

***Query Expansion Berbasis Lexical Database Untuk Cross
Language Information Retrieval (CLIR) Indonesia-Inggris***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada Jurusan Teknik
Informatika



Oleh:

Erwin Saputra

NIM: 09021281823170

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2023**

LEMBAR PENGESAHAN PROPOSAL SKRIPSI

***QUERY EXPANSION BERBASIS LEXICAL DATABASE
UNTUK CROSS LANGUAGE INFORMATION RETRIEVAL
(CLIR) INDONESIA-INGGRIS***

Oleh:

Erwin Saputra
NIM: 09021281823170

Indralaya, 05 Januari 2023

Pembimbing I,



Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

Pembimbing II,



Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Ahik Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Rabu tanggal 04 Januari 2023 telah dilaksanakan ujian sidang komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Erwin Saputra

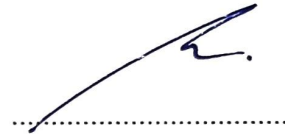
NIM : 09021281823170

Judul : Query Expansion Berbasis Lexical Database Untuk Cross Language Information Retrieval (CLIR) Indonesia-Inggris

Dan dinyatakan LULUS.

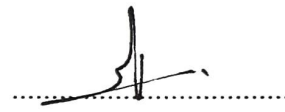
1. Ketua

Rizki Kurniati, M.T
NIP. 199107122019032016



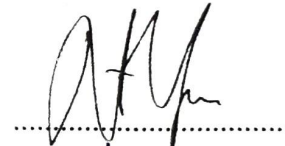
2. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005



3. Pembimbing II

Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001



4. Penguji

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan dibawah ini:

Nama : Erwin Saputra

NIM : 09021281823170

Program Studi : Teknik Informatika

Judul : Query Expansion Berbasis Lexical Database Untuk Cross
Language Information Retrieval (CLIR) Indonesia-Inggris

Hasil Pengecekan Software iThenticate/Turnitin: 8%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.



Indralaya, 27 Januari 2023



Erwin Saputra
NIM 09021281823170

MOTTO

“Allah tidak akan membebani seseorang melainkan sesuai dengan kemampuannya.”

– *QS. Al-Baqarah 286*

Kupersembahkan Skripsi ini kepada:

1. Allah SWT & Nabi Muhammad SAW
2. Kedua Orang tua dan Keluargaku
3. Teman Seperjuanganku

Query Expansion based on Lexical Database for Cross Language Information Retrieval (CLIR) Indonesian-English

Erwin Saputra (09021281823170)

Departement of Informatics, Faculty of Computer Science, Sriwijaya University

ABSTRACT

Query Expansion (QE) is an important process in Cross-Language Information Retrieval (CLIR) system, because it can make the document identification process better. The use of QE aims to see if there is an improvement in the results of the CLIR (Indonesian-English) system, compared to the CLIR Non QE system. In general, the use of QE in the CLIR system can provide an improvement in terms of recall, because the system can retrieve relevant documents that could not be retrieved before by the initial query. This research uses the lexical database method which is one of the expansion features analyzed in the linguistic approach using WordNet as its module. Based on the test results using 10 queries, the average precision value is 0.719 and recall is 0.785 on the QE system.

Keyword: *query expansion, cross-language information retrieval, lexical database*

Indralaya, January 5th 2023

Supervisor I,

Supervisor II,

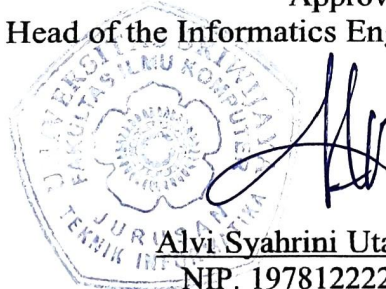


Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005



Novi Yuslian, S.Kom., M.T.
NIP. 198211082012122001

Approved,
Head of the Informatics Engineering Departement,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

Query Expansion Berbasis Lexical Database Untuk Cross Language Information Retrieval (CLIR) Indonesia-Inggris

Erwin Saputra (09021281823170)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya

ABSTRAK

Query Expansion (QE) merupakan proses penting dalam sistem *Cross-Language Information Retrieval* (CLIR), karena dapat membuat proses identifikasi dokumen menjadi lebih baik. Penggunaan QE memiliki tujuan untuk melihat apakah terdapat peningkatan terhadap hasil dari sistem CLIR (Indonesia-Inggris), dibandingkan dengan sistem CLIR Non QE. Pada umumnya, penggunaan QE dalam sistem CLIR dapat memberikan peningkatan dari sisi *recall*, karena sistem dapat mengembalikan dokumen-dokumen relevan yang tidak dapat dikembalikan sebelumnya oleh kueri awal. Penelitian ini menggunakan metode *lexical database* yang merupakan salah satu fitur ekspansi yang dianalisis dalam pendekatan *linguistic* dengan menggunakan WordNet sebagai modulnya. Berdasarkan hasil pengujian menggunakan 10 kueri, didapatkan nilai *precision* rata-rata sebesar 0.719 dan *recall* sebesar 0.785 pada sistem QE.

Kata kunci: *query expansion, cross-language information retrieval, lexical database*

Indralaya, 05 Januari 2023

Pembimbing I,



Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

Pembimbing II,



Novi Yusliani, S.Kom., M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

KATA PENGANTAR

Alhamdulillah, puji dan syukur kehadiran Allah SWT atas rahmat dan karunia-nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “*Query Expansion Berbasis Lexical Database Untuk Cross Language Information Retrieval (CLIR) Indonesia-Inggris*”. Adapun tugas akhir ini disusun untuk memenuhi salah satu persyaratan kelulusan tingkat sarjana pada jurusan Teknik Informatika Fakultas Ilmu Komputer Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih sebanyak-banyaknya kepada banyak pihak yang telah memberikan semangat, dukungan, motivasi, doa dan bimbingan dalam menyelesaikan tugas akhir ini, pihak tersebut antara lain :

1. Kedua orang tua saya beserta keluarga yang telah memberikan semangat dan motivasi dalam menyelesaikan tugas akhir ini.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
3. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya dan dosen penguji yang telah memberikan koreksi dan masukan untuk tugas akhir ini.
4. Bapak Dr. Abdiansah, M.Cs. selaku dosen pembimbing I dan Ibu Novi Yusliani, M.T. selaku pembimbing II yang telah membimbing dan mengarahkan penulis dalam proses perkuliahan dan pengerjaan tugas akhir
5. Ibu Rizki Kurniati, M.T. selaku dosen pembimbing akademik yang telah membimbing dan memberikan masukan sehingga penulis dapat menyelesaikan

tugas akhir ini dengan baik.

7. Wahyu Ramadhani, Valiant Reza, dan Argha Novan yang telah membantu dan menemani penulis dalam proses pengerjaan tugas akhir ini.
8. Teman-teman jurusan Teknik Informatika yang selalu berbagi informasi dan dukungan dalam menyelesaikan tugas akhir ini.

Penulis menyadari dalam penyusunan tugas akhir ini masih terdapat banyak kekurangan dan jauh dari kata sempurna. Untuk itu penulis mengharapkan kritikan dan saran yang membangun untuk kesempurnaan tugas akhir ini. Akhir kata semoga tugas akhir yang penulis buat memberikan manfaat bagi kita semua.

Indralaya, 05 Januari 2023



Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN PROPOSAL SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-3
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan	I-4
1.8 Kesimpulan	I-5
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 <i>Cross Language Information Retrieval (CLIR)</i>	II-1
2.2.2 <i>Query Expansion</i>	II-5
2.2.3 <i>Preprocessing</i>	II-8
2.2.4 <i>Lexical Database</i>	II-9
2.2.5 <i>Term Frequency – Inverse Document Frequency</i>	II-10

2.2.6	<i>Vector Space Model</i>	II-11
2.2.7	<i>Cosine Similarity</i>	II-12
2.2.8	Evaluasi Sistem.....	II-13
2.2.9	<i>Rational Unified Process (RUP)</i>	II-14
2.3	Penelitian Lain yang Relevan	II-16
2.4	Kesimpulan	II-18
BAB III	METODOLOGI PENELITIAN	III-1
3.1	Pendahuluan.....	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.2.2	Metode Pengumpulan Data.....	III-3
3.3	Tahapan Penelitian.....	III-3
3.3.1	Mengumpulkan Data.....	III-5
3.3.2	Kerangka Kerja Penelitian	III-5
3.3.3	Melakukan Pengujian Penelitian	III-8
3.3.4	Alat yang digunakan dalam Pelaksanaan Penelitian.....	III-10
3.3.5	Melakukan Analisis dan Kesimpulan Hasil Pengujian..	III-11
3.4	Metode Pengembangan Perangkat Lunak.....	III-12
3.4.1	Fase Insepsi.....	III-12
3.4.2	Fase Elaborasi	III-13
3.4.3	Fase Konstruksi.....	III-13
3.4.4	Fase Transisi	III-13
3.5	Kesimpulan	III-14
BAB IV	PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1	Pendahuluan.....	IV-1
4.2	Fase Insepsi.....	IV-1
4.2.1	Pemodelan Bisnis.....	IV-1
4.2.2	Kebutuhan.....	IV-2
4.2.3	Analisis dan Perancangan	IV-3
4.2.4	Implementasi.....	IV-11
4.3	Fase Elaborasi	IV-18

4.3.1	Pemodelan Bisnis.....	IV-19
4.3.2	Kebutuhan.....	IV-21
4.3.3	Analisis dan Perancangan	IV-22
4.4	Fase Konstruksi.....	IV-27
4.4.1	Kebutuhan.....	IV-27
4.4.2	Analisis dan Perancangan	IV-27
4.4.3	Implementasi.....	IV-28
4.5	Fase Transisi	IV-31
4.5.1	Pemodelan Bisnis.....	IV-31
4.5.2	Rencana Pengujian.....	IV-32
4.5.3	Implementasi.....	IV-32
4.6	Kesimpulan	IV-34
BAB V HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan.....	V-1
5.2	Hasil Program	V-1
5.3	Data Hasil Penelitian	V-5
5.3.1	Konfigurasi Percobaan.....	V-5
5.3.2	Skenario Pengujian Pertama	V-6
5.3.3	Skenario Pengujian Kedua.....	V-8
5.3.4	Skenario Pengujian Ketiga.....	V-12
5.3.5	Skenario Pengujian Keempat.....	V-15
5.3.6	Skenario Pengujian Kelima	V-18
5.3.7	Skenario Pengujian Keenam.....	V-21
5.3.8	Skenario Pengujian Ketujuh	V-25
5.3.9	Skenario Pengujian Kedelapan	V-28
5.3.10	Skenario Pengujian Kesembilan	V-31
5.3.11	Skenario Pengujian Kesepuluh	V-34
5.4	Analisis Hasil Penelitian.....	V-36
5.5	Kesimpulan	V-40
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Kesimpulan	VI-1

6.2 Saran	VI-2
DAFTAR PUSTAKA	xvii
LAMPIRAN.....	xx

DAFTAR TABEL

Tabel III-1. Contoh Berita yang Dikumpulkan	III-2
Tabel III-2. Hasil pengujian kueri.....	III-10
Tabel III-3. Hasil Pengujian Sistem QE dan Sistem Non QE.....	III-12
Tabel IV-1. Kebutuhan Fungsional Perangkat Lunak.....	IV-2
Tabel IV-2. Kebutuhan Non Fungsional Perangkat Lunak.....	IV-3
Tabel IV-3. TF-IDF pada kueri.....	IV-6
Tabel IV-4. Pembobotan TF-IDF Dokumen.....	IV-7
Tabel IV-5. Definisi Aktor.....	IV-12
Tabel IV-6. Definisi <i>Use Case</i>	IV-13
Tabel IV-7. Skenario <i>Use Case</i> Pencarian Dokumen	IV-14
Tabel IV-8. Skenario <i>Use Case</i> Menampilkan Hasil Salah Satu Kueri.....	IV-16
Tabel IV-9. Skenario <i>Use Case</i> Menampilkan Dokumen.....	IV-17
Tabel IV-10. Implementasi Kelas	IV-29
Tabel IV-11. Rencana Pengujian	IV-32
Tabel IV-12. Hasil Pengujian Black Box.....	IV-33
Tabel V-1. Hasil pengujian kueri “Gempa bumi” sistem Non QE.....	V-6
Tabel V-2. Hasil pengujian kueri “Gempa bumi” sistem QE	V-7
Tabel V-3. Hasil pengujian kueri “Feminisme” sistem Non QE	V-9
Tabel V-4. Hasil pengujian kueri “Feminisme” sistem QE	V-10
Tabel V-5. Hasil pengujian kueri “Kasus tak terpecahkan” sistem Non QE ...	V-12
Tabel V-6. Hasil pengujian kueri “Kasus tak terpecahkan” sistem QE.....	V-13
Tabel V-7. Hasil pengujian kueri “Perjalanan luar angkasa” sistem Non QE .	V-15
Tabel V-8. Hasil pengujian kueri “Perjalanan luar angkasa” sistem QE.....	V-16
Tabel V-9. Hasil pengujian kueri “Badai salju” sistem Non QE	V-18
Tabel V-10. Hasil pengujian kueri “Badai salju” sistem QE	V-19
Tabel V-11. Hasil pengujian kueri “Krisis makanan” sistem Non QE.....	V-22
Tabel V-12. Hasil pengujian kueri “Krisis makanan” sistem QE	V-23
Tabel V-13. Hasil pengujian kueri “Polusi” sistem Non QE.....	V-25
Tabel V-14. Hasil pengujian kueri “Polusi” sistem QE.....	V-26

Tabel V-15. Hasil pengujian kueri “Pemanasan Global” sistem Non QE	V-28
Tabel V-16. Hasil pengujian kueri “Pemanasan global” dengan QE.....	V-29
Tabel V-17. Hasil pengujian kueri “Olimpiade” sistem Non QE	V-31
Tabel V-18. Hasil pengujian kueri “Olimpiade” sistem QE	V-32
Tabel V-19. Hasil pengujian kueri “Piring terbang” sistem Non QE	V-34
Tabel V-20. Hasil pengujian kueri “Piring terbang” sistem QE	V-35
Tabel V-21. Hasil Pengujian Sistem QE dan Non QE.....	V-37

DAFTAR GAMBAR

Gambar II-1. Arsitektur umum dari Sistem CLIR	II-2
Gambar II-2. <i>Query Translation</i> pada Sistem CLIR.....	II-3
Gambar II-3. <i>Document Translation</i> pada Sistem CLIR	II-5
Gambar II-4. Representasi dokumen dan kueri pada ruang vektor.....	II-11
Gambar II-5. Arsitektur <i>Rational Unified Process</i>	II-15
Gambar III-1. Diagram Alir Tahapan Penelitian	III-3
Gambar III-2. Diagram Alur Proses Perangkat Lunak (Sistem QE).....	III-5
Gambar III-3. Diagram Alur Proses Perangkat Lunak (Sistem Non QE).....	III-6
Gambar IV-1. Flowchart Pra-pengolahan	IV-4
Gambar IV-2. <i>Use Case</i> Diagram Sistem CLIR	IV-12
Gambar IV-3. Bentuk format dokumen dalam file .txt.....	IV-19
Gambar IV-4. Rancangan Tampilan Awal Perangkat Lunak	IV-20
Gambar IV-5. Rancangan Tampilan untuk Membaca Dokumen.....	IV-21
Gambar IV-6. Diagram <i>Activity</i> Pencarian dokumen	IV-23
Gambar IV-7. Diagram <i>Activity</i> Menampilkan Hasil Salah Satu Kueri	IV-24
Gambar IV-8. Diagram <i>Activity</i> Menampilkan Dokumen	IV-24
Gambar IV-9. Diagram <i>Sequence</i> Pencarian Dokumen.....	IV-25
Gambar IV-10. Diagram <i>Sequence</i> Menampilkan Hasil Salah Satu Kueri	IV-26
Gambar IV-11. Diagram <i>Sequence</i> Menampilkan Dokumen	IV-26
Gambar IV-12. Diagram Kelas Perangkat Lunak	IV-28
Gambar IV-13. Tampilan Utama perangkat lunak.....	IV-30
Gambar IV-14. Tampilan Kedua (Mode Membaca) perangkat lunak	IV-31
Gambar V-1. Hasil Tampilan Awal perangkat lunak.....	V-2
Gambar V-2. Tampilan Hasil Dokumen yang didapatkan.....	V-3
Gambar V-3. Tampilan Tab Baru	V-4
Gambar V-4. Grafik perbandingan nilai <i>recall</i> sistem QE dan Non QE	V-38
Gambar V-5. Grafik perbandingan nilai <i>precision</i> sistem QE dan Non QE....	V-38

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian.

1.2 Latar Belakang

Query Expansion (QE) merupakan sebuah teknik yang digunakan dalam sistem *Cross-Language Information Retrieval* (CLIR) untuk meningkatkan kinerja pengembalian sistem terutama pada sisi *recall*. Tujuan penggunaan QE sendiri adalah karena sistem Non QE tidak dapat mengembalikan banyak dokumen relevan yang terdapat di dataset. Dengan menggunakan QE diharapkan sistem dapat meningkatkan hasil pengembalian dengan memformulasi ulang kueri, dan melakukan penelusuran kedua untuk mendapatkan dokumen relevan lainnya, yang tidak dapat diambil oleh kueri awal atau sistem Non QE. Sehingga peneliti dapat mengetahui apakah metode QE dapat memberikan peningkatan pada sistem CLIR (Indonesia-Inggris) atau sebaliknya.

Pada penelitian-penelitian sebelumnya, *Query Expansion* umumnya memiliki 2 pendekatan yaitu *Global Analysis* dan *Local Analysis*. Dalam *Global Analysis* terdapat pendekatan *linguistic*, *corpus-based*, *search log-based*, dan pendekatan *web-based* sedangkan dalam *Local Analysis* terdapat *relevance feedback* dan *pseudo-relevance feedback*. Penelitian ini

menggunakan metode *Lexical Database* yang merupakan salah satu fitur ekspansi yang dianalisis dalam pendekatan *linguistic*.

Corpus-based memiliki kelebihan yaitu dapat menemukan kata (*term*) yang relevan hanya berdasarkan kemunculan bersama pada kueri, tetapi memiliki kelemahan terhadap koleksi dokumen yang dinamis (Azad & Deepak, 2019). Salah satu kelebihan *search log-based* adalah secara implisit memasukkan umpan balik relevansi dengan ukuran implisit relatif baik, namun kinerjanya tidak selalu ideal untuk semua jenis pengguna dan tugas pencarian. (White et al., 2005). Kelebihan *web-based* yaitu dapat meningkatkan tradisional QE berdasarkan log kueri (Kraft & Zien, 2004), namun memanfaatkan data yang tidak selalu tersedia atau cocok untuk tugas pencarian informasi.

Pendekatan *linguistic* merupakan pendekatan yang menganalisa fitur ekspansi seperti leksikal, morfologi, semantik, dan hubungan kata sintaksis untuk memperluas kata (*term*) kueri awal (Azad & Deepak, 2019). Pendekatan ini menggunakan tesaurus, kamus, ontologi, *Linked Open Data (LOD) cloud* atau sumber pengetahuan serupa lainnya seperti WordNet atau ConceptNet.

Kelebihan dari pendekatan *linguistic* atau lebih tepatnya WordNet ini ialah dapat menyajikan beberapa kata ekspansi yang bagus, namun memiliki kelemahan yaitu dapat menyebabkan pergeseran topik (Azad & Deepak, 2019). Untuk mengatasi hal tersebut, domain pada sistem CLIR yang dibangun lebih berfokus pada spesifik domain untuk mengurangi pergeseran

topik.

Berdasarkan referensi pada penelitian sebelumnya, maka penelitian ini menggunakan WordNet untuk kueri ekspansinya dan menggunakan teknik mesin penerjemah dan *cosine similarity* pada sistem CLIR-nya. Sistem ini digunakan untuk mencari sebuah dokumen berbahasa inggris dari sebuah kueri berbahasa indonesia yang dimasukkan pengguna, yang harapannya dapat mempermudah pengguna dalam mencari informasi dalam bahasa yang berbeda (bahasa inggris).

1.3 Rumusan Masalah

Berdasarkan penjelasan dari latar belakang di atas, maka dapat dibuat rumusan masalah sebagai berikut:

1. Bagaimana membuat perangkat lunak CLIR Indonesia-Inggris menggunakan metode *Query Expansion* berbasis *Lexical Database*?
2. Bagaimana pengaruh *Query Expansion* terhadap hasil relevansi dokumen dari sistem CLIR Indonesia Inggris?

1.4 Tujuan Penelitian

Tujuan dari Penelitian ini adalah:

1. Menghasilkan perangkat lunak sistem CLIR Indonesia-Inggris menggunakan metode *Query Expansion* berbasis *Lexical Database*.
2. Mengetahui pengaruh penggunaan *Query Expansion* pada sistem CLIR Indonesia Inggris.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari Penelitian ini adalah:

1. Mendapatkan dokumen yang relevan dari kueri yang diberikan oleh pengguna terkait domain yang telah ditetapkan.
2. Hasil penelitian dapat dijadikan referensi mengenai penggunaan *Query Expansion* pada sistem CLIR Indonesia-Inggris.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Data yang digunakan merupakan kumpulan berita yang dipublikasikan oleh CNN tahun 2015-2017 yang diambil dari situs *Kaggle*¹.
2. *Dataset* yang dikumpulkan sebanyak 11324 berita berbahasa Inggris.
3. Sistem dibangun menggunakan bahasa pemrograman Python.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini akan membahas pokok pikiran yang melandasi penelitian, seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah atau ruang lingkup serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas seluruh dasar-dasar teori yang digunakan mulai dari definisi sistem, informasi mengenai domain, dan semua yang digunakan pada tahapan analisis, perancangan, dan implementasi.

¹ <https://www.kaggle.com/snapcrack/all-the-news>

BAB III. METODOLOGI PENELITIAN

Pada bab ini membahas mengenai alur yang akan dilaksanakan pada penelitian. Seperti pengumpulan data, analisis data dan perancangan pembangunan sistem. Serta setiap tahapan pada penelitian dijelaskan secara rinci berdasarkan kerangka kerja yang dibuat.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas analisa dan rancangan perangkat lunak yang akan dibangun. Dimulai dengan analisa kebutuhan, perancangan dan konstruksi perangkat lunak, serta diakhiri dengan melakukan pengujian yang bertujuan untuk memastikan pengembangan sistem sesuai rancangan dan kebutuhan yang telah dibuat.

BAB V. HASIL DAN ANALISA PENELITIAN

Pada bab ini akan dijelaskan hasil pengujian sesuai yang telah direncanakan. Tabel hasil pengujian dan analisisnya disajikan sebagai patokan dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini membahas mengenai kesimpulan dari semua uraian pada bab sebelumnya serta saran yang diuraikan dari hasil penelitian.

1.8 Kesimpulan

Pada bab ini telah dijelaskan pokok pikiran dalam penelitian yang akan dilakukan, seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

- Ayudita, I. M., & Indriati, P. P. A. (2018). Sistem Pencarian Jurnal Ilmiah Cross Language dengan Metode Vector Space Model (VSM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Azad, H. K., & Deepak, A. (2019). Query expansion techniques for information retrieval: a survey. *Information Processing & Management*, 56(5), 1698-1735.
- Crouch, C. J., and Yang, B. (1992). "Experiments in automatic statistical thesaurus construction," in *Proceedings of SIGIR Conference*, pp. 77–88. doi:10.1145/133160.133180
- Dwivedi, S. K., & Chandra, G. (2016). A Survey on Cross-Language Information Retrieval. *International Journal on Cybernetics & Informatics (IJCI)* Vol, 5.
- Franz, M., McCarley, J. S., and Roukos, S. (1999). "Ad hoc and multilingual information retrieval at IBM," in *Proceedings of TREC Conference*.
- Gaillard, B., Bouraoui, J. L., De Neef, E. G., & Boualem, M. (2010). Query expansion for cross language information retrieval improvement. In *2010 Fourth International Conference on Research Challenges in Information Science (RCIS)* (pp. 337-342). IEEE.
- Halim, A., Yusliani, N., & Sazaki, Y. (2018). *Query Expansion Pada Mesin Pencari Menggunakan Lexical Database Dan Vector Space Model* (Doctoral dissertation, Sriwijaya University).

- Imbar, R. V., Adelia, A., Ayub, M., & Rehatta, A. (2015). Implementasi *Cosine Similarity* dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*, 10(1).
- Karmayasa, O., Mahendra, I. B. 2012. Implementasi *Vector Space Model* dan Beberapa Notasi Metode *Term Frequency Inverse Document Frequency* (TF-IDF) pada Sistem Temu Kembali Informasi. *Jurnal Elektronik Ilmu Komputer Universitas Udayana*, Vol. 1, No. 1
- Kraft, R., & Zien, J. (2004). Mining anchor text for query refinement. In *Proceedings of the 13th international conference on World Wide Web* (pp. 666-674).
- Laxmi, M. D., & Indriati, M. A. F. (2018). *Query Expansion* Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme *Cosine Similarity* Berbasis WordNet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- McCarley, J. S. (1999). "Should we translate the documents or the queries in cross-language information retrieval," in *Proceedings of ACL Conference*, pp. 208–214. doi:10.3115/1034678.1034716
- Nie, J. Y. (2010). Cross-language information retrieval. *Synthesis Lectures on Human Language Technologies*, 3(1), 1-125.
- Pamungkas, Z. Y., & Indriati, A. R. (2015). *Query Expansion Pada Sistem Temu Kembali Informasi Dokumen Berbahasa Indonesia Menggunakan Pseudo Relevance Feedback* (Doctoral dissertation, Universitas Brawijaya).

- Qiu, Y., and Frei, H. P. (1993). "Concept query expansion," in *Proceedings of SIGIR Conference*, pp. 160–169. doi:10.1145/160688.160713
- Rahardjo, A. 2018. *Pengertian RUP (Rational Unified Process)*. (<https://medium.com/@andrerahardjo/pengertian-rup-rational-unified-process-1bec9c664458>, diakses 01 Januari 2022)
- Usharani, J., Iyakutti, K. 2013. A Genetic Algorithm Based on Cosine Similarity for Relevant Document Retrieval. *International Journal of Engineering Research & Technology (IJERT)*, Vol. 2, Issue 2.
- White, R. W., Ruthven, I., & Jose, J. M. (2005). A study of factors affecting the utility of implicit relevance feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 35-42).
- Xu, J., and Croft, W. B. (1996). "Query expansion using local and global document analysis," in *Proceedings of SIGIR Conference*, pp. 4–11. doi:10.1145/243199.243202
- Zhai, C., and Lafferty, J. (2001a). "Model-based feedback in the language modeling approach to information retrieval," in *Proceedings of CIKM Conference*, pp. 403–410. doi:10.1145/502585.502654
- Zhang, R. 2019. A Brief Introduction to Cross-Lingual Information Retrieval. (<https://medium.com/lily-lab/a-brief-introduction-to-cross-lingual-information-retrieval-eba767fa9af6>, diakses 14 Januari 2023)