BUILDING A SIMPLE COVID-19 KNOWLEDGE GRAPH IN BAHASA INDONESIA: A PRELIMINARY STUDY

Arief Purnama Muharram [1][2], Farhan Hilmi Taufikulhakim [1], Ayu Purwarianti [3]

[1] Faculty of Medicine, Universitas Indonesia, Central Jakarta, Indonesia
[2] Master Program of Informatics, School of Electrical Engineering and Informatics, Institut Teknologi Bandung
[3] School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia

OUTLINE

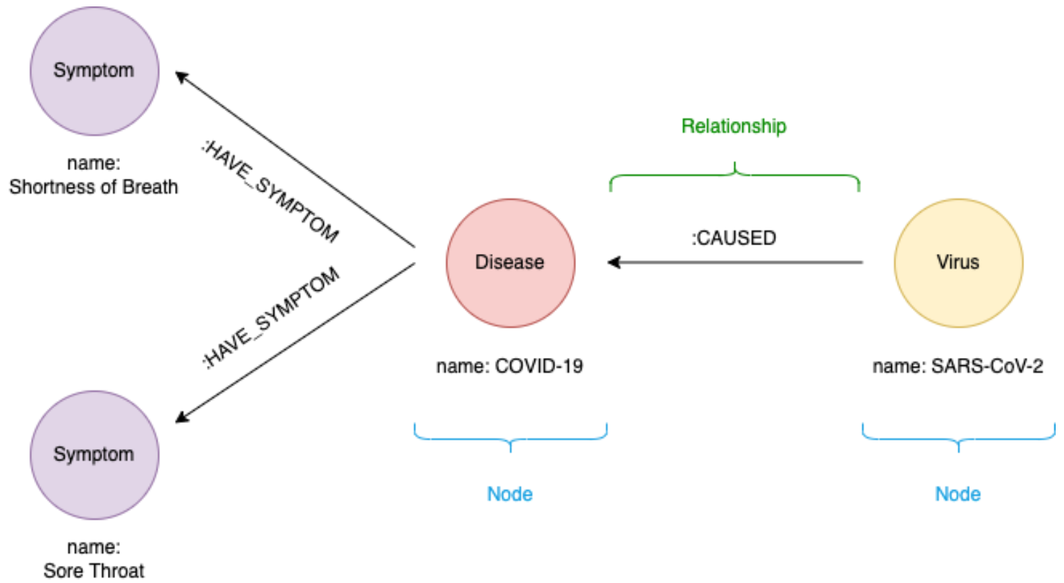
1. Introduction
2. Related Work
3. Method
4. Result
5. Conclusion

INTRODUCTION

- **Coronavirus disease 2019** (also known as **COVID-19**) is an acute respiratory infectious disease caused by the SARS- CoV-2 virus.
- **COVID-19** was first identified as an outbreak in the city of Wuhan, Hubei province, China and has since become a pandemic worldwide.
- To understand COVID-19, researchers and academics have conducted **intensive research on COVID-19**.
- The research findings provide information that shapes our current understanding of COVID-19.
- However, on the other hand, the abundance of information has resulted in a phenomenon known as "**information overload**".
- "**Information overload**" is a phenomenon that occurs due to the abundance of information, resulting in an explosion of information.
- To address the increasing volume of information and to better organize it, there is a growing interest in the development of **knowledge graphs** as a comprehensive means of organizing this information.

INTRODUCTION

- A **Knowledge Graph** (**KG**) is a graph data that contains "real-world knowledge", where **nodes** represent "real-world entities" and **edges** represent the "relationships" between those "entities".
- Nodes can also be referred to as **vertices**, while relationships can be called **edges**.
- In the medical context, an example of a node can be an object such as a disease or a symptom.
- On the other hand, a relationship example could be "HAVE SYMPTOM," which represents a relationship of a symptom of a disease.



OBJECTIVES

- Preliminary study on constructing a **COVID-19 KG using Bahasa Indonesia** (the Indonesia language).
- Our motivation arises from the interest in **organizing COVID-19 information** into a **comprehensible concept model** while promoting the use of **Bahasa Indonesia** in KG development.
- Main contributions:
  - The definition of the KG schema, including node and relationship labels and their interconnections
  - The KG that captures COVID- 19 information using the schema.

RELATED WORK

- Recently, there has been a growing interest in using KG in the biomedical field.
- The example of KG utilization in biomedicine can range widely, from creating intelligent systems to enhance healthcare services, developing models to comprehend diseases, exploring drug interactions, to providing personalized care for patients.
- **Hamed et al.** used the KG to predict various drug combinations for the COVID- 19 case.
  - The KG in their study to reconstruct knowledge regarding drugs, genes, and diseases information from various biomedical literature.
  - The KG used to identify the risks of drug interactions, thereby generating predictions of drug combinations that can be further considered by scientists in the further investigations.
- A study by **Chai** used KG to diagnose thyroid disease.
  - The results indicated that the use of KG and deep learning yielded better prediction results for diagnosing the disease.
- A study by **He et al.** used KG to visualize interactive dietary supplement information.

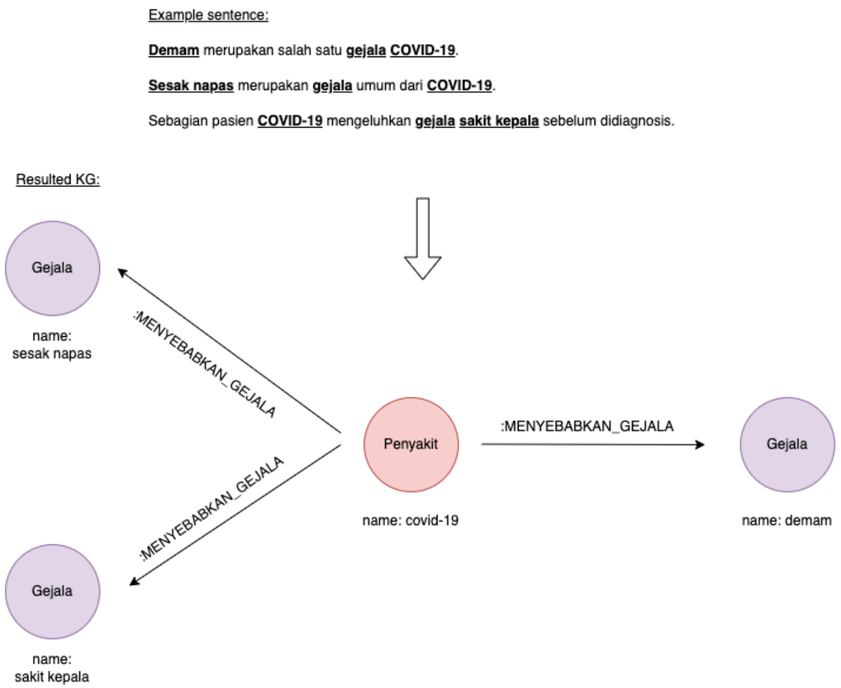
A. A. Hamed, J. Jonczyk, M. Z. Alam, E. Deelman, and B. S. Lee, "Mining Literature-Based Knowledge Graph for Predicting Combination Therapeutics: A COVID-19 Use Case," in 2022 IEEE International Conference on Knowledge Graph (ICKG), Orlando, FL, USA, 2022, pp. 79-86, doi: 10.1109/ICKG55886.2022.00018.
X. Chai, "Diagnosis Method of Thyroid Disease Combining Knowledge Graph and Deep Learning," in IEEE Access, vol. 8, pp. 149787-149795, 2020, doi: 10.1109/ACCESS.2020.3016676.
X. He et al., "Prototyping an Interactive Visualization of Dietary Supplement Knowledge Graph," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1649-1652, doi: 10.1109/BIBM.2018.8621340

RELATED WORK

- **Gu¨tebier et al.**'s study successfully created a "**CovidGraph**."
  - Serves as a large-scale COVID-19 KG encompassing information from publications, patents, clinical trials, as well as biomolecular and systems biology data related to COVID-19.
  - The CovidGraph comprises approximately 36 million nodes and 60 million relationships.
- **Wise et al.**'s research resulted in the creation of a COVID-19 Knowledge Graph (CKG) using the CORD19 Open Research Dataset,
  - The KG includes 336,887 entities and 3,332,151 relationships.
- **Domingo-Ferna´ndez et al.** constructing a COVID-19 KG using selected journals to establish a cause-and-effect knowledge model of COVID-19 pathophysiology.

L.Gu¨tebieretal.,"CovidGraph:aGraphtoFightCOVID-19,"inBioin- formatics, vol. 38, no. 20, pp. 4843-4845, Oct. 2022, doi: 10.1093/bioin- formatics/btac592.
C. Wise, V. Ioannidis, M. R. Calvo, X. Song, G. Price, N. Kulkani, R. Brand, P. Bhatia, and G. Karypis, "COVID-19 Knowledge Graph: Accelerating Information Retrieval and Discovery for Scientific Literature," in AACL-IJCNLP 2020 Workshop on Integrating Structured Knowledge and Neural Networks for NLP (KNLP), 2020. [Online]. Available: https://www.amazon.science/publications/covid-19- knowledge-graph-accelerating-information-retrieval-and-discovery-for- scientific-literature.
D. Domingo-Ferna´ndez et al., "COVID-19 Knowledge Graph: a Computable, Multi-Modal, Cause-and-Effect Knowledge Model of COVID-19 Pathophysiology," in Bioinformatics, vol. 37, no. 9, pp. 1332–1334, May 2021. [Online]. Available: https://doi.org/10.1093/bioinformatics/btaa834.

METHOD



- The KG modeling was conducted using a **domain expert-based approach**, which involved the participation of domain experts in discussions and analyses for creating both the KG schema and the KG itself.
- We use **Neo4J Graph Database** version 5 to implement our KG.
- For evaluation strategy, we conducted several tests to query the graph with the aims to ensure that the graph can be modeled well.
- **Problem limitations:**
  - The source of COVID-19 information is limited to the source selected by the authors.
  - Our KG will only cover the areas of etiology, pathogenesis, pathophysiology, risk factors, complications, and treatments of COVID-19 in general.
  - The annotation process for nodes and relationships is done manually by humans. Therefore, there is a potential for errors or biases in their determination.

METHOD

Authors	Title	Type of Reference
[13]	Pedoman tatalaksana covid-19	National guideline
[14]	Etiology of community-acquired and hospital-acquired pneumonia associated with COVID-19	Original article
[15]	A comprehensive review of various categories of face masks resistant to COVID-19	Review article
[16]	Differential diagnosis of coronavirus disease 2019 from community-acquired-pneumonia by computed tomography scan and follow up	Original article
[17]	Coronavirus disease 2019 (COVID-19): multisystem review of pathophysiology	Review article
[18]	COVID-19: current understanding of its pathophysiology, clinical presentation and treatment	Review article
[19]	Viral replication of SARS-CoV-2 could be self-limitative - the role of the renin-angiotensin system on COVID-19 pathophysiology	Review article
[20]	More than 50 long-term effects of COVID-19: a systematic review and meta analysis	Meta analysis
[21]	Efficacy and safety of three new oral antiviral treatment (molnupiravir, fluvocamine, and paxlovid) for COVID-19: a meta analysis	Meta analysis
[22]	Mechanism and adverse effects of COVID-19 drugs: a basic review	Review article

References Used to Build the COVID-19 KG

RESULT



- We have successfully built a simple COVID-19 KG using Bahasa Indonesia
- The KG consists of:
  - **15** node labels and 14 relationship labels type
  - **240** and **276** counted nodes and relationships, respectibely

RESULT

Node Label	Description	Examples	Count
Enzim	To describe the entity of enzyme molecules.	*angiotensin converting enzyme 2 (ace-2)*	1
Gejala	To describe entities such as conditions or events related to the symptoms of a disease.	*sesak napas*, *demam*	39
Hormon	To describe the entity of hormone molecules.	*endothelin-1 (et-1)*, *angiotensin ii*	2
Jaringan	To describe the entity of tissues of a specific body organ or body part.	*mukosa hidung*, *mukosa mata*	9
Organ	To describe the entity of a specific body organ or body part.	*paru-paru*, *hidung*	9
Patofisiologi	To describe the entity such as conditions or events related to the physiological process of a disease.	*pyroptosis*, *badai sitokin*	22
Pemeriksaan	To describe the entity such as methods or modalities related to the medical examination of a disease.	*x-ray toraks*, *ct-scan*	21
Penyakit	To describe the entity of diseases.	*covid-19*, *sepsis*	34
Preventif	To describe the entity such as methods or protocols related to the prevention of a disease	*memakai masker*, *mencuci tangan*	15
SitokinProInflamasi	To describe the entity of inflammation molecules related to a inflammation process caused by a disease.	*interleukin 1 beta (il-1 beta)*, *interleukin 6 (il-6)*	10
StrukturVirus	To describe the entity of proteins or parts related to the structure of a virus.	*protein membrane (m)*	4
Terapi	To describe the entity of methods, protocols, or things (such as drugs) related to the treatment of a disease.	*terapi oksigen*, *kortikosteroid*	51
Transmisi	To describe the entity of methods related the transmission of a disease.	*droplet udara*, *aerosol*	2
Vaksin	To describe the entity of specific vaccines of a disease.	*pfizer*, *moderna*	9
Virus	To describe the entity of viruses.	*delta*, *omicron*	12
Total			**240**

Node Labels

RESULT

Relationship Label	Description	Count
BERIKATAN_DENGAN	To describe the relationship between a molecule (such as a hormone or enzyme) and a specific receptor, tissue, organ, or other body part entity. **Example:** {(”protein spike (s)”, ”BERIKATAN_DENGAN”, ”angiotensin converting enzyme 2 (ace-2)”)}	1
DITULARKAN_MELALUI	To describe the relationship between a disease and method related to transmission of a disease entity. **Example:** {(”covid-19”, ”DITULARKAN_MELALUI”, ”droplet_udara”)}	2
MEMILIKI_DIAGNOSIS_ DIFERENSIAL	To describe the relationship between one disease and another disease entities that represent a differential diagnosis of a disease. **Example:** {(”covid-19”, ”MEMILIKI_DIAGNOSIS_DIFERENSIAL”, ”community acquired pneumonia”)}	2
MEMILIKI_EFEK_SAMPING	To describe the relationship between the entity of a method, protocol, or thing related to the treatment of a disease and an event or condition entity that represents a side effect of the treatment of a disease. **Example:** {(”azithromycin”, ”MEMILIKI_EFEK_SAMPING”, ”diare”)}	35
MEMILIKI_KOMPLIKASI	To describe the relationship between a disease entity and an event, condition, or another disease that represents a complication caused by the disease. **Example:** {(”covid-19”, ”MEMILIKI_KOMPLIKASI”, ”sindrom distres pernapasan akut”)}	21
MENYEBABKAN	To describe a ”cause” relationship between entities generally. **Example:** {(”badan sitokin”, ”MENYEBABKAN”, ”tesis”)}	19
MENYEBABKAN_GEJALA	To describe the relationship between a disease entity and an event or condition related to a symptom that represents a manifestation of the disease. **Example:** {(”covid-19”, ”MENYEBABKAN_GEJALA”, ”batuk kering”)}	39
MENYEBABKAN_ PENINGKATAN	To describe an ”increase” relationship between entities. **Example:** {(”severe acute respiratory syndrome coronavirus-2 (sars-cov-2)”, ”MENYEBABKAN_PENINGKATAN”, ”sitokin pro-inflamasi”)}	8
PADA_SAMPEL	To describe the relationship between the entity related to medical examination and the entity of tissue, organ, or body part that represent the sample used to conduct a medical examination. **Example:** {(”histopatologi”, ”PADA_SAMPEL”, ”ginjal”)}	7
STRATEGI_PEMERIKSAAN	To describe the relationship between the disease entity and the entity of methods or protocols related to the medical examination that represent the medical examination of a disease. **Example:** {(”covid-19”, ”STRATEGI_PEMERIKSAAN”, ”x-ray toraks”)}	7
STRATEGI_PREVENTIF	To describe the relationship between the disease entity and the entity of methods or protocols related to the prevention strategy that represent the prevention of a disease. **Example:** {(”covid-19”, ”STRATEGI_PREVENTIF”, ”isolasi”)}	4
STRATEGI_TERAPI	To describe the relationship between a disease and the methods, protocols, or treatments entities associated with it. **Example:** {(”covid-19”, ”STRATEGI_TERAPI”, ”terapi oksigen”)}	11
TERDAPAT_DI	To describe a “can be found at/in/on” relationship between entities. **Example:** {(”angiotensin converting enzyme 2 (ace-2)”, ”TERDAPAT_DI”, ”mukosa mulut”)}	12
TERDIRI_ATAS	To describe a “consist of” relationship between entities. **Example:** {(”obat antivirus”, ”TERDIRI_ATAS”, ”remdesivir”)}	108
Total		**276**

Relationship Labels

RESULT

Task	Cypher Query	Results
Show the symptoms of COVID-19 disease	MATCH(n:Penyakit{name: ”covid-19”}) -[r:MENYEBABKAN_GEJALA]->(m) RETURN *	nyeri perut, diare, anosmia, nyeri otot, hidung tersumbat, nyeri tenggorokan, sesak napas, batuk kering, demam, muntah, malaise, nyeri sendi
Show the complications caused by the COVID-19 disease	MATCH(n:Penyakit{name: ”covid-19”}) -[r:MEMILIKI_KOMPLIKASI]->(m) RETURN *	sepsis, aritmia, gagal jantung, stroke hemoragik, stroke iskemik, emboli paru, perikarditis, koagulasi intravaskular diseminata, pankreatitis, gagal hati akut, gagal ginjal akut, infark miokardia, sindrom inflamasi multisistem, gagal pernapasan akut, sindrom distres pernapasan akut, rhabdomyolysis, tamponade jantung, diabetes, miokarditis
Show the side effects of using dexamethasone as a COVID-19 treatment	MATCH(n:Terapi{name: ”deksametason”}) -[r:MEMILIKI_EFEK_SAMPING]->(m) RETURN *	pusing, insomnia, depresi, jerawat, menstruasi, gangguan pencernaan, muntah, nyeri kepala
Show the examination strategy of COVID-19	MATCH (n:Penyakit{name: ”covid-19”}) -[r:STRATEGI_PEMERIKSAAN]->(m) RETURN * LIMIT 5	x-ray toraks, ct-scan, pemeriksaan darah, pemeriksaan molekuler, pemeriksaan ekg
Show the therapy strategy of COVID-19	MATCH (n:Penyakit{name: ”covid-19”}) -[r:STRATEGI_TERAPI]->(m) RETURN * LIMIT 5	terapi oksigen, vitamin, obat antikoagulasi, obat imunomodulatori, obat antivirus

Knowledge Graph Evaluation

CONCLUSION

- We have successfully demonstrated a **simple COVID-19 KG using Bahasa Indonesia**.
- Our COVID-19 KG aims to model information about COVID-19 in a comprehensive graph way.
- With the KG, we can obtain clearer and more understandable information, thus reducing the risk of information overload.
- Since our study is a preliminary one, further research is needed to develop this KG into an even better.

THANK YOU