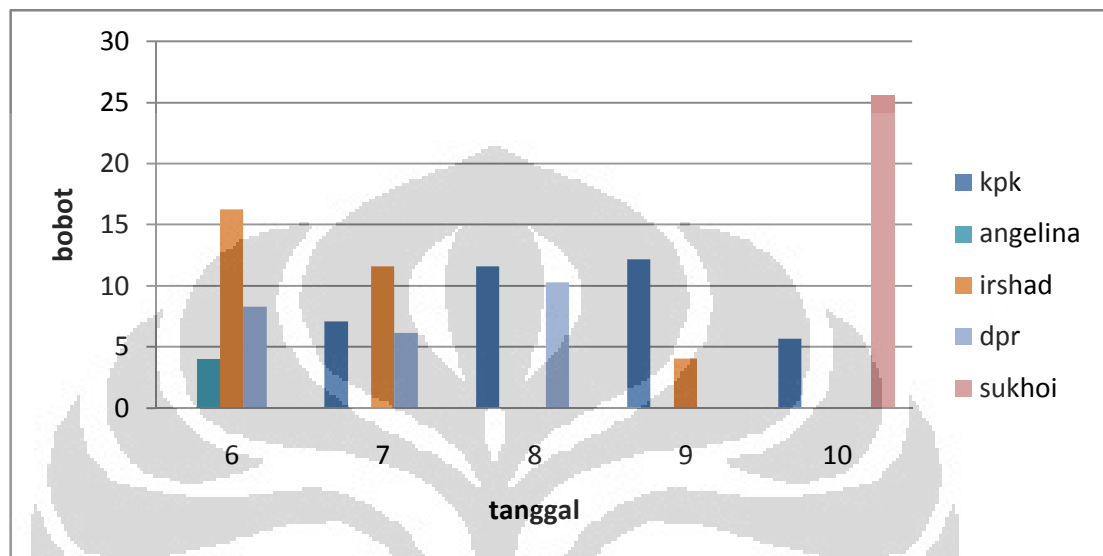
Topik ke	LSA (Kata-kata pendukung topik)	Interpretasi topik
1	'penumpang', 'luka', 'bentrokkan', 'solo', '380', 'cengkareng', 'kasus', 'orang', 'meninggal', 'dhana'	Penumpang luka dan meninggal akibat bentrokkan di solo
2	'kasus', 'ongen', 'indonesia', 'keluarga', 'terkait', 'munir', 'nba', 'kematian', 'komisi'	Kasus ongen terkait dengan keluarga munir
3	'kasus', 'ongen', 'kpk', 'angelina', 'angie', 'dhana', 'terkait', 'warga', 'solo', 'kematian'	Kpk terkait kasus ongen, angie dan dhana, kematian warga solo
4	'kasus', 'kpk', 'dhana', 'warga', 'jumat', 'indonesia', 'hukum', 'angelina', 'kejaksaan', 'angie'	kasus kpk terkait dhana dan Angelina
5	'nba', 'solo', 'bentrokkan', '2012', 'ketua', 'kota', 'ongen', 'terkait', 'sondakh', 'indonesia',	berita nba dan bentrokkan di solo

4.7.2 Tren dari Kata-Kata Pendukung Topik

Dari beberapa kata-kata pendukung topik yang dihasilkan oleh simulasi ekstraksi topik diatas, dapat dilihat tren dari topik berita pada portal berita *online* berbahasa Indonesia selama satu bulan. Berikut grafik yang menggambarkan tren dari topik berita pada tanggal 1 Mei sampai dengan 20 Mei 2012 :

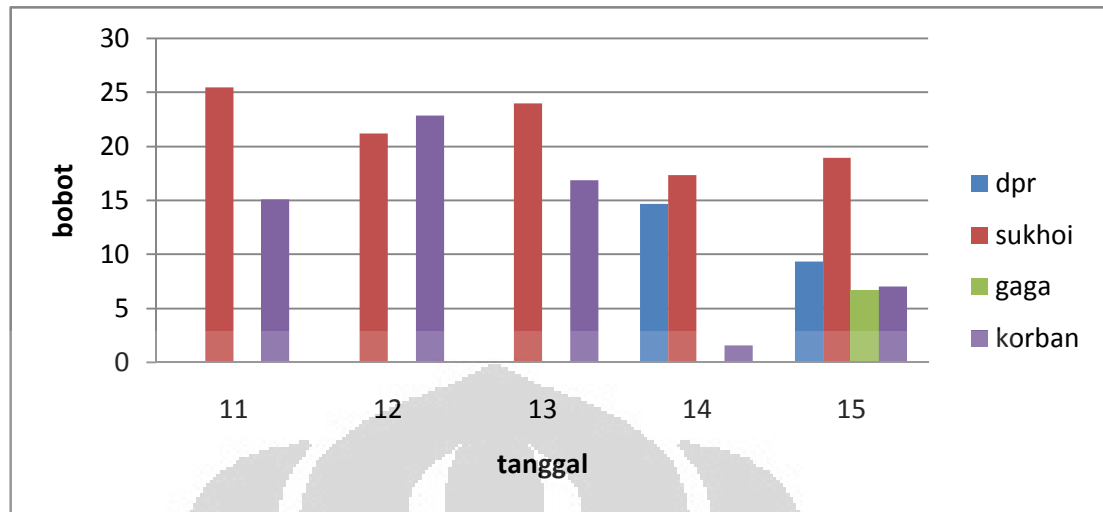
**Gambar 4.8** Tren topik berita pada tanggal 1-5 Mei 2012

Pada gambar diatas terlihat bahwa tanggal 1 Mei topik berita mengenai buruh sangat dominan, sedangkan berita dengan topik kpk terus hadir dalam lima hari tersebut. Pada tanggal 2 muncul berita dengan topik menkes dimana menjadi topik yang dominan pada hari ketiga.



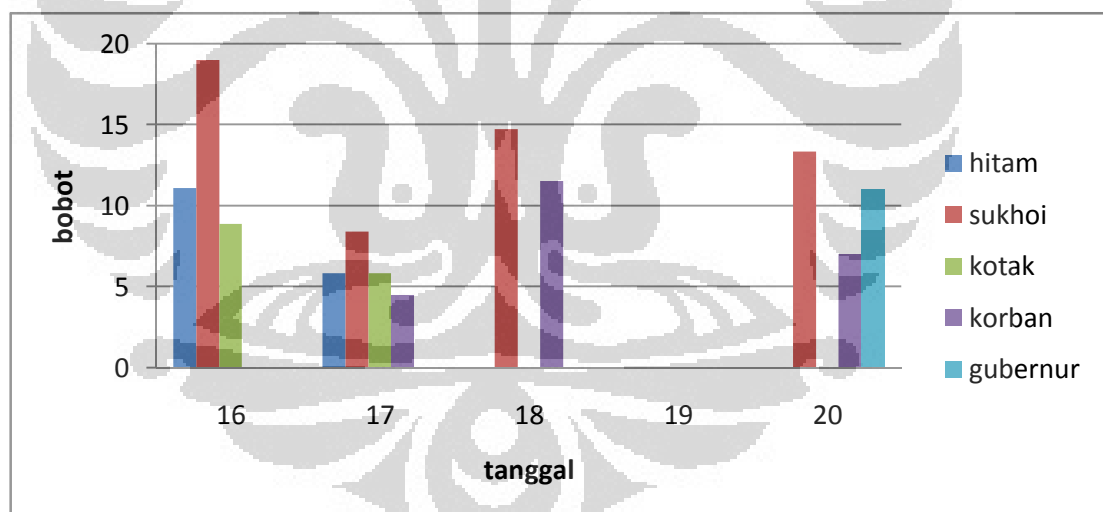
Gambar 4.9 Tren topik berita pada tanggal 6-10 Mei 2012

Pada tanggal 6 dan 7 Mei, topik berita mengenai irshad menjadi yang dominan yang kemudian sempat hilang pada tanggal 8 Mei kemudian muncul lagi pada tanggal 9. Pada tanggal 10 Mei, topik berita mengenai sukhoi sangat dominan sekali.



Gambar 4.10 Tren topik berita pada tanggal 11-15 Mei 2012

Dimulai dari tanggal 10 Mei, portal-portal berita online berbahasa Indonesia terus didominasi oleh berita mengenai kecelakaan pesawat sukhoi yang diwakili oleh kata sukhoi.



Gambar 4.11 Tren topik berita pada tanggal 16-20 Mei 2012

Masih dengan berita tentang kecelakaan sukhoi mendominasi topik berita pada tanggal 16 sampai tanggal 20. Dengan munculnya kata kotak dan hitam, dapat dikatakan berita tentang kotak hitam dari pesawat sukhoi yang menjadi topik yang dominan pada lima hari tersebut. Pada tanggal 19 tidak dapat dilakukan ekstraksi topik karena kesalahan pada pengumpulan data. Dengan

adanya analisa tren kata-kata pendukung topik ini, dari gambar gambar 4.10 dapat diprediksi bahwa pada tanggal 19 Mei masih akan didominasi oleh berita mengenai sukhoi.

4.8 Perbandingan Hasil dengan Pencarian Manual

Setelah melakukan simulasi dan interpretasi kalimat dari topik utama berita harian dari portal berita *online*, pada bagian ini akan dicoba menguji hasil simulasi yang dilakukan. Pengujian hasil ini dilakukan dengan membandingkan kalimat topik hasil interpretasi dari kata-kata pendukung dengan topik yang diekstrak secara manual. Pencarian topik secara manual dilakukan dengan membaca seluruh dokumen berita dan menyimpulkan beberapa topik secara intuitif oleh seorang relawan. Relawan merupakan seorang sarjana humaniora.

Pencarian topik oleh relawan dilakukan secara independen, dimana relawan tidak mengetahui hasil ekstraksi topik yang didapat dari metode NMF. Relawan diberikan dokumen yang sama dengan data masukan metode LSA. Relawan diminta menentukan beberapa topik yang disusun berdasarkan dominasinya. Topik pertama merupakan topik paling dominan. Topik kedua adalah topik dominan selanjutnya. Demikian hingga topik kelima.

Perbandingan topik hasil interpretasi keluaran dengan metode *nonnegative matrix factorization* dan pencarian secara manual oleh relawan dituliskan dalam tabel berikut :

Tabel 4.5 Perbandingan pada tanggal 1 Mei 2012

Topik ke	Interpretasi topik dengan NMF	Ekstraksi topik secara manual
1	aksi pada hari buruh nasional didepan istana dan gedung kpk	Demonstrasi warnai perayaan Hari Buruh
2		Kebakaran bus di Sumatra Barat tewaskan 13 penumpang
3	Massa pendemo di jalan merdeka pada hari buruh.	Dugaan keterlibatan Gubernur Riau dalam kasus suap PON
4	Hari buruh di Indonesia	Masalah Internal Golkar terkait pencalonan Ical
5	Aksi serikat pekerja pada hari buruh atau may day	Pemanggilan Anis Matta oleh KPK terkait kasus suap DPID

Dapat dilihat bahwa dominasi kata ‘buruh’ dan ‘hari’ pada tanggal 1 Mei dapat menyembunyikan topik-topik yang lain dari hasil ekstraksi menggunakan NMF.

Tabel 4.6 Perbandingan pada tanggal 2 Mei 2012

Topik ke	Interpretasi topik dengan NMF	Ekstraksi topik secara manual
1	Kesehatan menkes Endang Sedyaningsih	Kondisi kesehatan Menkes semakin menurun
2		Nazaruddin surati KPK terkait koordinasi pemulangan Neneng
3	Kasus Anas	Keterkaitan Anas dengan kasus Hambalang
4	Kasus korupsi anas	KPK tawari Angie sebagai Justice Collaborator
5	Kasus korupsi anggota kpk	Hukuman anggota Brimob Gorontalo dinilai terlalu ringan

Karena tidak ada kata yang terlalu dominan seperti pada tanggal 1 Mei, topik hasil ekstraksi, pada tanggal 2 Mei sudah memiliki sedikit perbedaan antar topik.

Tabel 4.7 Perbandingan pada tanggal 3 Mei 2012

Topik ke	Interpretasi topik dengan NMF	Ekstraksi topik secara manual
1		Pemakaman Menkes non-aktif Endang Rahayu
2	Kasus kpk yang melibatkan ketua umum partai nasdem dan demokrat anas	Dugaan keterlibatan Anis Matta dalam kasus PPID
3	Dugaan suap tersangka kasus korupsi dpr dan anas ketua partai	SBY memberikan penghormatan terakhir untuk Endang Rahayu
4	Jenazah Endang Rahayu Sedyaningsih	Meninggalnya saksi kunci kasus Munir
5	Jenazah endang rahayu sedyaningsih, kasus kpk yang melibatkan anis matta dan sby	Kawanan perampok bersenjata api rampok toko emas di Cilacap

Tabel 4.8 Perbandingan pada 4 Mei 2012

Topik ke	Interpretasi topik dengan NMF	Ekstraksi topik secara manual
1	Penumpang luka dan meninggal akibat bentrokan di solo	Kasus korupsi dan pencucian uang Dhana Widyatmika
2	Kasus ongen terkait dengan keluarga munir	Bentrok warga dan laskar militan di Solo
3	Kpk terkait kasus ongen, angie dan dhana, kematian warga solo	Kasus korupsi Angelina Sondakh
4	Kasus kpk terkait dhana dan Angelina	Meninggalnya saksi kunci kasus Munir
5	Berita nba dan bentrokan di solo	Sutan Bhatoegana kunjungi Angie di tahanan

Pada tanggal 3 dan 4 Mei hasil ekstraksi topik dengan menggunakan NMF terjadi kesalahan interpretasi pada topik yang bukan topik utama, hal ini disebabkan oleh tidak terlalu dominan kata-kata yang mendukung topik tersebut sehingga terjadi penggabungan topik yang mengakibatkan salah interpretasi. Secara umum terdapat paling tidak satu kesamaan topik hasil interpretasi topik menggunakan NMF dengan topik hasil ekstraksi secara manual.

BAB 5

KESIMPULAN DAN SARAN

Bagian akhir tugas ini berisikan kesimpulan dari proses penelitian yang telah penulis laksanakan dan saran untuk perbaikan penelitian lanjutan. Proses penelitian pada skripsi ini terdiri dari tinjauan pustaka terhadap teori-teori yang berkaitan dengan penelitian, kemudian menyiapkan data yang akan digunakan dalam penelitian. Data berupa file RSS dari portal berita *online* berbahasa Indonesia dikumpulkan selama satu bulan, kemudian ditransformasi menjadi matriks kata dokumen. Dari matriks kata dokumen ini dilakukan faktorisasi matriks dengan NMF, selanjutnya matriks hasil faktorisasi ini diinterpretasikan menjadi sebuah topik.

Pada penelitian ini ada beberapa *file* yang berhasil didapatkan, diantaranya :

1. Kumpulan *file-file* RSS dari beberapa portal berita Indonesia online selama bulan Mei 2012.
2. *Source code* untuk mengekstraksi topik berita dari kumpulan *file-file* RSS yang telah dikumpulkan.

5.1 Kesimpulan

Dari simulasi yang telah dilakukan oleh penulis, didapatkan beberapa kesimpulan, yaitu:

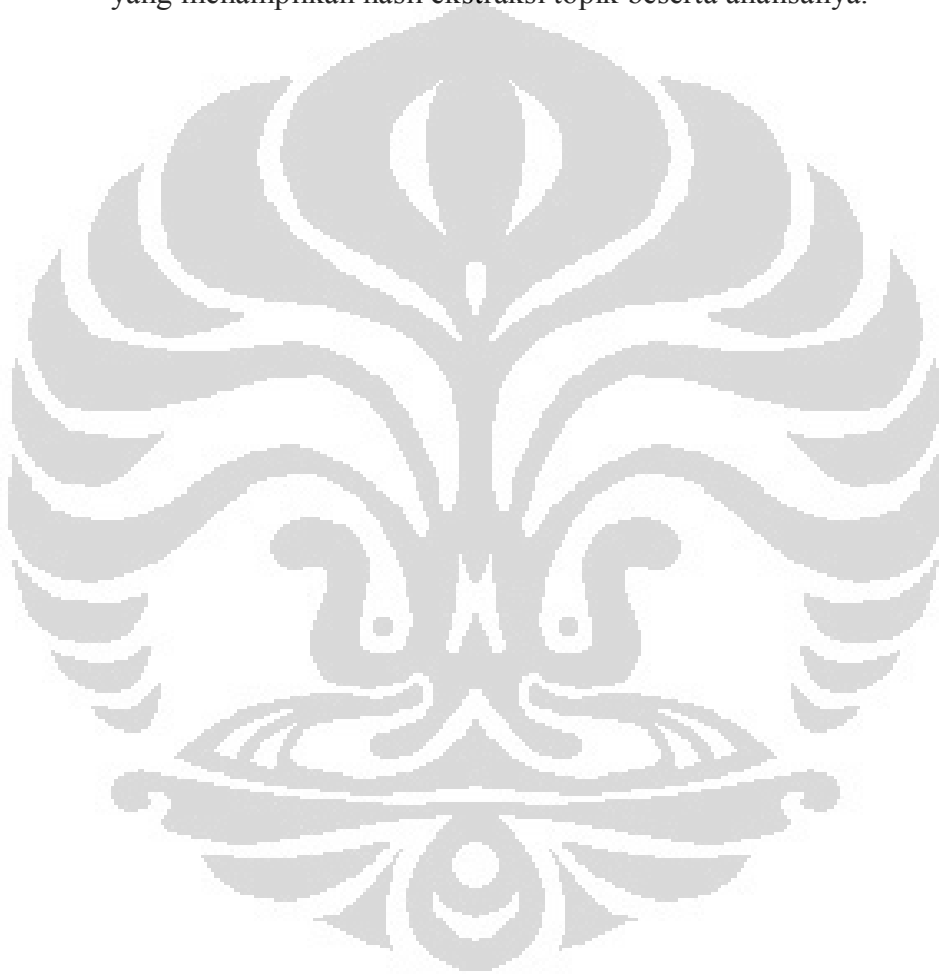
1. Metode *Latent Semantic Analysis* dengan menggunakan *Nonnegative Matrix factorization* dapat mengekstraksi kata-kata pendukung topik yang kemudian bisa diinterpretasikan menjadi sebuah topik.
2. Metode ini dapat menggambarkan tren dari topik berita pada portal berita *online* berbahasa Indonesia.

5.2 Saran

Setelah melakukan penelitian ini, ada beberapa hal yang dirasa perlu oleh penulis untuk dilakukan perbaikan pada masa mendatang:

1. Pengumpulan *file* RSS dapat dilakukan secara menyeluruh dalam satu hari.

2. Penggunaan metode lain yang dapat mengekstraksi topik utama harian dari portal berita online berbahasa Indonesia.
3. Membentuk aturan baku mengenai interpretasi kalimat topik dari kata-kata pendukung topik hasil ekstraksi.
4. Menambahkan analisa lanjutan dari kata-kata pendukung topik hasil ekstraksi.
5. Merancang program secara *online*, sehingga dapat membuat sebuah web yang menampilkan hasil ekstraksi topik beserta analisisnya.



DAFTAR PUSTAKA

- Berry, M. W., Browne, M., Langville, A. N., & Plemmons, R. J. (2007). Algorithm and applications for approximate nonnegative matrix factorization. *Computational Statistics and Data Analysis*, 15(1): 155-173.
- Laundauer, T. K., Dumais, S. T. (1997). A solution to plato's problem : The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, Vol. 104, No. 2, 211-240.
- Lee, D. D. & Seung, H. S. (1999). Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788-791
- Lee, D. D. & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *In Advances in Neural Information Processing System*, pages 556-562.
- Murfi, Hendri. (2010). *Machine Learning for Text Indexing*. Disertasi. Berlin: Von der Fakultat IV -- Elektrotechnik und Informatik der Technischen Universität.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.
- Pusat Bahasa Departemen Pendidikan Nasional (2008). *Kamus Bahasa Indonesia*. Jakarta: Departemen Pendidikan Nasional.
- Segaran, Toby. (2007). *Programming Collective Intelligence*. Sebastopol, CA: O'Reilly Media, Inc.

LAMPIRAN

Listing modul newsfeatures

Berikut adalah *listing* modul *newsfeatures.py* yang akan mengekstraksi kamus kata dari file RSS. Didalamnya juga terdapat fungsi untuk membentuk kata dokumen dan yang terakhir fungsi untuk menyimpan kata-kata dengan bobot terbesar pada masing-masing topik pada file txt.

```
import feedparser
import re
from numpy import *

feedlist = ['file:///D:/Skripsi/Data/Bulan Mei/26/detik.rss',
            'file:///D:/Skripsi/Data/Bulan Mei/26/republika.rss',
            'file:///D:/Skripsi/Data/Bulan Mei/26/mi.rss',
            'file:///D:/Skripsi/Data/Bulan Mei/26/viva.rss',
            'file:///D:/Skripsi/Data/Bulan Mei/26/kompas.rss',
            'file:///D:/Skripsi/Data/Bulan Mei/26/antara.rss',
            'file:///D:/Skripsi/Data/Bulan Mei/26/okezone.rss']

# Menghapus gambar dan markup dari artikel
def stripHTML(h):
    p=''
    s=0
    for c in h:
        if c=='<': s=1
        elif c=='>':
            s=0
            p+=' '
        elif s==0: p+=c
    return p

# Memisahkan kata
def separatewords(text):
    splitter=re.compile('\W*')
    return [s.lower() for s in splitter.split(text) if
len(s)>1]

# Mengambil kata dari artikel dan memisahkannya
def getarticlewords():
    allwords={}
    articlewords=[]
    articletitles=[]
    ec=0
    # Menyaring kata-kata yang tidak relevan
    filterwords=['saya', 'lebih', 'menjadi', 'di', 'dan', 'tidak',
```

```

        'ke','sudah','ini','itu','tak','bisa','saat',
'masih','belum','yang','akan','dari','dengan',
        'untuk','dalam','micom','jakarta','co','id']
fwords=[]
# Loop untuk setiap portal
for feed in feedlist:
    f=feedparser.parse(feed)

    #loop untuk setiap artikel
    for e in f.entries:

        # mengabaikan artikel yang identik
        if e.title in articletitles: continue

        # ekstraksi kata
        txt=e.title.encode('utf8')+'
'+stripHTML(e.description.encode('utf8'))
        words=separatewords(txt)
        articlewords.append({} and '/n')
        articletitles.append(e.title)

        # filtering kata
        for word in words :
            if word in filterwords : continue
            fwords.append(word)

        # menghitung jumlah kata, allwords (semua
kata pada portal), articlewords (semua kata pada artikel)
        for words in fwords:
            allwords.setdefault(words,0)
            allwords[words]+=1
        articlewords[ec].setdefault(words,0)
            articlewords[ec][words]+=1
            ec+=1
    return allwords,articlewords,articletitles

def makematrix(allw,artlew):
    wordvec=[]
    # mengambil kata yang diduga sebagai topik
    for w,c in allw.items():
        if c>3 :
            wordvec.append(w)
    # membentuk matriks kata
    l1=[[ (word in f and f[word] or 0) for word in wordvec] for
f in artlew]
    return l1, wordvec

```



```

def showfeatures(w,h,titles,wordvec,out='26(100).txt'):
    outfile=file(out,'w')
    pc,wc=shape(h)
    toppatterns=[] for i in range(len(titles))
    patternnames=[]

    for i in range(pc):
        slist=[]

        for j in range(wc):
            slist.append((h[i,j],wordvec[j]))

        slist.sort()
        slist.reverse()

        n=[s[1] for s in slist[0:10]]
        m=[s[0] for s in slist[0:10]]
        outfile.write(str(m)+'\n'+str(n)+'\n')
        patternnames.append(n)

        flist=[]
        for j in range(len(titles)):
            flist.append((w[j,i],i,titles[j]))
        toppatterns[j].append((w[j,i],i,titles[j]))

        flist.sort()
        flist.reverse()

        for f in flist[0:5]:
            outfile.write(str(f)+'\n')
            outfile.write('\n')

    outfile.close()

    return toppatterns,patternnames

```

***Listing* program faktorisasi matriks NMF**

Berikut *listing* program faktorisasi NMF yang dibangun berdasarkan algoritma pada table 3.2.

```

from numpy import *

def difcost(a,b):
    dif=0

```

```

for i in range(shape(a)[0]):
    for j in range(shape(a)[1]):
        # Euclidean Distance
        dif+=pow(a[i,j]-b[i,j],2)
return dif

def factorize(v,iter):
    ic=shape(v)[0]
    fc=shape(v)[1]
    pc=5

    # Initialize the weight and feature matrices with random values
    w=matrix([[random.random() for j in range(pc)] for i in
range(ic)])
    h=matrix([[random.random() for i in range(fc)] for i in
range(pc)])

    # Perform operation a maximum of iter times
    for i in range(iter):
        wh=w*h

        # Calculate the current difference
        cost=difcost(v,wh)
        if i%10==0: print cost
        # Terminate if the matrix has been fully factorized
        if cost==0: break

        # Update feature matrix
        hn=(transpose(w)*v)
        hd=(transpose(w)*w*h)+0.000000001
        h=matrix(array(h)*array(hn)/array(hd))

        # Update weights matrix
        wn=(v*transpose(h))
        wd=(w*h*transpose(h))+0.000000001
        w=matrix(array(w)*array(wn)/array(wd))
    return w,h

```

Listing program utama , file ‘main.py’.

Berikut listing program utama yang akan dieksekusi untuk mendapatkan hasil ekstraksi topik utama dari kumpulan file RSS dengan menggunakan NMF.

```
import newsfeatures
import nmf
from numpy import *

allw,artw,artt=newsfeatures.getarticlewords ()
wordmatrix,wordvecs= newsfeatures.makematrix(allw,artw)
v=matrix(wordmatrix)
w,f=nmf.factorize(v,iter=100)
topp,pn=newsfeatures.showfeatures(w,f,artt,wordvecs)
```

Dokumentasi Hasil

Berikut dokumentasi hasil simulasi ekstraksi topik utama harian portal berita *online* berbahasa Indonesia dengan menggunakan NMF.

1 Mei 2012

Topik 1

[31.58479613524273, 12.713074362926994, 8.3796642834101238, 6.3276948239847197, 6.2040788227243509, 6.16946327772036, 5.3965004099054594, 5.1382795459469914, 5.0779801602171073, 4.6568674126048464]

['buruh', 'hari', 'indonesia', 'selasa', 'istana', 'depan', 'kpk', 'aksi', 'nasional', 'polri']

Topik 2

[8.7113797291834132, 7.7845381666227063, 5.9585153090954961, 5.8984419241766908, 5.8240537029058874, 5.2326047306737919, 4.8460470195941552, 4.6325882240141016, 4.5764402876226695, 4.4918237514653665]

['buruh', 'kpk', 'dpr', 'bus', 'ketua', 'partai', 'jalan', 'presiden', '2012', 'istana']

Topik 3

[11.453988239037907, 7.7635960666811705, 6.7503474343143948, 6.0445767657495173, 5.4323033956934266, 5.3941503566760156, 4.8533530657147086, 4.7814717447947563, 4.43731321041402, 4.3658760300732]

['buruh', 'dpr', 'jalan', 'pendemo', 'istana', 'bus', 'massa', 'hari', 'sekitar', 'ketua']

Topik 4

[15.480723636425024, 6.0242384704035583, 5.9772814826252567, 5.3801813221487027, 4.5948013847341382, 4.307118983663182, 4.085510108487262, 3.7231575144258833, 3.7136514768378479, 3.6869430139736359]

['buruh', 'hari', 'indonesia', 'kpk', 'istana', 'jalan', 'tewas', 'presiden', 'polri', 'depan']

Topik 5

[28.526389494180847, 10.753439365696277, 6.2908735771352919, 5.1367456942942642, 5.0520381129206378, 4.9632165943093929, 4.8052855921879809, 4.7761553152536367, 4.7306804526799651, 4.6730760028067673]

['buruh', 'hari', 'selasa', 'may', 'istana', 'indonesia', 'pekerja', 'day', 'pada', 'serikat']

2 Mei 2012

Topik 1

[5.9854700934739826, 5.53254198402254, 4.5816206424073087, 4.3440857033361544, 4.1929411676124371, 3.7593946128424767, 3.4141141604436935, 3.3086489966605743, 3.2855082905975785, 3.1736618425724674]

['indonesia', 'kpk', 'menkes', 'kesehatan', 'kasus', 'endang', 'buruh', 'ktp', 'sedyaningsih', 'menteri']

Topik 2

[9.2462014050576204, 6.5288303955691998, 5.7940280530102353, 5.0886937634873162, 4.6815028927967424, 4.6196569126152447, 4.539192643677711, 3.9823670396889348, 3.7493868919821249, 3.5285765809351206]

['kpk', 'tni', 'anggota', 'dpr', 'ad', 'indonesia', 'neneng', 'kasus', 'palmerah', 'restoran']

Topik 3

[8.8394153150109673, 7.3599880126853288, 7.1378923138124648, 6.3382240634737679, 5.7342993902138142, 5.4734822213284566, 5.418519190972467, 5.2893972640867828, 4.8483771178346995, 4.7905645385450395]

['anas', 'kasus', 'indonesia', 'kpk', 'ketua', 'partai', 'buruh', 'kesehatan', 'demokrat', 'korupsi']

Topik 4

[8.9692977418179769, 6.4326023037737325, 6.0433806189414936, 5.021351939019894, 4.8593006027186627, 4.6528635761976886, 4.2926922718562359, 4.1162185725461713, 4.0278588080515787, 3.9540429580360463]

['kpk', 'menkes', 'kasus', 'anas', 'endang', 'nazaruddin', 'buruh', 'korupsi', 'angie', 'surat']

Topik 5

[6.9116743906918421, 5.0717542521870014, 4.8751014474894117, 3.9102430259677088, 3.7381783391185923, 3.0423284530488441, 2.8781843951057215, 2.8291667534768656, 2.8233369784122577, 2.6591220799491984]

['kpk', 'kasus', 'republika', 'korupsi', 'anggota', 'barat', 'ketua', 'dpr', 'partai', 'tni']

3 Mei 2012

Topik 1

[8.9339582522841905, 8.2906372349145183, 5.8442397541445885, 5.5734723748929387, 4.949988835472392, 4.9303397469282961, 4.7979446672065862, 4.6570623748268849, 4.5783918451095413, 4.2535966703554262]

['kpk', 'ketua', 'sby', 'api', 'emas', 'kasus', 'iswahyudi', 'sebagai', 'pemotor', 'patek']

Topik 2

[14.280815854507296, 13.91525759493244, 10.732723967916858, 8.8565725757735958, 8.6961090639311909, 7.4519453556306443, 7.15890884256613, 6.3164298668164776, 5.7456016009841031, 5.3335472162836179]

['partai', 'kpk', 'ketua', 'nasdem', 'endang', 'umum', 'anas', 'demokrat', 'kasus', 'rahayu']

Topik 3

[8.7791468584190184, 7.6322730424019332, 7.4059846094176631, 7.2358791822379152, 7.2211965533011364, 6.2223377202842709, 6.0168060796862068, 5.6138424504891455, 5.2670502906559342, 5.1666880139887272]

['kasus', 'korupsi', 'endang', 'tersangka', 'dpr', 'ketua', 'anas', 'partai', 'dugaan', 'suap']

Topik 4

15.063460894923045, 10.414053651882353, 10.210001346018249, 7.0349029382819568, 6.7488352166725978, 6.2064707444357143, 6.0348909210237212, 5.4139671395403477, 4.9965899429957474, 4.8497425972486665]

['endang', 'kpk', 'rahayu', 'sedyaningsih', 'jenazah', 'partai', 'kesehatan', 'terakhir', 'kamis', 'hills']

Topik 5

[13.992142595802811, 8.2954971479302966, 6.6791775743614217, 4.6525724444078662, 4.4135159181045669, 4.4091094510318687, 3.8713762295506915, 3.7899341235164057, 3.6939604838230347, 3.4765884848335418]

['endang', 'kpk', 'rahayu', 'kasus', 'anis', 'sedyaningsih', 'matta', 'ketua', 'sby', 'jenazah']

4 Mei 2012

Topik 1

[9.1950550063863119, 7.5242066336504667, 7.3904105864800549, 7.2800093089302962, 6.7891042342074703, 6.4527469334550114, 6.2104242554176219, 5.894869933069927, 5.2579546131126618, 4.9500119412765127]

['penumpang', 'luka', 'bentrokan', 'karena', 'solo', '380', 'cengkareng', 'terjadi', 'kasus', 'orang']

Topik 2

[13.6880686429755, 10.4375417752007, 5.8915465553426687, 4.9226087488354704, 4.9139961409229036, 4.7075294716587788, 4.3633462155524549, 4.2058637203960005, 4.1665811891183298, 4.1368846926898097]

['kasus', 'ongen', 'indonesia', 'keluarga', 'terkait', 'munir', 'nba', 'kematian', 'komisi', 'solo']

Topik 3

[11.018157450448657, 7.9518958456983473, 6.8850389914548247, 5.112995290445939, 5.011223541858997, 4.8941434258127208, 4.7235797803050046, 4.6120326814272978, 4.2176827363525575, 4.1671287083944755]

['kasus', 'ongen', 'kpk', 'angelina', 'angie', 'dhana', 'terkait', 'warga', 'solo', 'kematian']

Topik 4

[8.3578684332612934, 5.7068422447726359, 5.475024672813845, 5.4317959464121248, 5.3458204311358601, 5.1208805721404556, 4.8592031241304472, 4.6540478482201992, 4.529026113991752, 4.3035564918126008]

['kasus', 'kpk', 'dhana', 'warga', 'pada', 'jumat', 'indonesia', 'hukum', 'angelina', 'kejaksaan']

Topik 5

[9.2644428001357628, 8.5687861217785706, 8.5651591964900717, 8.0323599807606971, 7.9281295564975895, 7.9267731079136858, 5.7502248179449102, 5.7427116379632794, 5.6132774612234764, 5.302625561524092]

['nba', 'solo', 'bentrokan', '2012', 'ketua', 'kota', 'ongen', 'terkait', 'sondakh', 'indonesia']

5 Mei 2012

Topik 1

[8.1934073389561348, 7.1883293269829229, 6.2690129330380131, 5.9657935043739192, 5.6719190666225856, 5.5783088868895616, 5.2955287434914711, 4.1787350066703919, 4.1198509426730343, 4.0743023001637084]

['solo', 'masuk', 'bentrokan', 'demokrasi', 'jokowi', 'karena', 'artis', 'rs', 'kota', 'kepala']

Topik 2

[8.9270450064126319, 7.5868320168757943, 5.8234747036625754, 5.7559736261099088, 5.5107877919584247, 5.1658625933046336, 4.3261790522146137, 4.0794958344243621, 4.0327052236710159, 3.9652261443295798]

['kpk', 'kasus', 'korupsi', 'solo', 'warga', 'komisi', 'ongen', 'bentrok', 'ketua', 'diskusi']

Topik 3

[7.0403796084071102, 4.8519568503762542, 4.2626332290104818, 4.0572961807734407, 3.9953210197956546, 3.82660711096413, 3.8152053032003135, 3.7206097894882628, 3.6213119555023767, 3.4732971868424687]

['republika', 'as', 'solo', 'diskusi', 'terjadi', 'uu', 'karena', 'dinilai', 'salihara', 'kecelakaan']

Topik 4

[9.0212122090444797, 7.4919341555069279, 5.7125384922986431, 5.6540100819836399, 5.5482029797078063, 4.9627367632703567, 4.4291861895051996, 4.3608652563040158, 4.0000194226661927, 3.88349052576709]

['kpk', 'kasus', 'angelina', 'warga', 'indonesia', 'tersangka', 'jumat', 'dhana', 'korupsi', 'hukum']

Topik 5

[5.4012843542687676, 5.2197074145309799, 4.5405475857519715, 3.8927496360581455, 3.7471593984108766, 3.6448352607244781, 3.4470245790601344, 3.180397456353957, 3.0417032313885923, 3.0077496280069806]

['kpk', 'korupsi', 'komisi', 'jumat', 'pada', 'ketua', 'diskusi', 'salihara', 'besar', 'karena']

6 Mei 2012

Topik 1

[8.3541114365671127, 8.1659391652637439, 7.3903564043789753, 5.7141403935727837, 4.8357458117918526, 4.5912519232607503, 4.1797998308459201, 4.1240712791355012, 3.7881909621029419, 3.5859586319564483]

['dpr', 'senjata', 'api', 'polisi', 'diskusi', 'kasus', 'irshad', 'pada', 'komisi', 'anggota']

Topik 2

[16.987393585469476, 16.060476326067494, 14.409395574236049, 10.756942212550445, 7.1294227474006169, 5.9133446546251882, 5.7391822866649198, 5.093481833546484, 4.9561974833270543, 4.6888628511853891]

['api', 'senjata', 'dpr', 'anggota', 'didik', 'ktp', 'komisi', 'masyarakat', 'denda', 'ketua']

Topik 3

[10.83458508812301, 10.54349227881473, 10.058124654450804, 7.5563072445417596, 6.9766558759457631, 6.6358433316761518, 6.538105958677737, 5.176784422104113, 4.5950804942603822, 4.0679750636752008]

['dpr', 'api', 'senjata', 'anggota', 'ktp', 'republika', 'ketua', 'komisi', 'indonesia', 'didik']

Topik 4

[6.9888797073115656, 6.0825264770826388, 6.0711442770207542, 5.8921392919951208, 5.6809640515609141, 5.484263326959641, 5.4287679400476465, 4.8688110900384478, 4.7670508156508697, 4.7087776051664507]

['kpk', 'indonesia', 'diskusi', 'irshad', 'negara', 'kpk', 'kasus', 'manji', 'partai', 'sebagai', 'ketua']

Topik 5

[16.306073702238734, 16.033126053496623, 15.156650405843052, 9.0583533508196439, 7.1697347368241395, 6.6276441409993714, 6.034624493321572, 5.9564288299321735, 5.7047811358521487, 5.6394213616030306]

['irshad', 'manji', 'diskusi', 'buku', 'malam', 'selatan', 'pembubaran', 'salihara', 'indonesia']

7 Mei 2012

Topik 1

[[11.385823362509425, 11.123222357310857, 8.5847925416643562, 5.5480761714570184, 4.9619016332469892, 4.6472453501966671, 4.0903412043125096, 4.0039227431794409, 3.9602742277474374, 3.7029532222303074]

['manji', 'irshad', 'diskusi', 'buku', 'kpk', 'ada', 'malam', 'golkar', 'senjata', 'solo']

Topik 2

[7.0673001236054667, 6.8710382252238125, 6.7600451347633417, 6.18299560642531, 5.2588358469714844, 5.2434337066429082, 5.1098587459276406, 4.9481364519221778, 4.7540344542601245, 4.6542068231000178]

['indonesia', 'manji', 'irshad', 'kpk', 'sebagai', 'selatan', 'diskusi', 'partai', 'korupsi', 'kasus']

Topik 3

[10.828575471874373, 8.1398739799775672, 6.2081053862670554, 5.7830744333120521, 5.7448760937877665, 5.1669472933552134, 4.5124283164268952, 4.2709876524321757, 3.9051101972776081, 3.8454631196746836]

['senjata', 'api', 'dpr', 'foke', 'polisi', 'anggota', 'un', 'siswa', 'genangan', 'ada']

Topik 4

[7.1948878929172428, 5.7070694828165829, 4.6969748144930827, 4.170308831460491, 4.1659419034384921, 4.0466003236286783, 3.9574073674543193, 3.4998934105696735, 3.1608049609007107, 3.0613029940480025]

['kpk', 'api', 'senjata', 'neneng', 'masyarakat', 'korupsi', 'dpr', 'solo', 'polisi', 'nazaruddin']

Topik 5

[6.4706594398412625, 5.7265111861834725, 5.1661991392191853, 4.0063983942110717, 3.972366348085262, 3.9577858997895046, 3.7580284497189087, 3.6280963541066726, 3.6221021821956674, 3.5198080584360811]

['republika', 'senjata', 'kpk', 'taman', 'lalu', 'api', 'dua', '2012', 'siswa', 'ada']

8 Mei 2012

Topik 1

[12.198135085493082, 10.351710872580043, 7.5088986881345026, 6.9751941178591013, 5.9508607420946662, 5.1281993658892908, 4.9969858076664595, 4.9522970116130036, 4.8405453833578536, 4.6730414984757429]

['kasus', 'dpr', 'yulianis', 'kpk', 'anggota', 'oleh', 'daerah', 'tersangka', 'sebagai', 'pt']

Topik 2

[11.395620191237304, 9.2310374989492328, 8.1554140516608911, 7.9652306781170124, 7.795146644452565, 6.6466630625454162, 6.3751757064172319, 5.1125276112320774, 5.0726381194053385, 4.8754014455049273]

['kpk', 'keributan', 'salemba', 'eleven', 'kasus', 'dpr', 'anggota', 'terkait', 'pengunjung', 'reformasi']

Topik 3

[6.8076824752828875, 6.4450864782257193, 5.8293460424878907, 5.4467953485897693, 4.2252505185015172, 3.7103349007088227, 3.3844721812474448, 3.191115356631002, 3.1851458052171138, 3.1783629356547078]

['dpr', 'kpk', 'anggota', 'republika', 'kasus', 'saksi', 'terkait', 'ketua', 'dewan', 'keributan']

Topik 4

[11.642778685356392, 8.415598548234561, 6.5195739345068509, 6.4635948704805379, 6.1634750028231835, 5.0782359446726666, 4.9965720083983749, 4.8960847412189796, 4.8124339823516564, 4.3852629024249286]

['kpk', 'api', 'senjata', 'neneng', 'masyarakat', 'korupsi', 'dpr', 'solo', 'polisi', 'nazaruddin']

Topik 5

[6.4706594398412625, 5.7265111861834725, 5.1661991392191853, 4.0063983942110717, 3.972366348085262, 3.9577858997895046, 3.7580284497189087, 3.6280963541066726, 3.6221021821956674, 3.5198080584360811]

['kpk', 'kasus', 'dpr', 'ketua', 'komisi', 'korupsi', 'terkait', 'tersangka', 'yulianis', 'saksi']

9 Mei 2012

Topik 1

[14.134112447984975, 8.5297181108182478, 7.1650492147824201, 6.7545390679134245, 5.6127181928753069, 5.3311617442705099, 4.7068606286675321, 4.5639193631068125, 4.1011267614764275, 4.094002068107665]

['gorong', 'ugm', 'dpr', 'marzuki', 'besi', 'diskusi', 'penutup', 'anggota', 'manji', 'irshad']

Topik 2

[10.819788436265519, 7.5696230508530604, 6.9837140981661889, 6.7662541164160261, 6.5438627601344574, 6.4998328981491325, 5.9578363195209505, 5.0895137332246705, 4.9585061814578788, 4.9558780480714146]

['nunun', 'nbsp', 'gorong', 'republika', 'kasus', 'sidang', 'rabu', 'vonis', 'mantan', 'ugm']

Topik 3

[10.215975175436219, 8.5448919669900505, 7.1543143370696924, 6.6893901730795475, 6.3514671952374195, 6.1820236397858528, 5.3821856613019694, 5.3502789739075585, 5.2093875101358362, 4.9388941439086302]

['dpr', 'golkar', 'partai', 'indonesia', 'nunun', 'pada', 'akbar', 'rapat', 'ketua', 'kpk', 'polri']

Topik 4

[12.25024331665116, 8.5812567713571912, 7.1141764728041137, 6.6421486128601241, 6.1126390924147449, 5.9700531788462499, 5.2657536559880622, 4.6968372642953096, 4.5591462944562382, 4.5451746532951018]

['kpk', 'kasus', 'nunun', 'partai', 'tersangka', 'anggota', 'korupsi', 'izin', 'senjata', 'komisi']

Topik 5

[11.904833445373303, 9.5676611741621489, 8.5105175922207152, 8.4670602376764972, 5.0944293222928305, 4.9104126609980874, 4.8836703331575615, 4.8469880807744161, 4.4972221057510353, 4.4437924856116373]

['golkar', 'nunun', 'partai', 'kpk', 'sebagai', 'dpp', 'polri', 'ketua', 'korupsi', 'dewan']

10 Mei 2012

Topik 1

[8.1544800614040174, 7.8108649531104444, 7.1478411739798444, 6.4605376845021567, 5.7809141501628396, 5.7229200134045737, 5.7227255198412061, 5.6887833143291653, 5.5372330444336439, 5.5047583451798126]

['tiga', 'golkar', 'indonesia', 'rabu', 'kpk', 'wartawan', 'malaysia', 'salak', 'gunung', 'republika', 'menteri']

Topik 2

[34.628070749974107, 16.012735174182211, 12.966802367048906, 10.038208867532864, 9.050782061505668, 7.2258871395015731, 6.5688071088534823, 6.2982403724687099, 6.1908672344645232, 5.6584627620912524]

['sukhoi', 'pesawat', 'lokasi', 'tim', 'korban', '100', 'sar', 'gunung', 'jatuhnya', 'evakuasi']

Topik 3

[13.490062447621984, 7.8131411869829117, 7.793495445297757, 6.2750925183654305, 5.2316959396634264, 5.2299049908082402, 4.693189565143479, 4.4058946946982269, 4.1695109849917573, 4.1244728108372479]

['pesawat', 'presiden', 'sukhoi', 'kpk', 'gunung', 'indonesia', 'rabu', 'terkait', 'tebing', 'dewan']

Topik 4

[25.645586886549417, 11.301102505958257, 11.096989940963631, 10.662658358678099, 10.396908591453498, 8.8794744544879798, 7.3335608006288222, 7.2105367105815699, 5.9276993435238223, 5.7585856870381562]

['sukhoi', 'pesawat', 'korban', 'superjet', '100', 'tim', 'indonesia', 'penumpang', 'keluarga', 'halim']

Topik 5

[16.894106339929863, 15.058049544137669, 8.0522392205178619, 7.3495243692719407, 7.2594478536166616, 5.7757247001828462, 5.3687182353522545, 5.0207994382029604, 4.4923967913389342, 4.4794239593064633]

['sukhoi', 'pesawat', 'gunung', '100', 'salak', 'superjet', 'jatuhnya', 'lokasi', 'indonesia', 'kecelakaan']

11 Mei 2012

Topik 1

[13.889257007499042, 7.3862881274109347, 7.2594270990543599, 6.5775766989111171, 6.2850446443124373, 6.018408437282095, 5.9986117005013204, 5.8033468191202982, 5.5376655017628806, 5.3889222258640341]

['sukhoi', 'salemba', 'tim', 'bali', '100', 'sar', 'korban', 'superjet', 'aksi', 'uu']

Topik 2

[25.515165400338113, 12.992022317406485, 10.966505434520901, 8.4924259742671655, 7.2755100606642467, 5.9333077450705192, 5.2158158231705594, 4.9710903575983894, 4.8708508262954995, 3.7680227625899101]

['sukhoi', 'korban', 'pesawat', '100', 'superjet', 'tim', 'sar', 'salak', 'gunung', 'dua']

Topik 3

[22.338237842351059, 15.193588032257871, 14.067049047343962, 8.175535213468244, 7.7890571641975601, 7.5708201040078142, 6.9420237647483987, 5.7742696168913996, 5.5035532422424689, 5.3946782746124766]

['sukhoi', 'korban', 'presiden', 'kpk', 'tim', 'pesawat', 'rusia', '100', 'tersangka', 'indonesia']

Topik 4

[23.574365034865831, 13.123913262843358, 12.204200046342512, 10.291971562347632, 8.5913233913724874, 7.0873899889858443, 5.9278655940361764, 5.4431296537546778, 5.3438214668688442, 4.2608304375293224]

['sukhoi', 'presiden', 'pesawat', 'korban', '100', 'superjet', 'rusia', 'putin', 'tim', 'yudhoyono']

Topik 5

[18.947232932407616, 11.478805471398251, 10.328018631709007, 9.3457140259727662, 8.4304904656465887, 7.7975696311109122, 6.5475501139510319, 5.4098456928499123, 5.2569665130337464, 3.7184885982946114]

['sukhoi', 'korban', 'pesawat', '100', 'presiden', 'superjet', 'gunung', 'keluarga', 'salak', 'tni']

12 Mei 2012

Topik 1

[21.273784894055289, 16.064867215075086, 14.023797299869996, 12.600318670856865, 11.975902257759921, 11.270159815826823, 10.003913723264386, 9.4116303891459587, 7.8989099033433963, 7.2759163916524807]

['sukhoi', 'korban', 'pesawat', '100', 'jenazah', 'superjet', 'kecelakaan', 'evakuasi', 'kantong']

Topik 2

[24.201605288775866, 16.882889037682041, 13.664917838088984, 8.8784382852717272, 8.7007299385052654, 8.2861445145025492, 6.2234946959049147, 5.9361381991539073, 5.6760888090303983, 4.9969577845348541]

['sukhoi', 'korban', '100', 'superjet', 'pesawat', 'indonesia', 'identifikasi', 'pagi', 'tim']

Topik 3

[14.504311670144505, 13.496280784874493, 9.8813114996406437, 7.9688928989404637, 7.6000407731181943, 7.3328067589083057, 7.1287170867288276, 6.0381682266092698, 5.7934009608693637, 5.681491553260330]

['sukhoi', '100', 'korban', 'pesawat', 'ketua', 'superjet', 'dki', 'kecelakaan', 'republika', 'evakuasi']

Topik 4

[26.190465453997483, 22.958635841353615, 19.982362687282222, 14.091502124889255, 11.840282923849461, 10.16908892140466, 9.1524670212850392, 8.0698821561421319, 7.9332757123788467, 7.5618128588309421]

['jenazah', 'korban', 'sukhoi', 'pesawat', 'halim', 'evakuasi', 'dua', 'kantong', 'polri', 'rs']

Topik 5

[22.592435590118008, 15.657093043875184, 11.552058701871669, 11.090041263225102, 10.266865457814072, 8.519723465603585, 6.77036610899337, 6.0206514229466226, 5.8561454702273448, 5.7922053890863801]

['jenazah', 'korban', 'sukhoi', 'halim', 'pesawat', 'evakuasi', 'dua', 'polri', 'rs', 'kantong']

13 Mei 2012

Topik 1

[11.730125725145299, 9.9139268267324674, 9.281257424243325, 8.0669081216380807, 7.9498962796951753, 7.1680811134703246, 7.0497431748894925, 6.5193107908479426, 6.3503552992631018, 5.9933393180379317]

['sukhoi', 'tim', 'korban', 'superjet', '100', 'telah', 'presiden', 'jenazah', 'pesawat', 'republika']

Topik 2

[14.111147741445496, 10.718337414730305, 10.156690439652266, 9.3294185501703417, 9.3188090639760635, 7.5286113761418862, 7.0215376959639118, 7.013159819085697, 6.3741840773444718, 6.0008912651400017]

['sukhoi', 'tim', 'korban', '100', 'superjet', 'telah', 'salak', 'gunung', 'rusia', 'jenazah']

Topik 3

[17.018383539566404, 12.644416315508428, 9.243135758202504, 7.5525951978087065, 5.6099574495543258, 5.5707615726590127, 5.5245686529067726, 5.3166572784206219, 5.3138578489945925, 5.1395874458342536]

['sukhoi', 'indonesia', 'presiden', 'evakuasi', 'kpk', 'rusia', 'partai', 'ketua', 'komisi', 'korban']

Topik 4

[24.399556452642486, 16.97613285217534, 9.1328976313503372, 7.8347510790131256, 6.8894404612689542, 6.4464644038136676, 6.1481699138553134, 5.9552590090397279, 5.5915764391087253, 5.4785671750015359]

['sukhoi', 'korban', 'indonesia', 'identifikasi', 'rusia', 'presiden', 'pesawat', '100', 'evakuasi']

Topik 5

[11.345738208268267, 9.3905879558183987, 6.2556563989754572, 6.1864879778003736, 5.5426317539115022, 5.4064382936846282, 4.9574440193894, 4.3974723836454759, 4.0094743723498203, 3.9938571404700913]

['sukhoi', 'korban', 'superjet', '100', 'tim', 'pesawat', 'jenazah', 'presiden', 'indonesia', 'rusia']

14 Mei 2012

Topik 1

[17.293876261797021, 10.656615477624822, 8.4901689371505729, 7.024680475071774, 6.3488832161261186, 5.9591046081230763, 5.956436152913211, 5.5738231456695315, 5.5190162020025539, 5.1298058038201768]

['sukhoi', 'korban', 'evakuasi', 'presiden', 'dpr', '100', 'superjet', 'gunung', 'salak', 'pada']

Topik 2

[8.6428198010188453, 8.6395136134882051, 6.9133742548296047, 6.3092153173504721, 5.15614639407708, 5.0424995574084965, 4.7439203147322253, 4.5513981323281429, 4.4136177877350162, 4.3479343597755813]

['sukhoi', 'dpr', 'presiden', 'evakuasi', 'partai', 'dewan', 'rusia', 'masa', 'rapat', 'golkar']

Topik 3

[11.244924363184865, 8.5473217406709967, 8.3164911034904936, 6.4773722457704848, 5.9754754054729338, 5.4155214644323335, 5.3449859054538624, 4.7446415515391989, 4.4181933289297293, 4.008242805149039]

['indonesia', 'kpk', 'presiden', 'golkar', 'sukhoi', 'partai', 'sebagai', 'rusia', 'kasus', 'politik']

Topik 4

[8.8098880903669272, 7.6699117078352677, 6.4527958017497706, 6.3301605562586047, 6.1942182600526623, 5.8281459755122134, 5.7761928228067259, 5.612171123144746, 4.9852128859306974, 4.5377502461515329]

['dpr', 'anggota', 'republika', 'presiden', 'merpati', 'gunung', 'partai', 'golkar', 'masa', 'kasus']

Topik 5

[14.75390699658421, 9.8198088483941302, 6.1993416131134929, 5.6611359038044178, 4.9285930641176847, 4.6221268763114081, 4.5392871995052735, 4.5153623281221762, 4.2514533167960407, 4.2192277276712815]

['dpr', 'anggota', 'bk', 'tiang', 'masa', 'video', 'dki', 'pilkada', 'sidang', 'terkelupas', 'porno']

15 Mei 2012

Topik 1

[10.559585549022154, 9.3879641628078847, 7.4973949926502454, 6.2258196877730665, 6.1764329816880572, 5.5505355631853535, 5.3128310965830643, 5.0615495173094276, 4.8292993382513023, 4.5576767264332814]

['indonesia', 'dpr', 'sukhoi', 'anggota', 'golkar', 'kasus', 'agung', 'partai', 'kpk', 'komisi']

Topik 2

[12.379797603838639, 4.8788025962303898, 4.1379232028387412, 4.0399793637300609, 3.9260305648086651, 3.9226602669310502, 3.8969939753903327, 3.8623355770818959, 3.758869912601722, 3.637153022726954]

['sukhoi', 'kpk', 'lokasi', 'kasus', 'kecelakaan', 'indonesia', '100', 'pesawat', 'ada', 'komisi']

Topik 3

[8.2458680461215312, 6.737390873606893, 6.7069712012110996, 6.1196552771821517, 5.3719187064566132, 5.0899813949531971, 4.492748541940168, 4.3985152592701908, 4.0479707044680993, 3.8942809901899684]

['konser', 'gaga', 'lady', 'sukhoi', 'sby', 'polri', 'indonesia', 'presiden', 'pd', 'rusia', 'pertemuan']

Topik 4

[14.486462585376129, 7.0540430680402793, 6.6923846315690794, 6.3372897670367845, 5.8412078780267533, 5.6676320900967552, 5.639423483454121, 5.3964239102760745, 5.1589638012185697, 4.954220425952701]

['sukhoi', 'presiden', 'korban', 'selasa', '100', 'indonesia', 'kecelakaan', 'tim', 'kpk', 'sar']

Topik 5

[6.6168808516006488, 6.5760463830231117, 6.1347623222949821, 5.8916553013943833, 5.420538086136216, 5.3973266114885057, 5.263846792015495, 5.101137557409916, 5.0230320951418719, 4.69785041847011]

['republika', 'konser', 'lady', 'gaga', 'shelby', 'presiden', 'komisi', 'partai', 'polri', 'uang']

16 Mei 2012

Topik 1

[7.1831210963026821, 7.1344661396562552, 5.1707021253424985, 5.076500448357641, 4.9858930515269639, 4.9420458530184463, 4.9098640304501746, 4.2225023286757004, 4.007375446574259, 3.9858205959478976]

['black', 'box', 'karena', 'pks', 'artis', 'kpk', 'orang', 'kursi', 'partai', 'knkt', 'pesawat']

Topik 2

[11.142324319803382, 9.1956732083188051, 8.8843636675575937, 4.7319774456499957, 4.6799113457576151, 4.6604433571759509, 4.5181136967144075, 4.4859919280556868, 4.0806409173151463, 4.0703689225346471]

[hitam', 'sukhoi', 'kotak', 'ditemukan', 'partai', 'presiden', 'kpk', 'knkt', 'box', 'tim', 'ada']

Topik 3

[6.1603224549489113, 6.0924200140952074, 5.5941842755304743, 4.6869301917454393, 4.5122589390871388, 4.4504523909952045, 4.4109326861301339, 4.395009036649479, 3.7839143264910233, 3.6504368787264343]

[hitam', 'sukhoi', 'presiden', 'kotak', 'partai', 'nasional', 'pada', 'kpk', 'rabu', 'satu', '100']

Topik 4

[19.012852531649877, 12.448117928085054, 10.699225468156133, 7.5924324222401411, 7.5912163416865788, 6.6438676501760385, 6.1464339880655992, 5.7956597095442843, 5.2754522788811489, 4.5215135155579853]

[sukhoi', 'hitam', 'kotak', '100', 'ditemukan', 'pesawat', 'black', 'tim', 'superjet', 'box']

Topik 5

[8.478688122560616, 8.5099925511739531, 7.9305966472844229, 7.5945187035649706, 7.1230453487340082, 6.8581453227005387, 6.2483581431368203, 5.6037427958009154, 5.5121727164735361, 5.4629955417068699]

[presiden', 'partai', 'komisi', 'kpk', 'yudhoyono', 'korupsi', 'kotak', 'knkt', 'ketua', 'demokrat']

17 Mei 2012

Topik 1

[6.3674148905572894, 6.3346181376251538, 5.8047952461070835, 5.8016503899897574, 5.0578669777661194, 4.9762190829044153, 4.9195010030153759, 4.7678840050295932, 4.5374916647956702, 4.0796091294864025]

[pada', 'sukhoi', 'pesawat', 'presiden', 'rabu', 'indonesia', 'kpk', 'tahun', 'ketua', 'korban']

Topik 2

[5.8575091277606548, 5.8445296341872588, 5.4320019240638846, 5.3599915248725214, 5.1633424276166648, 5.0698079732718142, 4.9158581411137465, 4.869709048025963, 4.8361581579840909, 3.8528956900756053]

[box', 'black', 'kpk', 'pks', 'artis', 'karena', 'partai', 'kursi', 'orang', 'golkar', 'sukhoi']

Topik 3

[8.4691031659638565, 7.6199520009576966, 6.0400941749003936, 5.782263120923032, 5.2206034820944884, 5.2022522505999875, 5.0577385380255446, 4.2719900066068028, 4.1229895000213457, 3.835885021528556]

[sukhoi', 'kpk', 'rekening', 'partai', 'korupsi', 'kasus', 'polri', 'angie', 'presiden', 'korban']

Topik 4

[7.2137915353376938, 6.6296504034042893, 5.5849577934899202, 4.8562935428531402, 4.6847187342571175, 4.250213271452675, 3.9236104970784629, 3.8808972589483957, 3.8471764619653368, 3.7312076608966094]

['partai', 'kpk', 'sukhoi', 'artis', 'calon', 'komisi', 'korupsi', 'pesawat', 'harus', 'sebagai']

Topik 5

[8.2943398971766378, 7.4363284300000574, 6.7041899847239268, 6.5690553593833867, 6.445465981152541, 6.0662593847436872, 5.4173793618833042, 5.223540497064354, 4.5999695464785146, 3.9552060477743747]

['sukhoi', 'nbsp', 'republika', 'kpk', '100', 'pesawat', 'black', 'box', 'superjet', 'korban']

18 Mei 2012

Topik 1

[14.753193102187256, 11.548357691991972, 9.0934655138687699, 8.2775435572858758, 7.7360488339524318, 7.1543258871516224, 5.8483855596129883, 5.7020251811045082, 5.1290650115747463, 4.6942200593528831]

['sukhoi', 'korban', 'asuransi', 'tim', 'keluarga', 'gunung', 'hakim', 'ma', 'sby', 'timur', 'salak']

Topik 2

[11.520756607273553, 8.9742090417848424, 7.7331358131443508, 6.811114644404979, 5.0682539266826891, 5.0343760908576476, 4.6277050893900569, 4.2191975768479821, 3.9352454635457748, 3.8660919869284673]

['sukhoi', 'korban', 'tim', 'nbsp', 'indonesia', 'pesawat', '100', 'sar', 'presiden', 'kamis']

Topik 3

[10.660753588290213, 6.5262586486991019, 6.2488711354720001, 6.2079439218877521, 5.1524344375940636, 5.0360423265358527, 4.544418593223936, 4.4893211789212506, 4.139990377267365, 4.0795986301626517]

['sukhoi', 'korban', 'presiden', 'partai', 'kpk', 'demokrat', 'tim', 'polri', 'desa', 'keluarga']

Topik 4

[8.2077066205697413, 7.3277622712913573, 5.524747435124258, 5.1715471213103354, 4.4008711363293163, 4.2533670850241521, 3.7640359481496937, 3.6433791891015672, 3.5731819914924836, 3.2718477083580888]

['partai', 'sukhoi', 'keluarga', 'korban', 'tim', 'asuransi', 'kpk', 'korupsi', 'kasus', 'polri']

Topik 5

[6.3266148595733549, 6.322737265772874, 6.0794743671549432, 5.922787863306807, 4.9570715696956746, 4.6574235295539159, 4.4801149023755951, 4.4674616643584173, 4.0673500460145036, 4.0018538612617256]

['sukhoi', 'republika', 'kpk', 'nbsp', 'komisi', 'wakil', '100', 'tersebut', 'saja', 'sebagai']

20 Mei 2012

Topik 1

[11.728268173046542, 8.5416285266739607, 7.245636193491273, 6.7015680885428077, 6.114210072783413, 5.8620458113185601, 5.8124014898068488, 5.7125736594075667, 4.7351816526166504, 4.5037660177238008]

['sukhoi', 'presiden', 'korban', 'tim', 'timor', 'minggu', 'gubernur', 'leste', 'pada', '100']

Topik 2

[11.032102362614202, 10.925054945374386, 8.7767711236747044, 7.9126257606492212, 7.4530563837367456, 7.0990409475133713, 7.0871375727788903, 6.4190232133115686, 5.8551905785853551, 5.6836718117078799]

['gubernur', 'dpt', 'pemilih', 'tim', 'tetap', 'daftar', 'sukhoi', 'calon', 'penetapan', 'korban']

Topik 3

[9.1837165039330628, 8.0270596467867996, 6.7784562686634002, 4.9873515770206138, 4.7562222516774622, 4.3882280718059103, 4.1848556203331198, 3.8451996449060628, 3.8051737891419237, 3.571453256056035]

['presiden', 'tim', 'sby', 'ada', 'kpk', 'sukhoi', 'leste', 'fdr', 'kasus', 'korban', 'timor']

Topik 4

[9.8400581976509631, 5.8015797291430298, 5.1690902741631897, 5.0038539954070664, 4.6917076388058732, 4.5834027532558625, 4.5596827476969342, 4.5039194859771685, 4.3182962741862365, 3.9737820340748455]

['gubernur', 'nbsp', 'dpt', 'calon', 'presiden', 'pemilih', 'sukhoi', 'kpu', 'negara', 'tahun']

Topik 5

[13.367293423868428, 9.2216997582685352, 7.8948689314207465, 6.2528966536492376, 6.1945378468663845, 4.862435530043979, 3.7363546200411277, 3.698347514207748, 3.6887756469911257, 3.6231219668584123]

['sukhoi', 'gubernur', 'korban', 'tim', 'keluarga', 'dpt', 'dpr', 'pesawat', 'identifikasi', 'ada']

21 Mei 2012

Topik 1

[7.8787193594660154, 5.7268619070204174, 4.6047827659556075, 4.3742393336417118, 4.3637888896057895, 4.1335069401822953, 4.0216656486870574, 3.757103095369577, 3.6822325342943785, 3.4915780456287284]

['korban', 'sukhoi', '2012', 'terjadi', 'berhasil', 'jasad', 'matahari', 'pada', 'indonesia', 'pks']

Topik 2

[13.522357091490923, 13.377780867006948, 7.7034122700503902, 5.5924150466682203, 5.306474223729202, 5.0855431241144906, 4.6165849527917073, 4.6141007254050992, 4.4346143378983118, 4.2500353619913431]

['korban', 'sukhoi', 'tim', 'polri', 'keluarga', 'identifikasi', 'indonesia', 'dvi', 'kpk', 'presiden']

Topik 3

[14.482074696334063, 14.361451091564899, 6.4390583180461158, 5.7168736069699904, 4.6619728603033064, 4.4624620941370994, 4.3243221284154663, 4.2124418478587069, 3.4667687920935788, 3.2111639230892268]

['korban', 'sukhoi', 'keluarga', 'tim', 'identifikasi', 'polri', 'dvi', '100', '2012', 'jasad']

Topik 4

[9.8370907131495375, 7.8387096941744892, 5.5980018080673055, 5.0470596373329677, 5.00230274584169, 4.8373476183677546, 4.8190298674766883, 4.6203306786835778, 4.5274482030864895, 4.515906242312802]

['indonesia', 'presiden', 'korban', 'republika', 'ham', 'tahun', 'partai', 'hari', 'kpk', 'negara']

Topik 5

[7.6579355946431331, 7.5835410392318954, 5.6577075919844724, 4.8195670791600964, 4.776287592459564, 4.6687860603841758, 4.1819505232874867, 4.0051241704740699, 3.841696598814103, 3.2740338870411017]

['presiden', 'indonesia', 'pada', 'korban', 'hari', 'sukhoi', 'tim', 'kpk', 'kasus', 'selatan']

22 Mei 2012

Topik 1

[8.3789065301806804, 7.2743492044419433, 6.9029288757648724, 6.7241171997094069, 6.5369214564075779, 6.0084171865464766, 5.5119337023454067, 5.011873297298318, 4.3972754170611976, 4.3862754617295021]

['sukhoi', 'sebagai', 'korban', 'dpr', 'partai', 'capres', 'jenazah', 'ketua', 'anggota', 'komisi']

Topik 2

[6.1447990901333984, 5.7815542410857681, 5.6616251946810801, 5.2895044969839935, 5.1742511831542339, 5.0576715175622988, 5.0079804738520108, 4.8922781618213875, 4.7158211949402018, 4.6906016027328263]

['sebagai', 'dpr', 'kpk', 'konser', 'partai', 'lady', 'gaga', 'republika', 'komisi', 'ketua', 'indonesia']

Topik 3

[9.724048197646912, 9.6766440012999695, 9.5532618775416172, 6.5728278361103563, 6.5379449872646918, 6.1312143225454365, 5.728207239023595, 4.1264222783330764, 3.8792947735436831, 3.8247332577175035]

['lady', 'gaga', 'konser', 'korban', 'sukhoi', 'keluarga', 'jasad', 'anak', 'ada', '100', 'melihat']

Topik 4

[7.5898277269715955, 6.8508420703301889, 6.8479641164656311, 5.5175931470141446, 5.387768675931369, 4.800170438124006, 4.3181105774720967, 4.1291186288217236, 4.0108565846690301, 3.7007988195118817]

['sukhoi', 'keluarga', 'korban', 'ketua', 'indonesia', 'kpk', 'pesawat', 'partai', 'pada', 'dpr']

Topik 5

[10.461583514934778, 8.0902236783214203, 6.4293454552287903, 6.188751558888554, 6.158051132296638, 6.1071266683489087, 5.7893247052254448, 5.1378615347189562, 5.0312564816727932, 4.9661791498082888]

['sukhoi', 'korban', 'keluarga', 'lady', 'jenazah', 'gaga', 'partai', 'konser', 'dpr', 'ketua']

23 Mei 2012

Topik 1

[11.07739597925279, 10.545774185298963, 10.371264906807385, 9.9966719548097043, 8.0026328854485786, 7.8139559731279338, 7.6873748354968674, 7.063541897323824, 6.6765711438779451, 6.4612394490347569]

['indonesia', 'partai', 'demokrat', 'kasus', 'ketua', 'komisi', 'ada', 'korupsi', 'kpk', 'fraksi']

Topik 2

[9.1094367912625298, 6.9494017685381158, 6.0180732516293292, 5.4903097357783279, 5.1346276308963166, 5.0623642579155499, 4.6605483660200955, 4.4539567174532815, 4.2943749564983102, 4.1907928271788721]

['indonesia', 'corby', 'tim', 'grasi', 'sukhoi', 'kasus', 'ketua', 'tersangka', 'kpk', 'korupsi']

Topik 3

[8.9892525609169009, 7.5542958659274175, 6.8391292328848134, 6.7853176324271853, 6.7278890670798042, 6.7195086260927281, 6.2524742718105832, 5.7342123369060305, 5.3195098124779854, 4.8684950669429279]

['lady', 'gaga', 'konser', 'dki', 'dubes', 'ketua', 'gubernur', 'as', 'corby', 'tim', 'hari', 'dpr']

Topik 4

[13.785075078010419, 6.9648170369039715, 6.2280382654774185, 4.2900219150816072, 4.175915583977484, 3.9311105430215418, 3.6717562584649652, 3.4712891747755812, 3.4271758692138961, 3.371792864658778]

['sukhoi', 'tim', 'ode', 'wa', 'jenazah', 'indonesia', 'kasus', 'grasi', 'korban', 'presiden']

Topik 5

[9.7952112036784573, 9.5571044008597585, 8.6667428112559168, 8.3448648446866578, 6.0336832487843557, 5.3963742237175918, 5.2515042724105188, 5.2203760578210119, 5.1165264805564048, 5.1089087495925503]

['indonesia', 'republika', 'inter', 'partai', 'tim', 'kasus', 'ada', 'milan', 'demokrat', 'pada']

24 Mei 2012

Topik 1

[6.4055342729656672, 6.3214081057754195, 5.5403517427123221, 5.5004606753935956, 5.0409013604634909, 4.7441455150115885, 4.3306246040284719, 4.2894468750718522, 4.2861548121212394, 4.2786859837060218]

['surakarta', 'pihak', 'konser', 'pd', 'tahun', 'raja', 'republika', 'lady', 'tersebut', 'gaga']

Topik 2

[9.3460943907467158, 8.0703677945149046, 7.6782215067329549, 5.444319843295335, 5.0442297992082823, 4.5216775241100704, 4.2496736890313693, 3.9901527402973089, 3.9340305563814972, 3.5727558771705121]

['kpk', 'grasi', 'corby', 'kasus', 'indonesia', 'andi', 'sebagai', 'hambalang', 'ketua', 'australia']

Topik 3

[8.1588533260948033, 7.7114086375600532, 6.3989567412493935, 5.9956309010644073, 5.4435424574998423, 4.9735622045086618, 4.9270800540103004, 4.6549687414998324, 4.0840809680687151, 4.0072115181410259]

['grasi', 'corby', 'australia', 'kpk', 'kasus', 'sebagai', 'sby', 'presiden', 'pemberian', 'hukum']

Topik 4

[5.3001194595710084, 5.247382563905453, 5.0389389244226486, 4.890266561523033, 4.5758121789069328, 4.3028015509277546, 3.9532639232610185, 3.7249116658998611, 3.4330594081121109, 3.1501878165402153]

['grasi', 'kpk', 'corby', 'kasus', 'tahun', 'presiden', 'sebagai', 'indonesia', 'ketua', 'pemerintah']

Topik 5

[8.1523166893766952, 6.9172311089196423, 5.4136594665999338, 5.2431326946671257, 4.8997063785856332, 4.6981461680304797, 4.6002620855535055, 4.1947513138092809, 4.1548363836854216, 4.0419911458751461]

['kpk', 'kornel', 'corby', 'grasi', 'presiden', 'kasus', 'indonesia', 'sebagai', 'wakil', 'terkait']

25 Mei 2012

Topik 1

[9.6983093932621376, 6.5770810897117613, 6.4825845144919292, 6.4667670471129588, 6.202117794294014, 6.1868528735712358, 5.7964263800269595, 5.7302960261367852, 5.3008150630474642, 5.0591719716847221]

['grasi', 'corby', 'ham', 'partai', 'presiden', 'ketua', 'yudhoyono', 'dua', 'dewan', 'anas']

Topik 2

[9.3887210027732984, 5.5186266150420247, 5.5084393884393918, 5.3387036091449271, 4.9914367250737115, 4.4958306779540562, 4.1327747880376959, 4.1260572530295878, 3.9133356130829875, 3.8947571908995648]

['grasi', 'corby', 'komisi', 'dpr', 'ketua', 'indonesia', 'dewan', 'partai', 'rusia', 'sultan']

Topik 3

[5.7887508870657038, 5.5850754368226907, 5.2939264230937981, 4.5926536418146053, 4.3546950737899968, 4.2805864637446946, 4.2557867399935816, 4.1439503659432519, 4.1385680784771504, 4.1290832655607819]

['presiden', 'ketua', 'dpr', 'partai', 'grasi', 'komisi', 'demokrat', 'yudhoyono', 'daerah', 'dua']

Topik 4

[7.467535466381654, 6.1499455887020194, 5.5918584703521548, 5.2887502738512806, 5.2302871824762347, 5.0811344390475019, 4.8607420291914387, 4.7618052528782755, 4.4762814365594288, 4.449733210233676]

['orang', 'dpr', 'ketua', 'mobil', 'arus', 'australia', 'ipb', 'komisi', 'tewas', 'ani', 'satpam']

Topik 5

[87.6119505665766356, 7.3960263293950232, 5.6002989291101626, 5.5128390227975155, 5.1172796345696661, 4.7640877383020284, 4.6092118866145713, 4.533513518922919, 4.5178531116766916, 4.3696715109105853]

['dpr', 'republika', 'korupsi', 'presiden', 'yudhoyono', 'australia', 'corby', 'baik', 'demokrat']

26 Mei 2012

Topik 1

[6.2494096546936548, 5.6687328960185006, 5.3132411642513464, 4.5875165891644301, 4.2884186889729499, 4.0670839742302327, 3.8572436992006498, 3.7919777824098881, 3.7268231860877798, 3.5706211526077989]

['grasi', 'siswa', 'corby', 'presiden', 'demokrat', 'kelulusan', 'pada', 'ada', 'negara', 'sby']

Topik 2

[8.2019565929952396, 7.4728536857284444, 6.7318079597846294, 6.2318716524075626, 6.1168403421950837, 5.5869401933501441, 5.5523306832166055, 4.5811684981065941, 4.3764929549188141, 4.2357071659184564]

['presiden', 'grasi', 'corby', 'sby', 'kpk', 'korupsi', 'demokrat', 'buyung', 'partai', 'ketua']

Topik 3

[.566134914721153, 6.3710758063970818, 5.5803585132725475, 5.3726215782811542, 5.1603843969692162, 4.2482341787695175, 3.8070757781585125, 3.7508075841846531, 3.5545755371181511, 3.5238765774305354]

['grasi', 'corby', 'presiden', 'korupsi', 'sby', 'siswa', 'demokrat', 'kpk', 'pada', 'ketua', 'ada']

Topik 4

[9.0605855163173263, 8.0283657960197576, 5.3032218264193318, 5.1733143609882495, 5.1417626605813425, 4.9867808134150886, 4.7269119883355044, 4.6533552888427092, 4.5256126980125098, 4.4236406439827034]

[grasi', 'presiden', 'reformasi', 'korupsi', 'ketua', 'pemberian', 'partai', 'demokrat', 'corby']

Topik 5

[7.5883433138441587, 7.5677320259533305, 5.7894202521136613, 5.6530697347987928, 5.3733067799047127, 5.0123809376819697, 4.7429811217561006, 4.4203292853599621, 4.3517732098192354, 4.3208344070330122]

['siswa', 'lulus', 'polisi', '2012', 'pada', 'pelaku', 'motor', 'dua', 'smk', 'satpam', 'snmptn']

27 Mei 2012

Topik 1

[7.6388648438093751, 5.7132003277831336, 5.6180889877151206, 4.4780399971889748, 4.3138031851032288, 4.297029314933325, 4.1925390856386011, 4.0634811436801153, 3.8886111033896902, 3.7552804868847787]

['republika', 'demokrat', 'partai', 'ketua', 'pd', 'persija', 'bandara', 'presiden', 'persib']

Topik 2

[7.9874974113878716, 7.5207268077069331, 6.9183641646032825, 6.7015206069617452, 5.7504291140875274, 5.4966860387790764, 4.4324218021149093, 4.3572150438855708, 4.1579786684302729, 4.086946623993013]

['presiden', 'demokrat', 'sby', 'corby', 'kpk', 'grasi', 'yudhoyono', 'partai', 'pd', 'buyung']

Topik 3

[15.690224478958253, 5.701670849057769, 5.3704112370045554, 4.5499800039129248, 4.5471071889182557, 4.5198965595368916, 4.5038706365345638, 4.36434168027104, 4.3419326999332446, 4.04842176067047]

['pd', 'baru', 'bandara', 'dki', 'komwas', 'sby', 'penyerangan', 'kemacetan', 'ketua', 'nama']

Topik 4

[12.015095046236592, 11.844699938345183, 11.071404132532566, 10.048330117046833, 7.1248748987849533, 6.7808678939201652, 6.2029318835189127, 5.7534011718007516, 4.6784814250219675, 4.6040316284629039]

[partai', 'demokrat', 'sby', 'presiden', 'grasi', 'yudhoyono', 'corby', 'ketua', 'capres', 'negara']

Topik 5

[9.7811351185695852, 9.6079344683405594, 8.0739770021729065, 7.615079104709479, 7.5433857237572894, 7.505328407279543, 6.5514987634287971, 5.0583191737713173, 4.7222806719249126, 4.0813473853635509]

[partai', 'demokrat', 'presiden', 'corby', 'grasi', 'sby', 'ketua', 'yudhoyono', 'dewan', 'sebagai']

28 Mei 2012

Topik 1

[9.3973096623148145, 7.8935519681458173, 7.3448885897119771, 6.8858983171013763, 6.7357566327070595, 5.7875589865731545, 4.8015506859644024, 4.795460245262011, 4.7201528194706546, 4.3542988429358997]

[proyek', 'kasus', 'hambalang', 'sukhoi', 'ada', 'presiden', 'kpk', 'terkait', 'indonesia', 'demokrat']

Topik 2

[7.3797733935334024, 7.0818235490805517, 6.5971726973605591, 5.1263066786241547, 4.8492184922040149, 4.2246536389468146, 4.1929060154228077, 4.149121363062803, 3.9604164409963709, 3.8048675236221001]

['sukhoi', 'dpr', 'republika', 'komisi', 'sukabumi', 'korupsi', 'ada', 'pada', 'partai', 'hambalang']

Topik 3

[10.839394401018088, 5.3534431274700811, 4.5962867158478273, 4.1338098475174823, 4.1113294728322529, 4.0649933668017226, 4.0229448409261108, 3.8427141408552083, 3.8370447103920862, 3.8339692747516945]

['sukhoi', 'ada', 'terkait', 'surat', 'korban', 'kasus', 'polri', 'rusia', 'warga', 'karena', 'komisi']

Topik 4

[13.152586832821404, 10.429984776556445, 8.7158376553616126, 5.1270438195587547, 5.0165716404096061, 4.9872319032210042, 4.7729594131100157, 4.6563045905302598, 4.5577351751072026, 4.361119968503286]

['dpr', 'hambalang', 'komisi', 'korupsi', 'hatta', 'kpk', 'ketua', '2014', 'rajasa', 'sukhoi']

Topik 5

[9.6682964300426946, 8.2610543897433164, 6.0925023181164608, 5.536412610916984, 4.6521435879383244, 4.375057635136347, 4.0661816381272162, 3.8964793865498217, 3.8816351175821637, 3.3557496810673886]

['dpr', 'komisi', 'hambalang', 'sukhoi', 'kpk', 'warga', 'daerah', 'anak', 'kasus', 'ketua', 'korupsi']

29 Mei 2012

Topik 1

[6.95246546483971, 6.8236531225968493, 6.3924938987571851, 5.3444271199049922, 4.6676813612029751, 4.0605122938542566, 3.8628243248548126, 3.7128241126398254, 3.6781769140090983, 3.6641852422298649]

['kasus', 'komisi', 'dpr', 'grasi', 'anggota', 'polisi', 'tersebut', 'tersangka', 'ketua', 'rp']

Topik 2

[10.230649087564439, 5.4499655897574169, 5.2376404244788777, 4.8792053900333086, 4.4304919838805636, 4.3669522049611516, 4.3580442532585373, 4.2355718500006203, 3.7234661071203496, 3.4889993167855886]

['komisi', 'anggota', 'tersebut', 'hambalang', 'grasi', 'kasus', 'dpr', 'kpk', '2014', 'corby']

Topik 3

[7.373571905171886, 5.8809993111432926, 5.3148633276290687, 4.2261854621699459, 4.1445709739313026, 4.1294235759662099, 3.7463327390675367, 3.6657667761034052, 3.4901919160648078, 3.4773229520108453]

['republik', 'kasus', 'dpr', 'tahun', 'grasi', 'izin', 'presiden', 'sebagai', 'tobing', 'kepada']

Topik 4

[6.5449742546246146, 6.4668229715563843, 4.8021727082908985, 4.396289451960337, 4.3885354817618696, 4.3883473301023761, 4.2831590638703538, 4.0266212184905159, 3.9334183898375064, 3.6862584893999526]

['dpr', 'kpk', 'kasus', 'komisi', 'ada', 'tahun', 'tersebut', 'tersangka', 'demokrat', 'telah']

Topik 5

[7.1301790626511954, 4.6463228011224285, 4.445726480785857, 4.2279014646333675, 3.6806592458897422, 3.6746999837004992, 3.5450368880150824, 3.4742892105827416, 3.2668302375723903, 3.2429346739955416]

['dpr', 'demokrat', 'komisi', 'kasus', 'telah', 'tahun', 'anggota', 'pada', 'ada', 'kpk', 'partai']

30 Mei 2012

Topik 1

[4.737488469858286, 3.8971777381307278, 3.5699187963785963, 3.4984224773970438, 3.0634903062604764, 2.9836022913832987, 2.8984171513040269, 2.841544894499699, 2.822340097495073, 2.7923463934875636]

['hambalang', 'proyek', 'dpr', 'kpk', 'kasus', 'corby', 'korupsi', 'ketua', 'grasi', 'pembangunan']

Topik 2

[7.1066813854578736, 5.9676269231950254, 4.0736828944648948, 3.7003645862698979, 3.6748922871382153, 3.6521598437947307, 2.6916097918351052, 2.5548234667455478, 2.5269871943177136, 2.3864053001644279]

['hambalang', 'proyek', 'pembangunan', 'presiden', 'indonesia', 'daerah', 'kpk', 'olahraga', 'milik']

Topik 3

[8.7382914182561233, 6.7901111194349317, 4.2818645006471874, 4.2277667039678519, 4.1635846981591857, 4.0822234186851762, 4.082147907767383, 3.6708103324557997, 3.5128876061613421, 3.3824209359987081]

['hambalang', 'proyek', 'kpk', 'tahun', 'kasus', 'indonesia', 'dpr', 'presiden', 'daerah', 'corby']

Topik 4

[6.8357284446987219, 5.4047152829609413, 5.3148276523146389, 4.541714822905492, 4.0211166348435183, 3.870056920153564, 3.5281034829619249, 3.4868536490975552, 3.428660216399658, 3.1103090472272141]

['kpk', 'century', 'kasus', 'dpr', 'corby', 'pada', 'tahun', 'presiden', 'grasi', 'depan', 'bank']

Topik 5

[10.750725185682985, 5.3205969453701307, 5.1526415848135168, 4.0900242628418635, 3.8605987197149778, 3.7696623400369953, 3.6610428750725381, 3.2726246789577966, 3.243844802642573, 3.0287618348511174]

[republika', 'nbsp', 'jawa', 'pers', 'terhadap', 'orang', 'kekerasan', 'telah', 'tindakan', 'kebakaran']

31 Mei 2012

Topik 1

[8.6413260265958485, 6.9104407966904873, 6.7023575013680192, 5.7944196535045647, 5.6661605158579125, 5.1867208796489566, 5.1530109556022445, 5.1499405387688562, 5.0724385572653974, 5.0384319924351226]

['kasus', 'fdr', 'sukhoi', 'tni', 'proyek', 'pada', 'jaksa', 'korupsi', 'ditemukan', 'kpk', 'hambalang']

Topik 2

[7.0957363523241277, 5.3716984927031159, 5.2279016618438421, 4.6716556969504435, 4.502818129136692, 4.106497006687678, 4.0458309425429455, 3.9958663296421162, 3.8429902222684875, 3.7576405306699487]

['kasus', 'sukhoi', 'hukum', 'negara', 'kpk', 'negeri', 'korupsi', 'komisi', 'orang', 'telah']

Topik 3

[7.9943239667433259, 5.9125404507772679, 5.2980393309408296, 5.2171286810921433, 4.9004030658687281, 4.5210249891004546, 4.2108266162491299, 4.1674124705172888, 4.1546552878175307, 3.650242912828908]

['kasus', 'proyek', 'republika', 'tni', 'kpk', 'terhadap', 'sukhoi', 'inggris', 'hukum', 'angelina']

Topik 4

[8.4390522237017684, 5.6877744415916318, 5.6392824694643986, 5.2424408122859694, 4.8411587395335243, 4.8203507840115654, 4.8179634095679953, 4.7238823529396718, 4.2910651740474215, 4.2530026872006719]

['sukhoi', 'ketua', 'pohon', 'fdr', 'motor', 'rakyat', 'hujan', 'timbang', 'tni', 'namun', 'juga']

Topik 5

[11.155417041895376, 10.989430376804599, 9.2255589108934846, 6.1061995090394188, 5.6799144567858102, 5.4178500111237975, 4.1210603427463246, 4.1117512659212485, 3.9364523851320965, 3.8398315003184726]

['sukhoi', 'tni', 'fdr', 'panglima', 'kpk', 'dpr', 'ditemukan', 'pesawat', 'kasus', 'agung', 'bukan']