



Perancangan Kerangka Kerja
Focused Brand Exposure Monitoring
Menggunakan Pemodelan Topik
dan Formulasi Kueri

Focused Brand Exposure Monitoring

- Brief Overview -

Iwan Santosa

Universitas Kristen Maranatha

©2023/2024

Abstract

Brand exposure monitoring is a specialized form of media monitoring that aims to monitor brand exposure and brand mentions in media publications. This research proposes a framework model for monitoring brand exposure through analysis of media publications utilizing query formulation and topic modeling, aimed at extracting important topics contained in the company's internal media publications, and comparing them with external media publications or mass media.

Topic extraction is performed using Latent Dirichlet Allocation (LDA) method on text data in the form of SERP (Search Engine Results Page) snippets. Subsequent processing utilizes Jaccard index and cosine similarity calculations. The output of the framework includes visualizations of topics representing central themes of articles exposing the institution's brand, and measurement metrics that are useful for decision-making analysis related to media management and institutional communication.

I

PENDAHULUAN

Latar Belakang

- Proses *brand exposure monitoring* di Universitas studi kasus dilaksanakan secara manual dengan melakukan *web search* dan membaca teks SERP satu per satu, kemudian melakukan analisis untuk menemukan topik-topik yang perlu menjadi perhatian.
- Faktor subjektivitas tinggi karena mengandalkan kemampuan, kecakapan, dan kecermatan staf pemonitor.
- Tidak dapat menampilkan metrik yang konsisten, karena kesimpulan bergantung analisis staf pemonitor.

Brand exposure monitoring: bentuk khusus dari *media monitoring* yang bertujuan memantau paparan merek dan *brand mention* pada media-media publikasi.

Rumusan Masalah

1. Bagaimana *brand exposure monitoring* yang dilakukan melalui mesin pencari dapat didukung dan ditingkatkan agar dapat menghasilkan informasi yang mudah dicerna oleh *stakeholder* pengambil keputusan?
2. Bagaimana cara mendapatkan informasi topik *exposure* terhadap *brand* dengan cara yang lebih mudah dan objektif (*data driven*)?
3. Apakah topik pada *corporate website* dapat menjadi referensi informasi yang berguna dalam proses *brand exposure monitoring* bila dihubungkan dengan *website external media*?
4. Indikator apa yang dapat digunakan untuk menghubungkan *brand exposure* pada *corporate website* dengan *external media*?

Tujuan Penelitian

1. Merancang model kerangka kerja *brand exposure monitoring* untuk mendukung dan meningkatkan kinerja tim pemonitor di institusi studi kasus, yang dapat menghasilkan *report* yang mudah dipahami pengambil keputusan.
2. Mengolah informasi yang dihasilkan oleh *search engine* menggunakan metode-metode *machine learning* dan *natural language processing* (NLP) yang sesuai dengan kebutuhan kerangka kerja.
3. Mengekstraksi informasi hasil pencarian *corporate website* dan *external media* untuk menemukan topik yang terkandung di dalamnya.
4. Menentukan indikator dan metrik yang dibutuhkan sebagai salah satu bentuk output dari kerangka kerja.

||

KAJIAN LITERATUR

Landasan teori dan konsep yang mendasari penelitian

1. *Brand exposure*, pengelolaan media & komunikasi

(Cornwell *et al.* [5], Pitts & Slattery [6], Arthana *et al.* [7], Pomalaa *et al.* [8], Khoa [9], Pinar [10], Perwito *et al.* [11])

2. *Brand communication* di perguruan tinggi

(Gatfield *et al.* [11]; Gray *et al.* [12]; Pinar *et al.* [13]; Sujchaphong [14]; Noor *et al.* [15]; de Heer [16]; Khoa [9]; Pinar [10], Binnie [17], Chapleo [18], Gutiérrez [19])

3. *Brand & media monitoring*

(Jefkins [23]; Friedel & Lukman [24]; Habiba [25]; Astiti, *et al.* [26])

4. *Search engine & SERP snippet*

(Croft *et al.* [27]; Yang & Gerasoulis [28]; Kim *et al.*; Sachse *et al.*; Strzelecki & Rutecka)

5. *LDA topic model*

(Blei [34]; Maier *et al.* [35], Blei *et al.* [36]; Bhatia *et al.* [37])

6. Metrik similaritas (Jaccard index dan cosine similarity)

(Yang & Gerasoulis [28])

Penelitian sebelumnya mengenai penggunaan metode LDA

1. Liu *et al.* [22] menggunakan pendekatan *topic modeling* untuk analisis berita mengenai Covid-19 pada masa awal pandemi merebak di Tiongkok.
2. Abebe *et al.* [39] menggunakan LDA *topic modeling* untuk mengekstrak informasi hasil *search query* mengenai data kesehatan di Afrika.
3. Hidayatullah *et al.* [40] melakukan analisis data teks mengenai kondisi cuaca dan iklim di wilayah Jawa, dari *tweets* media sosial Twitter BMKG.
4. Karnalim *et al.* [41] menggunakan LDA untuk mengekstraksi wawasan informasi mengenai mitigasi plagiarisme pada bidang pendidikan dari hasil isian kuesioner survei *open-ended*.
5. Wahyudin [42] menggunakan LDA untuk analisis pemberitaan oleh portal berita *online* selama masa PSBB pertama.
6. Santoso *et al.* [43] menggunakan LDA untuk analisis topik tagar Covidindonesia pada media sosial Instagram.
7. Sagala & Toba [44] memanfaatkan LDA untuk menganalisis ulasan tekstual untuk menemukan aspek-aspek berpengaruh pada produk *smartphone*.
8. Kannitha *et al.* [45] menggunakan LDA pada media sosial Twitter untuk analisis keluhan pelanggan IndiHome dan First Media.
9. Kharisudin & Masri'an [46] menggunakan LDA untuk analisis *review* pengguna aplikasi WhatsApp pada Google Play.
10. Novarian *et al.* [47] menggunakan LDA untuk menemukan topik yang dibahas pada tugas akhir mahasiswa Institut Teknologi Telkom Purwokerto.

Kebaruan penelitian

1. Tujuan pemodelan topik dalam konteks kerangka kerja *brand exposure monitoring*.
2. Jenis *dataset* berupa SERP *snippet*.
3. Implementasi LDA tidak hanya untuk interpretasi topik, tetapi juga untuk mengukur metrik *brand exposure*.
4. Langkah dan parameter *preprocessing* dan *pipeline* keseluruhan yang *domain-specific*.

Kontribusi penelitian

Mengisi *research gap* dan memberikan wawasan baru dalam hal:

1. Implementasi LDA *topic modeling* untuk mengekstraksi informasi dari *dataset* berupa SERP *snippet*.
2. Metode *text preprocessing* dan evaluasi/validasi yang perlu dilakukan pada studi kasus dalam bidang terapan *brand exposure monitoring*.

III

METODE PENELITIAN

Hipotesis

1. Proses *monitoring* dapat ditingkatkan menggunakan *query* berupa *branded search* dengan parameter kondisional dan/atau *query terms* spesifik untuk menghasilkan SERP yang relevan.
2. SERP *snippet* mengandung data yang cukup representatif untuk proses ekstraksi informasi menggunakan metode pemodelan topik.
3. Teknik ekstraksi menggunakan *topic modeling* dapat diaplikasikan pada sumber berita media massa dan *corporate website* untuk mendeteksi topik-topik yang diekspos pada periode *monitoring*.
4. Topik-topik yang dihasilkan dari ekstraksi informasi pada kedua kategori media tersebut dapat dikomparasi dan dilakukan penghitungan untuk analisis pengambilan keputusan.

Desain Penelitian

Rancangan Kerangka Kerja

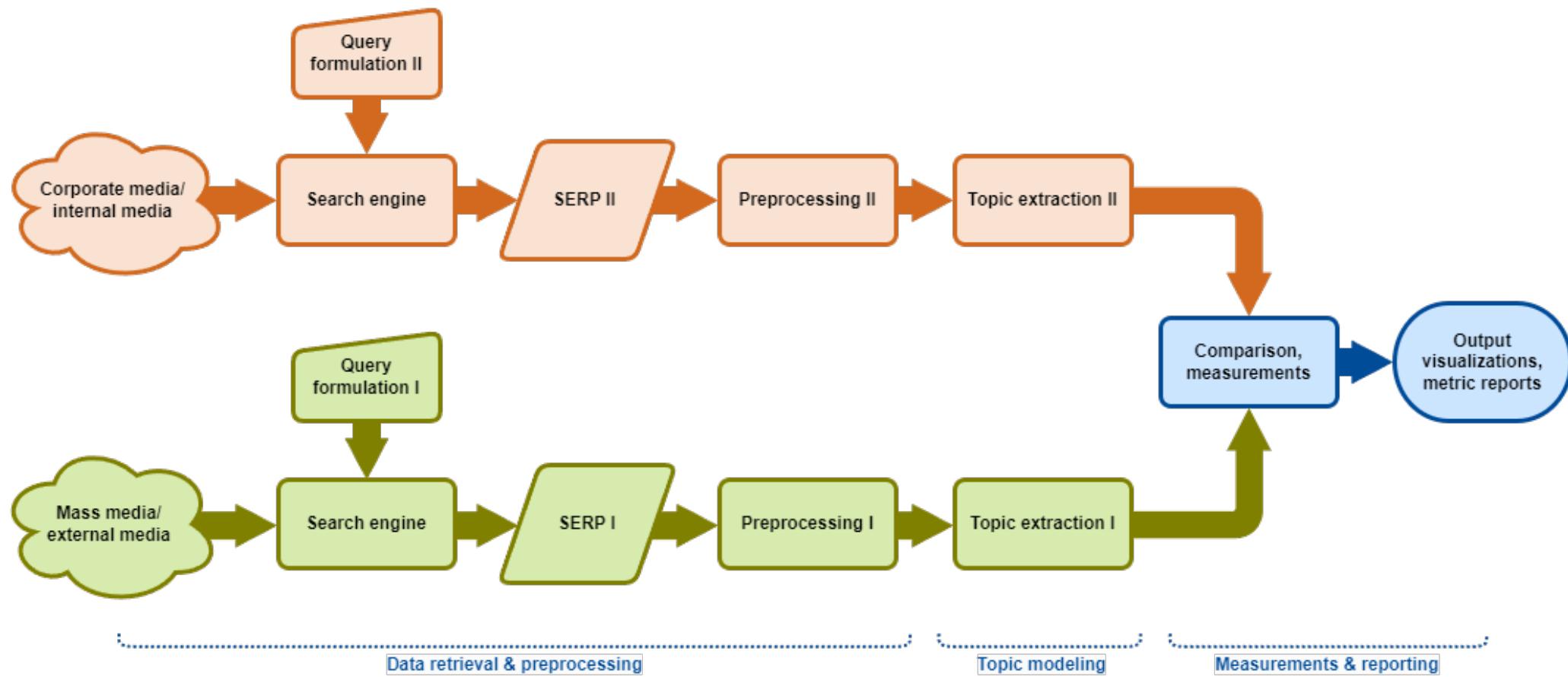
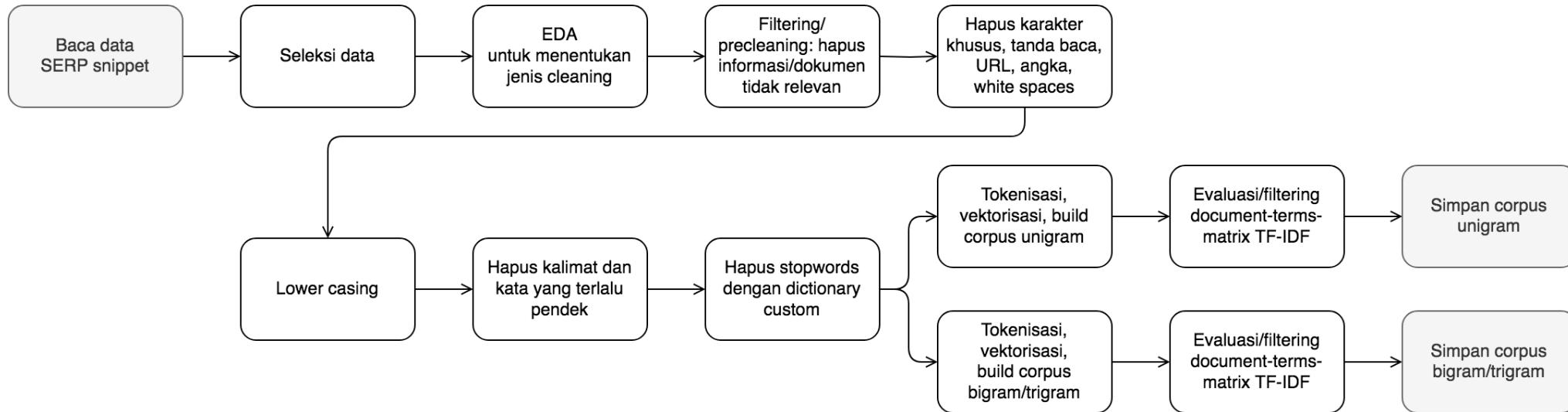


Diagram blok rancangan kerangka kerja “Focused Brand Exposure Monitoring” (FBxM)

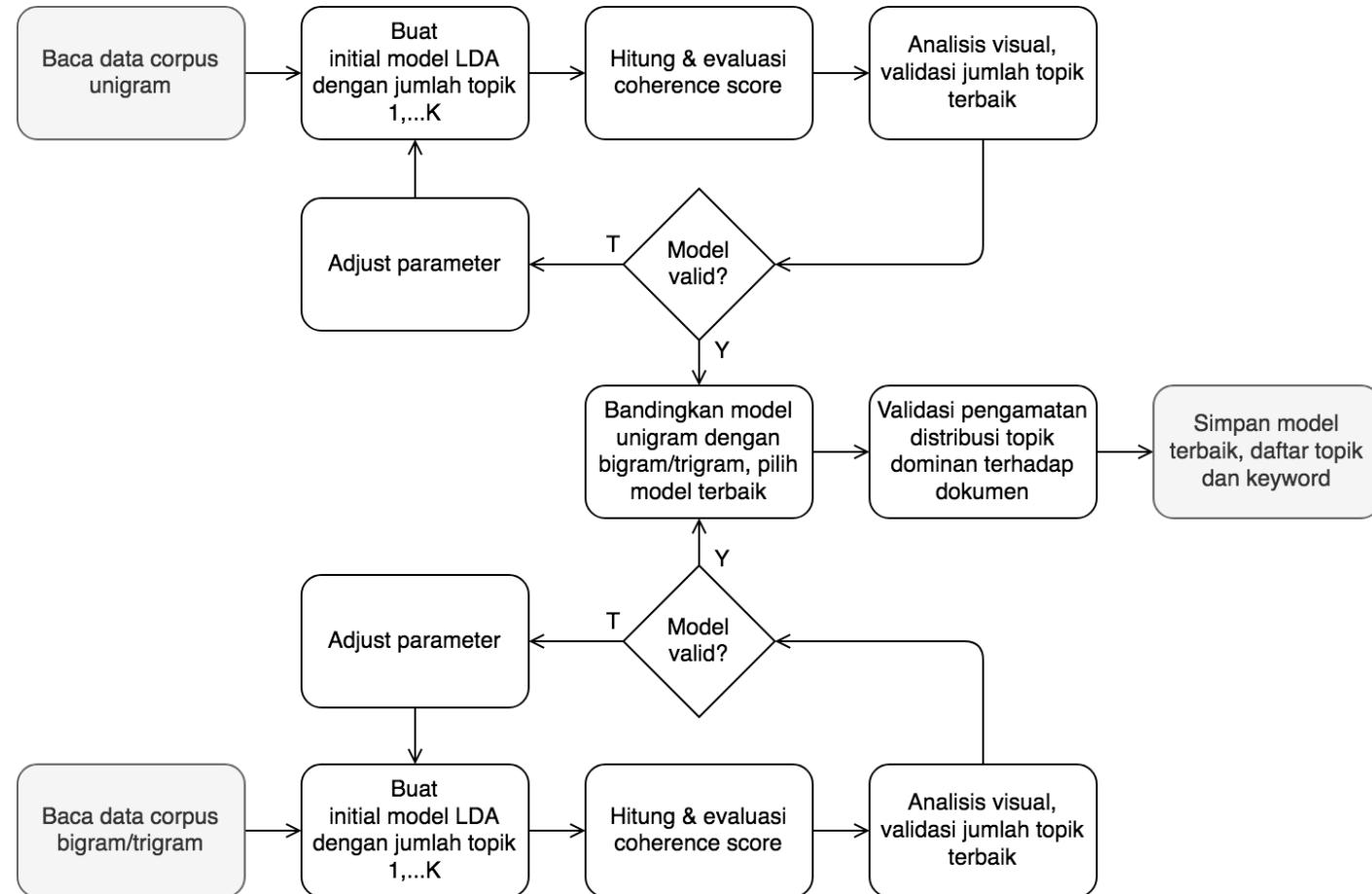
1. Data Retrieval dan Preprocessing



Tahapan langkah *data retrieval* dan *preprocessing*

Scraping menggunakan *extension SERP Snippet Extractor v2.2* pada *browser LibreWolf 107.0.1-1*

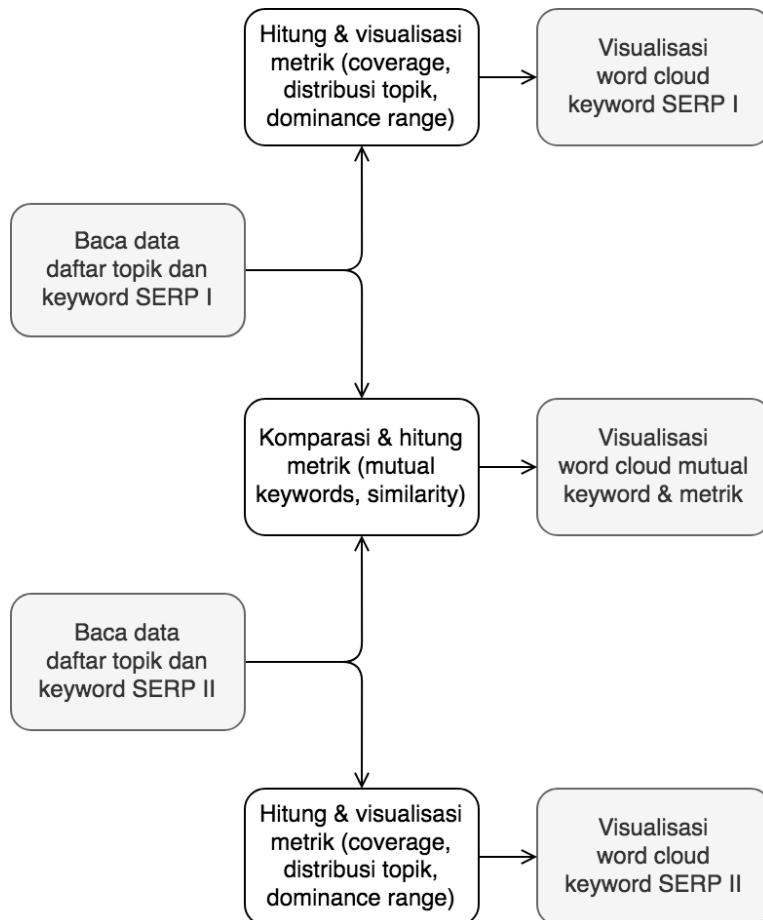
2. Pemodelan Topik



Tahapan langkah proses pemodelan topik

- Pemodelan topik LDA menggunakan *library* Gensim
- Visualisasi *intertopic distance map* menggunakan pyLDAvis

3. Measurements & Reporting



- Kelompok topik hasil ekstraksi divisualkan dalam bentuk *word cloud*
- *Mutual top keywords count* dihitung menggunakan *Jaccard score* (*Jaccard index*) pada hasil topik SERP I dan SERP II
- *Overall topics similarity* dihitung menggunakan *cosine similarity* antara gabungan semua topik yang terbentuk pada SERP I dan SERP II
- Visualisasi komparasi berupa *word cloud mutual keywords* (tanpa bobot) dan *metrics report* komparasi

Tahapan langkah *measurements & reporting*
(proses komparasi dan output visualisasi)

Pengujian Data Retrieval & Preprocessing SERP I

Data collection

Mengambil data dari *external website* (media massa dan situs lainnya) dengan cara melakukan *branded search* (pencarian dengan menggunakan *brand name* sebagai kata kunci utama) menggunakan Google *search engine*, kemudian dilakukan *scraping* menghasilkan data tabular.

Query formulation I

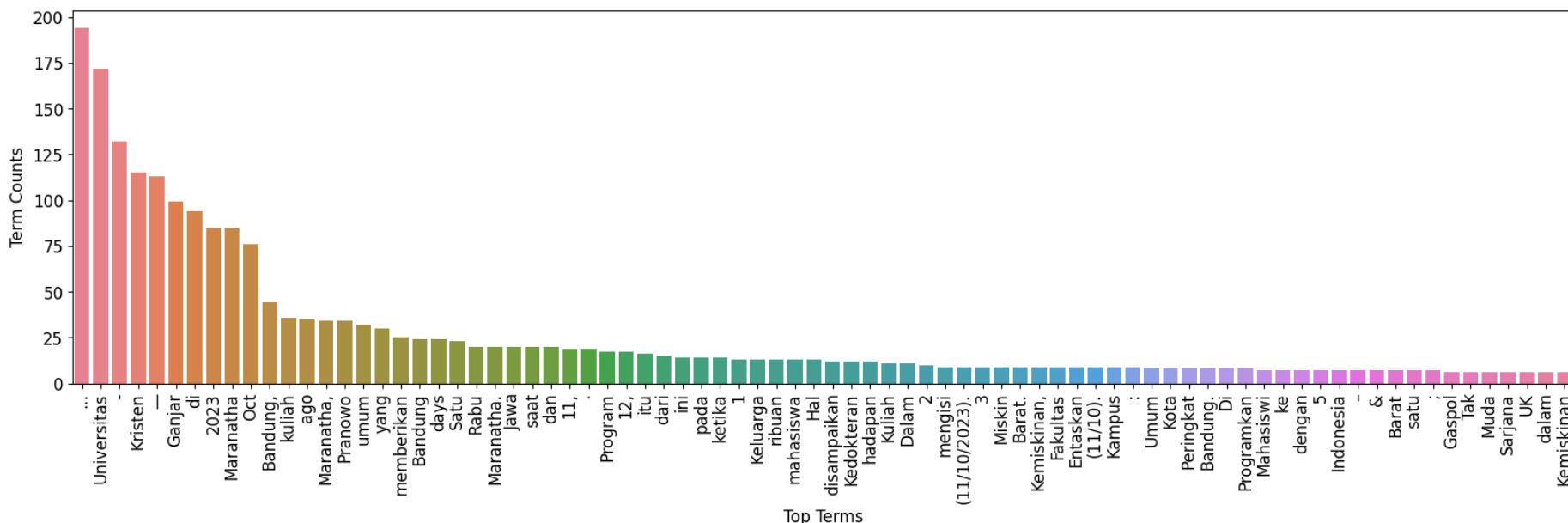
Brand Name	Operator	Search Parameter (exclude site)	Operator	Search Parameter (filter by text title)	Operator	Periode Tanggal Awal	Operator	Periode Tanggal Akhir
“brand name”	AND (+)	-site:[URL 1]	AND (+)	-intitle:[text 1]	AND (+)	after:yyyy-mm-dd	AND (+)	before:yyyy-mm-dd
OR		AND		AND				
“alias 1”		-site:[URL 2]		-intitle:[text 2]				
OR		AND		AND				
“alias 2”		-site:[URL 3]		-intitle:[text 3]				

Format dasar *query formulation* untuk tahap pencarian I

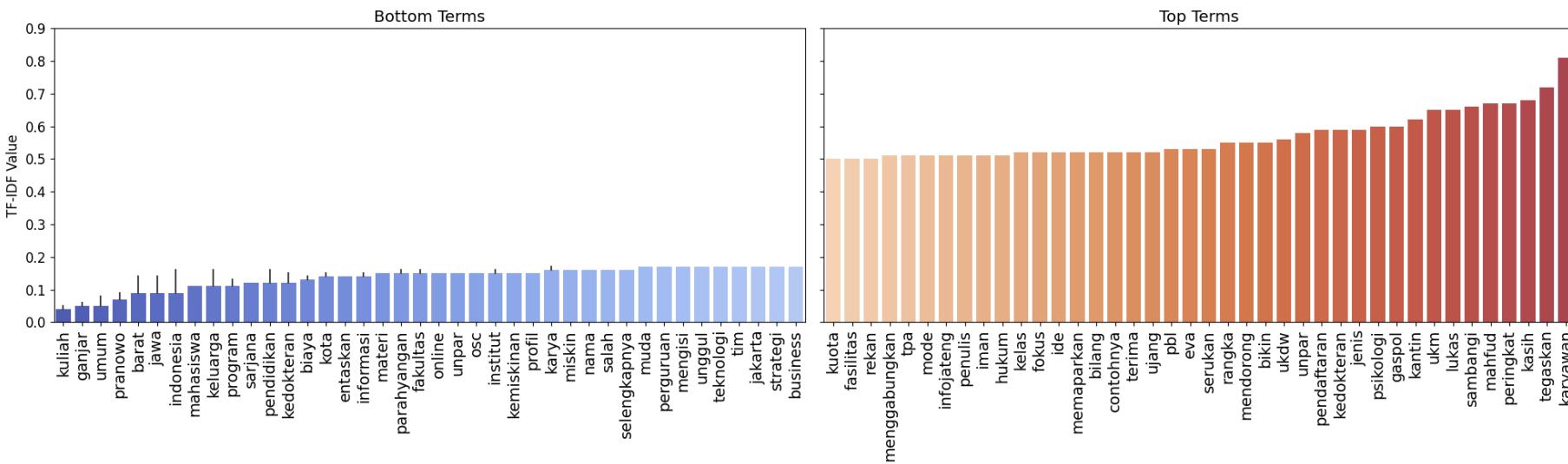
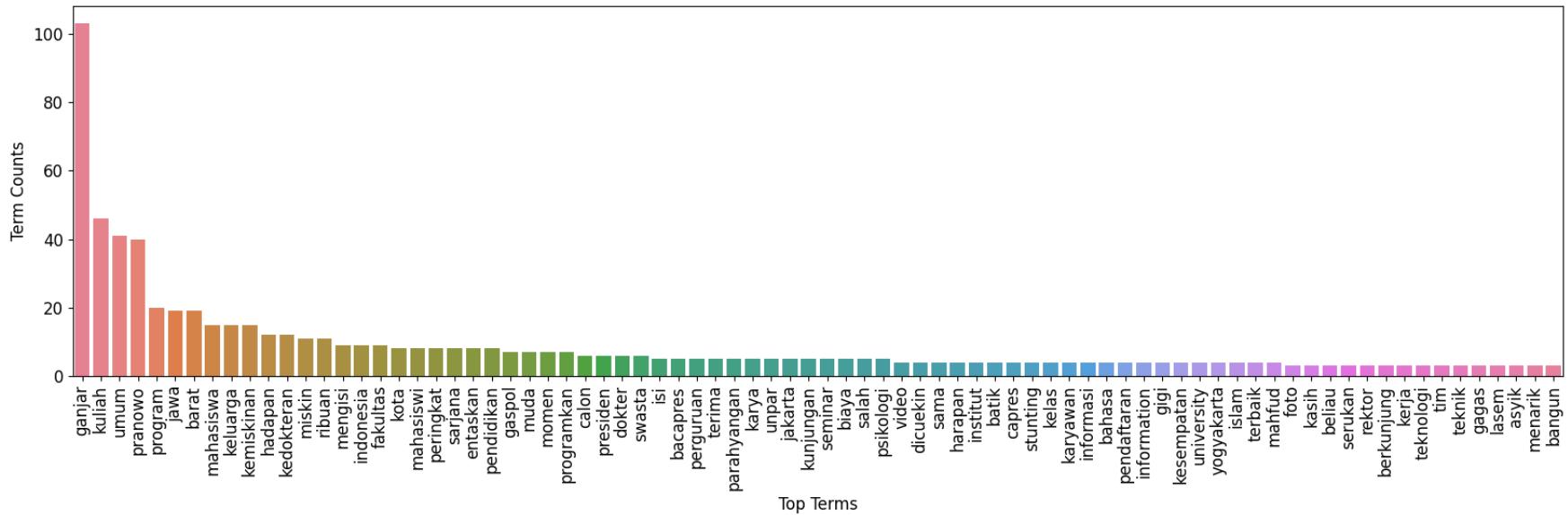
Precleaning

```
hapuskata = ['Berita Resmi Universitas', 'Archive – Maranatha News', 'Archives – Maranatha News', 'Popular News – Maranatha News', 'Contact – Maranatha News',
'Popular News', 'Popular Tags', 'Editor Pick', 'Untitled', 'Universitas Kristen Maranatha Archives',
'Archive – Page', 'Archives – Page', 'Kata Kunci – Maranatha News', 'Hasil Pencarian', 'Related Photo', 'More News', 'SERP Feature',
'Download', 'Unduh']
# List daftar kata berisi string tags navigasi dan penanda khas yang tidak berkontribusi makna pada pemodelan topik
# "SERP Feature" digenerate oleh Google Search Engine, yaitu hasil tambahan yang tidak dibutuhkan pada pemodelan topik
```

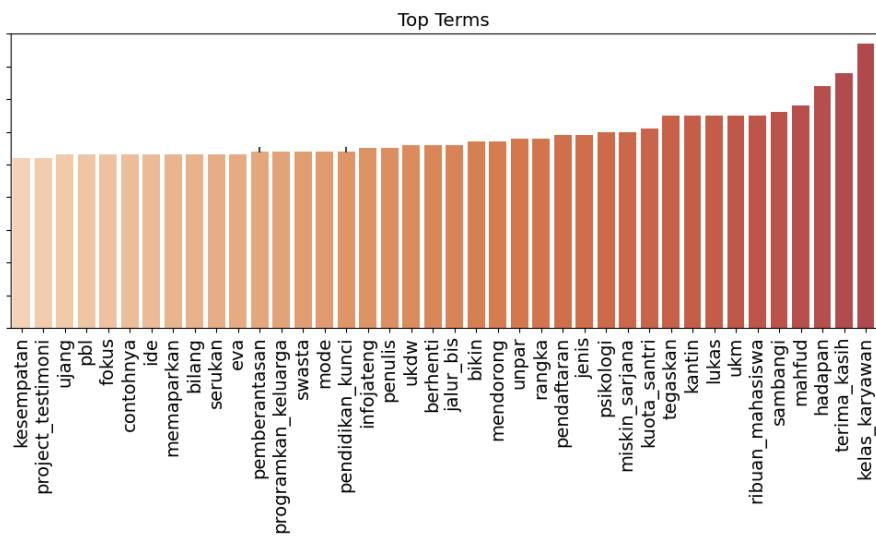
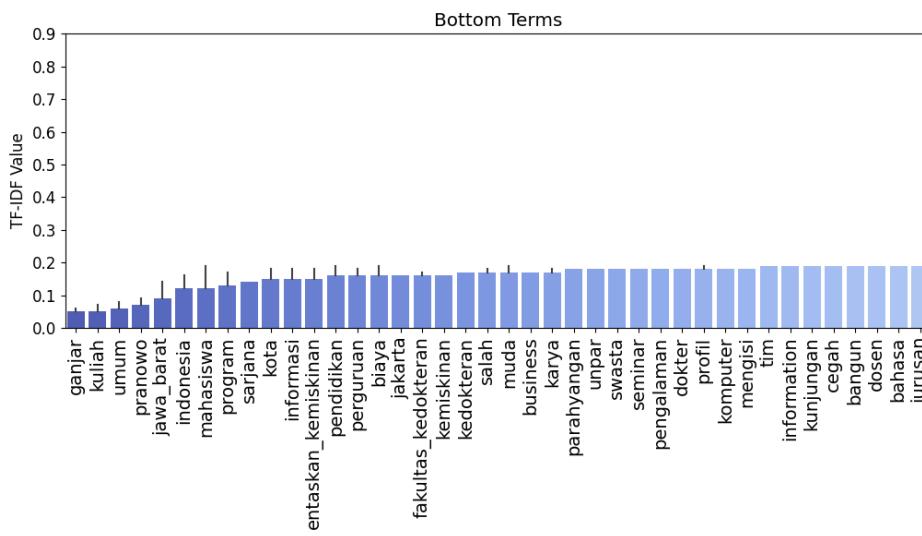
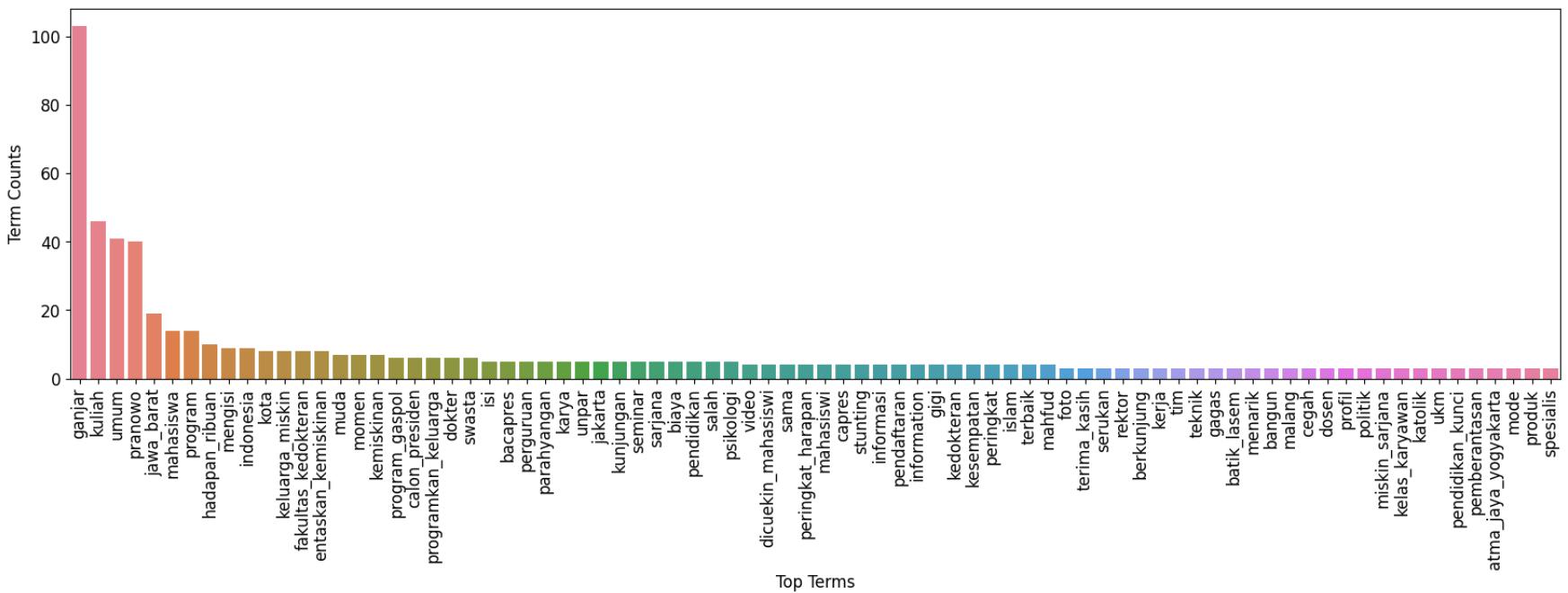
10 Siswa Teknik Otomasi Industri (TOI) SMK Negeri 1 Cimahi ... - 5 days ago - ... Universitas Kristen Maranatha, bekerja sama dengan Rexroth Bosch Company. Kompetisi yang berlangsung pada tanggal 25 ...
11 Ganjar Kenalkan Program Gaspol Saat Isi Kuliah Umum di ... - Oct 13, 2023 - Jakarta, Gesuri.id - Calon presiden dari PDI Perjuangan, Ganjar Pranowo mengenalkan program Gaspol saat mengisi kuliah u...
12 UKDW Terima Kunjungan dari Universitas Kristen ... - Oct 1, 2023 - Bertempat di Ruang Seminar Pdt. Dr. Harun Hadiwijono, Universitas Kristen Duta Wacana (UKDW) menerima kunjungan dari Universitas ...
13 AdminSAAC, Author at SAAC – Universitas Nusa Putra - Oct 17, 2023 - Program Studi Teknik Elektro Universitas Kristen Maranatha bersama Rexroth Bosch Company Mengadakan HaXathon Competition 2023. D...
14 Ganjar Gagas Program Satu Keluarga Miskin Satu Sarjana - Oct 12, 2023 - Gagasan itu disampaikan Ganjar saat dirinya mengisi kuliah umum di Universitas Kristen Maranatha, Bandung, Jawa Barat, Rabu, ...
15 Batik Lasem UK Maranatha Raih Penghargaan Seoul ... - 4 days ago - "The New Face of Batik Lasem Industry" karya tim UK Maranatha, Indonesia menempati posisi pemenang ketiga, bersama dengan lima ka...
16 Unbothered Queen, Mahasiswi yang Asyik Makan dan ... - Oct 14, 2023 - Ada yang menarik dari momen kunjungan bakal calon presiden Ganjar Pranowo ke Universitas Kristen Maranatha Bandung, Jawa Barat...
17 Browse Author Index - Oct 3, 2023 - S. Santi, Septhani Santi, Universitas Kristen Maranatha Saputra, Benny Ade, Tourism, art, culture. Saputri, Allen Nur, Universitas Kristen Maranatha Sari ...
18 Program Atasi Kemiskinan, Ganjar Ingin Bangun Boarding ... - 6 days ago - ... Capres Ganjar Pranowo saat memberikan kuliah umum di Universitas Kristen Maranatha Bandung./Radar Malang ...
19 Rahasia Melimpah Daun Kelor, Cegah Stunting Hingga ... - Oct 19, 2023 - Dosen Fakultas Kedokteran Universitas Maranatha Bandung dr Theresia Monica Rahardjo membeberkan formula jitu yang dapat meng...
20 DETAIL CERITA SMA TONIATTIC CIAMET RTVART - Oct 10, 2023 - UTETTACT KE INTIVEDCTTAC MARANATHA BANDUNG - MBIC SMA TONIATTIC CIAMET RTVART T A 2023 - 2024 // EXPLORING THE SPACE // ... Salongkanya



Distribusi kata unigram

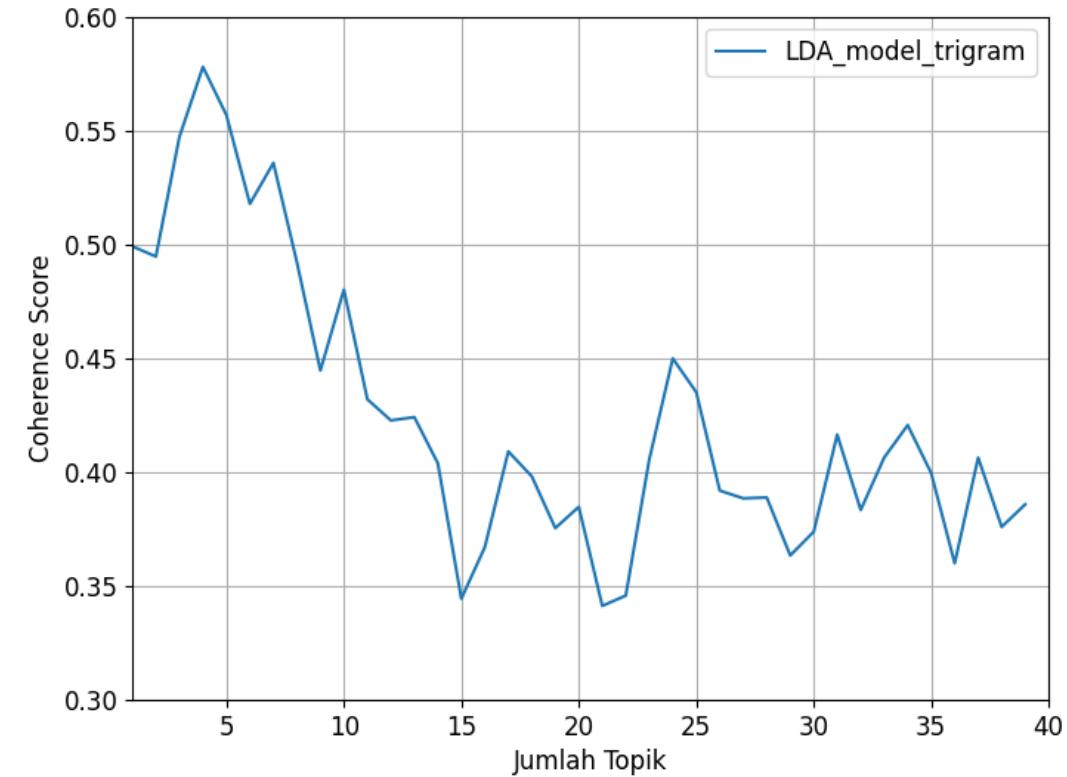
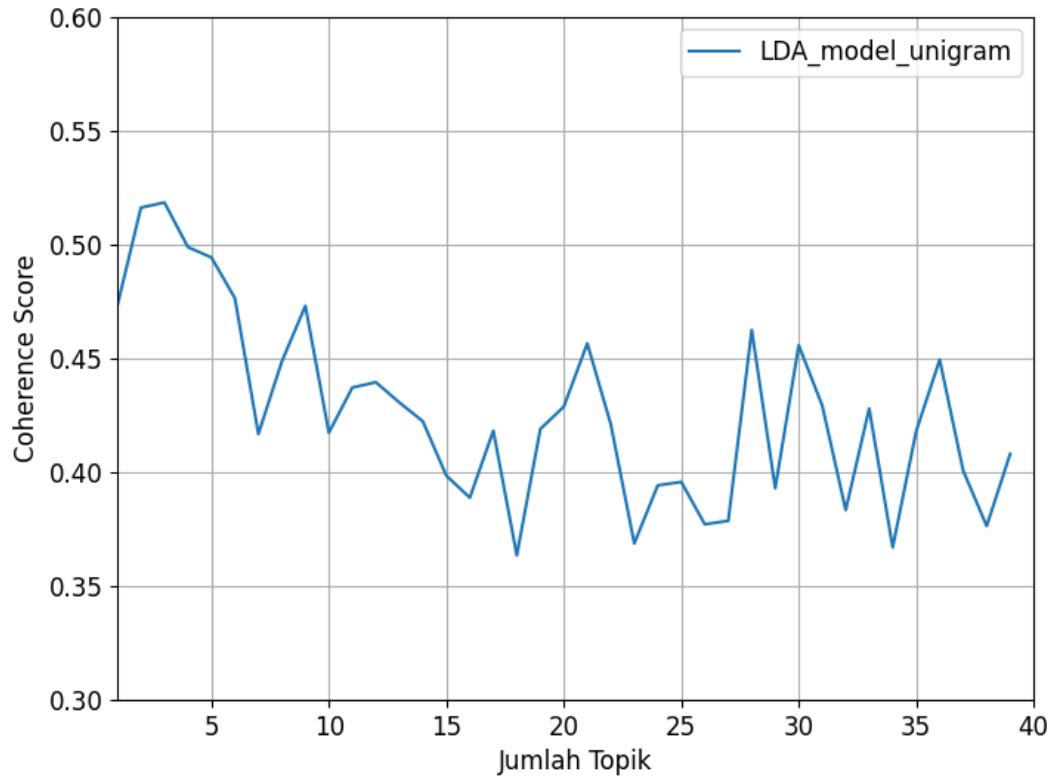


Distribusi kata bigram/trigram



Pengujian Pemodelan Topik SERP I

Buat 40 model initial, lalu hitung coherence score

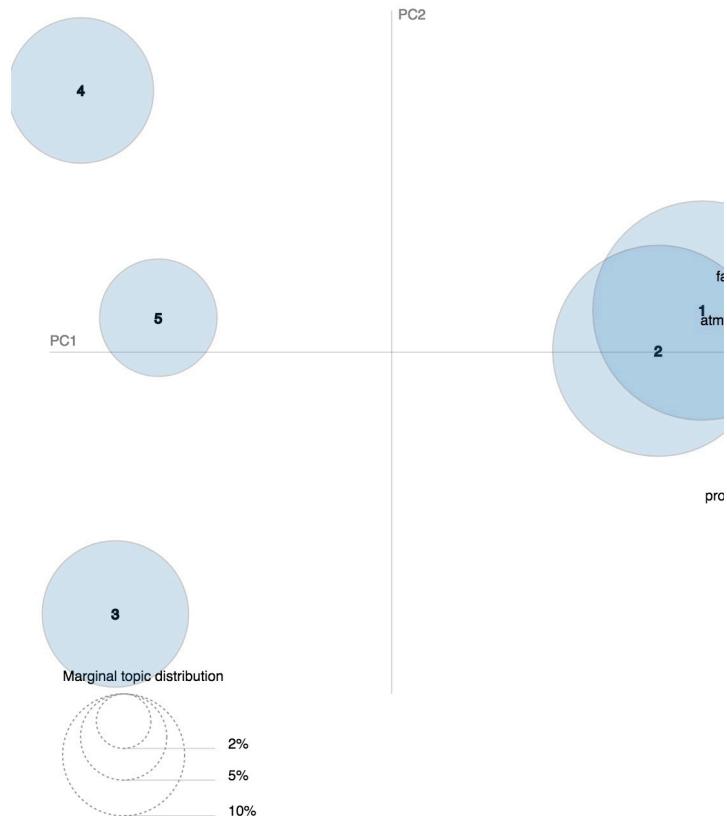


Analisis & validasi visual pyLDAvis (melihat keberhasilan model memisahkan topik)

```
vis5_t = pyLDAvis.gensim.prepare(ldamodel, corpus, dictionary)  
vis5_t
```

Selected Topic: 0

Intertopic Distance Map (via multidimensional scaling)



```
vis7_t = pyLDAvis.gensim.prepare(ldamodel, corpus, dictionary)  
vis7_t
```

Selected Topic: 0

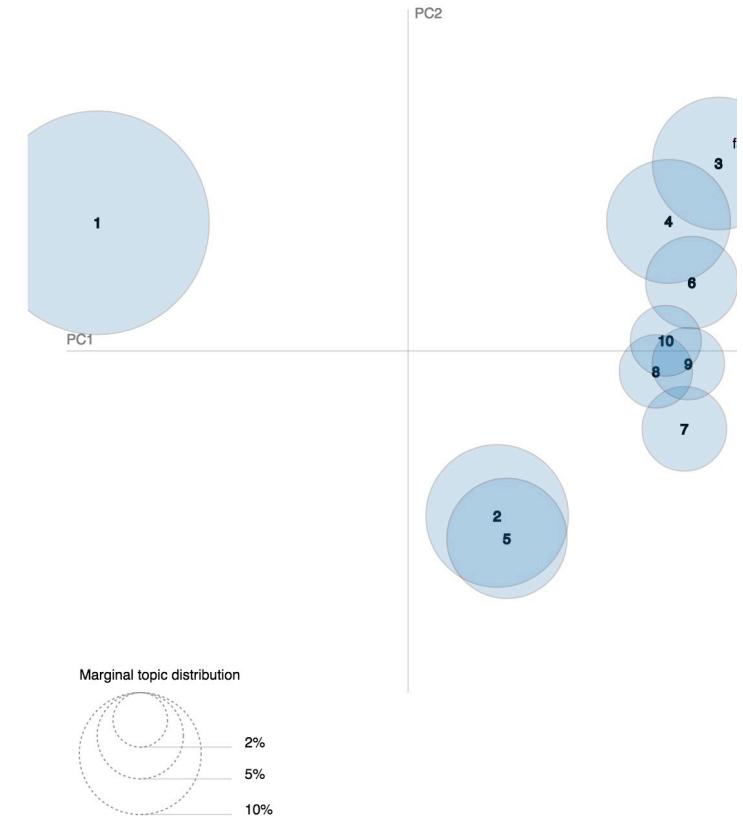
Intertopic Distance Map (via multidimensional scaling)



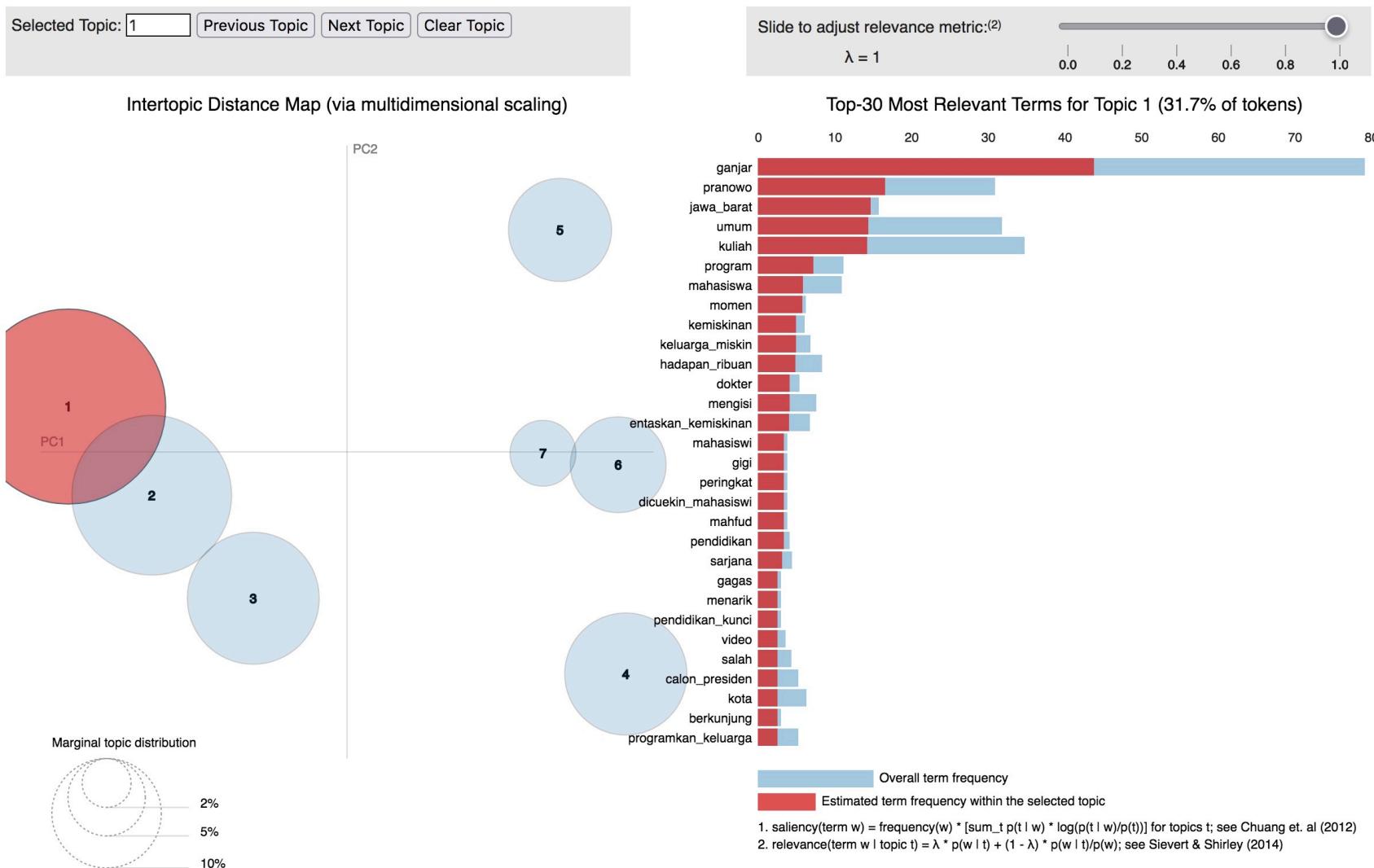
```
vis10_t = pyLDAvis.gensim.prepare(ldamodel, corpus, dictionary)  
vis10_t
```

Selected Topic: 0

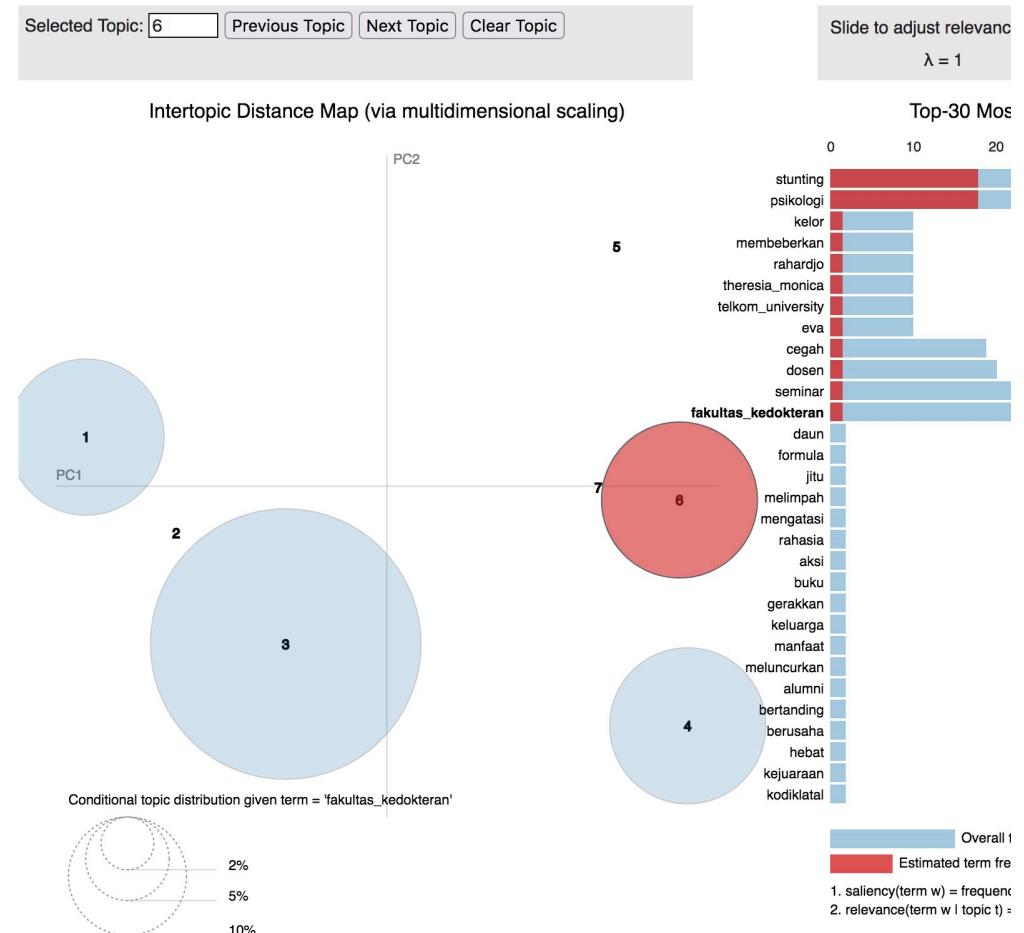
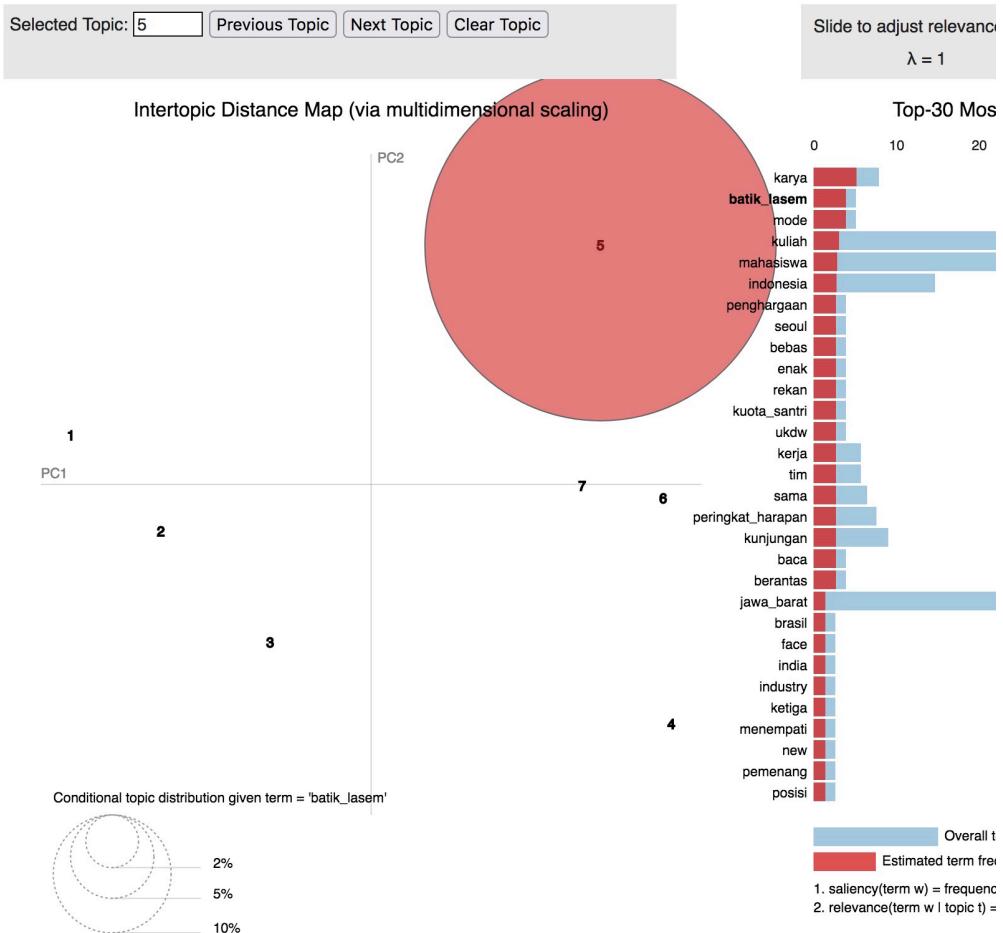
Intertopic Distance Map (via multidimensional scaling)



Analisis visual pyLDAvis (melihat distribusi most relevant terms pada topic 1)



Analisis visual pyLDAvis (melihat distribusi term "batik_lasem" dan "fakultas_kedokteran")



Distribusi terms pada topik (model pilihan adalah dengan jumlah topik = 7)

```

num_topics = 7
x = num_topics-1
optimal_model_lda_trigram = model_list_lda_trigram_high[x]
optimal_model_lda_trigram.show_topics(num_topics)

[(0,
 '0.078*"ganjar" + 0.049*"umum" + 0.049*"kuliah" + 0.026*"pranowo" + 0.014*"program_gaspol" + 0.014*"isi" + 0.011*"mengisi" + 0.010*"mahasiswa" + 0.008*"calon_presiden" + 0.008*"jakarta"'),
(1,
 '0.072*"ganjar" + 0.036*"pranowo" + 0.022*"kuliah" + 0.019*"umum" + 0.015*"muda" + 0.015*"swasta" + 0.015*"fakultas_kedokteran" + 0.011*"terima_kasih" + 0.008*"ajak" + 0.008*"politik"'),
(2,
 '0.020*"karya" + 0.015*"mode" + 0.015*"batik_lasem" + 0.012*"kuliah" + 0.011*"mahasiswa" + 0.011*"indonesia" + 0.010*"tim" + 0.010*"penghargaan" + 0.010*"seoul" + 0.010*"sama"'),
(3,
 '0.012*"lukas" + 0.012*"jenis" + 0.012*"pendaftaran" + 0.009*"program" + 0.009*"kuliah" + 0.008*"fasilitas" + 0.008*"lengkap" + 0.008*"hukum" + 0.008*"pengalaman" + 0.008*"information"'),
(4,
 '0.015*"tpa_berhitung" + 0.015*"bahasa_inggris" + 0.008*"ips" + 0.008*"osc" + 0.008*"online" + 0.008*"matematika" + 0.008*"test" + 0.008*"cek" + 0.008*"diujikan" + 0.008*"ipa"'),
(5,
 '0.021*"psikologi" + 0.021*"stunting" + 0.011*"rahardjo" + 0.011*"dosen" + 0.011*"fakultas_kedokteran" + 0.011*"membeberkan" + 0.011*"kelor" + 0.011*"theresia_monica" + 0.011*"cegah" +
0.011*"telkom_university"'),
(6,
 '0.106*"ganjar" + 0.040*"pranowo" + 0.036*"jawa_barat" + 0.035*"umum" + 0.035*"kuliah" + 0.018*"program" + 0.014*"mahasiswa" + 0.014*"momen" + 0.012*"keluarga_miskin" + 0.012*"kemiskinan"')
]

```

Tabel validasi & distribusi topik dominan pada dokumen

df_SERP1_validation_table										
Doc		Document_Text_Content			Keywords	Topic	Dominance	Mutual_Keywords	Mutual_Count	J_Score
0	0	['isi', 'kuliah', 'umum', 'ganjar', 'bacapres', 'ganjar', 'pranowo', 'mengisi', 'kuliah', 'umum', 'perguruan', 'kota']			ganjar, umum, kuliah, pranowo, program_gaspol, isi, mengisi, mahasiswa, calon_presiden, jakarta	0	0.9858	{ganjar, isi, mengisi, umum, pranowo, kuliah}	6	46.15
1	1	['ganjar', 'pranowo', 'program_gaspol', 'hadapan', 'ganjar', 'pranowo', 'program_gaspol', 'hadapan', 'mahasiswa', 'jawa_barat', 'foto']			ganjar, umum, kuliah, pranowo, program_gaspol, isi, mengisi, mahasiswa, calon_presiden, jakarta	0	0.7707	{program_gaspol, ganjar, mahasiswa, pranowo}	4	30.77
2	2	['kunjungi', 'ganjar', 'pranowo', 'calon_presiden', 'bacapres', 'ganjar', 'pranowo', 'mengenalkan', 'program_gaspol', 'mengisi', 'kuliah', 'umum', 'bertajuk']			ganjar, umum, kuliah, pranowo, program_gaspol, isi, mengisi, mahasiswa, calon_presiden, jakarta	0	0.9869	{program_gaspol, ganjar, kuliah, mengisi, umum, pranowo, calon_presiden}	7	50.00
3	3	['isi', 'kuliah', 'umum', 'ganjar', 'ingatkan', 'baca', 'faktor', 'ganjar', 'pemilu', 'video', 'rekомендasi', 'pst', 'bantargebang']			karya, mode, batik_lasem, kuliah, mahasiswa, indonesia, tim, penghargaan, seoul, sama	2	0.6273	{kuliah}	1	4.76
4	4	['terima_kasih', 'terima_kasih', 'terima_kasih', 'parahyangan', 'karya', 'wisata', 'episode', 'beliau', 'memaparakan', 'kegiatan']			ganjar, pranowo, kuliah, umum, muda, swasta, fakultas_kedokteran, terima_kasih, ajak, politik	1	0.9819	{terima_kasih}	1	5.88

Pengujian Data Retrieval & Preprocessing SERP II

Data collection

Mengambil data *corporate website* (news.maranatha.edu) dengan cara melakukan *web search* menggunakan Google *search engine*, kemudian dilakukan *scraping* menghasilkan data tabular.

Query formulation II

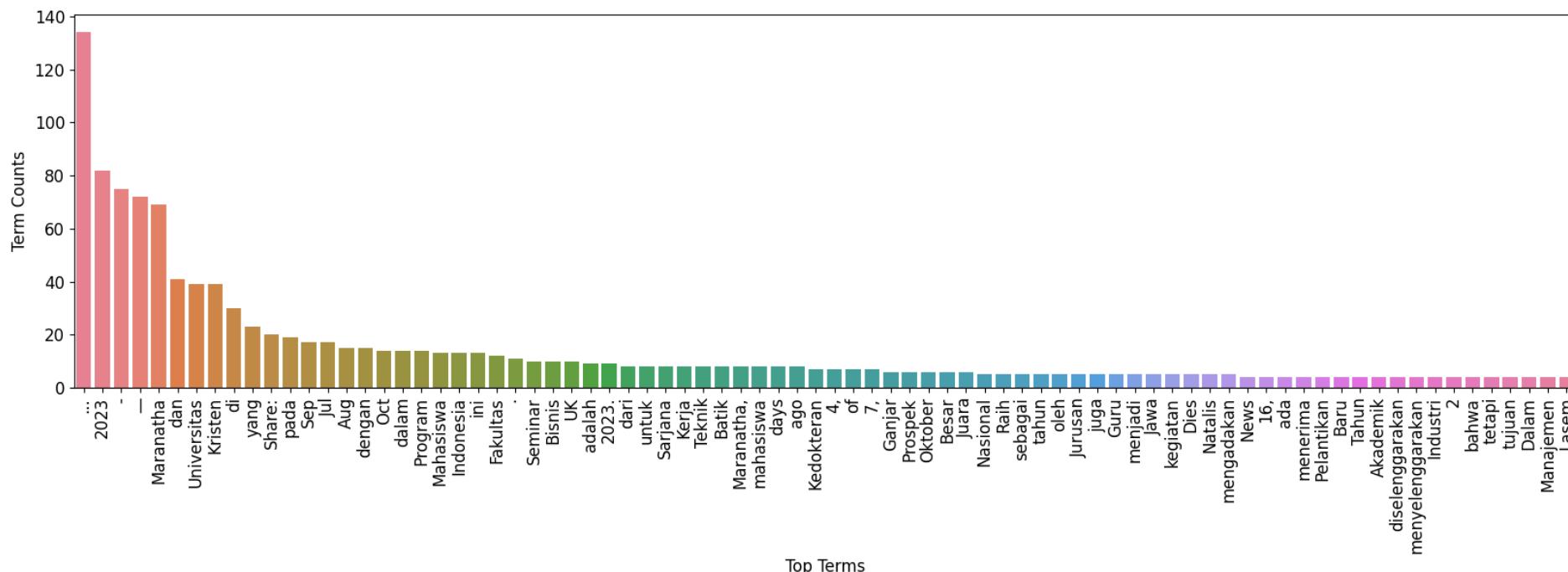
Search Parameter (specific site)	Operator	Periode Tanggal Awal	Operator	Periode Tanggal Akhir
site:[URL]	AND (+)	after:yyyy-mm-dd	AND (+)	before:yyyy-mm-dd

Format dasar *query formulation* untuk tahap pencarian II

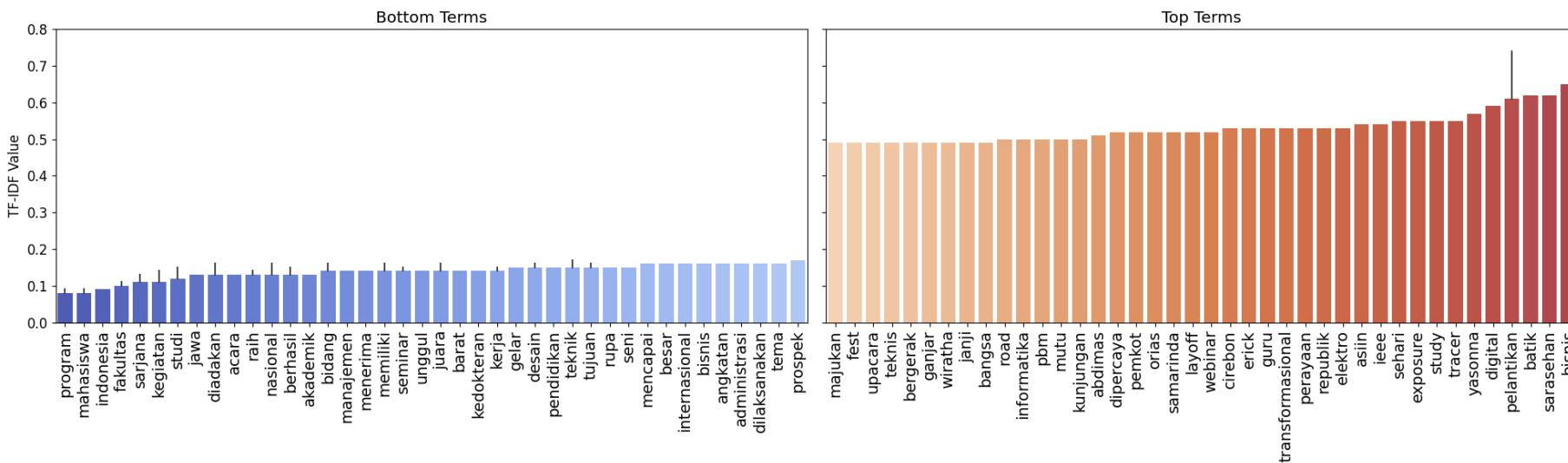
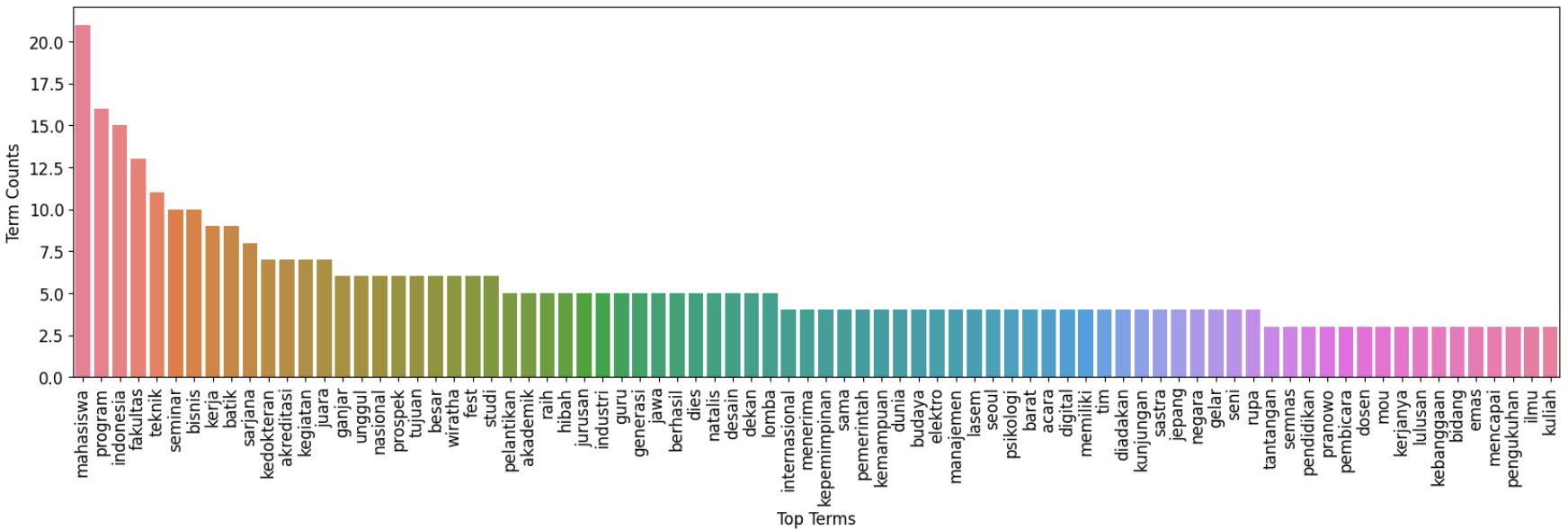
Precleaning

```
hapuskata = ['Berita Resmi Universitas', 'Archive – Maranatha News', 'Archives – Maranatha News', 'Popular News – Maranatha News', 'Contact – Maranatha News',  
'Popular News', 'Popular Tags', 'Editor Pick', 'Untitled',  
'Archive – Page', 'Archives – Page', 'Kata Kunci – Maranatha News', 'Hasil Pencarian', 'Related Photo', 'More News', 'SERP Feature']  
# List daftar kata berisi string tags navigasi dan penanda khas halaman website Maranatha News yang tidak berkontribusi makna pada pemodelan topik  
# "SERP Feature" digenerate oleh Google Search Engine, yaitu hasil tambahan yang tidak dibutuhkan pada pemodelan topik
```

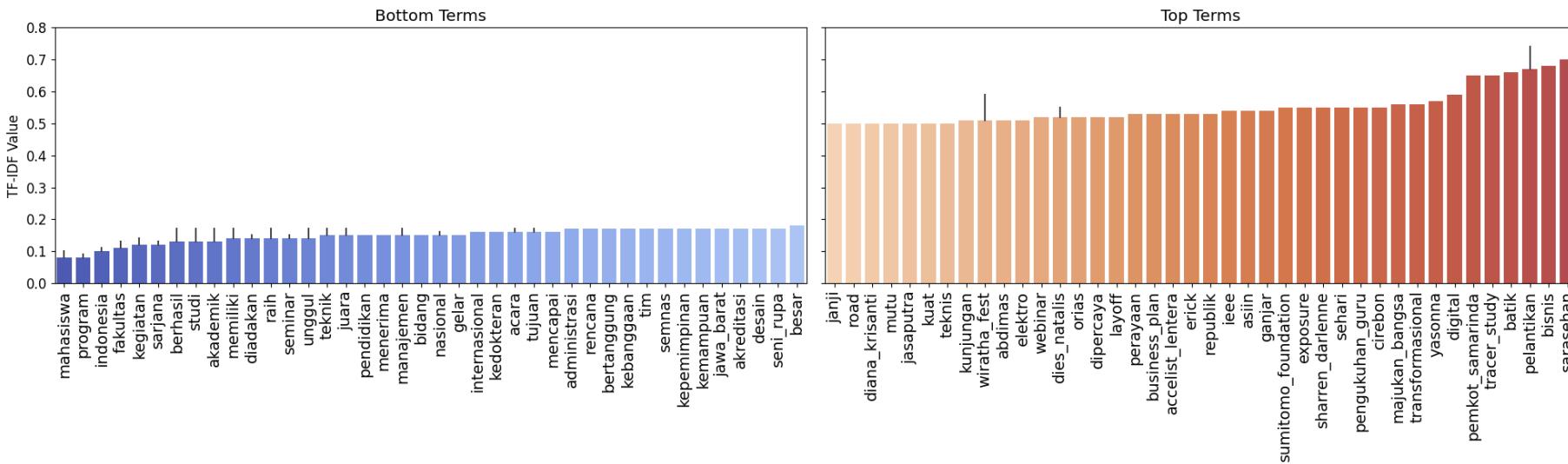
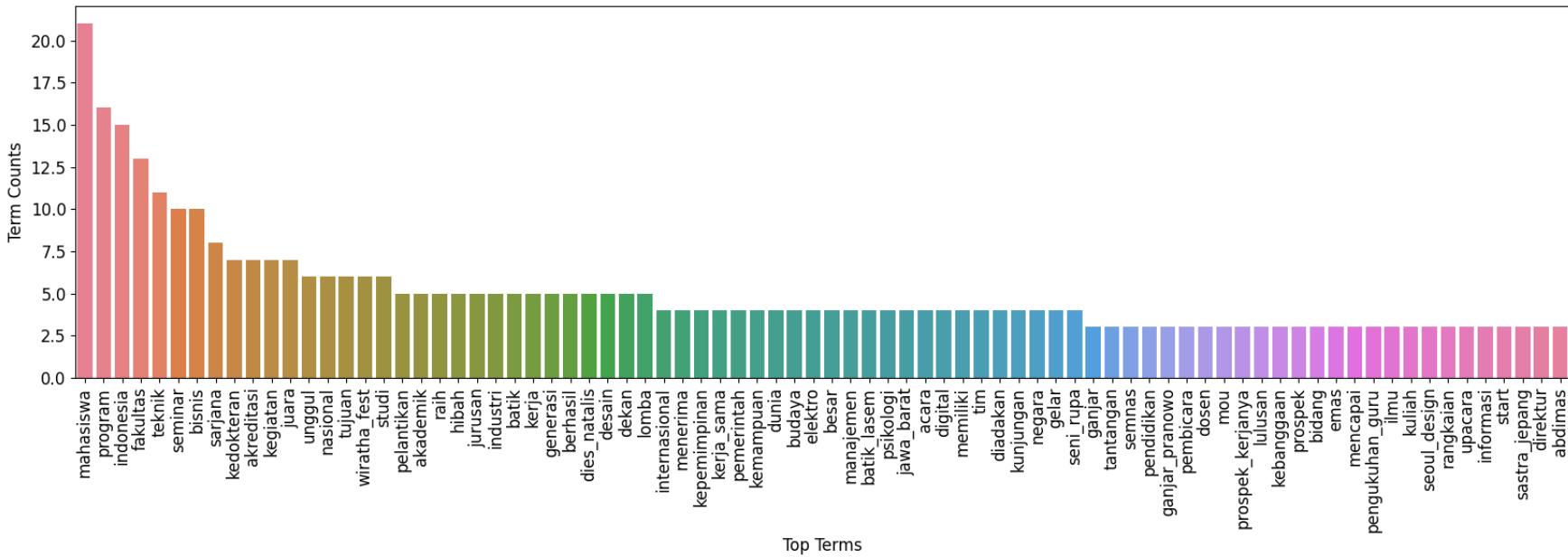
0 Belajar dari Ganjar – Maranatha News – Oct 16, 2023 – Ganjar tahu persis, mengupayakan keadilan dan kesejahteraan untuk semua adalah sebuah tantangan sekaligus harapan. Sebab, ada masalah dalam ha...
1 Semnas "Road to Unggul" (Foto, Video, Materi) – Aug 23, 2023 – Seminar Nasional "Road to Unggul" Universitas Kristen Maranatha berkolaborasi dengan Lembaga Layanan Pendidikan Tinggi (LLDIKTI) Wilia...
2 Kedokteran Maranatha Terakreditasi Internasional ASIIN – Aug 4, 2023 – Program Sarjana Kedokteran dan Profesi Dokter Universitas Kristen Maranatha menerima sertifikat akreditasi internasional ASII...
3 Pelantikan Mahasiswa Baru 2023/2024 – Maranatha News – Sep 16, 2023 – Sidang Terbuka Senat Universitas Kristen Maranatha dalam Rangka Pelantikan Mahasiswa Baru dan Pembukaan Tahun Akademik 2023/20...
4 Semnas "Kepemimpinan Transformasional" Ganjar Pranowo – Oct 11, 2023 – Universitas Kristen Maranatha menyelenggarakan Seminar Nasional 'Kepemimpinan Transformasional yang Menyejahterakan Wong Cili...
5 10 Penelitian Dosen Raih Hibah Kemendikbudristek 2023 – Jul 11, 2023 – Sepuluh penelitian dosen Universitas Kristen Maranatha terpilih sebagai penerima hibah dari Kemendikbudristek tahun 2023. Hib...
6 Maranatha Jalin Kerja Sama dengan Pemkot Samarinda – Sep 5, 2023 – Universitas Kristen Maranatha menandatangani Memorandum of Understanding (MoU) dengan Pemerintah Kota (Pemkot) Samarinda. Penanda...
7 Mengenal Jurusan Teknik Industri dan Prospek Kerjanya – Jul 7, 2023 – Sebagai sales engineer, lulusan teknik industri bertanggung jawab dalam menjual produk teknis dengan kemampuan teknis yang mer...
8 Bangkitnya Batik Wajah Indonesia di Hadapan Dunia – Oct 7, 2023 – Share: ... Batik adalah kebanggaan budaya Indonesia yang diakui dunia. UNESCO telah menetapkan batik sebagai warisan budaya nonben...
9 13 Prospek Kerja Teknik Elektro untuk Rencana Kariermu – Jul 17, 2023 – Prospek kerja teknik elektro selanjutnya adalah di bidang manufaktur. Sebagai ahli teknik elektro di sektor manufaktur, para...
10 Menkumham Yasonna H. Laoly Ungkap Tantangan Wuiudkan ... – Jul 12, 2023 – Yasonna menielaskan bahwa Indonesia Emas 2045 bukan sekadar impian, tetapi juga sebuah tujuan yang harus diwujudkan



Distribusi kata unigram

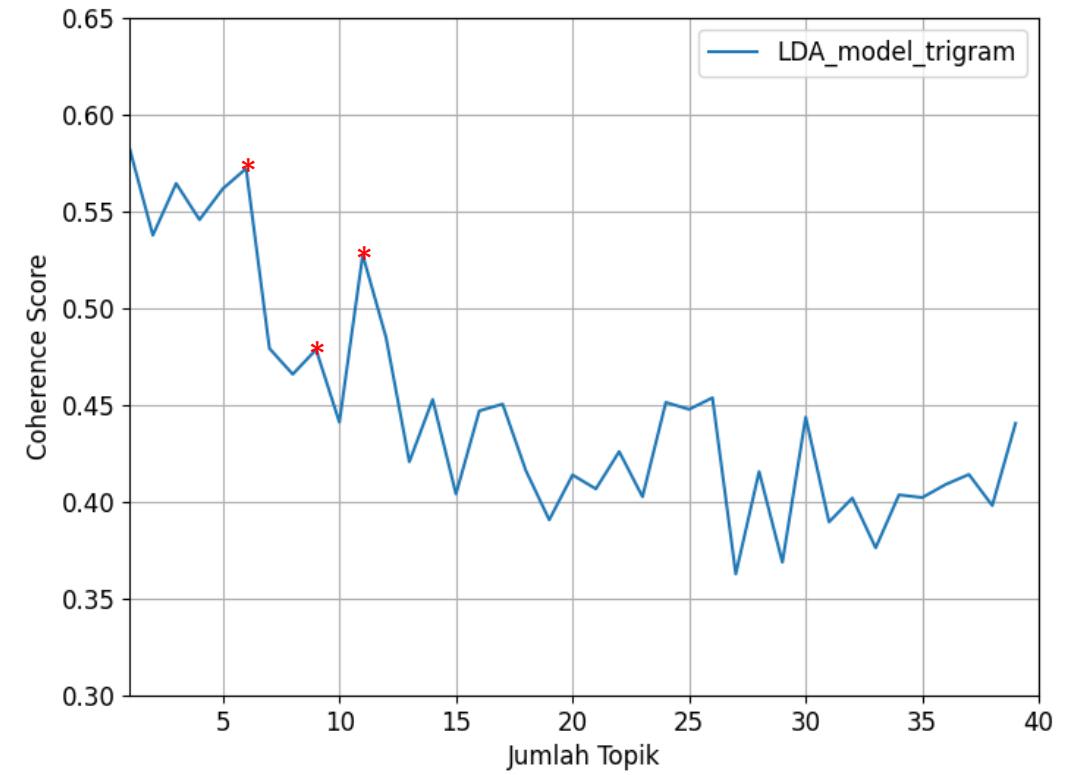
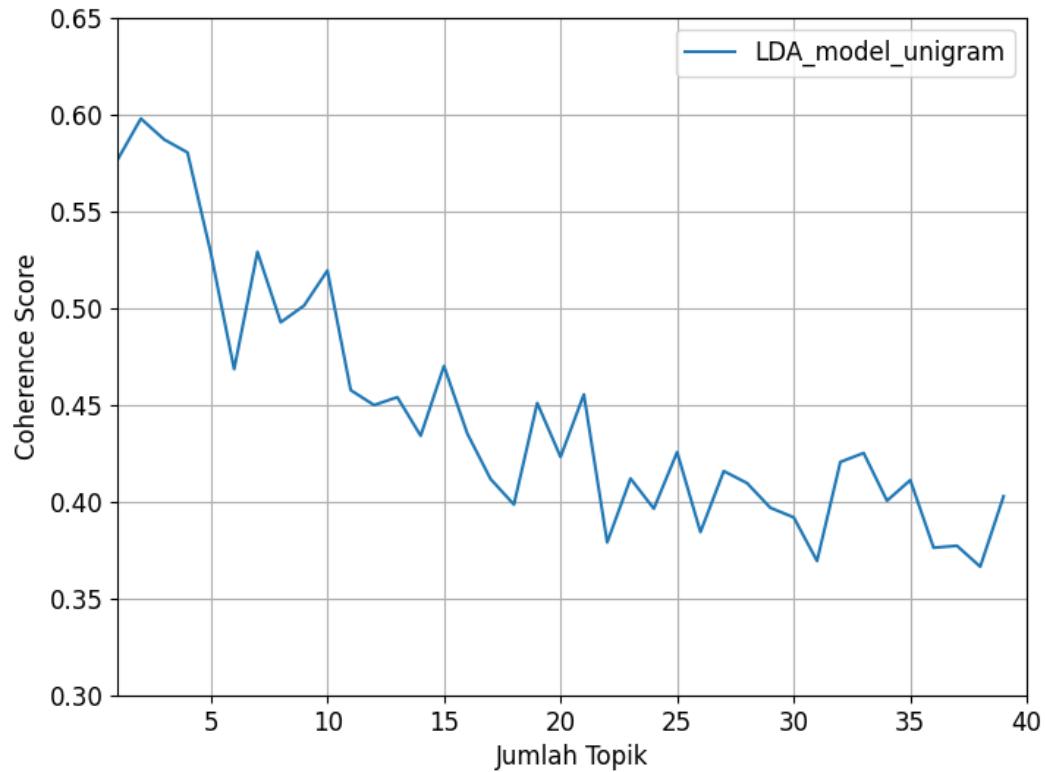


Distribusi kata bigram/trigram



Pengujian Pemodelan Topik SERP II

Buat 40 model initial, lalu hitung coherence score

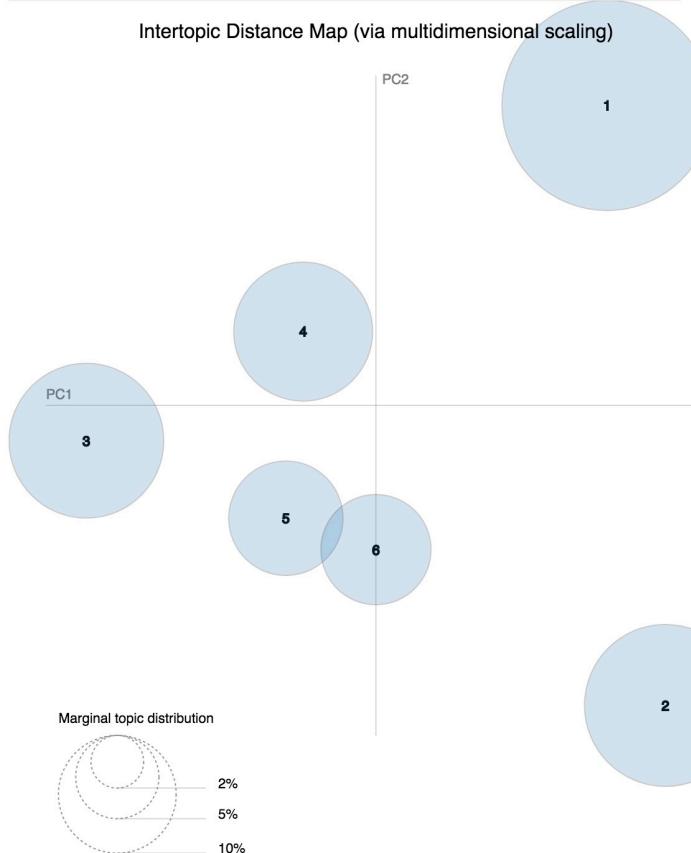


Analisis & validasi visual pyLDAvis (melihat keberhasilan model memisahkan topik)

vis6_t

Selected Topic:

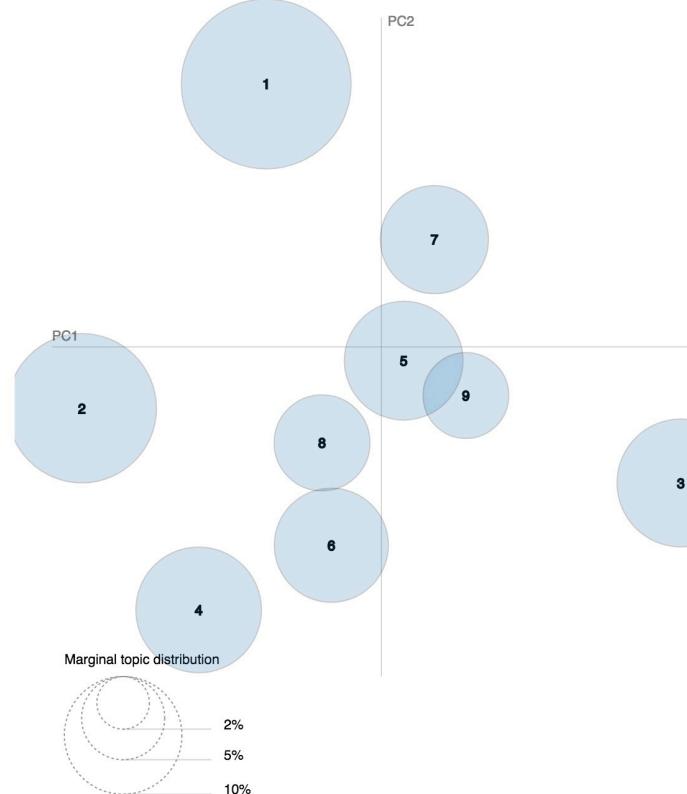
Intertopic Distance Map (via multidimensional scaling)



vis9_t

Selected Topic:

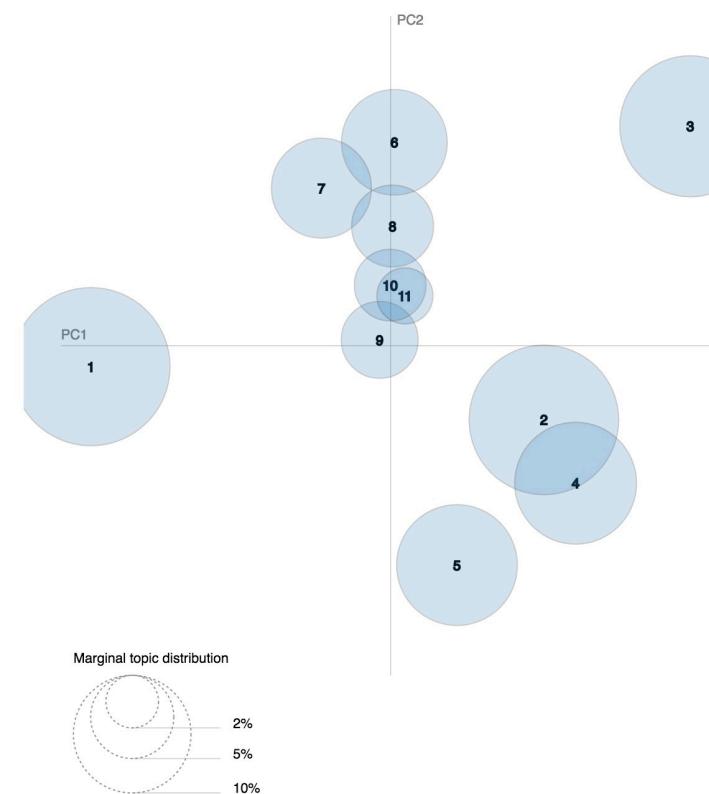
Intertopic Distance Map (via multidimensional scaling)



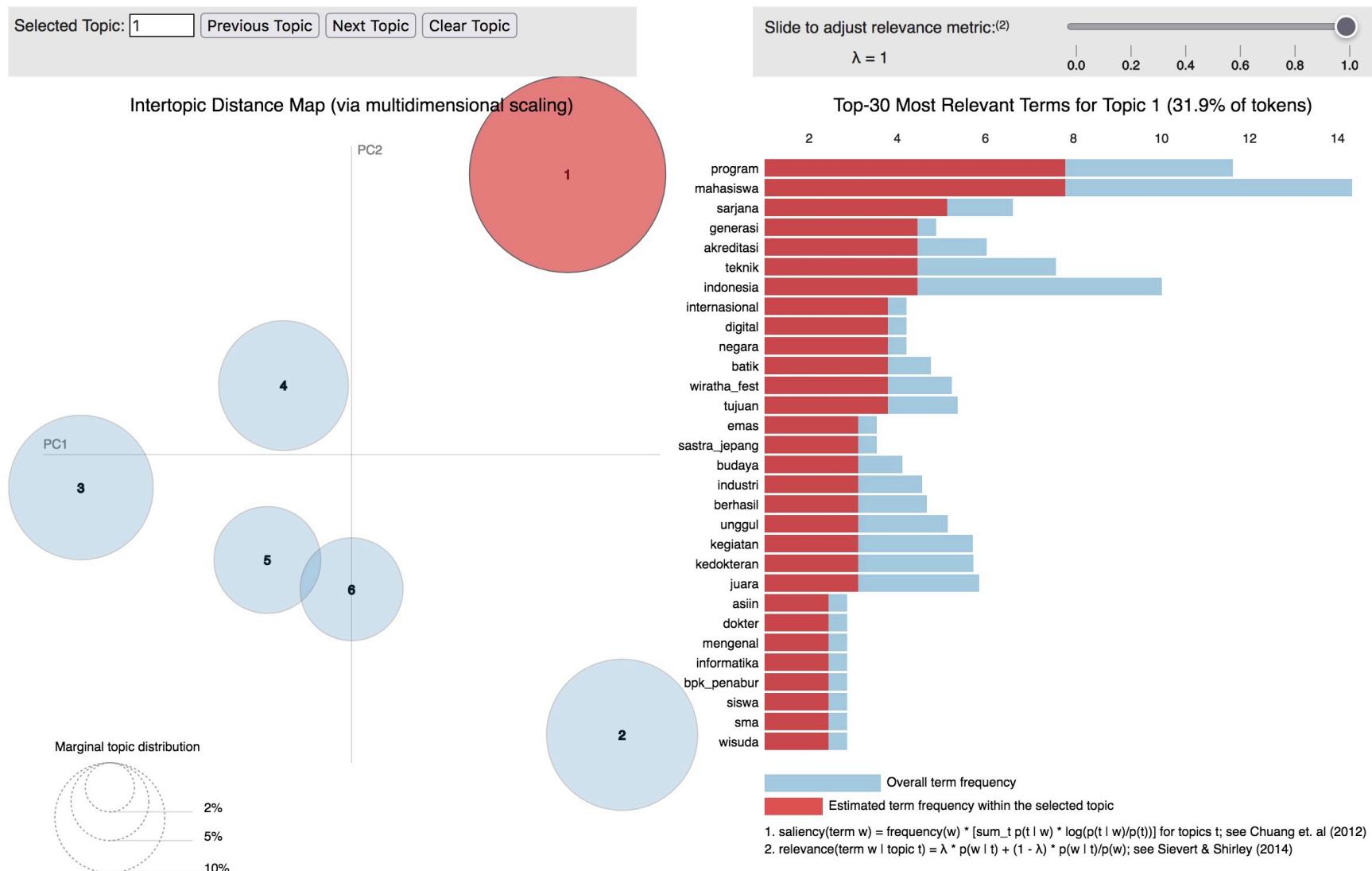
vis11_t

Selected Topic:

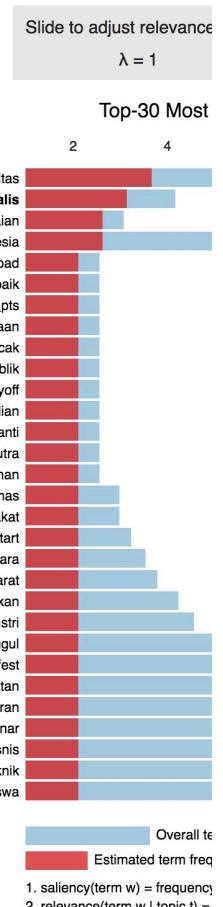
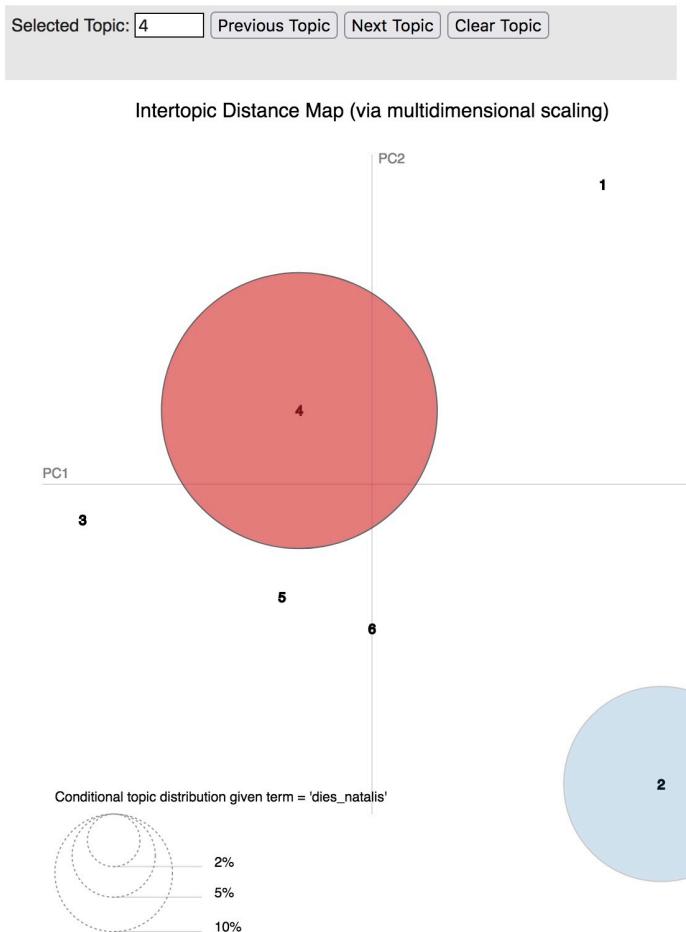
Intertopic Distance Map (via multidimensional scaling)



Analisis visual pyLDAvis (melihat distribusi most relevant terms pada topic 1)



Analisis visual pyLDAvis (melihat distribusi term "dies_natalis" dan "wiratha_fest")



Distribusi terms pada topic (model pilihan adalah dengan jumlah topik = 6)

```

num_topics = 6
x = num_topics-1
optimal_model_lda_trigram = model_list_lda_trigram_high[x]
optimal_model_lda_trigram.show_topics(num_topics)

[(0,
  '0.017*"elektro" + 0.017*"seminar" + 0.017*"pembicara" + 0.017*"kepemimpinan" + 0.017*"ganjar_pranowo" + 0.017*"hibah" + 0.017*"teknik" + 0.011*"kemendikbudristek" + 0.011*"kerja" +
  0.011*"penelitian"),
(1,
  '0.024*"mahasiswa" + 0.014*"studi" + 0.014*"juara" + 0.011*"mou" + 0.011*"seni_rupa" + 0.011*"kerja_sama" + 0.011*"desain" + 0.011*"kabupaten" + 0.011*"program" + 0.011*"pemerintah"),
(2,
  '0.021*"fakultas" + 0.017*"dies_natalis" + 0.013*"indonesia" + 0.013*"rangkaian" + 0.009*"start" + 0.009*"perayaan" + 0.009*"kedokteran" + 0.009*"layoff" + 0.009*"wiratha_fest" +
  0.009*"abdimas"),
(3,
  '0.023*"mahasiswa" + 0.023*"program" + 0.014*"sarjana" + 0.012*"generasi" + 0.012*"akreditasi" + 0.012*"teknik" + 0.012*"indonesia" + 0.010*"wiratha_fest" + 0.010*"internasional" +
  0.010*"batik"),
(4,
  '0.011*"batik_lasem" + 0.011*"indonesia" + 0.011*"penghargaan" + 0.011*"menerima" + 0.011*"sehari" + 0.011*"seminar" + 0.011*"exposure" + 0.011*"kunjungan" + 0.006*"pemenang" + 0.006*"seoul"),
(5,
  '0.029*"bisnis" + 0.011*"fakultas" + 0.011*"kuliah" + 0.011*"manajemen" + 0.011*"dekan" + 0.008*"jurusan" + 0.008*"bantu" + 0.008*"lomba" + 0.008*"prospek_kerjanya" + 0.008*"alat")]

```

Tabel validasi & distribusi topik dominan pada dokumen

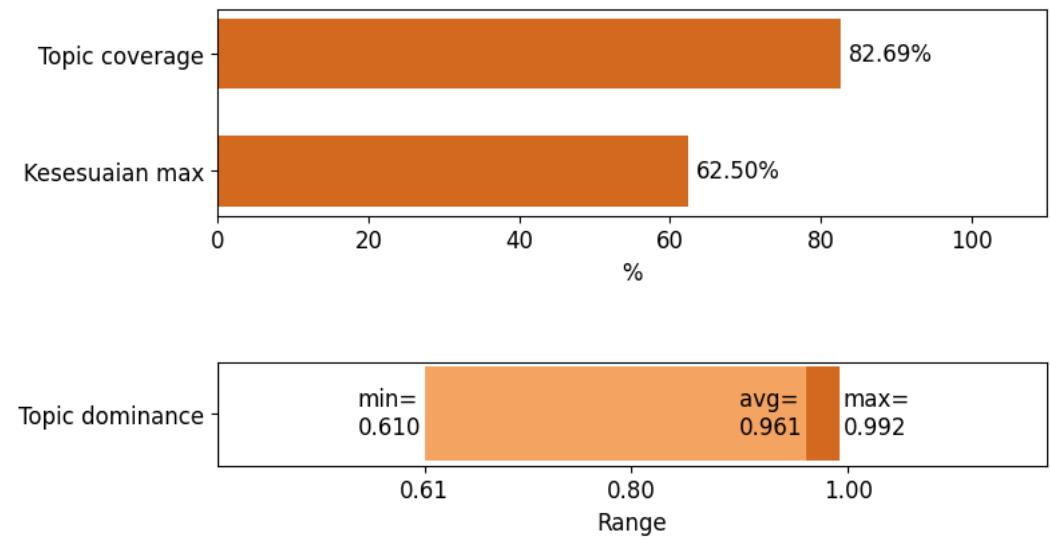
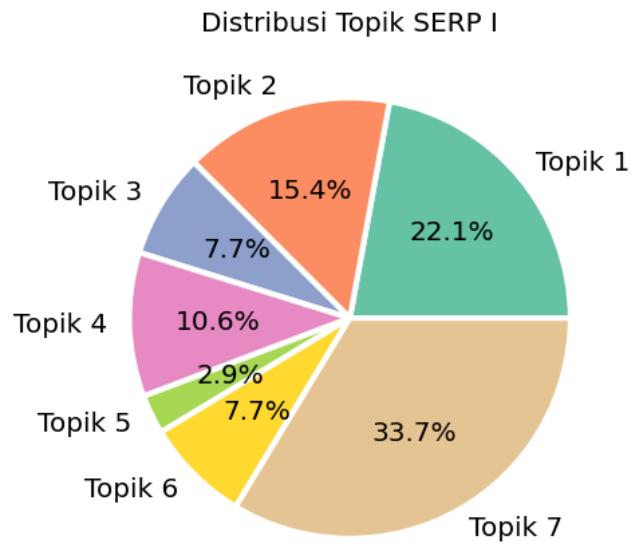
df_SERP2_validation_table									
Doc		Document_Text_Content	Keywords	Topic	Dominance	Mutual_Keywords	Mutual_Count	J_Score	
0	0	['belajar', 'ganjar', 'ganjar', 'persis', 'mengupayakan', 'keadilan', 'kesejahteraan', 'tantangan', 'harapan', 'regulasi']	bisnis, fakultas, kuliah, manajemen, dekan, jurusan, bantu, lomba, prospek_kerjanya, alat	5	0.9878		0	0.00	
1	1	['semnas', 'road', 'unggul', 'foto', 'video', 'materi', 'seminar', 'nasional', 'road', 'unggul', 'berkolaborasi', 'lembaga', 'layanan', 'pendidikan', 'Illdikti', 'wilayah', 'ban']	fakultas, dies_natalis, indonesia, rangkaian, start, perayaan, kedokteran, layoff, wiratha_fest, abdimas	2	0.9927		0	0.00	
2	2	['kedokteran', 'terakreditasi', 'internasional', 'asiin', 'program', 'sarjana', 'kedokteran', 'profesi', 'dokter', 'menerima', 'sertifikat', 'akreditasi', 'internasional', 'asin', 'tertanggal']	mahasiswa, program, sarjana, generasi, akreditasi, teknik, indonesia, wiratha_fest, internasional, batik	3	0.9929	{sarjana, internasional, program, akreditasi}	4	22.22	
3	3	['pelantikan', 'mahasiswa', 'sidang_terbuka', 'senat_rangka', 'pelantikan', 'mahasiswa', 'pembukaan', 'akademik']	mahasiswa, program, sarjana, generasi, akreditasi, teknik, indonesia, wiratha_fest, internasional, batik	3	0.9867	{mahasiswa}	1	6.67	
4	4	['semnas', 'kepemimpinan', 'transformasional', 'ganjar_pranowo', 'seminar', 'nasional', 'kepemimpinan', 'transformasional', 'menyejahterakan', 'wong', 'cilik', 'pembicara', 'ganjar']	elektro, seminar, pembicara, kepemimpinan, ganjar_pranowo, hibah, teknik, kemendikbudristek, kerja, penelitian	0	0.9901	{seminar, kepemimpinan, ganjar_pranowo, pembicara}	4	23.53	

Pengujian Komparasi dan Output Visualisasi

Visualisasi word cloud SERP I



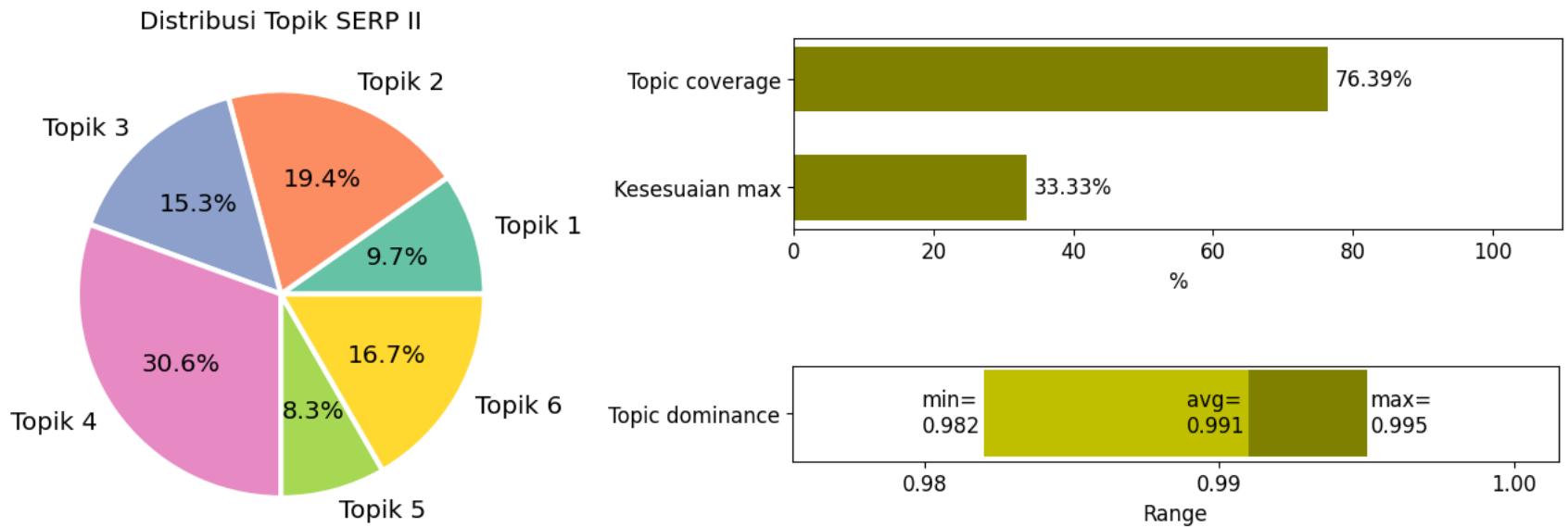
Metrics report SERP I



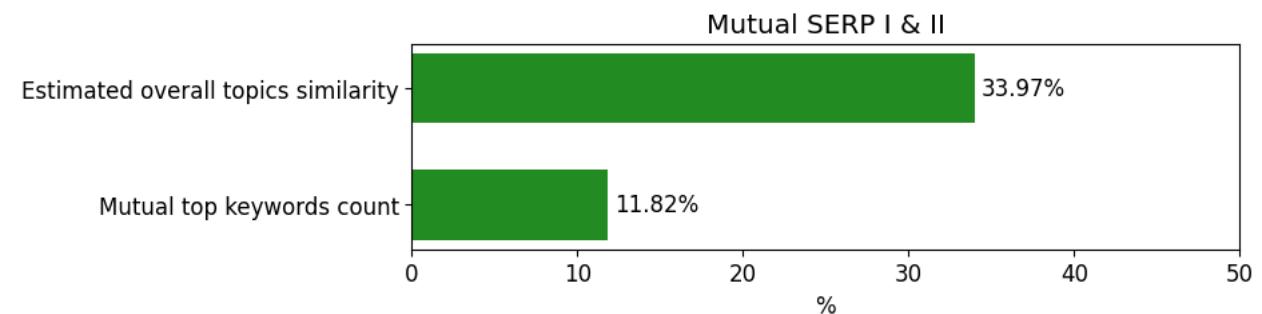
Visualisasi word cloud SERP II



Metrics report SERP II



Visualisasi mutual keywords & metrics report



Mutual top keywords (single terms) antara SERP I dan SERP II:

['program', 'kuliah', 'fakultas', 'indonesia', 'kedokteran', 'penghargaan', 'pranowo', 'mahasiswa', 'sama', 'seoul', 'ganjar', 'lasem', 'batik']

IV

HASIL & PEMBAHASAN

1.

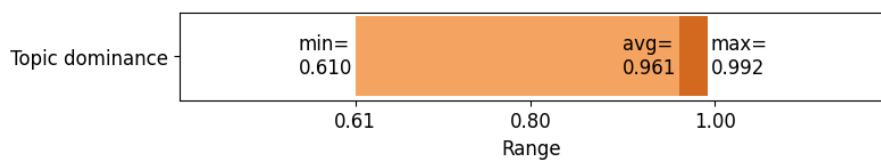
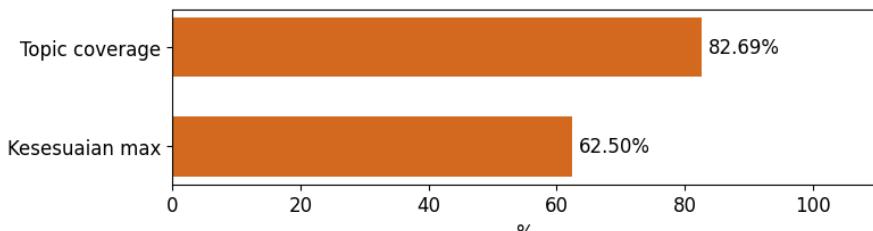
Analisis & Interpretasi Topik SERP I

Model topik SERP I menghasilkan 7 kelompok dominan

Grup Topik	Interpretasi Topik
Topik 1	kuliah umum Ganjar Pranowo mengenai program gaspol
Topik 2	kuliah umum Ganjar Pranowo berkaitan dengan politik dan generasi muda; kuliah fakultas kedokteran; perguruan tinggi swasta
Topik 3	penghargaan batik Lasem di Seoul; penghargaan karya mahasiswa; Program Desain Mode
Topik 4	pendaftaran kuliah; informasi fasilitas kampus; informasi Fakultas Hukum
Topik 5	materi ujian <i>online</i>
Topik 6	dosen Fakultas Kedokteran membeberkan pencegahan <i>stunting</i> ; informasi mengenai Fakultas Psikologi
Topik 7	kuliah umum Ganjar Pranowo mengenai program keluarga miskin di Jawa Barat

Interpretasi topik dominan pada SERP I

Analisis metrik SERP I



- *Topic coverage* → topik-topik yang dibentuk oleh model dapat mencakup 82,69 % dari keseluruhan topik yang terdapat pada SERP I.
- Sisanya (17,31%) → topik yang tidak tercakup (topik yang gagal diekstraksi, salah penempatan topik, dan dokumen yang frekuensinya sangat rendah).
- Kesesuaian maksimal → mengindikasikan adanya topik yang dapat merepresentasikan dokumen aslinya hingga sebesar 62,5% dari keseluruhan *terms* gabungan. Semakin besar angka kesesuaian, topik yang terbentuk cenderung semakin homogen (tidak terdiri dari banyak subtopik).
- Tingkat dominansi cenderung tinggi (rata-rata 0,961 dari skala 0 s.d. 1) → mengindikasikan lebih banyak dokumen yang direpresentasikan oleh topik yang dominan (berbobot tinggi).

Analisis tabel validasi SERP I

Topik Dominan	Berkesesuaian (% Doc-per-Topik)	Tidak Berkesesuaian (% Doc-per-Topik)	Doc-per-Topik Count (%)
Topik 1	73,91	26,09	22,12
Topik 2	93,75	6,25	15,38
Topik 3	87,50	12,50	7,69
Topik 4	90,91	9,09	10,58
Topik 5	33,33	66,67	2,88
Topik 6	75,00	25,00	7,69
Topik 7	85,71	14,29	33,65
Overall	82,69	17,31	100,00

Persentase rasio kesesuaian dokumen-per-topik-dominan untuk masing-masing topik dan keseluruhan topik (SERP I)

- Topik 2 memiliki nilai kesesuaian tertinggi di antara topik-topik lainnya (93,75%) → hampir seluruh dokumen dalam kelompok ini membahas tentang topik 2.
- Topik 5 memiliki nilai ketidaksesuaian tertinggi (66,67%) → dokumen-dokumen pada topik 5 cenderung tidak sesuai dengan representasi topik.
- Ketidaksesuaian yang cukup tinggi pada topik 5 ini secara keseluruhan berkontribusi rendah karena hanya terdiri dari 2,88% dokumen saja.

2.

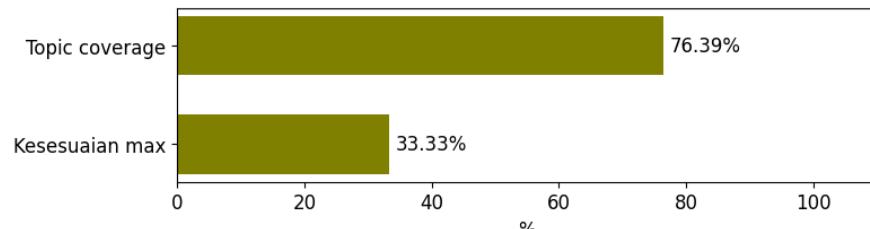
Analisis & Interpretasi Topik SERP II

Model topik SERP II menghasilkan 6 kelompok dominan

Grup Topik	Interpretasi Topik
Topik 1	seminar kepemimpinan dengan pembicara Ganjar Pranowo; peluang kerja teknik elektro; kegiatan penelitian hibah Kemendikbudristek
Topik 2	aktivitas mahasiswa Fakultas Seni Rupa dan Desain; perolehan juara; kerja sama dengan pemerintah
Topik 3	Rangkaian kegiatan dies natalis dan Wiratha Fest; kegiatan nasional; kegiatan pengabdian masyarakat; kegiatan Fakultas Kedokteran
Topik 4	kegiatan mahasiswa program sarjana; program akreditasi; tema mengenai Indonesia
Topik 5	karya batik Lasem menerima penghargaan di Seoul; kegiatan seminar dan kunjungan
Topik 6	prospek kerja dan kegiatan Fakultas Bisnis dan jurusan manajemen; kegiatan melibatkan dekan; kegiatan lomba

Interpretasi topik dominan pada SERP II

Analisis metrik SERP II



- Topic *coverage* → topik-topik yang dibentuk oleh model dapat mencakup 76,39% dari keseluruhan topik yang terdapat pada SERP II.
- Sisanya (23,61%) → topik yang tidak tercakup (topik yang gagal diekstraksi, salah penempatan topik, dan dokumen yang frekuensinya sangat rendah).
- Kesesuaian maksimal → mengindikasikan adanya topik yang dapat merepresentasikan dokumen aslinya hingga sebesar 33,33% dari keseluruhan *terms* gabungan. Semakin kecil angka kesesuaian, topik yang terbentuk cenderung semakin heterogen (terdiri dari banyak subtopik).
- Tingkat dominansi tinggi (0,982 sampai 0,995 dari skala 0 s.d. 1) → mengindikasikan hampir semua dokumen direpresentasikan oleh topik yang sangat dominan (berbobot tinggi).

Analisis tabel validasi SERP II

Topik Dominan	Berkesesuaian (% Doc-per-Topik)	Tidak Berkesesuaian (% Doc-per-Topik)	Doc-per-Topik Count (%)
Topik 1	57,14	42,86	9,72
Topik 2	64,29	35,71	19,44
Topik 3	81,82	18,18	15,28
Topik 4	95,45	4,55	30,56
Topik 5	66,67	33,33	8,33
Topik 6	66,67	33,33	16,67
Overall	76,39	23,61	100,00

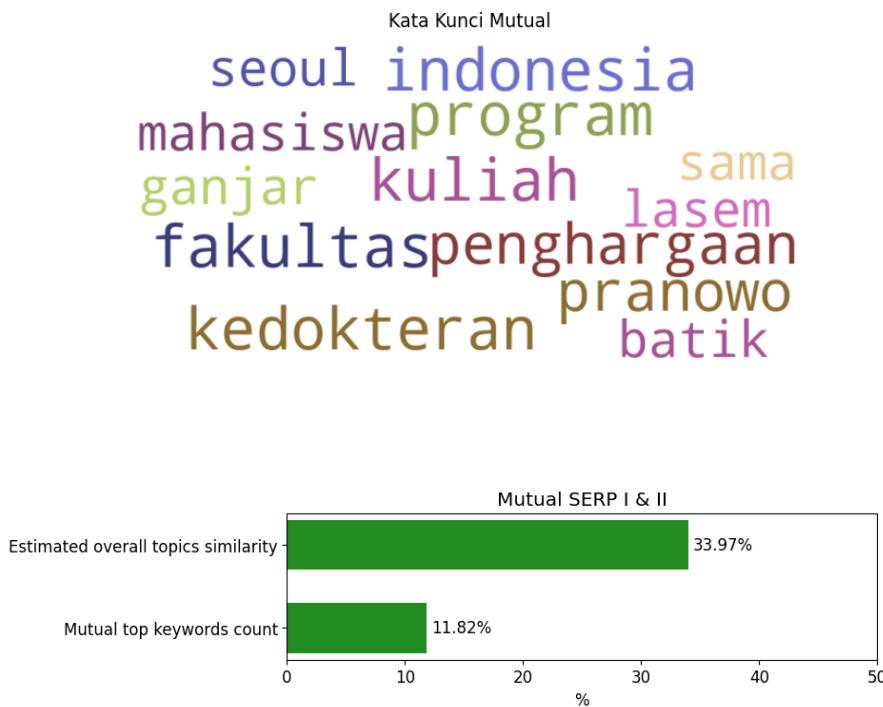
Percentase rasio kesesuaian dokumen-per-topik-dominan untuk masing-masing topik dan keseluruhan topik (SERP II)

- Topik 4 memiliki nilai kesesuaian tertinggi di antara topik-topik lainnya (95,45 %) → hampir seluruh dokumen dalam kelompok ini membahas tentang topik 4.
- Topik 1 memiliki nilai ketidaksesuaian tertinggi (42,86%) → dokumen-dokumen pada topik 1 cukup banyak yang tidak sesuai dengan representasi topik.
- Ketidaksesuaian yang cukup tinggi pada topik 1 ini secara keseluruhan berkontribusi rendah karena hanya terdiri dari 9,72% dokumen saja.

3.

Analisis Komparasi Topik SERP I & SERP II

Analisis komparasi topik

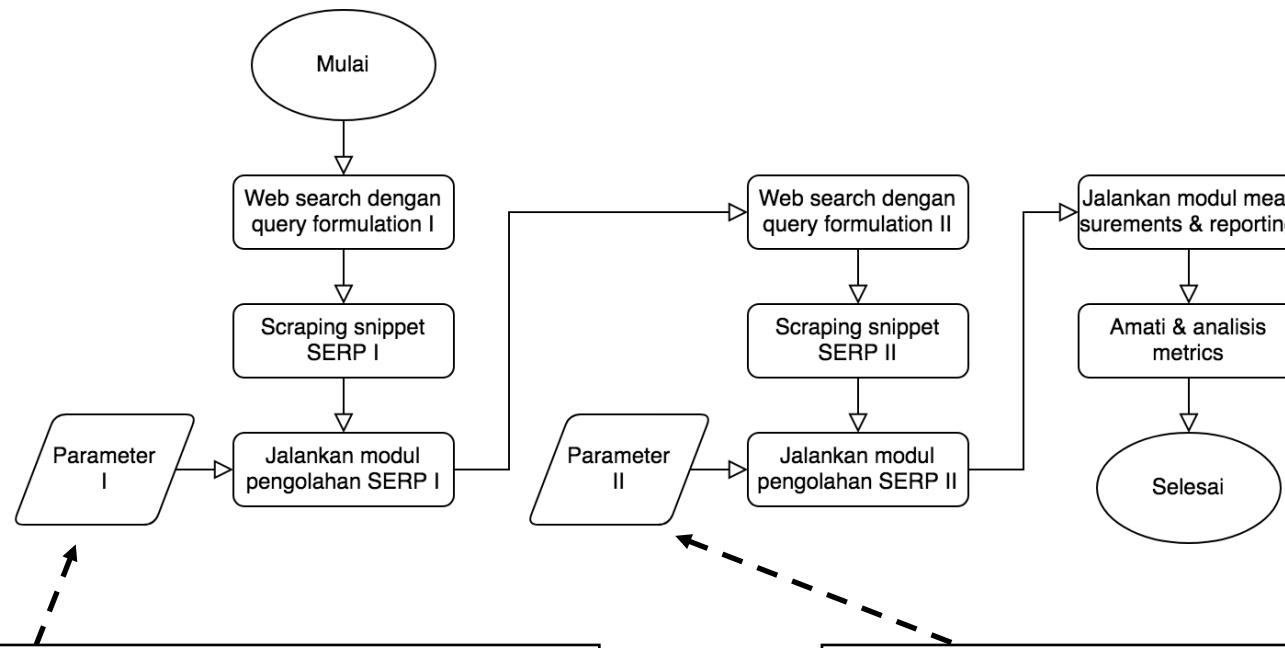


- Terdapat kesamaan isu pada situs Maranatha News dan situs media massa sebesar 33,97%.
- Terdapat kesamaan *keywords (mutual terms)* sebanyak 11,82%.
- Kesamaan topik paling dominan adalah mengenai Ganjar Pranowo, batik Lasem, dan Fakultas Kedokteran → mengindikasikan *brand exposure* yang lebih luas dari topik ini, paparannya mencakup kedua kategori media.
- Topik mengenai Ganjar Pranowo homogen pada topik 7 SERP I dengan distribusi tertinggi (33,7%), dan pada 2 topik lainnya; sedangkan di SERP II hanya ditampilkan pada topik 1 (bobot rendah) → mengindikasikan jumlah artikel Maranatha News yang membahas topik ini sangat sedikit.
- Ketidaksamaan di luar 33,97% → mengindikasikan masing-masing kategori media memiliki topik-topik khas yang berlainan.

4.

Ringkasan Alur Pengujian Keseluruhan

Langkah implementasi keseluruhan pengujian



- *Filtering*: dictionary untuk mengecualikan sumber-sumber data dari situs yang tidak relevan
- *Precleaning/cleaning*: dictionary untuk mengecualikan dokumen yang mengindikasikan artikel kurang relevan, berdasarkan pola kata-kata khas
- *Preprocessing*: transformasi *corpus* dalam bentuk *bigram/trigram*
- Pemodelan topik: $K=7$, dipilih dari kandidat topik antara $K=5$ hingga 15

- *Precleaning/cleaning*: dictionary untuk mengecualikan dokumen yang mengindikasikan halaman indeks dan/atau berisi *tag navigasi*, dan menghapus tanda-tanda penunjuk waktu terbit artikel dan *noise* lainnya
- *Preprocessing*: transformasi *corpus* dalam bentuk *bigram/trigram*
- Pemodelan topik: $K=6$, dipilih dari kandidat topik antara $K=5$ hingga 15

5. Diskusi

Poin-poin diskusi dan temuan penting

Keberhasilan rancangan kerangka kerja dapat dilihat dari dua aspek:

1. Aspek kebermanfaatan kerangka kerja bagi *stakeholder* institusi.
2. Aspek teknis kerangka kerja terkait penerapan metode-metode pengolahan data.

V

SIMPULAN & SARAN

Simpulan ketercapaian tujuan (1)

Tujuan pertama: merancang model kerangka kerja *brand exposure monitoring* untuk mendukung dan meningkatkan kinerja tim pemonitor di institusi studi kasus, yang dapat menghasilkan *report* yang mudah dipahami pengambil keputusan.

→ Tercapai dengan dibuatnya rancangan kerangka kerja FBxM yang telah diuji melalui eksperimen dan langkah-langkah validasi yang telah dipaparkan.

Simpulan ketercapaian tujuan (2)

Tujuan kedua: mengolah informasi yang dihasilkan oleh *search engine* menggunakan metode-metode *machine learning* dan *natural language processing* (NLP) yang sesuai dengan kebutuhan kerangka kerja.

→ Tercapai dengan pemanfaatan metode *query formulation*, *data retrieval*, *preprocessing*, dan pemodelan topik LDA pada *pipeline* pengolahan data dalam kerangka kerja FBxM.

Simpulan ketercapaian tujuan (3)

Tujuan ketiga: mengekstraksi informasi hasil pencarian *corporate website* dan *external media* untuk menemukan topik yang terkandung di dalamnya.

→ Tercapai dengan terbentuknya model topik yang berkinerja cukup baik, dan dapat menghasilkan output berupa kelompok-kelompok topik yang merepresentasikan dokumen berupa SERP *snippet* hasil pencarian menggunakan *search engine*.

Simpulan ketercapaian tujuan (4)

Tujuan keempat: menentukan indikator dan metrik yang dibutuhkan sebagai salah satu bentuk output dari kerangka kerja.

→ Tercapai dengan penerapan perhitungan similaritas, *mutual keywords*, dan metode visualisasi *metrics report* berupa diagram dan penyajian *word cloud* → berfungsi sebagai *dashboard* yang menampilkan indikator-indikator yang dibutuhkan untuk analisis selanjutnya pada tingkat pengambil keputusan.

Saran (1)

1. Metode pemodelan topik yang digunakan dalam kerangka kerja pada penelitian ini dapat digunakan sebagai *baseline* untuk penelitian-penelitian selanjutnya, yaitu menggunakan metode ekstraksi topik lainnya, guna memperoleh performa yang lebih baik.
2. Metode yang telah diuji pada penelitian ini memiliki keterbatasan, yaitu terbentuknya topik heterogen yang terdiri dari beberapa subtopik, sehingga ruang untuk mengatasi keterbatasan ini masih terbuka.
3. Penelitian selanjutnya dapat dilakukan untuk menemukan metode ekstraksi informasi atau pemodelan topik dengan kinerja yang lebih baik untuk mengatasi keterbatasan, sehingga kebermanfaatan kerangka kerja dapat lebih ditingkatkan.

Saran (2)

4. Penelitian lanjutan juga masih diperlukan untuk memperluas penerapan kerangka kerja FBxM agar dapat diterapkan juga di institusi selain tempat studi kasus, dengan penyesuaian-penyesuaian spesifik sesuai konteks dan kebutuhan.
5. Model prototipe kerangka kerja yang telah dirancang dapat dikembangkan lebih lanjut untuk implementasi berupa aplikasi yang *user friendly*, sehingga dapat dioperasikan secara mandiri oleh tim *monitoring* yang tidak memiliki latar belakang atau kemampuan teknis pemrograman.

LAMPIRAN

DAFTAR PUSTAKA

- [1] P. Kotler and K. L. Keller, *Marketing management*, 15 [edition]. Boston: Pearson, 2016.
- [2] D. A. Aaker, *Managing Brand Equity: Capitalizing on the Value of a Brand Name*. DIANE Publishing Company, 2000.
- [3] T. A. Shimp and J. C. Andrews, *Advertising, promotion, and other aspects of integrated marketing communications*, 9th ed. Mason, Ohio: Sout - Western Cengage Learning, 2013.
- [4] K. L. Keller, "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity," *Journal of Marketing*, vol. 57, no. 1, pp. 1–22, Jan. 1993, doi: 10.1177/002224299305700101.
- [5] T. B. Cornwell, G. E. Relyea, R. L. Irwin, and I. S. J. Maignan, "Understanding long-term effects of sports sponsorship: role of experience, involvement, enthusiasm and clutter," *International Journal of Sports Marketing and Sponsorship*, vol. 2, pp. 127–144, 2000.
- [6] B. Pitts and J. Slattery, "An examination of the effects of time on sponsorship awareness levels," *Sport Marketing Quarterly*, vol. 13, pp. 43–54, Jan. 2006.
- [7] R. Arthana, Noermijati, and C. Susilowati, "Pengaruh Brand Exposure Dan Brand Experience Terhadap Brand Trust Dan Brand Recall (Studi pada Produk Smartphone di Wilayah Kota Malang)," *Ekonomi Bisnis*, vol. 21, no. 2, Art. no. 2, Oct. 2016, Accessed: Mar. 19, 2023. [Online]. Available: <http://journal.um.ac.id/index.php/ekobis/article/view/7432>
- [8] L. Pomalaa, Y. T. Mursityo, and A. D. Herlambang, "Analisis Faktor Brand Awareness, Brand Exposure, Customer Engagement, Dan Electronic Word-of-Mouth Dalam Pemasaran Melalui Media Sosial Pada The Body Shop Indonesia".
- [9] K. T. Tran, P. V. Nguyen, H. T. S. Do, and L. T. Nguyen, "University students' insight on brand equity," 10.5267/j.msl, pp. 2053–2062, 2020, doi: 10.5267/j.msl.2020.2.006.
- [10] M. Pinar, "University Branding and Measuring Brand Equity: Conceptual Framework and Empirical Studies," *The Journal of International Scientific Researches*, vol. 5, no. Ek, pp. 1–5, Dec. 2020, doi: 10.23834/isrjournal.839937.
- [11] T. Gatfield, M. Barker, and P. Graham, "Measuring communication impact for university advertising materials," *Corporate communications : an international journal*, vol. 4, no. 2, 1999.
- [12] B. Gray, V. Llanes, and K. Fam, "Branding universities in Asian markets.," 2003, Accessed: Mar. 19, 2023. [Online]. Available: <https://repository.openpolytechnic.ac.nz/handle/11072/295>

- [13] M. Pinar, P. Trapp, T. Girard, and T. E. Boyt, "University brand equity: An empirical investigation of its dimensions," *International Journal of Educational Management*, vol. 28, no. 6, pp. 616–634, Aug. 2014, doi: 10.1108/IJEM-04-2013-0051.
- [14] N. Sujchaphong, B. Nguyen, and T. Melewar, "Towards a branding oriented higher education sector: An overview of the four perspectives on university marketing studies," *mark. rev.*, vol. 17, no. 1, pp. 87–116, May 2017, doi: 10.1362/146934717X14909733966128.
- [15] S. Noor, K. Abdul Manan, and H. M. Abdul Kuthoos, "Assessing Corporate Brand Equity of Public Universities," *JKMJC*, vol. 35, no. 3, pp. 283–299, Sep. 2019, doi: 10.17576/JKMJC-2019-3503-17.
- [16] F. de Heer, "Exploring The Understanding of University Brand Equity: Perspectives of Public Relations and Marketing Directors," *IOSR Journal of Business and Management (IOSR-JBM)*, vol. 22, no. 7, pp. 49–57, Jul. 2020.
- [17] C. Binnie, "Constructing University Brands Through University Research Magazines," University of Ottawa, 2008. Accessed: Mar. 19, 2023. [Online]. Available: https://www.academia.edu/6149604/constructing_university_brands_1_Constructing_University_Brands_Through_University_Research_Magazines
- [18] C. Chapleo, "What defines 'successful' university brands?," *International Journal of Public Sector Management*, vol. 23, no. 2, pp. 169–183, Mar. 2010, doi: 10.1108/09513551011022519.
- [19] B. Gutiérrez-Villar, P. Alcaide-Pulido, and M. Carbonero-Ruz, "Measuring a University's Image: Is Reputation an Influential Factor?," *Education Sciences*, vol. 12, no. 1, Art. no. 1, Jan. 2022, doi: 10.3390/educsci12010019.
- [20] J. M. T. Balmer and E. R. Gray, "Corporate identity and corporate communications: creating a competitive advantage," *Industrial and Commercial Training*, vol. 32, no. 7, pp. 256–262, Dec. 2000, doi: 10.1108/00197850010379811.
- [21] M. S. Dewi and I. I. Wahyuni, "Strategi Media Monitoring pada Pemberitaan Covid-19 di Media Daring," *MEDIALOG: Jurnal Ilmu Komunikasi*, vol. 5, no. 1, Art. no. 1, Apr. 2022, doi: 10.35326/medialog.v5i1.1226.
- [22] Q. Liu et al., "Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach," *J Med Internet Res*, vol. 22, no. 4, p. e19118, Apr. 2020, doi: 10.2196/19118.
- [23] F. Jenkins, *Public Relations*, 4th ed. Intl Ideas, 1992.

- [24] S. Lukman and T. Friedel, "Program Kliping Media Sebagai Kegiatan Media Monitoring Divisi Humas di PT Perusahaan Listrik Negara (PLN) Persero Distribusi Jawa Barat," *JRK Untirta*, vol. 10, no. 2, Feb. 2020, doi: 10.31506/jrk.v10i2.7261.
- [25] H. A. HABIBA, "Peran Media Monitoring dalam Penanganan Pemberitaan Limbah Pabrik pada Bagian Informasi dan Komunikasi Departemen Humas PT Pusri Palembang," *Universitas Sriwijaya*, 2020.
- [26] I. A. P. Astiti, E. Rusdiyana, and D. Beywiyarno, "Aktivitas Media Monitoring di PT Bisnis Indonesia Konsultan (Bisnis Indonesia Intelligence Unit)," SENASTER "Seminar Nasional Riset Teknologi Terapan," vol. 1, no. 1, Art. no. 1, Oct. 2020, Accessed: Mar. 19, 2023. [Online]. Available: <https://jurnal.untidar.ac.id/index.php/senaster/article/view/2630>
- [27] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines - Information Retrieval in Practice*. Pearson Education, Inc., 2015.
- [28] T. Yang and A. Gerasoulis, "Web search engines: Practice and experience," *Computing Handbook: Two-Volume Set*, pp. 1–24, May 2022, doi: 10.1201/b16812-57.
- [29] A. Aula, "Query Formulation in Web Information Search," presented at the International Conference WWW/Internet, 2003. Accessed: Mar. 19, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/Query-Formulation-in-Web-Information-Search-Aula/594a09ec87bd38c79b7480054385aa495d51cdab>
- [30] T. Russell-Rose and F. Shokraneh, "Designing the Structured Search Experience: Rethinking the Query-Builder Paradigm," *Weave: Journal of Library User Experience*, vol. 3, no. 1, Mar. 2020, doi: 10.3998/weave.12535642.0003.102.
- [31] K. A. Hambarde and H. Proenca, "Information Retrieval: Recent Advances and Beyond." *arXiv*, Jan. 20, 2023. Accessed: Mar. 19, 2023. [Online]. Available: <http://arxiv.org/abs/2301.08801>
- [32] Sachin and D. Patel, "Event-centric Query Suggestion for Online News." *arXiv*, Jan. 12, 2022. doi: 10.48550/arXiv.2201.04716.
- [33] P. Ghasiya and K. Okamura, "Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach," *IEEE Access*, vol. 9, pp. 36645–36656, 2021, doi: 10.1109/ACCESS.2021.3062875.
- [34] Blei, D.M. (2012) Probabilistic Topic Models. *Communications of the ACM*, 55, 77-84. <http://dx.doi.org/10.1145/2133806.2133826>.
- [35] Maier, Daniel & Waldherr, Annie & Miltner, P & Wiedemann, Gregor & Niekler, Andreas & Keinert, Alexa & Pfetsch, Barbara & Heyer, Gerhard & Reber, U & Häussler, Tom & Schmid-Petri, Hannah & Adam, Silke. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*. 12. 1-26. 10.1080/19312458.2018.1430754.

- [36] Blei, David & Ng, Andrew & Jordan, Michael. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*. 3. 601-608.
- [37] Bhatia, S., Lau, J.H., Baldwin, T., 2017. An Automatic Approach for Document-level Topic Model Evaluation. Conference on Computational Natural Language Learning 206–215.
- [38] Van Atteveldt, W., Trilling, D., & Calderón, C. A. (2022). Computational Analysis of Communication. Wiley Blackwell.
- [39] R. Abebe, S. Hill, J. W. Vaughan, P. M. Small, and H. A. Schwartz, “Using Search Queries to Understand Health Information Needs in Africa.” arXiv, Apr. 17, 2019. Accessed: Mar. 22, 2023. [Online]. Available: <http://arxiv.org/abs/1806.05740>.
- [40] Hidayatullah, Ahmad Fathan & Aditya, Silfa & Karimah, & Gardini, Syifa. (2019). Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA). *IOP Conference Series: Materials Science and Engineering*. 482. 012033. 10.1088/1757-899X/482/1/012033.
- [41] O. Karnalim, M. Ayub, and H. Toba, “Mitigating plagiarism and collusion in K-12 as initial knowledge for programming instructors in higher education,” *World Transactions on Engineering and Technology Education*, vol. 21, no. 1, pp. 44–49, 2023.
- [42] Wahyudin, “Aplikasi Topic Modeling pada Pemberitaan Portal Berita Online Selama Masa PSBB Pertama”, Seminar Nasional Official Statistics 2019: Pengembangan Official Statistics dalam mendukung Implementasi SDG’s, pp. 309-318, 2019.
- [43] Santoso, Kevin Rafi Adjie Putra & Husna, Asmaul & Putri, Nadia Widyawati & Rakhmawati, Nur Aini, “Analisis Topik Tagar Covidindonesia pada Instagram Menggunakan Latent Dirichlet Allocation”, *JISKA (Jurnal Informatika Sunan Kalijaga)* Vol. 7, No. 1, Januari 2022, pp. 1 – 9, 2022.
- [44] Sagala, Hetthroh & Toba, Hapnes, “Penentuan Aspek yang Berpengaruh Terhadap Produk *Smartphone* Berdasarkan Ulasan Berbasis Tekstual”, *Jurnal Teknik Informatika dan Sistem Informasi*, Volume 7 Nomor 1, April 2021, pp. 287-295, 2021, doi: 10.28932/jutisi.v7i1.3466.
- [45] Kannitha, Diandra Zakeshia Tiara & Mustafid & Kartikasari, Puspita, “Pemodelan Topik pada Keluhan Pelanggan Menggunakan Algoritma *Latent Dirichlet Allocation* dalam Media Sosial Twitter”, *Jurnal Gaussian*, Volume 11, Nomor 2, Tahun 2022, pp. 266 – 277, 2022.
- [46] Kharisudin, Iqbal & Masri'an, Hera, “Topic Modeling on WhatsApp User Reviews Using Latent Dirichlet Allocation”, *Scientific Journal of Informatics*, Vol. 9, No. 1, May 2022. pp. 51-62, 2022, doi: 10.15294/sji.v9i1.34941
- [47] Novarian, Nathanael & Khomsah, Siti & Arifa, Amalia Beladinna, “Topic Modeling Tugas Akhir Mahasiswa Fakultas Informatika Institut Teknologi Telkom Purwokerto Menggunakan Metode *Latent Dirichlet Allocation*”, *LEDGER: Journal Informatic and Information Technology*, Vol. 2, No. 1, Feb 2023, pp. 14-27, 2023.

<https://github.com/is004/FBxM>