



SKRIPSI

**OPTIMASI *TREE TRAVERSAL* BERBASIS
INTERLEAVING CHAIN-OF-THOUGHT PADA
ARSITEKTUR *RETRIEVAL-AUGMENTED*
GENERATION HIERARKIS**

MUKHAMAD KHAFID MAASSOBIRIN
NPM 22083010072

DOSEN PEMBIMBING

Wahyu Syaifullah J. S, S.Kom., M.Kom.
Andri Fauzan Adziima, M.Si.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

OPTIMASI TREE TRAVERSAL BERBASIS INTERLEAVING CHAIN-OF-THOUGHT PADA ARSITEKTUR RETRIEVAL-AUGMENTED GENERATION HIERARKIS

Oleh:
MUKHAMAD KHAFID MAASSOBIRIN
NPM. 22083010072

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 10 Juni 2026:

Menyetujui,

Wahyu Svaifullah J. S., S.Kom., M.Kom.
NIP. 19860825 202121 1 003



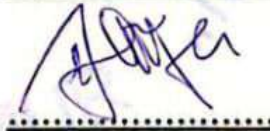
(Pembimbing I)

Andri Fauzan Adziima, M.Si.
NIP. 19950512 202406 1 001



(Pembimbing II)

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 19950723 202406 1 002



(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001



(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirna Hendrasarie, MT.

NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

OPTIMASI TREE TRAVERSAL BERBASIS INTERLEAVING CHAIN-OF-THOUGHT PADA ARSITEKTUR RETRIEVAL-AUGMENTED GENERATION HIERARKIS

Oleh:
MUKHAMAD KHAFID MAASSOBIRIN
NPM. 22083010072

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

**Plt. Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**



Dr. I Gede Susrama Mas Diyasa, ST., MT.
NIP. 19700619 2021211 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Mukhamad Khafid Maassobirin
NPM : 22083010072
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 26 Juni 2026
Yang Membuat Pernyataan,



MUKHAMAD KHAFID MAASSOBIRIN
NPM. 22083010072

ABSTRAK

Nama Mahasiswa / NPM : Mukhamad Khafid Maassobirin / 22083010072
Judul Skripsi : Optimasi *Tree Traversal* Berbasis *Interleaving Chain-of-Thought* pada Arsitektur *Retrieval-Augmented Generation* Hierarkis
Dosen Pembimbing : 1. Wahyu Syaifullah J, S.Kom., M.Kom.
2. Andri Fauzan Adziima, M.Si.

Sistem RAG hierarkis seperti RAPTOR dan HIRO memiliki keterbatasan pada mekanisme *traversal* yang hanya mengandalkan kemiripan semantik statis terhadap kueri awal, sehingga tidak mampu beradaptasi dengan kebutuhan penalaran yang berkembang selama proses penelusuran berlangsung. Padahal, pendekatan *retrieval* berbasis kemiripan statis terbukti tidak memadai untuk tugas yang membutuhkan penalaran mendalam pada dokumen panjang, dibuktikan oleh *benchmark* BRIGHT yang menunjukkan degradasi performa signifikan pada model *retrieval* terbaik sekalipun. Penelitian ini mengusulkan mekanisme *tree traversal* berbasis *Interleaving Chain-of-Thought* (IRCoT) pada arsitektur RAG hierarkis yang mengintegrasikan sinyal penalaran dinamis ke dalam algoritma *Depth-First Search*. Sistem yang dibangun mencakup tiga fase utama yaitu pembangunan pohon hierarkis melalui *embedding*, *clustering*, dan *summarization* berbasis LLM, kemudian *traversal* menggunakan fungsi skor gabungan dengan *adaptive dual-threshold pruning* yang diselingi pembangkitan penalaran di setiap langkah, serta fase generasi jawaban akhir. Inovasi utama sistem ini adalah sinyal penalaran yang terus berkembang secara dinamis sepanjang *traversal*, berbeda dari pendekatan HIRO yang menggunakan representasi kueri statis. Evaluasi komprehensif dilakukan pada empat *dataset benchmark* yaitu NarrativeQA, QASPER, QuALITY, dan TyDi QA menggunakan *multilingual-e5-large-instruct* sebagai model *embedding* dan *Qwen2.5-7B-Instruct-AWQ* sebagai model generasi. Hasil menunjukkan IRCoT meningkatkan ROUGE-L pada NarrativeQA dari 0,1144 menjadi 0,1275 dan *Answer F1* pada QASPER dari 0,3135 menjadi 0,3288, namun menurunkan akurasi pada QuALITY dari 54,98% menjadi 54,49% serta Token F1 pada TyDi QA dari 0,3922 menjadi 0,3862 dengan *overhead* waktu dua hingga empat kali lebih lambat. Analisis empiris mengidentifikasi bahwa sinyal penalaran efektif untuk dokumen dengan informasi tersebar, namun tidak menguntungkan untuk dokumen pendek maupun pertanyaan yang menuntut pemahaman holistik.

Kata Kunci: *Retrieval-Augmented Generation, Tree Traversal, Chain-of-Thought, Interleaving Reasoning, Hierarchical Retrieval*

ABSTRACT

Student Name / NPM : Mukhamad Khafid Maassobirin / 22083010072
Undergraduate thesis title : *Optimization of Tree Traversal Based on Interleaving Chain-of-Thought Hierarchical Retrieval-Augmented Generation Architecture*
Advisors : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom.
2. Andri Fauzan Adziima, M.Si.

Hierarchical RAG systems such as RAPTOR and HIRO are limited by traversal mechanisms that rely solely on static semantic similarity against the initial query, making them unable to adapt to evolving reasoning needs during the exploration process. Static similarity-based retrieval is insufficient for tasks requiring deep reasoning over long documents, as demonstrated by the BRIGHT benchmark which revealed significant performance degradation even among top-performing retrieval models. This study proposes an Interleaving Chain-of-Thought (IRCoT)-based tree traversal mechanism within a hierarchical RAG architecture that integrates dynamic reasoning signals into the Depth-First Search algorithm. The system encompasses three main phases: hierarchical tree construction through embedding, clustering, and LLM-based summarization, followed by traversal using a combined scoring function with adaptive dual-threshold pruning interleaved with reasoning generation at each step, and finally an answer generation phase. The key innovation lies in reasoning signals that evolve dynamically throughout traversal, distinguishing this approach from HIRO which relies on static query representations. Comprehensive evaluation was conducted on four benchmark datasets, namely NarrativeQA, QASPER, QuALITY, and TyDi QA, using multilingual-e5-large-instruct for embedding and Qwen2.5-7B-Instruct-AWQ for generation. Results show that IRCoT improves ROUGE-L on NarrativeQA from 0.1144 to 0.1275 and Answer F1 on QASPER from 0.3135 to 0.3288, but decreases accuracy on QuALITY from 54.98% to 54.49% and Token F1 on TyDi QA from 0.3922 to 0.3862, with computational overhead two to four times slower. Empirical analysis identifies that reasoning signals are effective for documents with dispersed information but offer limited advantages for shallow-structured documents or holistic-comprehension-demanding questions.

Keywords: *Retrieval-Augmented Generation, Tree Traversal, Chain-of-Thought, Interleaving Reasoning, Hierarchical Retrieval*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga proposal skripsi dengan judul **“Optimasi Tree Traversal Berbasis Interleaving Chain-of-Thought pada Arsitektur Retrieval-Augmented Generation Hierarkis”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Wahyu Syaifullah J. S, S.Kom., M.Kom. selaku Dosen Pembimbing I dan Bapak Andri Fauzan Adziima, M.Si. selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Selain itu, selama penyusunan skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Novirina Hendrasarie, S.T., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Bapak Dr. I Gede Susrama Mas Diyasa, ST., MT. selaku Plt. Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Dosen-dosen Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ayah dan Ibu tercinta yang selalu hadir melalui doa, dukungan, dan kasih sayang yang tiada henti, sehingga penulis dapat menyelesaikan perkuliahan dan skripsi ini dengan baik. Ayah dan Ibu merupakan sumber kekuatan dan alasan terbesar bagi penulis untuk terus berjuang dalam menghadapi setiap proses hingga mencapai tahap ini.
5. Adik-adik penulis yang selalu memberikan keceriaan, semangat, serta hiburan di sela-sela kejenuhan penulis selama menyusun tugas akhir ini. Kehadiran mereka di rumah selalu menjadi pengingat bagi penulis untuk memberikan contoh yang baik sebagai seorang kakak.
6. Ime yang telah kebersamai penulis dalam menyusun skripsi ini. Terima kasih atas segala dukungan, perhatian, serta kesabaran yang telah diberikan. Sosok yang menjadi motivasi penulis untuk terus melangkah ke depan.

7. Teman-teman seperjuangan Program Studi Sains Data, serta seluruh sahabat yang tidak dapat disebutkan satu per satu, terima kasih atas kebersamaan, diskusi, bantuan teknis, dan solidaritasnya selama masa perkuliahan.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, Juni 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL	ii
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
ABSTRAK	vi
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xvi
DAFTAR NOTASI	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	7
1.3. Batasan Masalah	7
1.4. Tujuan Penelitian	8
1.5. Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA	11
2.1. Penelitian Terdahulu	11
2.2. Kerangka Teori	18
2.2.1. <i>Information Retrieval</i>	18
2.2.2. <i>Large Language Model</i>	19
2.2.3. <i>Semantic Embedding dan Cosine Similarity</i>	21
2.2.4. <i>Retrieval-Augmented Generation</i>	23
2.2.5. <i>Hierarchical Retrieval</i>	24
2.2.6. <i>Chain-of-Thought Reasoning</i>	27
2.2.7. Reduksi Dimensi	29
2.2.8. <i>Clustering</i>	30

2.2.9.	<i>Vector Database</i>	32
2.2.10.	Metrik Evaluasi.....	33
2.2.11.	Uji Statistik.....	37
BAB III METODOLOGI PENELITIAN.....		41
3.1.	Variabel Penelitian dan Sumber Data.....	41
3.1.1.	NarrativeQA	42
3.1.2.	QASPER.....	44
3.1.3.	QuALITY	46
3.1.4.	TyDi QA.....	49
3.2.	Langkah Analisis	52
3.2.1.	Pengumpulan Data.....	55
3.2.2.	<i>Preprocessing Data</i>	55
3.2.3.	<i>Chunking</i>	60
3.2.4.	<i>Tree Construction</i>	62
3.2.5.	Penyimpanan ke Basis Data Vektor	70
3.2.6.	<i>Tree Traversal</i>	71
3.2.7.	<i>Answer Generation</i>	79
3.2.8.	Evaluasi	82
3.3.	Desain Sistem	83
BAB IV HASIL DAN PEMBAHASAN		85
4.1.	Hasil Penelitian.....	85
4.1.1.	<i>Preprocessing Data</i>	85
4.1.2.	<i>Chunking</i>	99
4.1.3.	<i>Tree Construction</i>	103
4.1.4.	<i>Tree Traversal</i>	119
4.1.5.	Studi Ablasi	134
4.1.6.	Evaluasi	158
4.2.	Tampilan Antar Muka (<i>Graphical User Interface</i>).....	176
BAB V PENUTUP		179
5.1.	Kesimpulan.....	179
5.2.	Saran Pengembangan.....	181

DAFTAR PUSTAKA.....	183
LAMPIRAN.....	189

DAFTAR GAMBAR

Gambar 2.1. Konsep <i>Vector Space Model</i> (VSM).....	19
Gambar 2.2. Arsitektur Transformer.....	20
Gambar 2.3. Konsep <i>Retrieval Augmented Generation</i>	24
Gambar 2.4. Konsep <i>Retrieval</i> pada RAPTOR.....	25
Gambar 2.5. Konsep <i>Traversal</i> pada HIRO.....	26
Gambar 2.6. Mekanisme <i>Interleaving CoT</i> dan <i>Retrieval</i> pada IRCOT	28
Gambar 3.1. Diagram Alir Langkah Analisis Sistem	53
Gambar 3.2. Arsitektur Sistem RAG Hierarkis Berbasis IRCOT	54
Gambar 3.3. Diagram Alir <i>Tree Construction</i>	62
Gambar 3.4. Ilustrasi Konsep <i>Tree Construction</i>	69
Gambar 3.5. Diagram Alir Utama <i>Pipeline Tree Traversal</i>	71
Gambar 3.6. Diagram Alir <i>Recursive Tree Traversal</i>	72
Gambar 3.7. Ilustrasi Mekanisme <i>Tree Traversal</i>	78
Gambar 3.8. Desain Sistem.....	83
Gambar 4.1. Tampilan Awal GUI.....	176
Gambar 4.2. Tampilan Contoh Percakapan	177
Gambar 4.3. Tampilan Contoh Hasil Statistik	178

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu	11
Tabel 3.1. Karakteristik <i>Dataset</i>	41
Tabel 3.2. Struktur <i>Dataset</i> NarrativeQA	43
Tabel 3.3. Contoh Data Mentah NarrativeQA	43
Tabel 3.4. Struktur <i>Dataset</i> QASPER.....	45
Tabel 3.5. Contoh Data Mentah QASPER.....	45
Tabel 3.6. Struktur <i>Dataset</i> QuALITY	47
Tabel 3.7. Contoh <i>Dataset</i> Mentah QuALITY	48
Tabel 3.8. Struktur <i>Dataset</i> TyDi QA.....	50
Tabel 3.9. Contoh <i>Dataset</i> Mentah TyDi QA.....	52
Tabel 3.10. <i>Template Prompt</i> pada Proses <i>Summarization</i>	68
Tabel 3.11. <i>Template Instruksi Embedding</i> Kueri	73
Tabel 3.12. <i>Template Prompt</i> Pembentukan Sinyal Penalaran IRCoT.....	75
Tabel 3.13. <i>Template Prompt</i> untuk <i>Dataset</i> Tipe Uraian.....	80
Tabel 3.14. <i>Template Prompt</i> untuk <i>Dataset</i> Tipe Pilihan Ganda.....	81
Tabel 4.1. Statistik Hasil <i>Preprocessing</i> Data	85
Tabel 4.2. Contoh Transformasi Teks <i>Preprocessing</i> NarrativeQA.....	88
Tabel 4.3. Contoh <i>df_ground_truth</i> NarrativeQA	90
Tabel 4.4. Contoh Transformasi Teks <i>Preprocessing</i> QASPER	92
Tabel 4.5. Contoh <i>df_ground_truth</i> QASPER.....	93
Tabel 4.6. Contoh Transformasi Teks <i>Preprocessing</i> QuALITY.....	95
Tabel 4.7. Contoh <i>df_ground_truth</i> QuALITY	96
Tabel 4.8. Contoh Transformasi Teks <i>Preprocessing</i> TyDi QA	98
Tabel 4.9. Contoh <i>df_ground_truth</i> TyDi QA.....	98
Tabel 4.10. Statistik Hasil <i>Chunking</i> per <i>Dataset</i>	100
Tabel 4.11. Contoh Hasil Tokenisasi dari Dokumen QASPER.....	103
Tabel 4.12. Statistik Hasil <i>Tree Construction</i> per <i>Dataset</i>	106
Tabel 4.13. Rata-rata Waktu Proses <i>Tree Construction</i> per Tahap (detik).....	109
Tabel 4.14. Hasil Konversi Token ke Token ID.....	110

Tabel 4.15. Nilai BIC untuk setiap kandidat jumlah <i>cluster</i> (K)	114
Tabel 4.16. Distribusi <i>chunk</i> per <i>global cluster</i>	114
Tabel 4.17. Hasil <i>local clustering</i> dan <i>multi-assignment</i>	115
Tabel 4.18. Statistik Agregat <i>Traversal (Reasoning)</i>	122
Tabel 4.19. Statistik Agregat <i>Traversal (Non-Reasoning)</i>	122
Tabel 4.20. Distribusi <i>Layer Node</i> yang Di- <i>retrieve (Reasoning)</i>	123
Tabel 4.21. Distribusi <i>Layer Node</i> yang Di- <i>retrieve (Non-Reasoning)</i>	124
Tabel 4.22. Waktu Rata-rata <i>Pipeline</i> per <i>Stage</i> dalam Detik (<i>Reasoning</i>).....	125
Tabel 4.23. Waktu Rata-rata <i>Pipeline</i> per <i>Stage</i> dalam Detik (<i>Non-Reasoning</i>)	125
Tabel 4.24. <i>Root Node</i> yang Lolos Filter pada Kueri Contoh.....	126
Tabel 4.25. Alur Keputusan <i>Traversal</i> per <i>Node</i>	127
Tabel 4.26. Contoh Lima <i>Node</i> Teratas Konteks Final.....	130
Tabel 4.27. <i>Root Node</i> yang Lolos Filter pada Kueri TyDi QA	131
Tabel 4.28. Alur Keputusan <i>Traversal</i> per <i>Node</i>	132
Tabel 4.29. Lima <i>Node</i> Teratas Konteks Final.....	133
Tabel 4.30. Statistik <i>Subset</i> Ablasi Setiap <i>Dataset</i>	135
Tabel 4.31. Hasil Ablasi α pada NarrativeQA	137
Tabel 4.32. Hasil Ablasi α pada QASPER.....	138
Tabel 4.33. Hasil Ablasi α pada QuALITY	139
Tabel 4.34. Hasil Ablasi α pada TyDi QA.....	140
Tabel 4.35. Hasil Ablasi τ pada NarrativeQA	142
Tabel 4.36. Hasil Ablasi τ pada QASPER.....	143
Tabel 4.37. Hasil Ablasi τ pada QuALITY	144
Tabel 4.38. Hasil Ablasi τ pada TyDi QA.....	145
Tabel 4.39. Hasil Ablasi δ pada NarrativeQA.....	146
Tabel 4.40. Hasil Ablasi δ pada QASPER	148
Tabel 4.41. Hasil Ablasi δ pada QuALITY.....	149
Tabel 4.42. Hasil Ablasi δ pada TyDi QA	150
Tabel 4.43. Hasil Ablasi top- k pada NarrativeQA	152
Tabel 4.44. Hasil Ablasi top- k pada QASPER.....	153
Tabel 4.45. Hasil Ablasi top- k pada QuALITY	154

Tabel 4.46. Hasil Ablasi top- k pada TyDi QA.....	155
Tabel 4.47. Konfigurasi Hiperparameter Optimal per <i>Dataset</i>	157
Tabel 4.48. Perbandingan Efektivitas ($Mean \pm Std$) pada NarrativeQA.....	159
Tabel 4.49. Perbandingan Efektivitas ($Mean \pm Std$) pada QASPER	160
Tabel 4.50. Perbandingan Efektivitas pada QuALITY	161
Tabel 4.51. Perbandingan Efektivitas ($Mean \pm Std$) pada TyDi QA	162
Tabel 4.52. Perbandingan Efisiensi pada NarrativeQA	164
Tabel 4.53. Perbandingan Efisiensi pada QASPER.....	165
Tabel 4.54. Perbandingan Efisiensi pada QuALITY	166
Tabel 4.55. Perbandingan Efisiensi pada TyDi QA.....	166
Tabel 4.56. Hasil Uji Normalitas pada NarrativeQA.....	169
Tabel 4.57. Hasil Uji <i>Wilcoxon Signed-Rank</i> pada NarrativeQA.....	170
Tabel 4.58. Hasil Uji Normalitas pada QASPER	171
Tabel 4.59. Hasil Uji <i>Wilcoxon Signed-Rank</i> pada QASPER	172
Tabel 4.60. Hasil Uji Normalitas pada TyDi QA	173
Tabel 4.61. Hasil Uji <i>Wilcoxon Signed-Rank</i> pada TyDi QA.....	174
Tabel 4.62. Hasil Uji <i>McNemar</i> pada QuALITY	175

DAFTAR LAMPIRAN

Lampiran 1. <i>Dataset</i>	189
Lampiran 2. <i>Kode Script</i>	190
Lampiran 3. <i>GUI</i>	191
Lampiran 4. <i>LoA Jurnal</i>	192

DAFTAR NOTASI

$S(q, r, n)$:	Skor relevansi gabungan dari <i>node</i> kandidat n
q	:	Vektor <i>embedding</i> dari kueri pengguna
r	:	Vektor <i>embedding</i> dari jejak penalaran IRCoT
n	:	Vektor <i>embedding</i> dari <i>node</i> yang sedang dievaluasi
α	:	Hiperparameter pembobot keseimbangan kueri dan sinyal penalaran
τ	:	Ambang batas kemiripan untuk pruning dan filter <i>root node</i>
δ	:	Ambang batas <i>delta score</i> untuk keputusan <i>backtracking</i>
ΔS	:	Selisih skor relevansi antara <i>child node</i> dan <i>parent node</i>
K	:	Jumlah <i>cluster</i> pada GMM
K^*	:	Jumlah <i>cluster</i> optimal hasil minimal BIC
K_{max}	:	Batas atas jumlah <i>cluster</i> kandidat
P_{ij}	:	Probabilitas keanggotaan <i>node</i> ke- i pada <i>cluster</i> ke- j
$\tau_{cluster}$:	Ambang batas <i>soft assignment</i> pada GMM (0,1)
π_k	:	Bobot komponen Gaussian ke- k
μ_k	:	Vektor rata-rata distribusi Gaussian ke- k
Σ_k	:	Matriks kovarians distribusi Gaussian ke- k
γ_{ik}	:	Probabilitas kontribusi komponen Gaussian ke- k pada data x_i
N	:	Jumlah total data
N_k	:	Jumlah efektif data pada komponen Gaussian ke- k
p	:	Jumlah parameter bebas model
w_{ij}	:	<i>Similarity weight</i> antara titik data x_i dan x_j pada UMAP
w'_{ij}	:	<i>Similarity weight</i> antara x_i dan x_j pada ruang berdimensi rendah
C	:	Nilai fungsi <i>loss</i> UMAP