

Penerapan Knowledge Graph dalam Chatbot Berbasis Inteligensi Artifisial

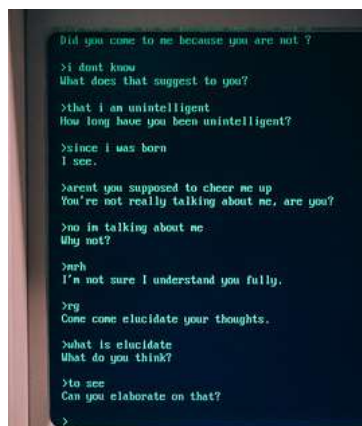
Reza Pahlevi Ubaidillah - 13521165¹
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
¹13521165@std.stei.itb.ac.id

Abstract—Knowledge graph merupakan jenis struktur data graf yang menjelaskan hubungan satu entitas dengan entitas lain melalui relasi yang logis. Struktur data ini dapat digunakan untuk memodelkan chatbot berbasis teknologi inteligensi artifisial.

Keywords—Knowledge graph, chatbot, artificial intelligence, computational linguistic

I. PENDAHULUAN

Dewasa ini, kerap terdengar ledakan demi ledakan riset dan pengembangan dalam bidang *Artificial Intelligence* 'Inteligensi Artifisial', sebuah kecerdasan—sebagaimana manusia—yang diwujudkan oleh komputer. Dari sini, penulis akan merujuk inteligensi artifisial sebagai AI (*Artificial Intelligence*) untuk mempermudah. AI sering disalahmakan sebagai esensi dari bot, padahal bot hanyalah *interface* 'antarmuka' dari suatu layanan, sebagaimana aplikasi ojek online hanyalah antarmuka dari layanan ojek, berarti tidak semua bot memanfaatkan teknologi AI.



Gambar 1.1. ELIZA Chatbot

(Sumber: <https://www.flickr.com/photos/mwichary/44129332940/>)

Salah satu antarmuka yang menerapkan teknologi AI ialah chatbot. Chatbot adalah aplikasi yang melakukan percakapan melalui media teks, didesain sedemikian rupa sehingga seolah-olah perbincangan tersebut dilakukan oleh seorang manusia. Meskipun demikian, saat ini tak ada satu pun chatbot yang mampu lulus uji turing, sebuah ujian untuk menentukan apakah

suatu mesin tak dapat dibedakan dengan manusia biasa.

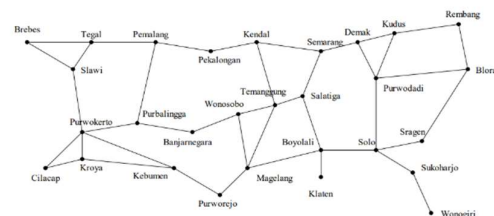
Metode operasi chatbot pada awal mula pengembangannya, di antaranya ELIZA (1966) dan PARRY (1972), berkaitan dengan mengenali kata kunci atau frasa tertentu pada input yang diberikan, kemudian chatbot akan memberikan output berupa respons bersesuaian yang sudah diprogram sebelumnya. Hal inilah yang membedakan chatbot tradisional dengan chatbot yang memanfaatkan teknologi AI.

Pendekatan teknologi AI dapat dibagi menjadi dua kategori utama: non-simbolik dan simbolik. Kategori non-simbolik melakukan pendekatan data training dengan teknik seperti machine learning dan neural network untuk memproses dan menganalisis data. Sementara dalam kategori simbolik, AI belajar dari representasi simbolik dan aturan logika, tanpa adanya data training, di sini AI didesain untuk menjiplak kemampuan manusia dalam pengambilan keputusan.

Dalam makalah ini, akan dibahas pendekatan secara simbolik dengan representasi knowledge graph, di mana terdapat relasi logis yang menghubungkan satu data dengan data lain. Contoh sederhananya adalah tentang ITB, dapat dibuat model dengan fakta-fakta: ITB adalah kampus, ITB terletak di Kota Bandung, Kota Bandung juga dijuluki sebagai kota kembang. Kita bisa menanyakan "Apa julukan dari kota tempat ITB terletak?" tanpa perlu melatihnya dengan pendekatan pembelajaran mesin.

II. TEORI DASAR

A. Graf



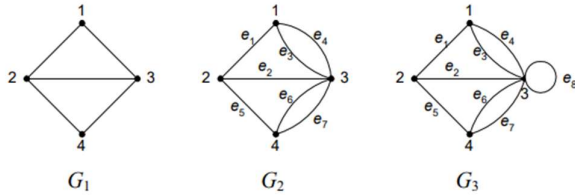
Gambar 2.1. Graf peta jaringan jalan raya di Jawa Tengah
(Sumber: <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/>)

Graf dalam matematika diskrit adalah data terstruktur yang tersusun atas himpunan objek-objek diskrit yang saling berpasangan dengan hubungan tertentu. Objek-objek itu disebut sebagai simpul, *node*, atau *vertices*, dan hubungan

satu objek dengan objek lain disebut dengan sisi, *edges*, atau *arcs*. Secara geometri, graf digambarkan sebagai himpunan lingkaran atau titik representasi simpul dan dihubungkan oleh garis representasi sisi.

Graf dapat dikelompokkan berdasarkan beberapa sudut pandang:

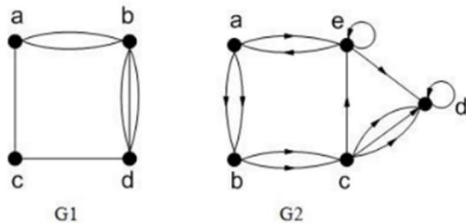
1. Berdasarkan ada tidaknya gelang atau sisi ganda



Gambar 2.2. Tiga buah graf: sederhana, ganda, dan semu (Sumber: <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/>)

- a. Graf sederhana (*simple graph*)
Graf yang tidak mengandung gelang maupun sisi-ganda.
- b. Graf tak-sederhana (*unsimple-graph*)
Graf dengan sisi ganda atau gelang. Ada dua macam, yaitu graf ganda dan graf semu. Graf ganda mengandung sisi yang menghubungkan pasangan simpul lebih dari satu buah. Graf semu mengandung gelang (termasuk bila mengandung sisi ganda).

2. Berdasarkan orientasi arah pada sisi



Gambar 2.3. Graf tak berarah dan berarah (Sumber: <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/>)

- a. Graf tak-berarah (*undirected graph*)
Graf yang sisinya tidak mempunyai orientasi arah.
- b. Graf berarah (*directed graph* atau *digraph*)
Graf yang setiap sisinya memiliki orientasi arah.

B. Linguistik Komputasi

Linguistik adalah studi mengenai bahasa, termasuk grammar, semantik, dan fonetiknya. Linguistik klasik terlibat dalam penurunan dan penilaian aturan-aturan dari suatu bahasa. Masalahnya adalah bagaimana cara membuat aturan-aturan yang dapat dimengerti oleh mesin.

Linguistik komputasi adalah studi yang mengkaji sistem komputer sedemikian rupa sehingga mereka dapat memahami dan menghasilkan bahasa alami sebagaimana manusia. Memahami bahasa alami tentunya membutuhkan pengetahuan luas tentang morfologi, sintak, semantik, pragmatik, serta pengetahuan umum tentang dunia.

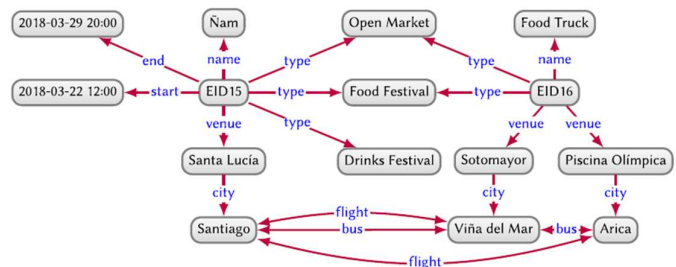
Tanpanya, tak mungkin seseorang bisa mengembangkan sistem bahasa yang efektif dan kokoh. Di balik itu, linguistik komputasi memiliki sisi *engineering* yang disebut dengan *Natural Language Processing*.

Natural Language Processing, berkaitan erat dengan bidang linguistik dan inteligensia artifisial, terfokus pada interaksi komputer dengan bahasa manusia (alami). Pengembangannya melibatkan algoritma dan model komputasional yang bisa memahami, menginterpretasikan, dan menghasilkan bahasa manusia. NLP biasa digunakan untuk menyelesaikan masalah-masalah praktis seperti terjemahan, ringkasan, menjawab pertanyaan, dan sebagainya.

C. Knowledge Graph

Knowledge graph adalah model data berbasis graf yang menyimpan informasi berupa entitas (orang, tempat, atau hal-hal lain) dan hubungan di antara mereka. Knowledge graph dapat memberikan konteks dan informasi latar belakang yang bisa membantu sistem *Natural Language Processing* dalam memahami *intent* dari pengguna.

Knowledge ‘pengetahuan’ sendiri memiliki berbagai definisi yang sudah dikemukakan. Di sini penulis akan menggunakan definisi *explicit knowledge* yang digagas oleh Nonaka dan Takeuchi, yaitu sesuatu yang diketahui dan dapat ditulis. Pengetahuan dapat terdiri dari pernyataan sederhana, seperti “Bandung adalah ibukota Provinsi Jawa Barat”, atau pernyataan terkuantifikasi, seperti “Semua ibukota adalah kota”. Kemudian metode deduktif digunakan untuk mengumpulkan pengetahuan lebih luas lagi, seperti “Bandung adalah kota”.



Gambar 2.4. Directed-edge labelled graph yang mendeskripsikan event dan venue

(Sumber: Diadaptasi dari [4])

Salah satu model yang kerap digunakan ialah *directed edge-labelled graph* (juga dikenal sebagai *multi-relational graph*), atau *del graph*, yang didefinisikan sebagai kumpulan simpul dan sisi berarah yang memiliki label di antara simpulnya. Pemodelan dengan cara ini mempermudah penambahan data baru, cukup dengan menambahkan simpul dan sisi baru. Dibandingkan dengan tree, data dalam graf tidak perlu disusun secara hierarkis. Selain itu, graf juga memungkinkan adanya siklus antara simpul-simpulnya.

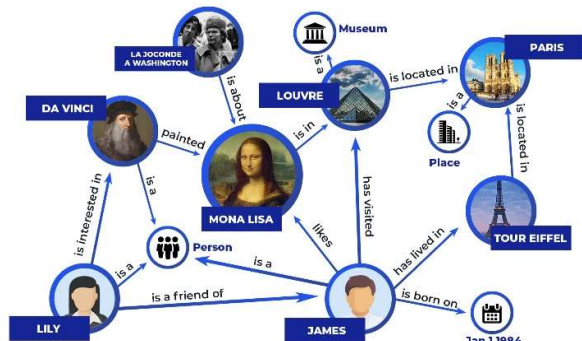
Model-model seperti inilah yang kemudian dikembangkan berdasarkan informasi terstruktur sehingga terbentuk pengetahuan berupa fakta-fakta yang benar dalam konteks tertentu yang direpresentasikan secara eksplisit melalui hubungan antara entitasnya. Misalnya, jika kita mempunyai

knowledge graph yang merepresentasikan informasi tentang negara-negara dan bagaimana hubungan antara satu sama lain, konteks dari suatu negara dapat berupa letak geografis, sistem politik, dan hubungannya dengan negara lain. Dalam knowledge graph, konteks ini direpresentasikan secara eksplisit melalui hubungan antara entitasnya, seperti satu negara dengan negara lain dihubungkan melalui hubungan “borders” yang mana akan menjelaskan konteks geografisnya.

Pada dasarnya, knowledge graph adalah kunci utama bagi mesin untuk memahami makna dan hubungan satu kata dengan kata lain, sebagaimana manusia. Untuk itu, diperlukan sebuah database leksikal lengkap dengan hubungan semantik antara katanya, di antaranya adalah WordNet. WordNet mengklasifikasikan kata menjadi nomina, verba, dan adjektiva kemudian mengumpulkannya dalam himpunan sinonim (*synset*) dan menyambungkan satu sama lain menjadi sebuah konsep.

III. IMPLEMENTASI

A. Membangun Model Knowledge Graph



Gambar 3.1. Knowledge graph sederhana
(Sumber: <https://yashueth.wordpress.com/2019/10/08/introduction-question-answering-knowledge-graphs-kqqa/>)

Akan sangat sulit untuk membangun knowledge graph yang dapat menampung semua cakupan bahasa. Berarti, langkah awal yang perlu dipertimbangkan adalah menentukan use-case dari model yang ingin kita bangun, apakah itu tentang pariwisata, negara, atau apa saja.

Knowledge graph dapat dibangun berdasarkan data terstruktur berupa fakta-fakta mengenai konteks yang berada dalam cakupan kita. Untuk mempermudah pembuatan graf, perlu dilakukan pendaftaran entitas dan hubungan satu entitas dengan entitas lain sesuai dengan cakupan yang sudah diberikan sebelumnya. Hal ini perlu dilakukan untuk menghindari tumpang tindih pada fakta-fakta yang didefinisikan.

Kemudian, perlu dilakukan pembersihan data secara berkala untuk memastikan kualitas model yang dibangun. Hal ini bisa berarti menghapus entri yang tidak valid atau tak berarti, menyesuaikan entri yang mungkin akan bertumpang tindih, memperbaiki ketidakkonsistenan relasi atau fakta, dan lain-lain.

Selanjutnya, model data semantik dapat dibangun dengan menganalisis kerangka utama sehingga dapat tercipta data yang harmonis, di antaranya adalah model del graph.

Penggunaan dataset seperti yang disediakan oleh WordNet akan sangat mempermudah pembuatan model knowledge graph yang efektif dan kokoh.

Perlu diperhatikan bahwa satu knowledge graph yang memuat satu atau beberapa konteks saja tak akan cukup untuk membangun sistem chatbot yang mangkus. Oleh karena itu, beberapa knowledge graph dengan cakupan beragam perlu dibangun. Dalam penyatuan beberapa graf, terkadang ada beberapa cakupan yang memiliki irisan sehingga perencanaan arsitektur utama dari model juga perlu dipertimbangkan.

Untuk memvisualisasikan knowledge graph yang sudah disusun, dapat digunakan kakas seperti GraphViz atau Neo4j sehingga dapat dilihat bagaimana entitas dan hubungan-hubungannya tersusun. Hal ini juga akan sangat mempermudah kita dalam mengidentifikasi kesalahan dalam model yang dibuat.

Natural Language Processing juga dapat dimanfaatkan untuk mengekstrak entitas dan hubungannya dari teks sehingga knowledge graph akan terisi secara otomatis. Hal ini akan sangat menghemat waktu dan tenaga, serta berguna untuk membuat knowledge graph yang lebih komprehensif.

Dalam praktiknya, model dari knowledge graph perlu secara terus menerus diperbarui dan dipelihara sehingga model yang dibuat tetap relevan dan berguna di masa yang akan datang.

B. Membangun Chatbot

Salah satu hal yang paling penting dalam membangun chatbot adalah adanya alur pembicaraan. Hal tersebut dapat dicapai dengan model knowledge graph yang sudah dibangun.

Kemudian, perlu dilakukan tes dan evaluasi terhadap bot untuk memastikan bahwa respons yang diberikan selalu sesuai dengan apa yang diharapkan.

Dalam hal ini, akan selalu diperlukan adanya penyesuaian sehingga dapat bot dapat berkembang.

III. KESIMPULAN

Chatbot menjadi semakin relevan pada masa kini. Namun, chatbot tak bisa dibuat dengan sekadar kumpulan jika-maka, diperlukan inteligensial artifisial untuk membangun chatbot yang komprehensif. Dalam pengembangannya, dapat dilakukan pendekatan dengan knowledge graph sehingga chatbot dapat melakukan pengambilan keputusan seolah-olah manusia.

IV. UCAPAN TERIMA KASIH

Dengan ini, penulis ingin menyampaikan rasa syukur kehadiran Allah SWT atas kehendak-Nya yang telah memberikan kemudahan dalam penulisan makalah ini. Penulis juga ingin berterima kasih kepada Ibu Fariska Zakhralatifa Ruskanda atas bimbingannya dalam mata kuliah Matematika Diskrit. Selanjutnya, penulis juga ingin mengucapkan terima kasih sebanyak-banyaknya kepada keluarga dan teman-teman yang sudah memberikan dukungan dan doa selama proses penulisan makalah ini.

REFERENCES

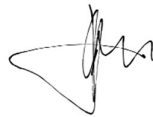
- [1] A. Shevat, *Designing Bots: Creating Conversational Experiences*. Sebastopol, CA: O'Reilly Media, 2017.
- [2] M. Z. Kurdi, *Natural Language Processing and Computational Linguistics 1: Speech, Morphology, Syntax*. Hoboken, NJ: Wiley, 2016.

- [3] I. Nonaka and H. Takeuchi, *The Knowledge-Creating Company*. Oxford University, 1995.
- [4] A. Hogan et al., "Knowledge Graphs", *ACM Computing Surveys*, vol. 54, no. 4, article 71, June 2021.
- [5] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2022-2023/matdis22-23.htm> . Diakses pada 11 Desember 2022 pukul 21.00
- [6]

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 3 Desember 2020



Reza Pahlevi Ubaidillah (13521165)