

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Perjanjian perdagangan internasional, khususnya dalam bentuk *Trade Agreements* (TA), telah mengalami pertumbuhan signifikan. Data yang dikumpulkan oleh Burri dan Polanco (2020) menunjukkan bahwa dua pertiga anggota *World Trade Organization* (WTO) kini terikat dalam kesepakatan yang mengatur perdagangan *digital*. Sejak tahun 2000, 77% dari perjanjian perdagangan dunia bersifat bilateral, dengan 83% di antaranya mencakup ketentuan *digital trade*. Namun peningkatan kuantitas ini diikuti oleh lonjakan *volume* dan variasi isi dokumen. Dalam penelitiannya, Burri dan Polanco menyoroti bahwa rata-rata jumlah kata dalam bab perdagangan elektronik meningkat dari 835 kata menjadi 1.476 kata dalam kurun waktu lima tahun terakhir. Bahkan, jumlah tersebut mencapai 5.346 kata yang terbagi dalam 22 pasal pada perjanjian *Japan-US Digital Trade Agreement*, melebihi perjanjian *Australia–Singapore Free Trade Agreement* (SAFTA) yang memiliki 2.997 kata dalam 19 pasal. Menurut peneliti Burri dan Polanco, kondisi ini mengakibatkan kerangka kerja yang terbentuk menjadi terbagi-bagi dan tidak teratur [1]. Dengan volume dokumen yang besar dan berbagai bahasa hukum yang beragam memicu hambatan informasi bagi aktor ekonomi.

Dalam WTO, TA dikategorikan sebagai *Regional Trade Agreement* (RTA), juga tercermin pada sistem pencarian yang tersedia di *platform* legal.tina.trade, sebuah *platform* yang dikembangkan oleh *United Nations Economic and Social Commission for Asia and the Pacific* (UN ESCAP) sebagai antarmuka akses terhadap dokumen-dokumen *Regional Trade Agreement* (RTA). Dataset yang digunakan dalam penelitian ini diperoleh dari *platform* tersebut, yang mencakup 700 data dokumen RTA. Mekanisme pencarian yang tersedia pada *platform* ini masih berbasis kata kunci, yang hanya mampu mencocokkan kemunculan karakter teks tanpa memahami konteks maknanya [2]. Keterbatasan ini memicu munculnya *vocabulary gap*, di mana sistem gagal mengenali relevansi antara kueri dengan dokumen apabila terdapat variasi sinonim atau terminologi hukum yang berbeda secara leksikal namun identik secara semantik [3]. Akibatnya, dua ketentuan yang secara substantif memiliki makna serupa namun menggunakan terminologi berbeda tidak akan teridentifikasi sebagai dokumen yang relevan oleh sistem pencarian

tersebut [4].

Upaya untuk mengatasi keterbatasan pencarian berbasis kata kunci pada dokumen legal telah menjadi perhatian dalam beberapa penelitian sebelumnya yang memanfaatkan pendekatan berbasis *embedding* untuk tugas pencarian dan analisis dokumen. Kim et al. (2023) dalam penelitian berjudul “Content-Based Korean Journal Recommendation System Using Sentence BERT” mengimplementasikan *Sentence Bidirectional Encoder Representations from Transformers* (SBERT) untuk pencarian berbasis konteks pada dokumen akademik Korea dan berhasil mencapai akurasi Top-10 sebesar 76,6% [5]. Zhao et al. (2024) dalam penelitian berjudul “Key Information *Retrieval* to Classify the Unstructured Data Content of Preferential Trade Agreements” menerapkan *information retrieval* berbasis *embedding* BERT pada dokumen *Preferential Trade Agreement* (PTA) untuk mengekstraksi ketentuan perdagangan yang relevan. Namun, penerapan tersebut hanya berfungsi sebagai tahap *preprocessing* untuk klasifikasi ketentuan dan bukan sebagai sistem pencarian semantik yang dapat dikueri secara mandiri [6]. Sementara itu, Ozturk dan Mesut (2024) dalam penelitian berjudul “Performance Analysis of Chroma, Qdrant, and Faiss Databases” melakukan analisis komparatif performa tiga *vector database* menggunakan dataset umum [7]. Penelitian-penelitian tersebut menunjukkan bahwa penerapan pendekatan berbasis *embedding* pada domain dokumen RTA dengan memanfaatkan *vector database* belum pernah dilakukan secara bersamaan.

Berdasarkan hal tersebut, penelitian ini mengusulkan penerapan pendekatan *vector search* pada dokumen RTA yang bersumber dari *platform* legal.tina.trade. Pendekatan ini merepresentasikan teks dokumen ke dalam bentuk *vector* numerik berukuran tetap (*fixed-size vectors*) menggunakan model SBERT, sehingga pencarian dapat dilakukan berdasarkan kemiripan semantik antar dokumen tanpa perlu memproses ulang teks pada setiap kueri [8]. Pencarian kemiripan secara eksak (*exact nearest neighbor*) pada koleksi dokumen berukuran besar membutuhkan komputasi yang tidak sebanding dengan skala data, sehingga penelitian ini mengadopsi pendekatan *Approximate Nearest Neighbor* (ANN) [9]. Untuk mengimplementasikan ANN tersebut, penelitian ini menggunakan Qdrant sebagai sistem manajemen basis data *vector* yang mendukung pencarian kemiripan *vector* serta pemfilteran metadata [7, 10].

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan arsitektur *vector search* yang terintegrasi dengan model SBERT dan *vector database management system* Qdrant untuk melakukan pencarian provisi pada dokumen *Regional Trade Agreement* (RTA)?
2. Bagaimana akurasi yang didapatkan dari hasil implementasi *vector search* pada RTA menggunakan *Normalized Discounted Cumulative Gain at K* (NDCG@K)?

## 1.3 Batasan Permasalahan

Penelitian ini berfokus pada implementasi *vector search* untuk dokumen *regional trade agreement*. Agar penelitian ini lebih terfokus sesuai dengan rumusan masalah yang telah dibahas, maka batasan-batasan masalahnya sebagai berikut:

1. *Dataset* yang digunakan dalam penelitian ini merupakan *dataset regional trade agreement* yang berasal dari legal.tina.trade. Sehingga, data yang digunakan hanya data berbasis bahasa inggris.
2. Sistem tidak dapat menggunakan lintas bahasa untuk melakukan pencarian *query*.
3. Evaluasi akurasi menggunakan metrik NDCG@K; metrik lain seperti *Precision@K*, *Recall@K*, atau *F1-Score* tidak digunakan dalam penelitian ini.

## 1.4 Tujuan Penelitian

Pada bagian ini disampaikan tujuan yang hendak dicapai dari penelitian ini adalah:

1. Mengimplementasikan sebuah arsitektur *vector search* yang mengintegrasikan model SBERT dengan *Vector Database Management System* (VDBMS) Qdrant untuk melakukan pencarian *regional trade agreement* berdasarkan provisi pada *platform* legal.tina.trade.

2. Mengevaluasi akurasi hasil pencarian menggunakan metrik *Normalized Discounted Cumulative Gain at K* (NDCG@K) untuk mengukur kualitas *ranking* hasil *retrieval*.

### 1.5 Manfaat Penelitian

Pada bagian ini diuraikan manfaat yang akan diperoleh dari penelitian ini adalah:

1. Menghasilkan sebuah *prototype* sistem pencarian yang dapat membantu pengguna untuk menemukan informasi spesifik dalam dokumen *regional trade agreement* secara efisien berdasarkan konteks, bukan hanya kata kunci.
2. Mengatasi masalah *vocabulary gap* terhadap studi kasus dokumen *regional trade agreement*.

### 1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

#### a. Bab 1 PENDAHULUAN

Bab ini berisi latar belakang masalah mengenai kompleksitas dokumen *trade agreement* dan kelemahan pencarian kata kunci tradisional. Bab ini juga merumuskan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penulisan.

#### b. Bab 2 LANDASAN TEORI

Bab ini memaparkan landasan teori yang mendukung penelitian, meliputi konsep *Information Retrieval*, model SBERT untuk ekstraksi fitur semantik, mekanisme kerja *vector database* Qdrant, serta teori metrik evaluasi.

#### c. Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai analisis kebutuhan dan perancangan sistem pencarian semantik. Di dalamnya mencakup arsitektur sistem, alur *preprocessing* dokumen *Regional Trade Agreements* (RTA), hingga perancangan skema indeks vektor pada Qdrant.

d. Bab 4 HASIL DAN DISKUSI

Bab ini berisi detail implementasi perangkat lunak dan lingkungan pengembangan yang digunakan. Selain itu, bab ini memaparkan hasil pengujian sistem menggunakan metrik NDCG@K untuk mengevaluasi kemampuan sistem dalam memberikan urutan hasil pencarian yang relevan berdasarkan konteks hukum RTA.

e. Bab 5 KESIMPULAN DAN SARAN

Bab ini menyajikan simpulan dari seluruh hasil penelitian yang telah dilakukan serta memberikan saran-saran untuk pengembangan sistem pencarian dokumen perjanjian perdagangan di masa yang akan datang.

