

***AUTOMATIC TAGGING DENGAN MENGGUNAKAN
ALGORITMA BIPARTITE GRAPH PARTITION DAN TWO WAY
POISSON MIXTURE MODEL***

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer**



*Mencerdaskan dan
Memartabatkan Bangsa*

**Oleh:
Muhammad Zhafran Bahij
1313619012**

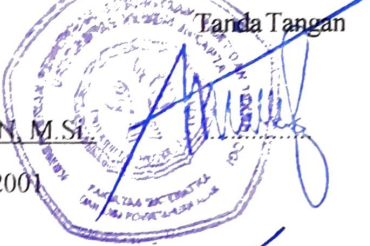
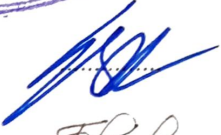




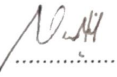
**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA**

2023

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI
AUTOMATIC TAGGING DENGAN MENGGUNAKAN
ALGORITMA BIPARTITE GRAPH PARTITION DAN TWO WAY
POISSON MIXTURE MODEL

Nama : **Muhammad Zhafran Bahij**

No. Registrasi : **1313619012**

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Prof. Dr. Muktiningsih N. M. Si</u> NIP. 196405111989032001		01-08-2023
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002		31-08-2023
Ketua	: <u>Ir. Fariani Hermin Indiyah, MT.</u> NIP. 196605171994031003		29-08-2023
Sekretaris	: <u>Ari Hendarno, S.Pd, M.Kom</u> NIP. 198811022022031002		23-08-2023
Penguji	: <u>Dr. Ria Arafiyah, M.Si.</u> NIP. 197511212005012004		24-08-2023
Pembimbing I	: <u>Muhammad Eka Suryana, M.Kom.</u> NIP. 198512232012121002		24-08-2023
Pembimbing II	: <u>Med Irzal, M.Kom.</u> NIP. 197706152003121001		24-08-2023

Dinyatakan lulus ujian skripsi tanggal: 22 Agustus 2023

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul *Automatic Tagging dengan Menggunakan Algoritma Bipartite Graph Partition dan Two Way Poisson Mixture Model* yang disusun sebagai syarat untuk memperoleh gelar Sarjana komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 10 Agustus 2023



Muhammad Zhanan Bahij



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Muhammad Zhoran Bahij
NIM : 1313619012
Fakultas/Prodi : MIPA / Ilmu Komputer
Alamat email : muhammadzhoranbahij@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Automatic Tagging dengan Menggunakan Algoritma Bipartite Graph Partition dan
Two Way Poisson Mixture Model

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 4 September 2023

Penulis

(M. Zhoran Bahij)
nama dan tanda tangan

KATA PENGANTAR

Ungkapan Puji dan Syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Adapun jenis penelitian yang dengan judul *Automatic Tagging dengan Menggunakan Algoritma Bipartite Graph Partition dan Two Way Poisson Mixture Model*

Dalam menyelesaikan skripsi ini, penulis selalu mendapat dorongan dan bantuan. Oleh karena itu, penulis menyampaikan terima kasih kepada:

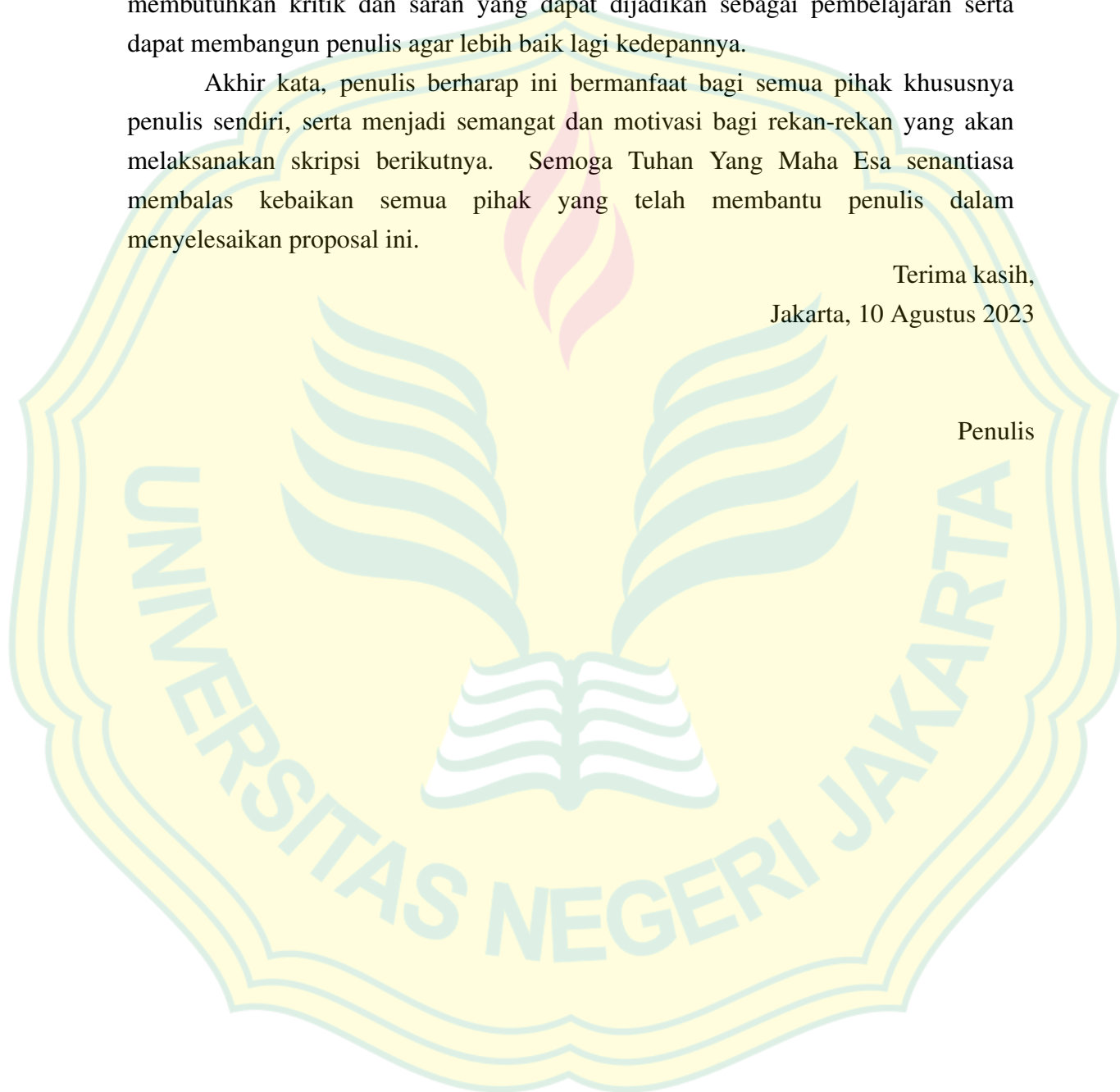
1. Para petinggi di lingkungan FMIPA Universitas Negeri Jakarta.
2. Ibu Dr. Ria Arafiyah, M.Si. selaku Koordinator Program Studi Ilmu Komputer.
3. Bapak Muhammad Eka Suryana, M.Kom. selaku Dosen Pembimbing I yang telah membimbing, mengarahkan, serta memberikan saran dan koreksi terhadap skripsi ini.
4. Bapak Med Irzal M.Kom. selaku Dosen Pembimbing II yang telah membimbing, mengarahkan, serta memberikan saran dan koreksi terhadap skripsi ini.
5. Ayah dan Ibu penulis yang selama ini telah mendukung penulis dalam menyelesaikan skripsi ini dalam berbagai hal.
6. Sepupu penulis yang telah menginspirasi penulis untuk menjalani kehidupan di Program Studi Ilmu Komputer.
7. Teman-teman Ilmu Komputer 2019 yang telah menyemangati penulis dalam penulisan skripsi.
8. Rekan-rekan PT Fhadira Inovasi Teknologi yang telah mendukung penulis untuk menyelesaikan skripsi.
9. Teman-teman KANAU yang secara tidak langsung melatih penulis agar terbiasa melakukan penulisan.
10. Kawan-kawan penulis yang tidak bisa disebutkan satu persatu.

Dalam penulisan skripsi ini, penulis menyadari bahwa dengan keterbatasan ilmu dan pengetahuan penulis, skripsi ini masih jauh dari sempurna, baik dari segi penulisan, penyajian materi, maupun bahasa. Oleh karena itu, penulis sangat membutuhkan kritik dan saran yang dapat dijadikan sebagai pembelajaran serta dapat membangun penulis agar lebih baik lagi kedepannya.

Akhir kata, penulis berharap ini bermanfaat bagi semua pihak khususnya penulis sendiri, serta menjadi semangat dan motivasi bagi rekan-rekan yang akan melaksanakan skripsi berikutnya. Semoga Tuhan Yang Maha Esa senantiasa membalas kebaikan semua pihak yang telah membantu penulis dalam menyelesaikan proposal ini.

Terima kasih,
Jakarta, 10 Agustus 2023

Penulis



ABSTRAK

MUHAMMAD ZHAFRAN BAHIJ. *Automatic Tagging* dengan Menggunakan Algoritma *Bipartite Graph Partition* dan *Two Way Poisson Mixture Model*. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2023. Di bawah bimbingan Muhammad Eka Suryana, M. Kom. dan Med Irzal, M. Kom.

Automatic Tagging merupakan hal yang dilakukan untuk menentukan suatu kata kunci atau frasa kunci yang relevan pada suatu artikel, dokumen, gambar, atau video secara otomatis. Penelitian ini merupakan salah satu payung penelitian dari mesin pencari atau *search engine* *Telusuri*. Agar bisa melakukan pencarian dengan lebih efisien, salah satu caranya adalah pencarian melalui *tag*. Sebelum mencapai tahap pencarian melalui *tag*, langkah awal yang diperlukan adalah membuat program *Automatic Tagging*. Dalam penelitian ini, algoritma yang digunakan adalah *Bipartite Graph Partition* dan *Two Way Poisson Mixture Model* dengan menggunakan data latih dari suatu *website thehill.com*. Proses pembentukan algoritma tersebut menggunakan bahasa pemrograman *Python*. Hasil akhir dari penelitian ini adalah mampu memberikan enam *tag* dengan ketepatan akurasi sebesar 37% dengan menggunakan data 229 dokumen atau artikel dan 723 *tag*. Hal ini terjadi karena program yang telah dibuat tidak bisa membagi *Bipartite Graph Partition* sebanyak K lebih dari dua.

Kata kunci : *tagging, database, bipartite graph, Poisson Mixture Model*

ABSTRACT

MUHAMMAD ZHAFRAN BAHIJ. *Automatic Tagging Using Bipartite Graph Partition and Two Way Poisson Mixture Model. Thesis. Faculty of Mathematics and Natural Sciences, State University of Jakarta. 2021. Supervised by Muhammad Eka Suryana, M. Kom. and Med Irzal, M. Kom.*

Automatic Tagging is finding relevant key words or key phrases in article, document, image, and video automatically. This research is under search engine Telusuri's research. Another way to increase efficiency of Telusuri is searching with tag. Before do searching with tag, the first step is create an Automatic Tagging. Two algorithm to build this program are Bipartite Graph Partition and Two Way Poisson Mixture Model with data from thehill.com . This program created by Python language. The final result of this research is Automatic Tagging can give six tag with 37% accuracy with 229 document and 723 tag. The accuracy is too low because the program cannot partition the Bipartite Grap Partition with K more than two.

Keywords: *tagging, database, bipartite graph, Poisson Mixture Model*

DAFTAR ISI

KATA PENGANTAR	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
II KAJIAN PUSTAKA	8
2.1 Representasi Bipartite Graph	8
2.1.1 Normalisasi dan Aproksimasi	8
2.1.2 Bipartite Graph Partitioning	10
2.1.3 Dengan Cluster Node Ranking	10
2.2 Online Tag Recommendation	12
2.2.1 Two-Way Poisson Mixture Model	15
2.2.2 Tag Recommendation for New Documents	18
2.3 Mixture Model	18
III DESAIN MODEL	21
3.1 Tahapan Penelitian	21
3.2 Algoritma <i>Automatic Tag</i>	21
3.3 Flowchart <i>Automatic Tag</i>	23
3.4 Alat dan Bahan Penelitian	23
3.5 Tahapan Penelitian <i>Automatic Tag</i>	24
3.5.1 Penginputan	24

3.5.2	Menentukan Matriks W	24
3.5.3	Menghitung <i>Low Rank Approximation Matrix</i> menggunakan algoritma Lanczos	26
3.5.4	Melakukan partisi \hat{W} ke dalam klaster K menggunakan <i>SRE</i>	28
3.5.5	Melakukan pelabelan setiap dokumen	29
3.5.6	Menghitung <i>Node Rank</i> $Rank(T)$ untuk Setiap Tag	29
3.5.7	Membuat Two Way Poisson Mixture Model	30
3.5.8	Rekomendasi Tag Untuk Dokumen Baru	30
3.5.9	Rekomendasi Tag Berdasarkan Ranks Tag	30
3.6	Skenario Pengujian	30
IV HASIL DAN PEMBAHASAN		32
4.1	Implementasi	32
4.1.1	Hasil Crawling	32
4.1.2	Pengambilan Data dari Database	34
4.1.3	Penginputan	34
4.1.4	Mengolah Data Menjadi Matriks	34
4.1.5	Menghitung <i>Low Rank Approximation Matrix</i> Menggunakan Algoritma Lanczos	36
4.1.6	Melakukan Partisi W ke dalam Klaster K	37
4.1.7	Melakukan Pelabelan Setiap Dokumen	37
4.1.8	Menghitung $Rank(T)$ untuk setiap Tag	38
4.1.9	Membuat Two Way Poisson Mixture Model	39
4.1.10	Rekomendasi Tag	41
4.2	Hasil Pengujian	41
4.3	Hasil Analisa	42
V KESIMPULAN DAN SARAN		44
5.1	Kesimpulan	44
5.2	Saran	44
DAFTAR PUSTAKA		46
LAMPIRAN		47
A main.py		47

B	data_from_database.py	51
C	input_processing.py	52
D	matrix_processing.py	56
E	low_rank_approximation_matrix.py	58
F	spectral_recursive_embedding.py	60
G	assign_label.py	64
H	node_rank_t.py	67
I	word_count_in_matrix.py	71
J	word_count_in_list.py	73
K	two_way_poisson_mixture_model.py	75
L	tag_recommendation_for_new_document.py	85
M	top_k_accuracy.py	87



DAFTAR GAMBAR

Gambar 1.1	Penggunaan search engine terpopuler (Christ, 2022)	1
Gambar 1.2	High Level Google Architecture (Brin & Page, 1998)	2
Gambar 1.3	Bagian <i>Automatic Tagging</i> pada <i>Indexer</i>	3
Gambar 1.4	Relasi antara user, tag, dan dokumen (Song et al., 2011)	4
Gambar 1.5	Skema <i>Mashup</i> , <i>Tag</i> , dan <i>API</i> (Shi et al., 2016)	6
Gambar 2.1	Suatu <i>bipartite graph</i> X dan Y Song et al. (2008)	8
Gambar 2.2	Smoothed Ranking Function Song et al. (2008)	12
Gambar 2.3	Dua bipartite graph dari dokumen-dokumen, kumpulan kata, dan kumpulan tag. Song et al. (2008)	13
Gambar 2.4	Distribusi Poisson dalam dua klaster. Bagian atas menggambarkan histogram dari <i>mixture components</i> . Bagian bawah menggambarkan hasil dari klasifikasi <i>mixture model</i> . Bagian (a) <i>three component mixtures</i> dan bagian (b) <i>two component mixtures</i> Song et al. (2008)	16
Gambar 2.5	Perbedaan antara Distribusi Gaussian dengan Distribusi Poisson Rzeszotarski (1999)	20
Gambar 3.1	Flowchart alur penelitian	21
Gambar 3.2	Diagram alir untuk tahap <i>offline computation</i>	23
Gambar 3.3	Melakukan <i>online recommendation</i> berdasarkan hasil data training	23
Gambar 3.4	Contoh simpel dua <i>bipartite graph</i>	25
Gambar 4.1	Contoh <i>tag</i> dari artikel thehill.com	32
Gambar 4.2	Dataset yang digunakan	33
Gambar 4.3	Dataset pada tabel <i>page_tags</i>	33
Gambar 4.4	Dataset pada tabel <i>page_informations</i>	34
Gambar 4.5	Dataset yang digunakan	34
Gambar 4.6	Matrix Tag Document dan Matrix Document Word	35
Gambar 4.7	Matrix W	35
Gambar 4.8	Matriks Q	36
Gambar 4.9	Matriks T	36
Gambar 4.10	Matriks \tilde{W}	37

Gambar 4.11	Matriks hasil partisi	37
Gambar 4.12	Pelabelan dokumen	38
Gambar 4.13	Rank(T)	38
Gambar 4.14	π_m	39
Gambar 4.15	Daftar kata dengan nilai $\tilde{\lambda}_m$	40
Gambar 4.16	Nilai $p_{i,m}$ pada setiap dokumen	40
Gambar 4.17	Contoh dari hasil rekomendasi <i>Tag</i>	41
Gambar 4.18	Contoh dari hasil rekomendasi <i>Tag</i>	42
Gambar 4.19	Nilai $\theta(d(i, j) \tilde{\lambda}_{m,i,j}^{(t)})$	42

