Pontifícia Universidade Católica
do Rio de Janeiro

Veronica dos Santos

Context Augmented Knowledge Graphs for Decision-making Scenarios

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências - Informática.

Advisor : Prof. Sérgio Lifschitz
Co-advisor: Dr. Daniel Schwabe

Rio de Janeiro
November 2023

Pontifícia Universidade Católica
do Rio de Janeiro

Veronica dos Santos

Context Augmented Knowledge Graphs for Decision-making Scenarios

Thesis presented to the Programa de Pós-graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências - Informática. Approved by the Examination Committee:

Prof. Sérgio Lifschitz
Advisor
Departamento de Informática – PUC-Rio

Dr. Daniel Schwabe
Co-advisor
Pesquisador Autônomo

Prof. Edward Hermann Haeusler
PUC-Rio

Profa. Fernanda Araujo Baião
PUC-Rio

Prof. Altigran Soares da Silva
UFAM

Prof. Daniel Cardoso Moraes de Oliveira
UFF

Rio de Janeiro, November 16th, 2023

Veronica dos Santos

Bachelor's degree in Informatics from the State University of Rio de Janeiro (UERJ - 1998), MBA in Software Engineering from UFRJ (2007), and Master's degree in IT from UNIRIO (2011). Certified Associate in Project Management (CAPM)® by PMI.

Bibliographic data

Santos, Veronica

Context Augmented Knowledge Graphs for Decision-making Scenarios / Veronica dos Santos; advisor: Sérgio Lifschitz; co-advisor: Daniel Schwabe. – 2023.

132 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2023.

Inclui bibliografia

1. Informática – Teses. 2. Grafos de Conhecimento. 3. Busca Exploratória. 4. Modelagem de Contexto. I. Lifschitz, Sérgio. II. Schwabe, Daniel. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Those who pass by us do not go alone and do not leave us alone.
They leave a bit of themselves and take a little of us.
The Little Prince - Antoine de Saint-Exupéry

Acknowledgments

To my advisor, Professor Sérgio Lifschitz, and my co-advisor, Daniel Schwabe, for encouraging reflections, conversations and feedback over the years to carry out this work. In the end it was all right.

To CNPq, IBGE, and PUC-Rio, thank you for the support granted, without which this work could not have been carried out.

To my co-workers and superiors at IBGE, especially Jose Luiz Thomaselli Nogueira and Paulo David de Jesus Tostes dos Santos, for all their support, patience, and understanding in granting my licenses. I also thank the Qualification and Training Committee (CCT), especially the Committee's executive secretary, colleagues Fabiane Lucena Cavalcante and Ana Cristina Von Calmbach, for all their support throughout the process.

To all colleagues from PUC-Rio. In particular, to my research group colleagues Ema, Antony, and Mariana, and my classmates such as Bruno, Leonardo, and Cezar. I would also like to thank a great friend I gained on this journey, Luciana Perciliano.

To the professors who participated in the examining committee for their suggestions. In particular, I thank Professor Hermann for his collaboration with this research regarding its philosophical aspects.

To all the professors in the Department for their teachings. The secretariat and IT support team were very helpful at all times.

My family for their encouragement and for understanding my absence. To my dogs, Gaza and Zulu, thank you for their company. To my body mentor Isabela Callado for the support with embodiment therapy.

For two years, we have been experiencing a health emergency across the planet. I thank everyone who in some way acted to ensure that everyone around me and I were safe and healthy.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Santos, Veronica; Lifschitz, Sérgio (Advisor); Schwabe, Daniel (Co-Advisor). **Context Augmented Knowledge Graphs for Decision-making Scenarios**. Rio de Janeiro, 2023. 132p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

In decision-making scenarios, an information need arises when an agent, human, or machine needs more knowledge to decide due to a knowledge gap. Users can consciously take the initiative to acquire knowledge to fill this gap through information search tasks. User queries can be incomplete, inaccurate, and ambiguous. It occurs because part of the information needed is implicit or because the user does not fully understand the domain or the task that motivates the search. This condition is foreseen within the exploratory search approaches. Although Knowledge Graphs (KG) are recognized as information sources with great potential for data integration and exploratory search, they are incomplete by nature. Besides, Crowdsourced KGs, or KGs constructed by integrating several different information sources of varying quality, need a Trust Layer to be effective. The evaluation of knowledge truthfulness depends upon the contexts of claims and tasks being carried out or intended (purpose). This research aims to prepare and query KGs to support context-aware exploration in decision-making scenarios. The contributions include a framework for Context Augmented Knowledge Graphs-based Decision Support Systems composed of a Decision Layer, a Trust Layer, and a Knowledge Layer that operates under a Dual Open World Assumption. The Knowledge Layer comprises a Context Augmented KG (CoaKG) and a CoaKG Query Engine. CoaKG contains contextual mappings to identify explicit context and rules to infer implicit context. CoaKG Query Engine is designed as a query-answering approach that retrieves all contextualized (possible answers) from the CoaKG. Wikidata is the object of a Proof of Concept to evaluate the effectiveness of the Knowledge Layer.

Keywords

Knowledge Graphs; Exploratory Search; Context Modeling.

Resumo

Santos, Veronica; Lifschitz, Sérgio; Schwabe, Daniel. **Grafos de Conhecimento Enriquecidos de Contexto para Cenários de Tomada de Decisão**. Rio de Janeiro, 2023. 132p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Em cenários de tomada de decisão, quando um agente, humano ou máquina, necessita de mais conhecimento para decidir devido a uma lacuna de conhecimento, surge uma necessidade de informação. Os usuários podem conscientemente tomar a iniciativa de adquirir conhecimento para preencher essa lacuna através de tarefas de buscas por informação. As consultas do usuário podem ser incompletas, imprecisas e ambíguas. Isso ocorre porque parte da informação necessária está implícita ou porque o usuário não compreende totalmente o domínio ou a tarefa que motiva a busca. Esta condição está prevista nas abordagens de busca exploratória. Embora os Grafos de Conhecimento (KG) sejam reconhecidos como fontes de informação com grande potencial para integração de dados e busca exploratória, eles são incompletos por natureza. Além disso, KGs Crowdsourced, ou KGs construídos pela integração de diversas fontes de informação de qualidade variável, precisam de uma Camada de Confiança para serem eficazes no suporte a processos de tomada de decisão. A avaliação da veracidade do conhecimento depende dos contextos das alegações e das tarefas a serem realizadas ou pretendidas (propósito). Esta pesquisa tem como objetivo preparar e consultar KGs para apoiar a exploração ciente de contexto em cenários de tomada de decisão. As contribuições incluem uma arquitetura para sistemas de apoio à decisão, composta por uma Camada de Decisão, uma Camada de Confiança e uma Camada de Conhecimento que opera sob a hipótese de Mundo Aberto Dual. A Camada de Conhecimento é composta por um Grafo de Conhecimento enriquecido de Contexto (CoaKG) e uma Máquina de Consulta baseada em CoaKG. CoaKG estende um KG padrão com mapeamentos de contexto para identificar o contexto explicitamente representado e regras para inferir o contexto implícito. A máquina de Consulta baseada em CoaKG foi projetada como uma abordagem de resposta a consultas que recupera todas as respostas contextualizadas (possíveis). A Wikidata é objeto de uma Prova de Conceito para avaliar a eficácia da Camada de Conhecimento.

Palavras-chave

Grafos de Conhecimento; Busca Exploratória; Modelagem de Contexto.

Table of contents

List of figures

List of tables

List of Abreviations

BGP – Basic Graph Patterns

CoaKG – Context Augmented KG

CGP – Complex Graph Patterns

CWA – Closed World Assumption

DSS – Decision Support Systems

GQL – Graph Query Language

ILA – Information-literate actions

KBDSS – Knowledge-based Decision Support Systems

KG – Knowledge Graph

LLMs – Large Language Models

NGP – Navigational Graph Patterns

OWA – Open World Assumption

PoC - Proof of Concept

RPQ - Regular Path Query

WD – Wikidata

Six blind men are brought to examine an elephant that has come to their village.
The first man touches the trunk and says the elephant is like a thick snake.
The second man touches the tusk and says the elephant is like a spear.
The third man touches the ear and says the elephant is like a fan.
The fourth man touches the leg and says the elephant is like a tree.
The fifth man touches the side and says the elephant is like a wall.
The sixth man touches the tail and says the elephant is like a rope.
Each blind man is convinced he is right and everyone else is wrong.

The Blind Men and The Elephant, *Tittha Sutta - 500 B.C.*

1
Introduction

Trust judgments are made by humans based on their prior knowledge about the domain, common sense, and even past personal experience and beliefs. Such judgments allow people to act under uncertainty and deal with the risk of negative consequences. However, when an agent, human or machine, has inadequate knowledge to achieve a goal due to a knowledge gap (a difference between an ideal state and the actual state of knowledge), an information need situation arises [1].

1.0.1
Searching for Information

Information needs can range from basic information used in short-term actions, such as the weekly weather forecast, to information that explains a broader phenomenon, such as the relation between dollar exchange rate variation and gasoline prices. Continuous knowledge accumulation that experts undertake throughout their professional lives is also motivated by information needs. Similarly, decisions involving health and happiness, such as getting vaccines or having kids, are also influenced by information-seeking behaviors. Search and decision are part of daily lives [1].

Different information needs require different search strategies that specialized computational tools can support. Lookups searches, where discrete and well-structured objects are returned as search results, are well supported by database management systems (DBMS) and Web search engines using fact/information retrieval, query answering, navigation, and known item search approaches [2].

Learning and investigating using information from the web and social networks, with multiple, distributed, and even contradictory sources, requires considering contextual information for trustworthiness evaluation. The post-truth[1] phenomena and the current spread of disinformation show that Truth

[1] According to the Oxford Dictionary, the Word of the Year 2016 is defined as *"relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief"* `https://languages.oup.com/word-of-the-year/2016/`

depends on what is accepted by a community or even by the individual in charge of a decision [4].

Users must decide what is relevant or not, what is reliable, and which source of information they trust to consider the information accurate and helpful in carrying out the task one has in mind. Taking a claim as fact depends on the contextual information that qualifies it, the context constraints of the task in which that claim will be applied, and the individual, organization, or community's trust policies. Although this additional meta information is necessary, it is possible that the user does not include it in the initial formulation of the query..

Since **Context** is an overloaded concept, this thesis adopts a definition proposed by Hogan et al., 2022 [3]: "*By context, we herein refer to the scope of truth, and thus talk about the context in which some data are held to be true*". The scope circumscribes limits to interpreting information in time, place, and according to its origin. For example, in the case of Brazil's capitals, the start and end dates define when the relationship was valid. Therefore, even if the original query does not specify the context information to be added, it must be retrieved to interpret the answers correctly. Here, the notion of Truth concerns trusting some information to the point of taking an action based on it. This is the perspective of Pragmatic Theories of Truth from philosophy [4].

Trust can be defined as "*knowledge-based reliance on received information*" ([5] *apud* [6]); that is, someone decides to trust (or not) the knowledge acquired to the point that this decision to trust implies the decision to carry out some action, based on the truth of information received or already known [6].

In this thesis, we assumed that Figure 1.1 represents how an agent relies on Trust to take action as proposed by Schwabe et al., 2020 [7]. The inputs of the Trust Process are: data/information composed by data itself, general metadata, context metadata, and trust policies retrieved from the Knowledge Repository. Trust policies are agent and task-dependent and are applied using general and context metadata. The agent selects the appropriate policies to be applied, guiding the process. The output of the Trust Process is Trusted Data/Information, and the user decides to act or not based on it. Decisions generate additional information stored in the Knowledge Repository to be used in upcoming decisions.

Analysing this information flows, some questions were identified as follows. These questions serve as the base for defining the research problem enunciated in Chapter 3.

– If the repository consists of entities and claims with respective meta infor-

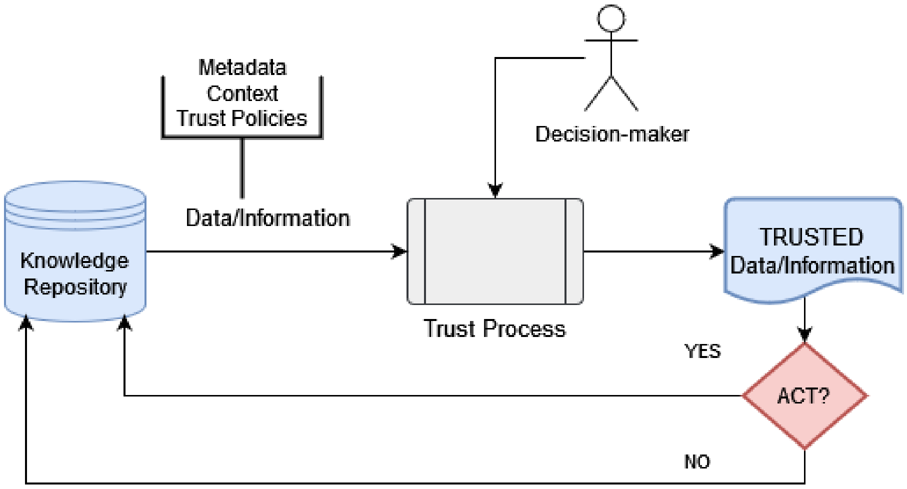


Figure 1.1: Information flows to support Agent to take Action

mation regarding the context, how can these elements be differentiated?

- How do we identify the different types of context?
- How do we identify the lack of information about a relevant context for decision-making?
- If queries do not specify context, can they be modified to retrieve the relevant context?
- If context information is implicit and not materialized in the repository, would it be possible to infer it using the retrieved information?
- If the same repository contains all these components of information, how can we identify each part of it to generate contextualized responses to be used by the Trust Process?

There is currently a strong emphasis on conversational Artificial Intelligence (AI) that makes it part of information-seeking tools. Although large language models (LLM) are trained to incorporate language patterns, grammar, and semantics effectively, this does not make them reliable sources of information. The appearance of conversational fluency may enhance the perception of trustworthiness, but this does not guarantee their reliability [8]. Regardless, users must know that the response provided by LLMs and machine learning (ML) methods is the most likely (probable) answer. However, it cannot guarantee accuracy [9]. One of the main drawbacks is that LLMs do not guarantee information about the origin and reliability of the data when asked. Verifying the correctness of the information becomes difficult, and information without context can be misinterpreted or misapplied.

Next, this thesis will present some examples where the context of the information is also critical to understanding the information needed and evaluating the truthfulness of the answer.

1.0.2 Motivating Examples

Consider the two following claims:
(i) Water boils at $100 \circ$C;
(ii) Salvador is the capital city of Brazil.
Which is one true, and which is false?

If you are cooking at sea level, the boiling water temperature is $100 \circ$C, but if you are at the top of Mount Everest, Himalayas, water should boil at $75 \circ$C. The first claim can be valid depending on the location where you are cooking. Nevertheless, does everyone know that water's boiling point depends on atmospheric pressure? As the altitude rises, air pressure drops, and lower atmospheric pressure decreases the water's boiling temperature.

If one asks any Large Language Model (LLM), *e.g.* ChatGPT, and any online search engine, *e.g.* Google Search: "*What is the capital of Brazil?*" they would receive answers like Figure 1.2 Observe that both assume the default temporal context of the question as current. However, the query did not mention it. Anyone familiar with the domain knows that the "*capital of*" relationship has an intrinsic temporal context (time-variant geopolitical phenomenon). Nevertheless, if the user did not know, it would be possible to learn this aspect from the responses received from both systems since the context is presented.

Would this answer make the second claim "*Salvador is the capital city of Brazil*" false? Would responding this way still work in all cases? No. The answer's usefulness and validity depend on the circumstances of the task that motivated the information search.

If the user carried out this query motivated by the need to decide where his country should install an embassy in Brazil or if one plans a trip to Brazil's capital next month, this answer would satisfy him. However, in other cases, for example, suppose the user is a historian and has found a document written in the Brazilian Colonial Period that references the capital but does not name the city. To understand the context of this document, it is necessary to find out what the capital was at this time. Since the first capital city was Salvador, between 1549 and 1763, claim (ii) is valid considering the scope of the search.

As a last example, consider the situation where a person wants to claim some inheritance from Anitta Garibaldi by virtue of being a descendant of

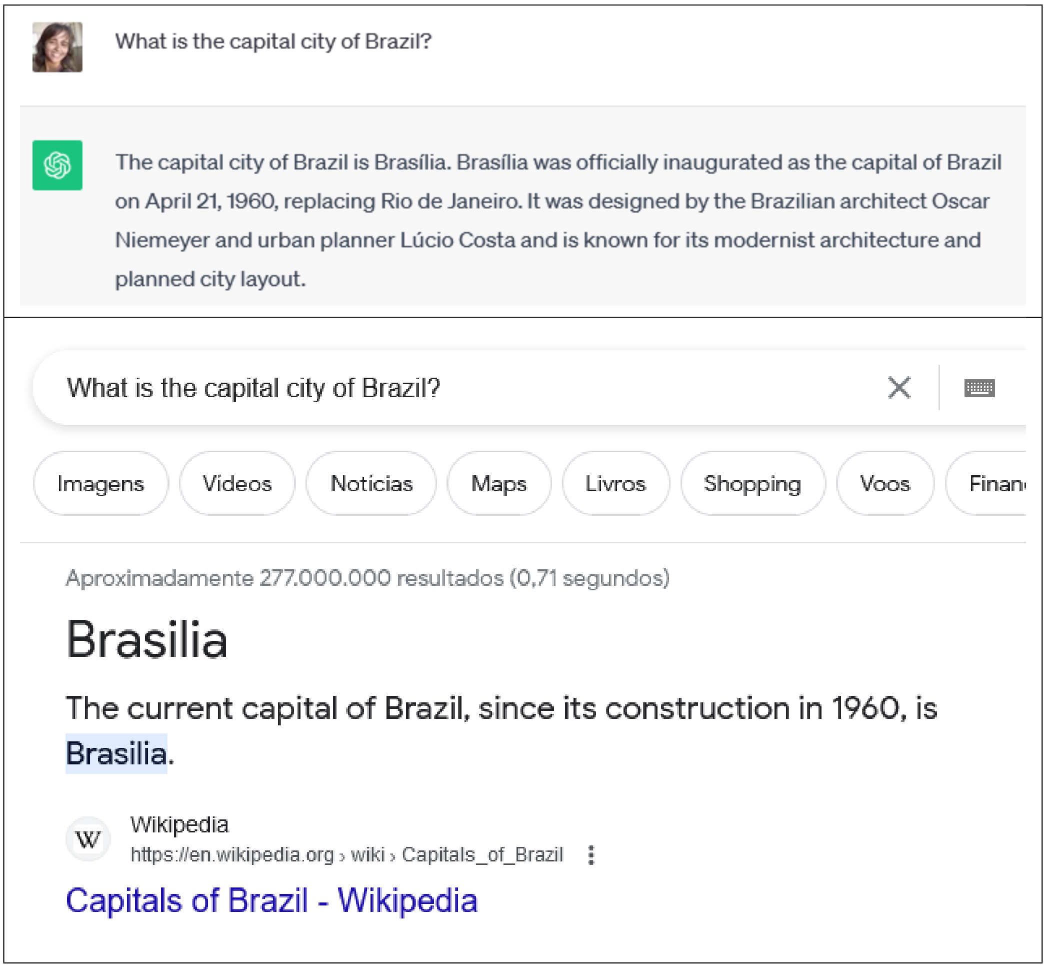


Figure 1.2: Answers from ChatGPT and Google Search

Menotti Garibaldi, the son of the famous revolutionary Giuseppe Garibaldi. Investigating parenthood using Wikidata (WD), there are two claims for who is Menotti Garibaldi's mother (see Figure 1.3). Observe that the WD interface has a sign close to each mother's name. It explains that the "*mother*" relationship violates the *single-best-value* constraint due to missing additional information to clarify multiple values. Which one is true? Can a person have two mothers? If yes, under which circumstances?

Analyzing other statements from WD, it is possible to identify that Menotti had Brazilian nationality and was born in Mostardas City so that the information seeker can search for his birth record in Brazilian registry offices. There is a late registration of Domenico Menotti Garibaldi's birth, son of Ana Maria de Jesus Ribeiro and Giuseppe Garibaldi, born in Rio Grande do Sul on September 16, 1840 (Figure 1.4 obtained from Legislative Memorial of the Legislative Assembly of Rio Grande do Sul). Only on June 12, 2007, he received his Brazilian citizenship due to legal action promoted by the Public Ministry and granted by the Mostardas City Court. Ana Maria de Jesus Ribeiro is the birth name of Anita Garibaldi, so this document confirms the first mother's

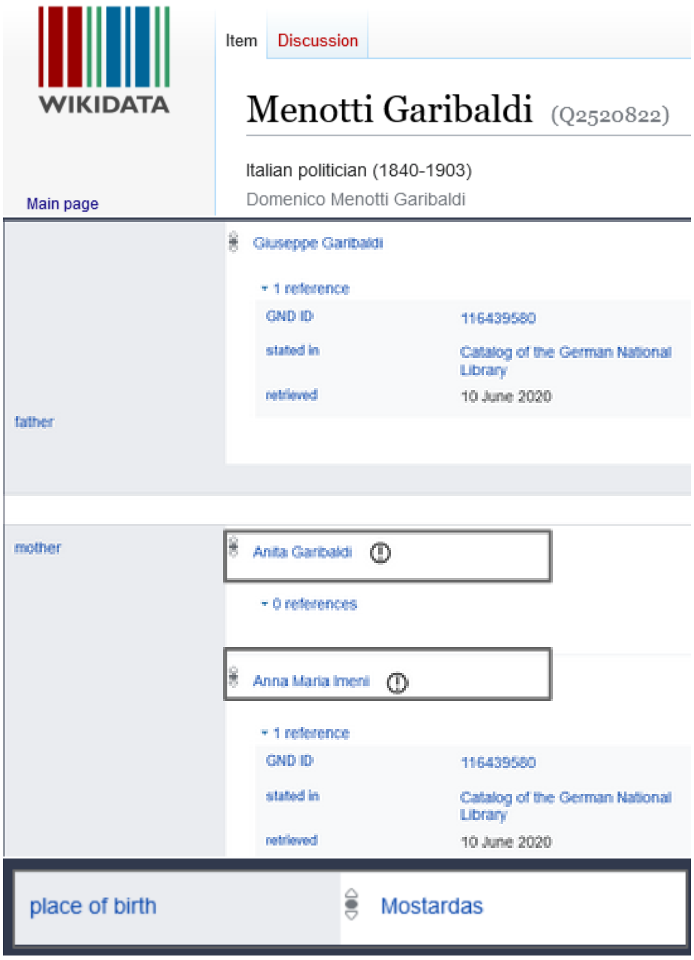
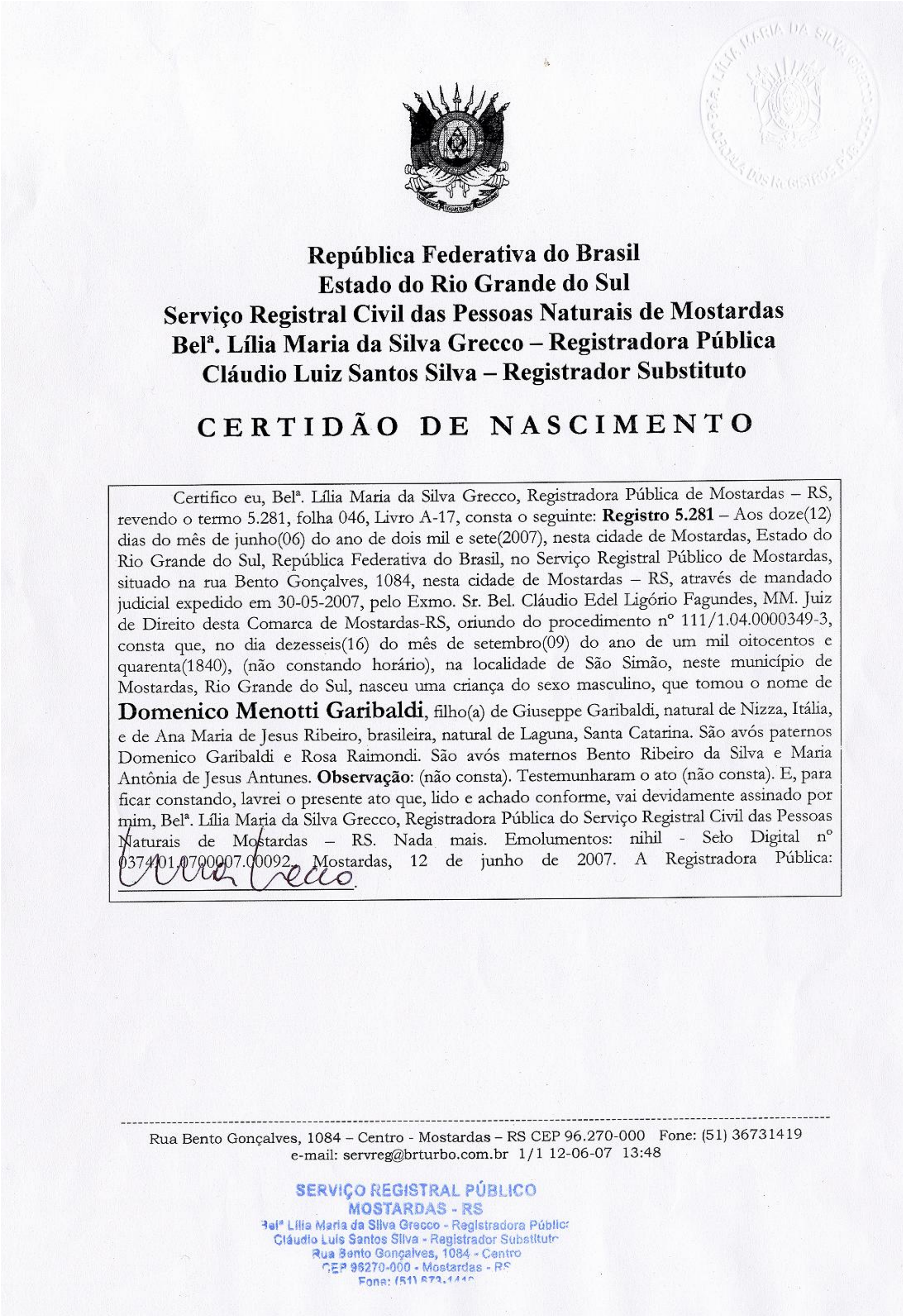


Figure 1.3: Menotti Garibaldi (Q2520822)

name. And who is Anna Maria Imeni? According to the German National Library, used as a reference for the second name, this is another name of Anita Garibaldi, but considering her birth date and parents' name, she is probably his sister. So, to reach this conclusion, it was necessary to access sources external to WD.

All these examples show that, in addition to the information itself, the contextual information is also critical to understanding the information needed and evaluating the truthfulness of the answer. Most search cases can be fulfilled by a default context such as temporal as current time and location as current user location. Such default context are assumed by search engines, but there are still other scenarios in the long tail that need attention to provide the most appropriate answer. Especially when the query does not fully specify the context of interest for the information needed. Efficacy of lookups and exploratory search, as in the berry-picking [10] example of Garibaldi parenthood, can be enhanced with the appropriate additional information.

Searching has evolved beyond merely retrieving pertinent information



República Federativa do Brasil
Estado do Rio Grande do Sul
Serviço Registral Civil das Pessoas Naturais de Mostardas
Belª. Lília Maria da Silva Grecco – Registradora Pública
Cláudio Luiz Santos Silva – Registrador Substituto

CERTIDÃO DE NASCIMENTO

Certifico eu, Belª. Lília Maria da Silva Grecco, Registradora Pública de Mostardas – RS, revendo o termo 5.281, folha 046, Livro A-17, consta o seguinte: **Registro 5.281** – Aos doze(12) dias do mês de junho(06) do ano de dois mil e sete(2007), nesta cidade de Mostardas, Estado do Rio Grande do Sul, República Federativa do Brasil, no Serviço Registral Público de Mostardas, situado na rua Bento Gonçalves, 1084, nesta cidade de Mostardas – RS, através de mandado judicial expedido em 30-05-2007, pelo Exmo. Sr. Bel. Cláudio Edel Ligório Fagundes, MM. Juiz de Direito desta Comarca de Mostardas-RS, oriundo do procedimento nº 111/1.04.0000349-3, consta que, no dia dezesseis(16) do mês de setembro(09) do ano de um mil oitocentos e quarenta(1840), (não constando horário), na localidade de São Simão, neste município de Mostardas, Rio Grande do Sul, nasceu uma criança do sexo masculino, que tomou o nome de **Domenico Menotti Garibaldi**, filho(a) de Giuseppe Garibaldi, natural de Nizza, Itália, e de Ana Maria de Jesus Ribeiro, brasileira, natural de Laguna, Santa Catarina. São avós paternos Domenico Garibaldi e Rosa Raimondi. São avós maternos Bento Ribeiro da Silva e Maria Antônia de Jesus Antunes. **Observação**: (não consta). Testemunharam o ato (não consta). E, para ficar constando, lavrei o presente ato que, lido e achado conforme, vai devidamente assinado por mim, Belª. Lília Maria da Silva Grecco, Registradora Pública do Serviço Registral Civil das Pessoas Naturais de Mostardas – RS. Nada mais. Emolumentos: nihil - Selo Digital nº 0374401.0700007.00092. Mostardas, 12 de junho de 2007. A Registradora Pública: ________________.

Rua Bento Gonçalves, 1084 – Centro - Mostardas – RS CEP 96.270-000 Fone: (51) 36731419
e-mail: servreg@brturbo.com.br 1/1 12-06-07 13:48

SERVIÇO REGISTRAL PÚBLICO
MOSTARDAS - RS
Belª Lília Maria da Silva Grecco - Registradora Públic
Cláudio Luis Santos Silva - Registrador Substitut
Rua Bento Gonçalves, 1084 - Centro
CEP 96270-000 - Mostardas - RS

Figure 1.4: Birth Certificate of Menotti Garibaldi

from a limited set of sources [8]. With nearly anyone's ability to generate and distribute information, it has become progressively more critical to discern the creators of data and their underlying motivation. This discernment is vital to identifying useful and reliable information to be considered **Knowledge** within a particular scope of use.

1.0.3 Searching in Knowledge Graphs

At the same time, Knowledge Graphs (KG) have emerged as practical tools for knowledge sharing and discovery due to their ability to represent complex domain data and long-tail scenarios coverage. A significant use case where KGs have become a key asset is web search engines [11]. KGs offer data modeling and integration flexibility since a predefined schema is not mandatory.

Due to advances in ML techniques, it is feasible to automatically construct large-scale KGs, transforming unstructured content into statements on entities, representing their attributes and relationships between them [12], [13]. Initiatives such as WD enable modeling and sharing knowledge that reflects different perspectives through a crowdsourced multi-lingual and large-scale KG [14].

Therefore, this research focuses on supporting users in the decision process by providing the most contextualized information extracted from KGs. This thesis will cover mapping context information already available in KGs. Mappings are identified during KG Engineering based on the KG Profiling or the KG schema, if available. Competency Questions (CQ) can also be used since it is used in developing ontologies to define a domain's scope and requirements. CQ is a question designed to elicit information about the knowledge and identify the key concepts, relationships, constraints, and rules that must be represented in the ontology. This thesis will also use an appropriate graph representation structure, multi-layer graphs [15], since it has the expressive power to represent contextual information of any element of a KG. Graph queries and the KG can be context incomplete, so the query-expansion approach proposed by this thesis enables their completion with context information using mappings and semantic rules.

1.0.4 Document Organization

The remaining content of this thesis is divided into eight chapters and one appendix:

1. Chapter 2 presents background concepts about Knowledge Graphs, Decision-making, and Information Seeking.

2. In Chapter 3, a general problem and a research problem about KG usage in Decision-making scenarios are enunciated and justified, and the main contributions of this research are highlighted.

3. Chapter 4 presents the Context Augmented KG (CoaKG) definition and schema characterization, which are necessary for modeling explicit contextual information and extracting implicit contextual information from a standard KG. These contributions are illustrated with a realistic example based on the Brazilian geopolitical history domain.

4. In Chapter 5, the query-answering approach that retrieves All Contextualized Answers from a CoaKG within an exploratory search process is described.

5. Chapter 6 discusses how an existing KG can be enriched to become a CoaKG. It describes a Proof of Concept over a CoaKG within an exploratory search process. The CoaKG was built using a subset of WD as a standard KG whose Context Types are specified based on the result of WD Profiling and Competency Questions.

6. Related work is discussed in Chapter 7.

7. Finally, Chapter 8 concludes the thesis by discussing results, summarizing contributions, and pointing to future works about CoaKG.

8. Ongoing work covering experiment design, focusing on trustworthiness for Decision-making, and questionnaire preliminary analysis using LLM are described in Appendix A.

9. Appendix B contains the complete version of tables mentioned in Chapters 4, 5 and 6.

Additionally, a GitHub repository[2] is also available with datasets, scripts, and other research materials.

[2] `https://github.com/versant2612/CKG_UseCases`

2
Foundations

The main background concepts used in this research are introduced in this chapter. This chapter covers Information Seeking, Decision-making, and Knowledge Representation. WD, as an important and open KG, is also covered in more detail, as it is used to build a proof-of-concept (PoC) KG.

2.0.1
Information Seeking

Information seeking encompasses the cognitive process undertaken by individuals to modify their final knowledge state. When an agent, human or machine, has inadequate decision-making knowledge due to a knowledge gap, an information need arises [1]. People can consciously acquire information in response to this need to fill the gap in their knowledge. Any information retrieved by search tasks is potential knowledge. The search will finish when this agent deems that the knowledge gap has been resolved or when the process has reached some restriction in the task context, such as the deadline for making the decision, or systems restriction, such as the access limit, to the data.

Different information needs require different search strategies supported by specialized computational tools. According to Marchionini, 2006 [2], there are three kinds (groups) of information-seeking activities: lookup searches, learn, and investigate. Lookup searches can be associated with "*fact-retrieval*", "*question-answering*", and "*known item search*" tasks. For more straightforward information needs, as in the example of capital cities, one or two results retrieved in a web search engine are enough to bring the expected answers.

Exploratory search comprehends learning and investigation where cognitive activities such as comparison, interpretation, analysis, and synthesis are often combined with lookup activities during an interactive and iterative process. Learning requires multiple iterations to retrieve information submitted to cognitive processing and interpretation to acquire new knowledge. Planning, forecasting, and researching engage users in investigation, transforming existing knowledge into new knowledge and discovering knowledge gaps. An exploratory search is applicable in more complex cases involving uncertainty, as in the example of inheritance in the Garibaldi family. The need for information was met by finding a document whose confidence allowed the decision-making that motivated the search..

User queries can be incomplete, inaccurate, and ambiguous. This occurs because part of the information is implicit, such as the context, or because the user needs help understanding the problem motivating the search. This condition is foreseen within the exploratory search approach through an interactive query refinement and reformulation as the level of uncertainty about the domain decreases. In this process, new queries about the answers should be formulated. This additional information retrieved, known as meta-information, is essential for interpreting and integrating responses. In the case of capital cities, meta-information about validity periods is an example of temporal context. In the case of inheritance, the provenance context of the parenthood relationship corresponds to a birth certificate, a document that is valid for legal issues.

As stated by Smith and Rieh, 2019 [16], if knowledge context is available for searchers during information seeking, they will engage with information more actively and critically, and it can also positively influence users' trust in the information delivered. They defined "*knowledge context*" as the meta-information individuals can use to interpret information on a search engine results page. Such hypotheses motivate experimental research designs target to reflect how information-seeking behavior is affected by knowledge context availability instead of focusing on improving retrieval system performance in terms of precision and recall.

Information-literate actions (ILA) are information-seeking behaviors observed during exploratory search approaches since it is related to learning and investigation activities. ILA [16], part of sense-making, extrapolates users' ability to locate, access, and use information since it engages users in asking questions about the retrieved information, comparing, evaluating, and differentiating information sources. Information-literate searchers typically pose new questions about search results, such as "*Who made this statement?*" "*When was it made?*" "*When was it valid?*" "*Where was it valid?*" "*How have others used it?*" and "*How does it compare with statements from other sources?*"

Shah and Bender, 2022 [8] argued that search systems should support users in increasing their information literacy rather than solely retrieving information using pre-programmed and universal definitions of relevant results. Efficiency in response time, precision, and recall does not translate into effectiveness in satisfying the user's need for information. Furthermore, contextualized responses can avoid the negative consequences of recognizing patterns in datasets that reflect dangerous social biases. Individuals employ search engines not solely to locate specific information but also to engage in learning, exploration, and decision-making processes.

Sense-making is an essential cognitive process that underlies exploratory search, and both are components of the larger information-seeking scope. Sense-making involves interpretation, integration, and hypothesis generation in situations characterized by high complexity or uncertainty [10]. Regarding search, decision support tools can assist users in selecting the most effective paths through the information space. These systems also aggregate data from various sources to equip users with the information necessary to make informed decisions regarding the current task.

At the same time, when faced with a large volume of information to analyze, it is necessary to establish filter criteria and applicability to the intended use of this information. Abductive reasoning can be applied to decision-making with uncertainty to generate possible actions for the target problem and justify the actions taken. Schurz defines abduction [17] apud [18] as "*a search strategy which leads us, for a given kind of scenario, in a reasonable time, to a most promising explanatory conjecture which is then subject to further test*".

2.0.2 Trust in Decisions

Trust can be defined as "*knowledge-based reliance on received information*"; that is, someone decides to trust (or not) the knowledge acquired to the point that this decision to trust implies the decision to carry out some action based on the truth of information received (or already known) [6].

The pragmatic theory of truth redirects the focus from determining the universal criteria for a statement's truth to understanding the individual's intentions and actions when trusting the statement. These theories emphasize how truth is utilized and serves a purpose. The pragmatic theory of Truth typically regards it as a result of individuals' procedures and commitments when solving problems, formulating claims, or conducting scientific investigations [4].

Based on theories of Truth, any symbolic representation of knowledge (truth-bearers) can be either true or false. In comparison, truth-makers are the entities or concepts outside truth-bearers repositories that confer truth upon statements [19]. Some examples involving concrete and abstract truth-bearers are shown below.

Pragmatic approaches to confirming the veracity of those representations center on the process by which truth is established through language, human

ABSTRACT TRUTH-BEARER: "*Innovation drives business competitiveness.*"
TRUTH-MAKER: Notorious companies with a competitive advantage in their market due to technological innovations, product design, and business processes.

ABSTRACT TRUTH-BEARER: "*Education is the key to personal growth and development.*"
TRUTH-MAKER: Countries that prioritize investments in educational systems, access to knowledge, and learning opportunities have a high index of social development and low crime rate.

CONCRETE TRUTH-BEARER: "*The glass is half full.*"
CONCRETE TRUTH-MAKER: A glass that is physically filled to its halfway point.

CONCRETE TRUTH-BEARER: "*The water boiling point is 100 degrees Celsius.*"
CONCRETE TRUTH-MAKER: A thermometer measuring the temperature 100 degrees Celsius inside a container with water on top of a lit stove burner from a house located near the beach.

behaviors (decisions, actions), and contextual considerations.

Uncertainty is an important concept both in information-seeking and decision-making. Whether in search of learning or to make a decision, people are engaged in trying to reduce uncertainty. However, consuming more information is not guaranteed to reduce uncertainty [1]. From Philosophical decision theory, decision-making under uncertainty involves concepts related to expected utility and maximizing rational choice regarding the likelihood of different outcomes arising from a particular choice [20]. An essential concern in Decision Support Systems (DSS) is Trust since the decision-maker has to trust the veracity of the information to be used in order to make an informed decision.

Decision Support Systems (DSS) were developed in the early 1970s for semi-structured or unstructured decision-making activities. A DSS that supports strategic decision-makers using trustable expert knowledge is known as Knowledge-Based DSS (KBDSS). Such systems enable knowledge storage, retrieval, transfer, and application, providing pertinent information in an efficient and easy-to-access manner to allow users to make more informed decisions. The main components of a KBDSS are (i) a knowledge base (with an inference engine) and (ii) a decision support shell [21].

In the field of AI, expert systems were developed as computer systems that aim to replicate the decision-making capabilities of human experts in solving problems within a well-defined, structured, and narrow domain. An expert system comprises the Knowledge Base and Inference Engine subsystem. The Knowledge Base contains facts and rules, while the Inference Engine employs these rules to deduce or infer new facts from the known facts [22]. Expert systems often provide explanations for their recommendations. Users can understand why a particular decision or solution was reached, which enhances their trust in the system.

A KBDSS serves as a bridge between traditional decision support systems and expert systems. It leverages the advantages of knowledge-based reasoning while still allowing for user-driven decision-making.

2.0.3
Knowledge Bases

A database instance is a structured and complete collection of data representing facts that requires a computational theory to be organized and stored for efficient retrieval, management, and manipulation. Here, *complete* is regarding a model of the real-world phenomena the database is designed to represent. Database queries operate under the Closed World Assumption (CWA) hypothesis. Thus, it is possible to ask a Human Resources relational database if: "*Is there a tuple x such that Q(x)*?" and the system can only answer YES or NO since the subject of the query pertains to the stored data structure. In databases, if a tuple does not exist as an instance of a relation T, then it exists as an instance in the complement T [22]. CWA employs Classical Logic, which is non-monotonic. Consequently, adding new facts to a database instance creates a new instance and can influence prior inferences (responses).

A Knowledge Base (KB), which represents propositions some agent believes about the outside world, is incomplete and operates under the Open World Assumption (OWA). The query "*Is there an employee x such that P(x)?*" can generate the answers YES, NO, or DO NOT KNOW since it engages the user to a particular view of the world and requires a semantic theory to interpret their contents. The absence of KB information only means that this knowledge was not made explicit [22]; its truth value is unknown.

Query answers are only those that can be logically proven; if it is not possible to prove a claim, it cannot be assumed that the negation of that claim is true (fact). OWA relies on classical logic and first-order logic, which are monotonic. Under OWA, acquiring new facts does not invalidate previously established knowledge. The Semantic Web, including ontologies and languages like OWL (Web Ontology Language), adopts OWA, and reasoners are used to deduce new facts based on the available information.

One fundamental principle in KB's design is to store only positive statements, *i.e.*, facts known as true. However, there are situations where explicitly stating negative assertions is valuable, such as statements that are known not to hold, even if they contradict common belief or have other significant implications. Decision-makers should make decisions despite the lack of complete information but being aware of it. As an old quote by Dr. Carl Sagan stated in [23]: "*the absence of evidence is not evidence of absence*".

2.0.3.1 Knowledge Graphs (KG)

Knowledge can also be structured as a Knowledge Graph (KG). Several KG definitions can be found in the literature, some related to graph data structures (RDF and Labeled Property Graph) and others focusing on what should be represented. Some definitions can lead to the interpretation that KG is any graph-based knowledge representation, assuming that the term "*knowledge graph*" is synonymous with "*knowledge base*", which itself is often used interchangeably with "*ontology*" [24].

One older definition, dating from the 1980s [25], introduced the concept of a KG to formally depict their knowledge-based system, which integrates knowledge from various sources. Their proposed KGs for representing natural language featured limited word relationships and concentrated on qualitative modeling involving human interaction.

In 2012, the term gained popularity since Google introduced the Knowledge Graph as a semantic improvement to Google's search function, allowing searches for real-world objects rather than just string matching[1]. According to Weikum [11], KGs are KBs modeled as a graph since relationships are the focus of analysis (semantic networks). But they consider the term KG a misnomer since KGs are not limited to binary relationships; they also encompass higher-arity tuples and intensional data through constraints and rules.

Various KGs have been automatically constructed from web data: DBPedia, YAGO (Yet Another Great Ontology), Freebase, WD, Facebook's entity graph, Liquid (LinkedIn), and Microsoft Academic Graph (MAG). These KGs aim to incorporate knowledge from external sources as statements, expanding the concept from a purely KB system to one that embodies integration systems, and these statements are interconnected.

Nevertheless, it is important to differ KGs from KBs in some aspects [12]:

[1]Introducing the Knowledge Graph: things, not strings written by Singhal, A `https://www.blog.google/products/search/introducing-knowledge-graph-things-not/`

(i) KB emphasizes complex logical inference, while modern KGs focus on supporting analytics operations through scalable graph algorithms and neurosymbolic reasoning.

(ii) KB development follows a top-down schema design with manual knowledge engineering. The vast amount of available data sources and ML technology led to a bottom-up methodology for creating KGs. ML extracts named entities, relationships, and properties from text corpora and completes the KG through triple prediction.

(iii) KGs are also built or augmented using human-centric knowledge engineering techniques at large scale such as crowdsourcing;

(iv) KGs are usually bigger than KB in scale.

According to Marx et al., 2017 [26], even though there is no universally accepted concept definition until then, what differentiates a KG from a dataset organized in a graph data structure is the need to enrich this data with information context. Context attributes such as temporal and spatial data added to statements result in higher-arity tuples. Graph data structures for KGs address this using composite objects and qualification of statements.

Another contextual information critical to KG consumption is provenance since a KG can represent information from different perspectives and purposes. Preserving knowledge provenance through recording its source, methods, and extraction time is essential to managing and curating the KG as its content evolves over extended periods [11]. Sikos and Philp, 2020 identified six levels of provenance granularity in RDF KGs [26]:

– Dataset-level,
– RDF document-level (each repository),
– Graph-level (named graphs),
– Molecule-level provenance (subgraph, subset of triples),
– Statement-level (triples reification mechanisms), and
– Element-level (subjects, predicates, and objects of RDF triples).

Statement and element-level provenance help represent diverse claims of disputed or uncertain information from various sources. In the Garibaldi family inheritance example, the claim about his mother being Anita Garibaldi had no provenance context in WD. Therefore, it was necessary to access an external source to locate the birth certificate and confirm the information.

2.0.3.2 KG Data Structures

There are several proposals for KG data structures, some more straightforward, like RDF (directed graphs with labeled edges) and LPG (labeled and directed property graphs), and others more complex and abstract, such as the multi-layer graph (graphs with higher-arity relationships and with identifiers on the edges) [15]. Every edge corresponds to a statement in a KG.

The KG data structure can be mathematically defined as labeled multi-graphs, a graph where multiple edges can connect two nodes, composed of:

– V is a finite set of vertices (or nodes) that represents concepts, entity types, and instances;

– L is a finite set of vertices that represents literals corresponding to property values;

– R is a finite set of binary, directed or undirected, relationship types represented by their labels;

– E is a finite set of edges representing relationships based on a relationship type R between two vertices from V;

– P is a finite set of property (or attribute) types, and

– pV is a finite set of edges representing relationships, based on a property P, between one vertice from V and another from L.

Such graph data structure can be materialized using RDF, a W3C graph representation specification, in which edges are represented by triples without unique identifiers. It aims to portray a directed multi-graph with labeled edges and attributed nodes as a collection of triples in the form (subject, predicate, object).

However, rich knowledge modeling requires capturing temporal, spatial, provenance, and other context attributes, which usually implies having higher-arity tuples. Decomposing them into binary relations (as triples) may cause information loss [11]. Triple-based KGs do not address these requirements natively, which motivated some proposals through RDF reification approaches (RDF-Star, Named Graphs). However, reification and similar techniques make querying more challenging, requiring additional joins and considering paths instead of single edges when dealing with compound nodes in the graph model [23], [31].

Labeled Property Graphs (LPG) is another graph data structure for labeled multi-graphs, used in native GraphDBs such as Neo4J, which enables nodes and edges to have property values as key-value pairs [32]. However, property values are limited to literals. On the other hand, a hyper-relational graph data structure directly attaches key-value pairs to edges whose property values can be another node from V [33], [34]. Besides, key-value pairs are also used to represent n-ary relationships.

A hyper-relational graph structure comprises a tuple containing a uniquely identified triple E or pV associated with a set of 0 to n key-value qualifiers. It adds to the labeled multi-graph the following sets:

– Q is a finite set of qualifier keys represented by their labels;
– qE is a finite set of edges representing relationships, based on a qualifier Q, between one edge from E and a node from V or L, and
– qP is a finite set of edges representing relationships, based on a qualifier Q, between one edge from pV and a node from V or L.

These key-value qualifiers qE allow differentiating instances of relationships E when they involve the same pair of entities from V and the same type of relation from R. The key-value qualifiers qP also differentiate relationship occurrences pV between an entity V and different property values L of the same property P. Table 2.1 shows an example of a KG represented as a hyper-relational graph structure.

<table>
<tr><td colspan="2">V = {Brasilia, Brazil, Salvador, Mundo Educação}</td></tr>
<tr><td colspan="2">L ={April/1960, 1549, 1763, “https://mundoeducacao.uol.com.br/”}</td></tr>
<tr><td colspan="2">R = {CapitalCityOf}</td></tr>
<tr><td colspan="2">P = {url}</td></tr>
<tr><td rowspan="2">E = {e1, e2}</td><td>e1 = (Brasilia, CapitalCityOf , Brazil)</td></tr>
<tr><td>e2 = (Salvador, CapitalCityOf , Brazil)</td></tr>
<tr><td>pV = {p1}</td><td>p1 = (Mundo Educação, url,
“https://mundoeducacao.uol.com.br/”</td></tr>
<tr><td colspan="2">Q = {StartDate, EndDate, Source}</td></tr>
<tr><td rowspan="2">qE = {qe11, qe12,
qe21, qe22, qe23}</td><td>qe11 = (e1, StartDate, April 1960)
qe12 = (e1, Source, Mundo Educação)</td></tr>
<tr><td>qe21 = (e2, StartDate, 1549)
qe22 = (e2, EndDate, 1763)
qe23 = (e2, Source, Mundo Educação)</td></tr>
</table>

Table 2.1: Hyper-relational KG

Multi-layer graph data structure [15] is an abstract and concise model that can naturally support LPG, RDF, and hyper-relational graph data structures. This model avoids reification when modeling n-ary relationships and adding qualification to edges. A multi-layer graph H is an n-layer graph, where each layer results from the nested use of edge IDs, and n represents the highest layer associated with an edge identifier within H. A multilayer graph, H = (O, γ), represents a hyper-relational (multi)graph where edges on edges can exist. In any concrete data model based on multi-layer graphs, O can be divided into different types of possibly disjoint elements, and γ corresponds to directed, labeled, and identified edges between elements. O contains V, L, R, P, and Q, while γ contains E, pV, qE, and qP.

Such syntactic alternatives enable context representation but do not establish how context semantics should be interpreted. Users' queries/searches are often incomplete and implicitly contextualized. However, users generally expect systems, when providing an answer, to be cooperative in the sense of assuming the same context they thought while querying,

Including contextual information to establish the scope of entities and statements improves the data quality. However, how to contextualize query answers is not undoubtedly defined and turns out to be application-dependent. In tasks that require interpretation and analysis of data to support decision-making, the usefulness of context can be perceived. Knowledge contexts can expand searchers' perspectives and stimulate ILA, facilitating learning, critical thinking, and creativity when available in search systems.

2.0.3.3 KG Engineering

Knowledge Graph engineering refers to the process of automatically creating extensive KG and involves refining unstructured and noisy information from various sources (including the Internet) and converting it into clear and precise statements about entities, their characteristics, and their connections [11].

However, unlike traditional databases, KGs often lack a fixed schema to which all data must comply. KGs in practice usually follow a "*pay-as-you-go*" approach, continuously enhancing and refining their content by introducing new types, attributes, and relationships [23]. They are also referred to as *schemaless* or *schema-on-read*. This contrasts with Data Warehouses (DW), a prevalent physical data integration approach, employing relatively fixed schemas optimized for specific multi-dimensional data analysis [13].

It is essential to consider that KG construction is not a one-time task; instead, it should be viewed as an ongoing, "*never-ending*" task, adhering to an Open-World Assumption (OWA) [23]. KGs are long-term data assets necessitating continuous maintenance, as shown in Figure 2.1 extracted from [13]. Their life-cycle includes error correction, adding new entities and statements, marking outdated statements with temporal annotations, expanding attribute and relation schemas, and implementing various quality assurance measures.

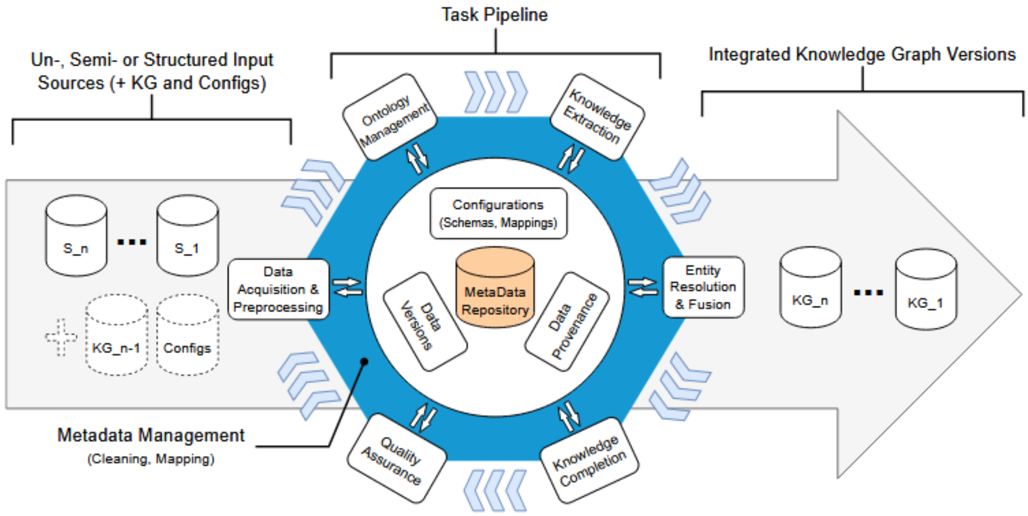


Figure 2.1: KG Generation and Versioning Pipeline

Quality assurance in KG generation requires data cleansing efforts to identify and remove errors and inconsistencies within the input sources during the import process to prevent the propagation of incorrect or low-quality data in the KG. Data cleansing typically involves several sub-tasks to address these issues, including *data profiling* to identify quality problems, *data repair* to rectify recognized problems, *data transformation* to standardize data representations, and *data deduplication* to eliminate duplicate information. KG construction involves a trade-off between precision and recall. The more entities and facts it includes, the greater the chance that some statements may be incorrect [11].

Merging multiple pieces of information about a real-world entity into a single, coherent, and clean representation, known as Data (or Entity) Fusion, constitutes a fundamental stage in data integration, as it consolidates information from entities of different sources into an enhanced entity. It addresses data inconsistencies or conflicts when an entity's records contain apparent attribute-value or relationship discrepancies. Strategies to manage such issues were initially developed for relational data [27] but can also be applied in KG construction:

- **Conflict Ignorance:** This approach does not address the conflict directly, and different attribute values may be retained, or the issue can be postponed for resolution by the user application.
- **Conflict Avoidance:** It applies a uniform strategy to all data, often prioritizing data from trusted sources over others.
- **Conflict Resolution:** This approach comprehensively considers all data and metadata before deciding on a specific strategy, such as selecting the

most frequent, most recent, or randomly chosen value.

KG completion tasks can also be tackled with LLMs [28] or without external sources using link prediction and rule mining. AMIE is a rule mining proposal that does not use schema, only instances. Rules generated could help complete instances and the definition of constraints. AMIE generates closed IF-THEN (Horn) rules that apply only to binary relationships [29].

Link prediction is the task of predicting an entity from V related to another given entity from V based on a specific relation type r from R, i. e., a missing triple $(?,r,t)$ or $(h,r,?)$ where h and t are the head and tail entities. Link prediction can use embeddings to deduce new statements based on existing statements [30]. KG embedding techniques aim to represent entities and predicates as low-dimensional vectors, preserving the original KG structures and relations. Each embedding techniques keep relationship types (symmetry, antisymmetry, inversion, composition, transitivity, etc.), and they can be grouped into [31]:

1. Geometric or translational distance models use geometric operations like distance functions in the embedding space to assess the plausibility of triples (*e.g.*, TransE, TransH, TransD, RotatE).

2. Semantic matching or tensor decomposition models that compute similarity based on latent features using inner product formulations (*e.g.*, RESCAL, DistMult, Tucker).

3. Neural network-based models typically employ convolutional neural networks (CNNs) to predict triple plausibility (*e.g.*, ConvE, ConvKB) or graph neural networks (GNNs) to capture multi-hop relations in node neighborhoods (*e.g.*, RGCN, CompGCN, KBAT).

4. Rule-based models, which incorporate logical rules into the embedding learning process (*e.g.*, ComplEx-NNE-AER, IterE).

According to Hofer et al., 2023 [13], it is also essential that KG construction comprehensively addresses the representation, management, and usability of various types of metadata, such as metadata for each data source (schema and access specifications), each processing step (including inputs, configuration, outputs, log files, and reports), intermediate results, and, of course, the KG and its versions. Furthermore, for each element (entity, relationship, property value), there should be at least provenance about the origin of data (also called deep or statement-level provenance).

Deep provenance includes data creation dates, confidence scores (associated with the extraction method), or the original text paragraph from which the information was extracted. Metadata can be embedded with the data items (embedded metadata) or stored separately and referenced using unique IDs (associated metadata). The specific implementations for statement-level metadata depend on the chosen graph data structure.

2.0.3.4 KG Querying

Graph workloads can be generally categorized into two groups [35] apud [31]:

- Online graph queries encompass ad hoc graph traversals and pattern matching, exploring a limited portion of the graph and demanding lower response times.
- Offline graph analytics involve iterative and batch processing across the complete graph to accomplish tasks like PageRank calculations, clustering, community detection, and the execution of ML algorithms.

Online graph queries are supported by modern graph query languages (GQL): SPARQL for RDF graphs [36], Cypher, and Gremlin for LPG graphs [37]. GQL is subject to an initiative to design an International Standard (DIS) with features like graph updates, query across multiple graphs, and return a graph result instead of a binding table[2]. Meanwhile, ISO/IEC 9075-16 (Database languages – SQL) was released this year (2023) with Property Graph Queries (SQL/PGQ). There are other GQL in the literature and commercial Graph DBMS such as GSQL (TigerGraph) and PGQL (Oracle) [31].

Ali et al., 2022 [36] and Angles et al., 2017 [37] classified online graph queries supported by GQL into two groups: graph pattern matching, subdivided into basic graph patterns (BGP) and complex graph patterns (CGP), and graph navigation, subdivided into regular path query (RPQ) and navigational graph patterns (NGP).

A BGP query is specified as a graph (or triple) pattern and must be compared with the objective graph to find an isomorphic subgraph. Regarding relational operations, BGPs cover natural join based on vertices V and selection based on types of relationships R, properties P, and literal values L. Constants and variables can be embedded in BGP queries for Lookup operations.

Graph patterns centered on a specific entity of V are a particular case of BGP known as Star Joins since the join between the edges occurs through

[2]GQL ISO initiative, 2023 `https://www.gqlstandards.org/home`

a common vertex, origin, and/or destination of the edges [36]. Furthermore, BGP can be combined with other operators (similar to relational databases) to perform other operations such as projection, union, difference, optional (also known as left-outer-join), and filter (which complements the selection), generating CGP queries.

Angles et al., 2017 [37] define a path query as having the general form K = vi (p)$\rightarrow$ vj, where *p* specifies conditions on the paths the query has to retrieve while *vi* and *vj* denote the start and end nodes from V of the path. Specifying the conditions in the paths allows for using types of relations R or properties P called predicates. For the predicates, user can specify the occurrence of zero or more predicates $p*$, one or more predicates $p+$, the disjunction of predicates *p1/p2*, the concatenation of predicates *p1.p2* in that order, and the inversion of the edge direction $p-$.

Fixed-length path queries specified using regular expressions are commonly known as Regular Path Queries (RPQ). A variable-length path query gives rise to navigation graph patterns (NGPs). No separated group classification was found for graph queries that use key-value qualifiers in supported operations for BGPs, CGPs, RPQ, and NGP queries. Nevertheless, context knowledge can help filter, aggregate, combine, and find correlations during exploratory search.

KGs can be used by knowledge workers like (data) journalists, business and media analysts, health experts, and others in more complex operations than finding entities or looking up their properties. Such users often need to filter, compare, group, aggregate, and rank their search results to derive new knowledge, and KGs are suitable for such complex searches [11]. So graph query languages (GQL) should also support context retrieval and manipulation related to sub-graph answers.

There are more user-friendly techniques than structured queries to consume KG, such as keyword search, graph query-by-example, faceted search, visual interface navigation, natural language questions (Q&A), and interactive methods for query refinement. Deep learning techniques can also be applied to evaluate input graph query patterns and deal with KG incompleteness and query versus KG schema mismatch [31], [30]. StarQE [38] is a recently proposed method for query answering over KGs that relies on a specialized graph neural network to encode hyper-relational graphs to obtain an embedding of hyper-relational query. It was evaluated using a query dataset called WD50K-QE, generated by querying a hyper-relational subset of WD using various specialized logical query patterns.

Querying with meaning [39] covers different aspects of the search process.

It can range from understanding the query rather than simply seeking matches among its components within the data to comprehending the data itself rather than solely scanning it for these matches until presenting knowledge in a format conducive to meaningful retrieval. Semantic search can be applied to text, knowledge bases (and KG), and combined data using different search paradigms: keyword, structured query language, and natural language as questions. KGs have also been used as the background of web search engines to distinguish ambiguous entities, enabling more precise and concise answers [39].

KG exploration is an incremental analysis of its content to (i) understand the KG structure and domain, (ii) identify its usefulness to meet information needs, and (iii) retrieve relevant sub-graphs. Those sub-graphs must be linked to an information need that is inherently vague and difficult to materialize in a single query. State-of-the-art approaches for KG exploration can be classified into three areas: summarization and profiling, exploratory search, and exploratory analytics [40]. Exploratory search approaches over KG aid in retrieving specific pieces of information of interest by answering graph sub-pattern queries.

Summarization and profiling reveal a concise representation and salient features from KG structure and contents facilitating understanding that is valuable mainly in the initial exploratory stages [41]. The objective is to distill meaning from data while reducing its overall size. In some cases, summarization results in the generation of an additional KG, which can enhance the initial stages of KG exploration, making specific tasks more efficient. The summary may even construct a KG schema, similar to a reverse engineering approach, which may be absent in the original data graph. Rule mining methods also fall under the umbrella of summarization, as they employ mining techniques to uncover patterns within the data.

Statistical techniques, used in profiling, provide quantitative summaries of the contents within the data graph. These methods focus on quantifying occurrences, such as counting class instances or constructing value histograms per class, property, and value type. Other quantitative measures encompass the frequency of specific property usage, vocabularies, and the average length of string literals, among others. One primary motivation for statistical summarization (profiling) work is addressing the source selection problem. These methods offer quantitative statistics about the KG's content, assisting in determining its relevance for specific use cases.

2.0.4 Wikidata

Wikidata (WD) [14] is currently one of the most extensive publicly available Knowledge Graphs (KGs). WD is an open KG constructed collaboratively as a multilingual structured repository of items. A fundamental principle in crowdsourced KGs, like WD, is to allow for different perspectives [23]. Thus, WD refers to statements as "*claims*". These claims may not necessarily represent a single, universally agreed-upon perspective on the world, a ground truth.

In WD parlance, items are a thing, an entity, a concept, or whatever can be described through its properties, like labels, aliases, identifiers, etc. WD items are Entities and Properties. An Item ID (aka QNode) uniquely identifies Entities and Property ID (aka PNode), properties (data properties and object properties in Ontologies).

To allow a more precise terminology and avoid ambiguity, this thesis refers to:

1. the role of a property in a (main value) statement as Predicate,
2. a pair <property, value> associated with a statement as a Qualification,
3. the role of the property in a qualification as Qualifier; and
4. a property appearing in a <property value> pair within a reference associated with a statement as Referrer.

Figure 2.2 represents a claim composed of a statement (inside the green rectangle) with qualifications and references about entity Brazil (Q155).

Each statement has at least a property and its value (a value can be another Entity or a literal). WD overcomes the limitations of triple-based knowledge representation by employing reification and qualifiers to enhance subject-predicate-object triples. Qualifiers add context to statements, including information about sources, dates, reasons, and more [23], as Entity or literal. Statements are composed of qualifiers and their values, references, and rank (normal, preferred, and deprecated) to compose a claim. However, WD qualifiers are not explicitly identified as context or additive information for n-ary relations [42]

Given an item, WD allows it to be the subject of multiple statements with different values for the same Predicate, *i.e.*, expressing apparent conflict or different perspectives about a subject even if they contradict one another. While this may be appropriate in some cases (*e.g.*, spouse, head of government,

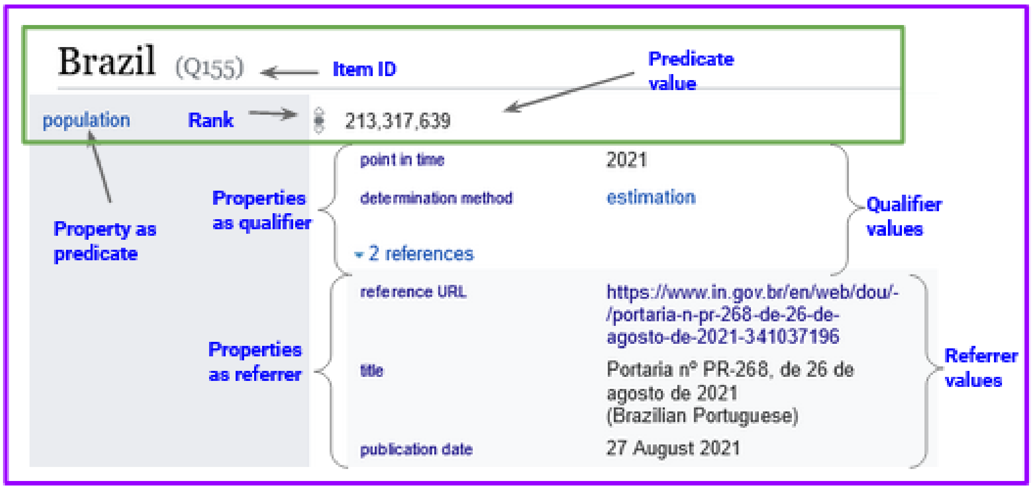


Figure 2.2: Brazil (Q155) population in WD

etc.), it is not for others, such as inverse functional properties (*e.g.*, mother, date of birth, etc.), configured as single-value and single-best-value constraints.

By design, WD contains statements based on claims about items that represent references for facts (instead of facts itself). WD contains inconsistent and contradictory claims representing the diversity of knowledge about a given entity. In principle, there is no guarantee of the truthfulness of claims. Both References and Qualifications can contain provenance information.

The values for specific predicates are sometimes time-dependent (*e.g.*, position held, capital), location-dependent (*e.g.*, boiling point), or, more generally, context-dependent. Qualifiers are used in WD to represent such dependencies[3], which provide contextual information for statements. Constraints specifications over these properties characterize the intended semantics for property and formulate restrictions on using these properties within statements in WD.

WD data are frequently available through some dump files. Truthy dump excludes any qualification so there is no contextual information about statements. They also do not deliver statements marked with a deprecated rank. In cases where there is more than one statement for the same predicate P with both normal and preferred rank statements, only the ones with a preferred rank are included in the dump without its qualifiers. However, if there is no preferred statement for property P, then all statements with a normal rank for P are regarded as truthy.

Websites are regularly using WD content for different purposes: services (search engines, personal assistants, libraries, and museums [43]), applications (Daimler, Lufthansa, Novartis, data journalists, . . .), and research projects

[3]Wikidata Foundation, Accessed in 2023 `https://www.wikidata.org/wiki/Help:Qualifiers`

[44]. WD is considered by many as part of the Semantic Web ecosystem (*e.g.*, [45]).

In some applications using KGs [46], one needs to obtain the value of some property of an entity (item in WD parlance) to make some computations. Examples of such situations are finding the capital city of a country or province, obtaining a person's birth or death date, determining the author of some artifact, and obtaining some physical property of a known substance.

In this chapter, the thesis presented important concepts such as KG data structures, KG query languages, KG engineering phases. Exploratory search and Decision-makings were also covered and the importance of Trust were highlighted. WD was detailed since this important open source KG was used in the Proof of Concept (PoC). Next chapter will present the research problem and proposal overview. The two following chapters will cover the proposed features of the Knowledge Layer making use of these background concepts.

3
Research Problem

This chapter starts by discussing issues related to trust in systems, based on information retrieved from KGs, to formulate and justify the problem statement. Next, the Research contributions are presented, considering the research objectives and premises described.

3.0.1
Problem Scope

Trust judgments are made by humans based on their prior knowledge about the domain, the source's apparent credibility, commonsense, and even past personal experience and beliefs. Such judgments allow people to act under uncertainty and deal with the risk of negative consequences. However, it is worth modeling trust in systems that support tasks with some level of risk and possibility of deception [47], such as decision-making, question-answering (Q&A), information retrieval (IR), fact-checking, and recommendation systems (RS).

Even though the trust process applies to statements that assert a single value for a property of an item, it is more crucial in situations where there are multiple statements about a property of an item, each with a different value, and a single value is expected to perform some computation. Suppose a person has a biological mother and a foster (legally defined) mother. The relevant information to evaluate health issues is about the biological mother. Like Garibaldi's example, the relevant information to assess inheritance issues is about the legally defined mother. In such cases, the user is further confronted with the additional decision of which value to choose beyond applying the trust process to each asserted value.

Crowdsourced KGs such as WD, or KGs constructed by integrating several different information sources of varying quality, must be used via a Trust Layer [6], [7]. Trust is usually evaluated using policies (by hard evidence) or reputation (by estimation). These approaches should also consider the context of the information and the circumstances and associations of the goal of the trust decision[47]. Furthermore, to provide all the information this layer requires, the knowledge retrieved from the KG should be explicitly contextualized, at least in terms of Provenance, Temporal, and Location dimensions. The Trust Layer transforms information into trusted information by applying task and user-dependent trust policies. Given the goal of carrying out some action, decision-making occurs above the Trust Layer, after evaluating the veracity of each relevant claim retrieved from the underlying KG.

After analyzing the current status of WD [48], it was confirmed that given the characteristics of WD data structure (such as the lack of references for qualifiers, incorrect specification of property scope in addition to the absence and ambiguities of definitions in constraints) and in the data (violations of constraints that generate incompleteness and possible interpretation errors) it is necessary to have an additional and separate layer.

3.0.2 Problem Statement

This thesis proposes the following problem statements, considering the scope previously discussed.

Given that contextualized information supports trust in knowledge, how can we effectively model and retrieve contextual information to be employed to implement trust criteria?
(General Problem Statement)

$\Downarrow$

Given that contextualized information improves the support to KG usage (for decision making), how to effectively model and retrieve contextual information to be employed to implement trust criteria?
(Instance Problem Statement)

The hypothesis raised is that context-aware exploratory search for decision-making over KGs requires explicit representation and semantic interpretation of contextual dimensions of KG claims. Trust policies take Contextualized knowledge as input and generate Trusted Knowledge that satisfies the task and agent criteria. Enabling explicit context representation, retrieval, and interpretation enhances the support for decision-making since context is essential for identifying the scope within which a claim can be considered true and useful.

3.0.3 Expected Contributions

Based on the problem statement, this research expects to produce the following contributions to address it. All contributions have the same level of importance.

Context Augmented KG-based Decision Support Systems Recapping some points from the literature:

1. the main differences between KB and KG stated by Chaudhri et al., 2022 [12],
2. the limitations of KGs to be used in DSS [49]: "*Not trustworthy*", "*Usually Incorrect*" and "*Incomplete*", and
3. the trust definition as "*knowledge-based reliance on received information*".

This thesis proposes an architecture of Context Augmented Knowledge Graph-based Decision Support Systems (CoaKG4DSS) that incorporates a Trust Layer positioned between Knowledge and Decision Layers, as shown in Figure 3.1.

The Trust Layer services should provide a feasible method to verify that a claim is valid through trust policies and rules, and these services rely on Knowledge Layer services to retrieve as complete contextualized information as possible. Using this input, reasoners from the Trust Layer can assess which of the many pieces of information available, sometimes contradicting one another, are more helpful in answering a question.

With some generalizations, a Trust Layer has already been proposed regarding statements in WD and, more generally, about the Semantic Web [50]. Polyvocal KGs [51], which incorporate multiple voices (or points of view) about a subject, have also recognized a demand for a Trust Layer.

The Knowledge Layer is composed of a Context Augmented Query Engine that uses a Graph Query Language (GQL) to query the underlying KG to retrieve both trust policies and Contextualized Knowledge (*i.e.*, claims and metadata), Codomain Algebra functions associated with each type of Context, and an Inference Engine to execute Domain and Contextual Rules. Mappings are added to the standard KG to identify contextual information among the claims and metadata.

Dual Open World Assumption (DOWA) The architecture assumes that the Knowledge Layer operates under DOWA, defined as:

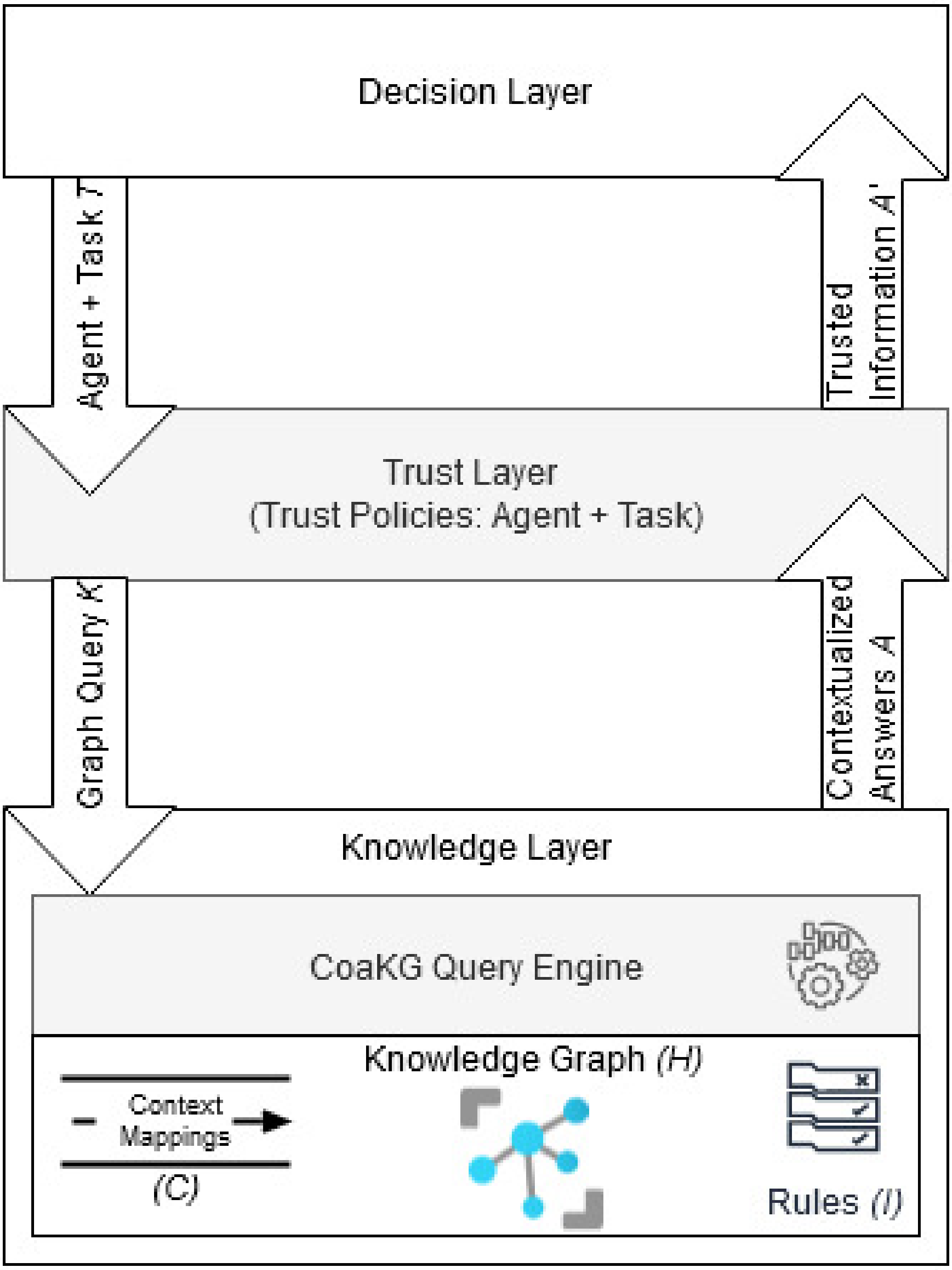


Figure 3.1: Context Augmented Knowledge Graph-based Decision Support Systems

Definition 3.1

DOWA is a variant of the traditional Open World Assumption (OWA). Under DOWA, the presence of a claim in a KG does not automatically imply that it is a fact (true proposition). Instead, truthfulness evaluation depends upon the contexts of claims and tasks being carried out or intended (purpose). It is also subjective to the user who decides to act based on contextualized information.

Such a definition, proposed in this thesis, was constructed based upon, mainly, the Pragmatic Theory of Truth [4]. This theory of truth focuses on the practical consequences or utility when a truth-bearer is considered valid. According to this view, a statement (claim, proposition, utterance, etc) is true if it works or helps achieve desired goals or outcomes. Two commonly accepted constraints on truth and falsehood apply to truth-bearers: (i) Every proposition is true OR false, and (ii) No proposition is both true AND false.

The existing literature contains state-of-the-art data fusion methods [27] that aim at conflict resolution to establish a single truth. They are applicable in Web data integration approaches that operate under OWA or CWA [52]. Under DOWA, this architecture does not resolve potential conflicts in the Knowledge Layer, adopting a Conflict Ignorance approach when they are present. The Knowledge Layer answers provide the contextualized data to the Trust Layer, which is responsible for applying the trust rules and policies of the agent for the task at hand. This layer, in turn, supplies a Decision Layer with trusted information to use its methods and criteria to decide about an action being contemplated.

Context Augmentation of the Knowledge Layer The following contribution is a framework for the Knowledge Layer that retrieves all contextualized answers from CoaKGs to support the Trust Layer, thus enhancing decision-making trustworthiness. Our proposal comprises a Conceptual KG Schema and a Query-Answering approach and is designed considering two definitions - Context Augmented KG (CoaKG) and All Contextualized Answers.

An Context Augmented KG (CoaKG) definition and schema characterize the necessary elements for modeling explicit contextual information and extracting implicit contextual information from a standard KG. A common way to contextualize claims (statements) is by adding property-value pairs as qualifiers. We must distinguish between additive qualifiers, which represent n-ary relationships and do not affect the assessment of the claim's truthfulness, and contextual qualifiers, which can restrict the contexts in which the underlying claim is considered true and may modify the claim itself [42].

Since explicit contextual information is part of the KG, context mappings specify how it is represented as qualifiers and predicates among the claims and metadata. According to the proposed Conceptual Schema, context mappings added to a standard KG transform it into a CoaKG. CoaKGs are repositories of truth-bearers (contextualized claims, propositions, hypotheses, etc.). The Trust Layer evaluates truth-bearers to transform them into trusted knowledge using trust policies and rules of the decision-maker and the task.

As part of this contribution, the thesis proposes a query-answering approach that retrieves claims, as contextualized as possible, from a CoaKG within an exploratory search process for decision-making. Anticipating potential additional queries about the answers and retrieving contextual information already available enables an exploratory search with fewer interactions. Graph queries generated during exploration by the Trust Layer/User Interface can be complete or incomplete regarding context information. The degree of incompleteness can be assessed by executing graph queries against the context mappings in the CoaKG. Besides, the CoaKG can be contextually complete or incomplete regarding the instances. The original graph query K is expanded to retrieve the Answer Set A. Rules can be used to extract implicit context.

Given a graph query K, generated to partially represent a task T, as input and a contextualized answer A as output of the Knowledge Layer, the Trust Layer will infer the trustworthiness level of each contextualized claim Si to support the user to select which best answers K. A subset of A can be considered satisfactory/good enough by the Trust Layer after applying trust policies defined by the user and related to the task T. An ordered set A' can also be generated by the Trust Layer after applying ranking functions related to the task T. In the Decision Layer, based on decision criteria, one or more claims will be considered true to decide the action to be taken.

3.0.4 Research Premises

The thesis adopts the Dual Open World Assumption (DOWA) considering the Pragmatic theory of Truth. The truth of a statement is context-dependent and agent-dependent. This decision justifies that statements in a KG are referenced as claims instead of facts (true statements).

Any open KG constructed collaboratively by contributors with different levels of domain knowledge and world views can suffer from wrong, biased, outdated, incomplete, and inconsistent content. Besides, KG construction relies on information sources and knowledge extraction processes where it is impossible to guarantee that every claim represents a fact.

Additionally, KGs can incorporate claims that represent different perspectives that are apparently contradictory or deny each other. So, the decision-makers must rely on their trust policies to decide what is true and useful according to the context of their information needs and tasks. Moreover, the knowledge retrieved from a KG must be more contextualized to support the trust process.

The thesis also assumes that both the Query and the KG can be incomplete, mainly regarding contextual information. User questions and queries can be incomplete, inaccurate, and ambiguous, often because crucial details, such as the context, are implicit or because users may not fully comprehend their underlying information needs within a given domain. KG incompleteness is addressed through semantic rules and Codomain Algebra, inferring implicit context. Codomain, in mathematical and ontological terms, typically refers to the set of all possible values that a function or relation can output. Similarly, the Codomain is the set of values that could correspond to Context Values, which is part of the contextual information. Context Values of Entities and Claims from the same Context Type can be implicitly related, and Codomain Algebra operators associated with the Context can be used to reveal such relationships. Contextualized knowledge also indicates missing contextual information for the Trust Layer.

These conditions highlight the importance of considering explicitly contextualized information to address these challenges.

3.0.5 What is out of scope?

Search interfaces usually offer various mechanisms, from traditional keyword search to more sophisticated faceted and semantic search capabilities. Natural language translation to structured graph queries using a proper graph query language (GQL) enables users to seamlessly interact with a KG by posing queries in their native language. This thesis starts with graph queries as input, assuming the query interface performed this translation.

Another aspect outside this thesis is search results visualization. Visualization is a crucial aspect related to information consumption and interpretation. The thesis focuses on supporting users with contextualized information, and it is highly recommended that information and its context be presented in a readily available and user-friendly manner.

KGs enriched with contextual information raise concerns about efficient query performance. It is not the scope of this research to develop techniques and optimizations to ensure a responsive user experience. Given the proposal's effectiveness, the efficiency issue can be better studied in future works.

The Trust Layer, with trust rules and policies established according to users and task requirements, is essential in ensuring the accuracy and reliability of the information provided by KGs. Nevertheless, its design and implementation are also outside of the scope.

3.0.6
What are the research goals?

The goal of this research is to design and implement a framework for the Knowledge Layer that is able to:

1. Answer potentially incomplete graph queries concerning context information.

2. Support the Trust Layer with explicitly represented and inferred contextual information retrieved from potentially incomplete KGs.

3. Enrich any standard KG to become a Context Augmented KG (CoaKG) through the addition of context mappings based on a proposed Conceptual Schema.

In this chapter, a general problem and a research problem about KG usage in Decision-making scenarios were enunciated and justified. Also, the main contributions of this research were summarized. The next chapter will detail the Context Augmented KG (CoaKG) definition and schema.

4
Contextual Schema for Knowledge Graphs

Context-aware exploratory search over KGs requires explicit representation and interpretation of contextual information to help users evaluate whether retrieved KG claims are true and useful for their decision-making tasks. A Context Augmented KG (CoaKG) definition and Conceptual schema proposed by this thesis are described in this chapter.

4.0.1
Context Augmented KG (CoaKG)

Definition 4.1

CoaKG H' $= <$ H, C, I $>$

A multi-layer KG H with entities, claims, and context information; the context mappings C between KG elements and context information; and interpretations I as rules to extract implicit context. A Contextual KG extends a standard KG with resources that enable it to explicitly represent contextual information and infer implicit context.

Each context type Ci contains a set of relations from R, properties from P, and qualifiers from Q associated with entity types, concepts, and claims. Context types represent perspectives from which entities, concepts, and claims can be filtered, aggregated, interpreted, analyzed, compared, and correlated. Making the context explicit allows for the interpretation of information from different perspectives.

An interpretation through I is a head-if-body semantic rule composed of a header hj and a body bj. Each semantic rule establishes a possible interpretation of a subgraph pattern according to the associated context that allows additional contextualized information to be derived that is not explicitly represented in CoaKG.. Rules help to complete KG semantics, and it consists of if-then declarations (bj $\rightarrow$ hj) where if some condition holds (body), then some entailment holds (head). Both head and body are expressed as graph patterns [53].

Similarly to Contextualized Ontologies [54], in CoaKGs, mappings establish the role of each object, *i.e.*, whether it functions as an entity in the standard KG or as context information. The context specification C works as a layer meta schema over the KG containing meta-information.

Context representation and rules enable reasoning with the knowledge and its associated contextual information. This requirement introduces a new concern in knowledge engineering. Besides entities, concepts, properties, and their relations, engineers must also identify the set of context types and how to interpret knowledge considering these contexts. However, it is possible that the representation of a claim remains incomplete and needs to be augmented to be fully contextualized.

All entities, claims, contexts, and mappings are components of the CoaKG itself. The KG engineer plays a crucial role in identifying context information. When a KG schema is absent, KG profiling should be employed to extract latent structures from the KG instances. For each relationship type, property, or qualifier, one should identify if it belongs to any context C, specifying the corresponding mappings and adding them to the KG.

Geopolitical History of Brazil CoaKG (H4) A real-world use case was created for this thesis to illustrate the benefits of our approach. CoaKG, identified as H4, focusing on the Brazilian Geopolitical History domain, was built from scratch[1]. Geography-based political events and dynamics studies help to illustrate the complexities of internal and external relations among and within countries. These relations impact organizations and people's lives. Spatial conflicts, territorial unit demography, and economic and cultural aspects can be explained based on geopolitical factors and history.

This KG was built using data extracted from various websites that provide educational content and entrance exam (ENEM) preparation materials for elementary and high school students. Figure 4.1 shows a partial view. Rectangles represent concepts and entities from V, ellipses represent literal values from L, diamonds represent qualified edges from E, pV, qE, and qP, and lines represent non-qualified edges from the same sets. It has two disconnected components but shares implicit relationships associated with the temporal context that will be revealed at query time. The CoaKG H4 was built from scratch, considering all context types necessary to represent claims and entities.

[1] `https://github.com/versant2612/CKG_UseCases/blob/53bb930d2a86d4a74f36cbb77c5c6c2bd7088aad/H4`

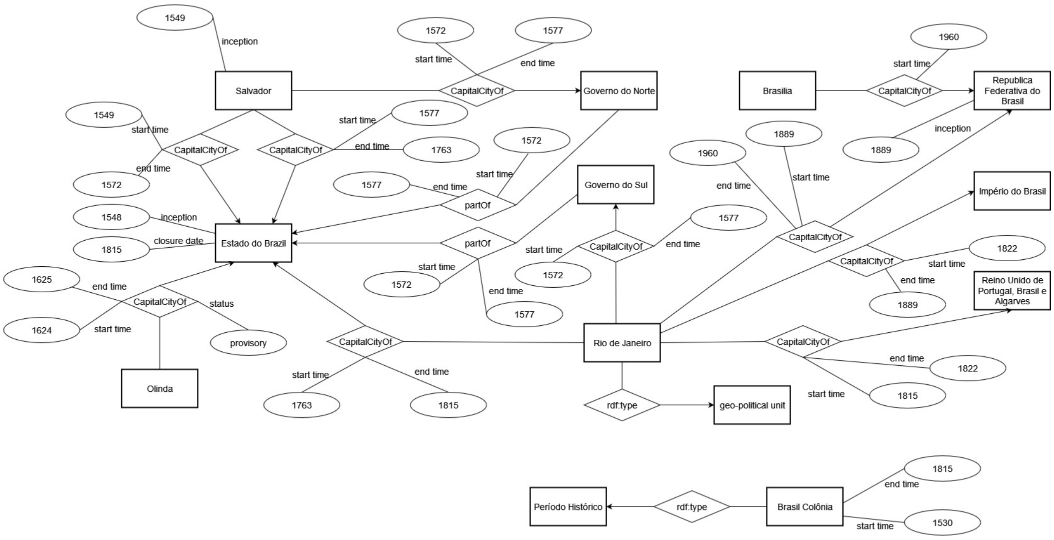


Figure 4.1: Fragment of CoaKG H4 - Brazilian Geopolitical History

The Context Augmented KG (CoaKG) schema that defines the necessary elements for modeling contextual information is presented in Figure 4.2. Such schema represents the explicit and implicit contextual information that can be retrieved from KGs.

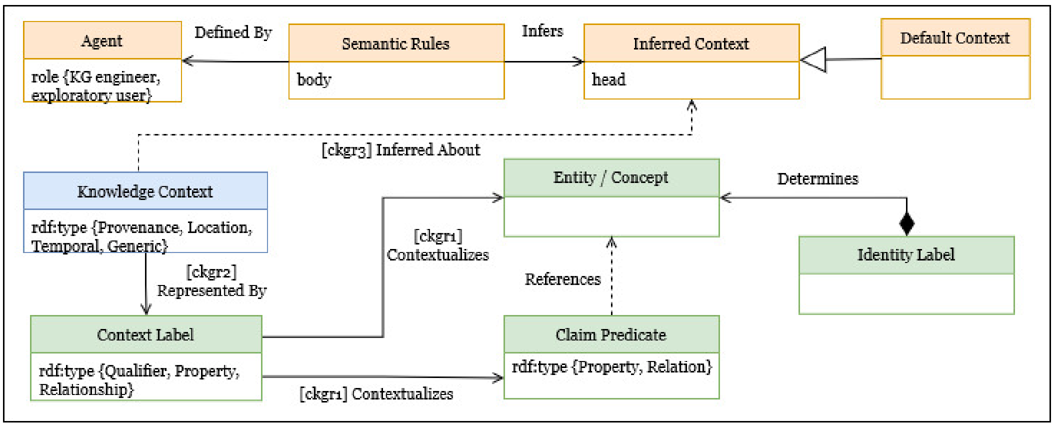


Figure 4.2: CoaKG Schema for Context Mappings and Interpretation Rules

4.0.2
Explicitly represented context

Explicit context, represented as green entities, is associated with Context Types that initially belong to the three types most found in the literature and are common to several application domains. New domain-specific context types can be added to the schema.

- **Temporal**: corresponds to dates, periods, events, and so on. This meta-information associated with a claim, an entity, and a concept enables time-related interpretations. It is possible to differentiate a claim, an entity, or a concept from the others according to the period that it is valid. For example, Start Date, End Date, Occurrence Date, Point in Time, etc. With this context, WHEN-type questions that may arise during the exploratory search can be answered. Besides, it can also help reveal temporal relations among claims, entities, or concepts not explicitly represented in the KG.

- **Location**: corresponds to places, coordinates, regions, spatial shapes, etc. This meta-information associated with a claim, an entity, and a concept enables spatial-related interpretations. Using location context values, it is possible to differentiate a claim, an entity, or a concept from others about the place where it occurred or exists. For example, Geographic Coordinates, Altitude, Country, University, etc. The place can be an entity that represents an object in space (e.g., PUC-Rio Campus Gávea) or a class of places (e.g., university). With this dimension, it is possible to answer WHERE-type questions that may arise during the exploratory search. Besides, it can also help reveal spatial relations among claims, entities, or concepts not explicitly represented in the KG.
- **Provenance**: corresponds to the information source, agents, processes, methods, and so on. This meta-information associated with a claim, an entity, and a concept enables lineage-related interpretations. Using provenance context values, it is possible to differentiate a claim, an entity, or a concept from others regarding its source, origin, or how it was produced, obtained, and calculated. For example, the dataset, system, the person who stated/attested, the process that generated it, etc. Besides identifying the source, the user may wish to obtain more information about this source, exploring aspects that may contribute to assessing the veracity and usefulness of the answers. With this contextual information, it is possible to answer questions such as WHO stated, WHEN stated, WHERE stated, HOW the claim was generated, etc., that may arise during the exploratory search.

Identity labels are also helpful as contextual meta-information in cases of ambiguity among entities involved in claims. It allows the user to identify each entity involved in the claim. For example, the entity type, the entity's name, a unique identifier, any Inverse Functional Property of an entity, and properties that present a human-readable version. This contextual information makes it possible to answer WHAT-type or WHO-type questions that may arise during the exploratory search to clarify ambiguity.

The data structure of a standard KG has no explicit notion of knowledge context as this thesis is proposing. A "*native*" CoaKG can be constructed directly as an instance of our CoaKG schema, mapping the KG schema/ontology properties to its context types. KG built using shared Ontologies allow interoperability and are precise regarding the semantics of the predicates and classes of their instances.

Ontologies are formal representations that can be used to interpret and deduce new information. A set of Upper Ontologies, data, object properties, and classes of domain Ontologies associated with contextual types can help the KG engineer model a Contextualized KG. To assist KG engineers, some useful vocabularies (complete Ontologies and Standards or Properties and Classes from ontologies) that can be used to model context dimensions (see Table 4.1) were identified.

Context	Vocabulary Elements
Location	GeoSPARQL Ontology location - Schema.org Property; geo - Schema.org Property; Place - Schema.org Type
Temporal	Time Ontology in OWL ISO 8601 standard duration - Schema.org Property; startDate - Schema.org Property; temporalCoverage - Schema.org Property
Provenance	PROV-O: The PROV Ontology Open Provenance Model Vocabulary Specification dc:contributor; dc:creator, dc:publisher
Identity	alternateName - Schema.org Property; description - Schema.org Property; identifier - Schema.org Property; name - Schema.org Property; rdfs:label; rdfs:comment;

Table 4.1: Contextual Vocabularies

Context Mapping for H4 The schema of KG H4 is described in Table 4.2. For the entire dataset, see CKG-H4.tsv at GitHub [2]. Using the KG schema, the KG Engineering can identify each context that is represented in the KG and build the mappings according to the CoaKG schema 4.2. These mappings are specified as instances of the blue and green entities in the CoaKG conceptual schema (Figure 4.2). Three context types were identified for geo-political unit entity type (h4c1) associated with four properties and one relationship, as

[2] `https://github.com/versant2612/CKG_UseCases/blob/53bb930d2a86d4a74f36cbb77c5c6c2bd7088aad/H4`

shown in figure 4.3. Table 4.3 shows a sample of Context Mappings of H4. The complete mappings can be found in Appendix B.

Entity Types (sub set from V)	*V'* = {geo-political unity (h4c1), Período Histórico (h4c2)}
Relationships	*R* = {form of government (h4r1), geoshape (h4r2), partOf (h4r3), colonyOf (h4r4), createdBy (h4r5), capitalCityOf (h4r7), proclaim (h4r8), headOfState of (h4r9)}
Properties	*P* = {arrival place (h4p1), inception (h4p3), closure date (h4p4), first name (h4p5), reference URL (h4p6), alias (h4p7), image (h4p8)}
Qualifiers	*Q* = {start time (h4q1), end time (h4q2), stated in (h4q3), point in time (h4q4), headed by (h4q5), local (h4q6)}

Table 4.2: KG Schema of H4

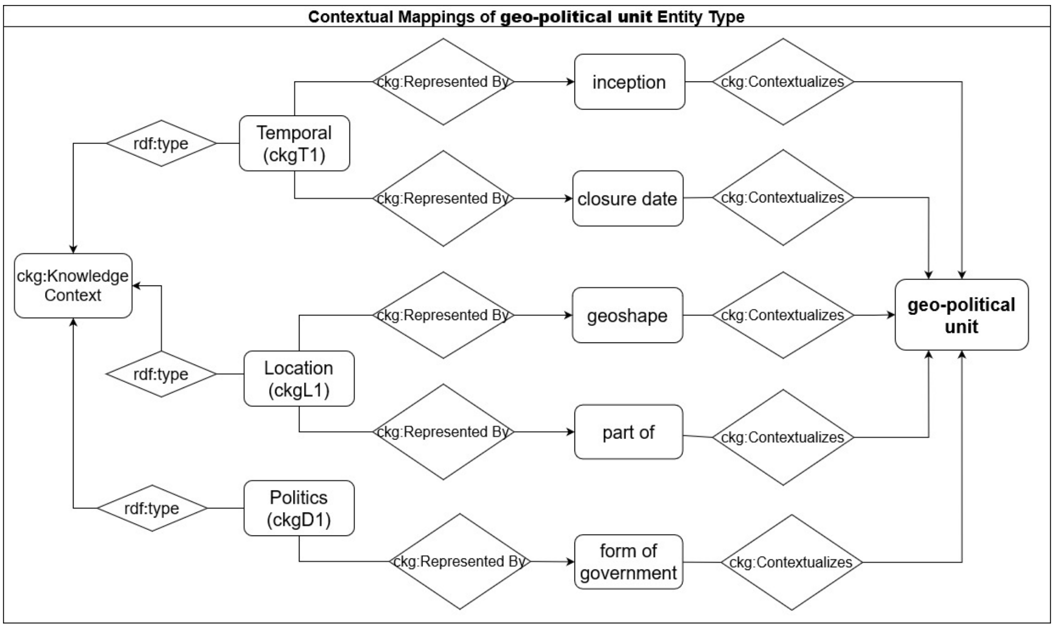


Figure 4.3: Contextual Mappings of geo-political unit Entity Type

4.0.3 Implicit inferred context

Semantic Rules can be used to infer implicit context. These rules can be defined manually by knowledge engineers or suggested by the users who explore the KG. Rules are an implication between an antecedent (or IF clause or BODY rule) and a consequent (or THEN or HEAD rule). Semantic Rules consist of a set of claims, its explicit context, and operators ($\wedge$, v, $\nexists$) as antecedents that infer a relative context as consequent. See Table 4.4 for some examples of Semantic rules utterances.

NODE1	EDGE	NODE2
Provenance (ckgP1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Provenance (ckgP1)	ckg:Represented By (ckgr1)	stated in (h4q3)
stated in (h4q3)	ckg:Contextualizes (ckgr2)	form of government (h4r1)
stated in (h4q3)	ckg:Contextualizes (ckgr2)	capital city of (h4r7)
Politics (ckgD1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Politics (ckgD1)	ckg:Represented By (ckgr1)	form of government (h4r1)
form of government (h4r1)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Location (ckgL1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Location (ckgL1)	ckg:Represented By (ckgr1)	geoshape (h4r2)
geoshape (h4r2)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Temporal (ckgT1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Temporal (ckgT1)	ckg:Inferred Context (ckgr3)	Current (ckgl1)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	start time (h4q1)
start time (h4q1)	ckg:Contextualizes (ckgr2)	capital city of (h4r7)
start time (h4q1)	ckg:Contextualizes (ckgr2)	Período Histórico (h4c2)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	inception (h4p3)
inception (h4p3)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)

Table 4.3: Contextual Mappings of H4

The default context is a special kind of context. It can be used, for example, to specify the context of interest of an application, or it can be specified by KG Engineering based on KG patterns as the most frequent claims' context or according to the query log analysis to identify the most frequent query's context. The user can also define the default context according to their preferences or the context of the task that motivated the search.

Semantic Rules for H4 Table 4.5 shows two examples of relative context from the Temporal dimension. Rule *I1* indicates that every claim that has a start date (h4q1) qualifier but does not have an end date (h4q2) qualifier is related to the Current time in temporal context. Rule *I2* indicates that the geoshape claim corresponding to the maximum point in time (h4q4) qualifier, among all geoshape claims associated with the same origin node that has a point in time (h4q4), is related to the Current time in temporal context.

4.0.4 CoaKG Engineering

A CoaKG can be built from scratch when contextual modeling is a concern of the KG purpose as H4. However, our CoaKG schema can also be superimposed

Location Context	
IF	Water Boils At 100℃ **At the Altitude of** 0 meter
THEN	Water Boils At 100℃ At the Sea Level
IF	Water Boils At 100℃ **At the Altitude of** 0 meter AND Water Boils At 100℃ **At** Rio de Janeiro
THEN	Rio de Janeiro is located At the Sea Level
IF	Water Boils At 100℃ **At the Sea Level**
THEN	Water Boils At 100℃ is **DEFAULT context**
Provenance and Location Context	
IF	**According to his birth certificate**, Barack Obama was **born at** the Kapiolani Maternity and Gynecological Hospital **is located in** Honolulu AND Honolulu **is located in** Hawaii AND Hawaii **is part of** the USA AND Anyone **born in** the USA is an American citizen.
THEN	Barack Obama is an American citizen.
Provenance, Location, and Temporal Context	
IF	**According to his identity card**, João José da Silva **was born on** December 1, 2004 AND **According to his birth certificate**, João José da Silva **was born on** December 1, 2005 AND The age of criminal responsibility **in** Brazil corresponds to 18 years of age AND The age for criminal responsibility is calculated based on **Birth Certificate** AND The crime **occurred on** November 1, 2023
THEN	João José da Silva is under age **on the date of the crime**.

Table 4.4: Context Rule examples

ID	Body / IF / antecedent	Head / THEN / consequent
I1	(?v1)-[?p1]→(?v2). (?p1)-[:start time]→(?v3). WHERE NOT EXISTS {(?p1)-[:end time]→(?v4).}	(?p1)-[:Inferred Context]→(Current)
I2	(?v1)-[?p1:geoshape]→(). (?p1)-[:point in time]→(?v2). WHERE ?v2 = { max(?v3) FROM (?v1)→[?p2:geoshape]-(). (?p2)-[:point in time]→(?v3). }	(?p1)-[:Inferred Context]→(Current)

Table 4.5: Temporal Context Interpretation of Geopolitical Claims

on a given KG (thus generating a corresponding CoaKG), and a fundamental point, therefore, is to identify existing contextual predicates and types as well as semantic rules to complete missing contextual information. Figure 4.4 shows the CoaKG Engineering phases that must be executed to build a CoaKG.

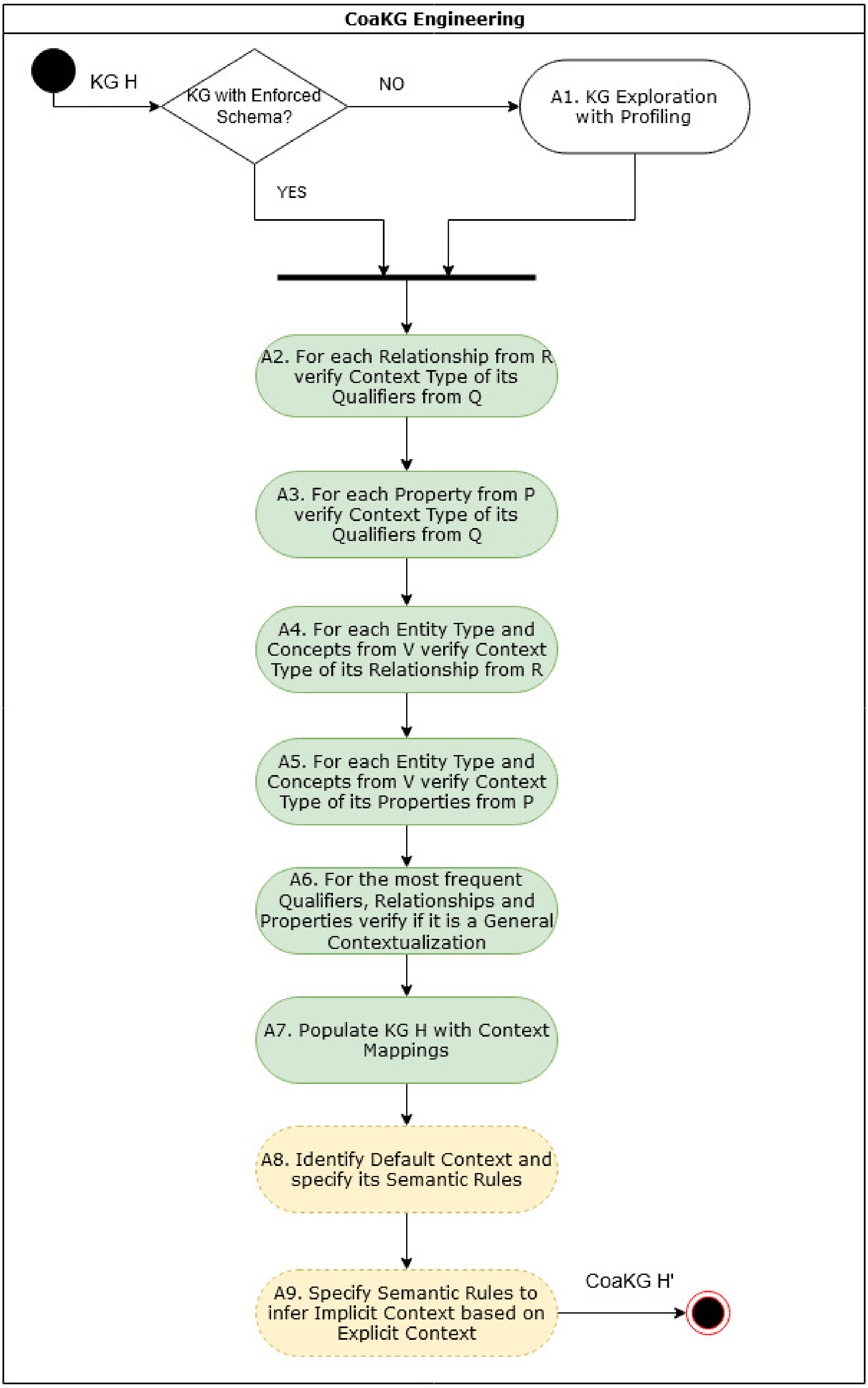


Figure 4.4: CoaKG Engineering steps

A standard KG without a formal schema should first be submitted to KG Summarization & Profiling [40] to identify its latent schema. The process requires no specific domain knowledge, uses general graph algorithms, and returns a high-level overview of KG content. This result will assist knowledge engineers in evaluating context dimensions present in the KG and help formulate semantic rules interpretations for qualifiers' existence, absence, and possible claims contradictions.

Context types and elements identification are a challenge since any information (Qualifications, References, and Rank) associated with an element (Entity, Concept, or Claim) can, in principle, be considered context. Knowledge engineers should consider each component's relevance to context-aware applications' purpose [71]. However, noticing that an existent KG instance may not contain all the context information that allows a more informative response (degenerate CoaKG) is essential.

For each relation R, property P, and their qualifiers Q from the KG schema (extracted through Profiling or modeled with), the knowledge engineer would identify which context type it belongs to. Furthermore, contexts that apply to all elements can also be identified. Steps A2 upto A6, according to Figure 4.4, detail this activities. Knowledge engineers, at step A7, should build on top of an existing KG a set of mappings and rules that make contextual dimensions explicit about transforming a standard KG into a Contextualized KG (CoaKG). The links between the KG schema, the KG instance data, and the CoaKG schema represent the Contextual Mappings C and are part of the final CoaKG. Contextual Mappings allow knowledge-context retrieval and application manipulation in a loosely coupled way. Finally, the KG engineer can identify rules for default context. For example, using more frequent context values in the graph. As well as rules for inferring implicit context according to the context values already represented. For example, rules for interpreting values absences as in the case of `CapitalCityOf` relationship without end date (h4q2) qualifier corresponding to the current time in the temporal context.

This chapter presented the details of the Context Augmented KG (CoaKG) definition and conceptual schema exemplified with a KG about the Geopolitical History of Brazil. Additionally, ontologies about Temporal, Location, and Provenance contexts were identified to help KG Engineerings build a CoaKG from scratch. In the next chapter, this thesis describes the query-answering approach that retrieves All Contextualized Answers. from a CoaKG within an exploratory search process, exemplified using the CoaKG presented in this chapter.

5
Contextualized Answers for Exploratory Search

The query-answering approach over a CoaKG within an exploratory search process is described in this Chapter. Before that, a definition of All Contextualized Answers is presented and justified.

5.0.1
All Contextualized Answers

Definition 5.1

$A = \{S1, S2, \cdot \quad \cdot \quad \cdot, Si\} \simeq K$

A is composed of a set of zero, one or more fully contextualized claims S that potentially meet the user's information need. All Contextualized Possible Answers result from a graph query K over a CoaKG H, considering gaps in the user's knowledge and KG and query incompleteness. The CoaKG Query Engine uses context mappings C to complete the graph query K and interpretation rules I to complete the answers concerning implicit context.

Query K is expressed as graph pattern (BGP or CGP) matching where at least one element must be specified as a constant. Query context is optional; it can be implicit, and the default context associated with the KG is applied or can be explicitly specified in the query as it derives from the information needed.

User queries are typically incomplete, inaccurate, and even ambiguous, often because crucial information, such as the context, is implicit or because users need higher comprehension of their underlying information needs within a given domain. Keyword-based search engines often assume that users are primarily interested in current or local information, implicitly adopting a specific default (temporal or geographical) context, which neglects so-called long tail scenarios. Based on such observations, this thesis highlights the importance of considering contextual information in exploratory search approaches to address such challenges.

The All Contextualized Possible Answer implies retrieving contextualized claims, especially when there would be conflicts and ambiguity between the claims, allowing the user to analyze and decide which one is considered true. All Contextualized Possible Answer is a complete (as possible) answer over integrated and contextualized data from KG [27]. This thesis considers the All Contextualized Possible Answer as a KG query expansion approach. When query context can be unknown, incomplete, or implicit, the results of a query must turn contextual information explicit. In this way, the user can detect if the system has assumed a different context from the one the user was expecting.

All Contextualized Possible Answers can also restrict the claims retrieved if the user is specific about contextual dimensions of interest. For example, suppose the user is only interested in recent claims (temporal context), claims associated with proximity to their location or a specific region (location context), claims from sources they consider reliable (context of provenance), claims about a particular subject (thematic context), or, claims with accuracy above a specific value (uncertainty context), etc. Codomain Algebra can be applied to filter the results based on Contextual relations. If no claims are retrieved as exact answers due to missing context, another version of the expanded query is set up to retrieve approximate answers.

If they exist, semantic rules must be applied to identify how to interpret the context from retrieved claims and finally infer implicit context to generate answer A. If no context was specified in the query, a default context can be applied and used in query processing to indicate the most probable answer.

5.0.2
CoaKG Query Engine

Figure 5.1 shows the steps the CoaKG Query Engine takes to transform a K graph query, potentially contextually incomplete, into a contextually complete query. The query is considered contextually complete since all mappings created by the KG Engineer that reference elements mentioned in the original query are used to retrieve all context information that is represented in the KG.

Furthermore, the query can also be modified to add hidden relationships using specific Codomain Algebra operations for the value of each context (e.g., Dates, Geometries, Integers, and so on) to infer relationships not directly materialized in the KG, such as claims co-occurrence in time or entities overlapping in space or the ordering of information sources based on ranking.

The CoaKG query engine generates two versions of the expanded K: Exact and Approximated, to be executed in steps B5 and B6. The Approximated queries deal with KG incompleteness and optional (not mandatory) context. Three types of modification can be applied: (i) OPTIONAL clause for context qualifiers, properties, and relations, (ii) change constant by variables for context values, and (iii) test context values for NULL to substitute by DEFAULT VALUES in WHERE clause.

After executing the queries, it is possible to apply semantic rules defined by the KG engineer or by the user performing the exploration. To produce all possible contextualized responses, the rules are executed to infer implicit context among the set of responses. These steps will be explained using a query example with CoaKG H4.

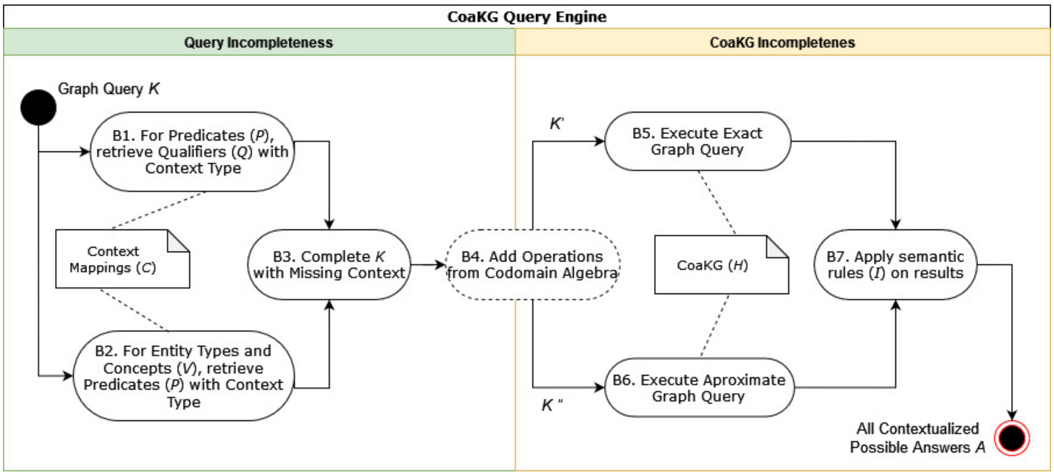


Figure 5.1: CoaKG Query Engine for Query and CoaKG Incompleteness

An exploratory search scenario about Geopolitical History of Brazil Suppose a historian found a historical document written in the Colonial Brazil Period that references the capital of Brazil but does not name the city. To understand the context of this document, the historian needs to discover what the capital was at this time. Then he performs a query to answer the following question: "*What were the capital cities of Brazil DURING the Brasil Colônia period?*", and the query interface translates into the following graph query *K3*:

K3 Capital cities of Brazil during Brazil Colony (of Portugal) period
(?v1) -[?r1:capital city of]→ (?v2)
Brasil Colônia) -[?p1:is_a]→ (Período Histórico)

To verify if *K3* is contextually incomplete, the engine parses the query to identify predicates, qualifiers, and entity types. For each element, a graph query is generated and executed. Table 5.1 shows the general templates and instances of graph queries over Context Mappings *(C)* for the example. Predicates, qualifiers, entity types labels, and identifiers from the H4 schema can be found in Table 4.3. In step *B1*, the query engine evaluates *K3* completeness using query *ck1* over the relation *capital city of* (h4r7) and retrieved three context qualifiers, two Temporal (*start time* - h4q1 and *end time* - h4q2), and one for Provenance (*stated in* - h4q3). The same verification is done using *is_a* predicate. Regarding Entity types and Concepts, query *ck3*, executed in step B2 for entity type *Periodo Historico* (h4v20), retrieves two Temporal (*start time* - h4q1 and *end time* - h4q2) and one for Provenance (*stated in* - h4q3).

ID	Query type	Graph Query on Context Mappings (C)
gck1	ALL elements	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(ALL)
gck2	ALL Predicates	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(ALLPred)
gck3	ALL Entity Types	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(ALLEntTp)
mck	one element (template)	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(?Element)'
ck1	Capital City Of (predicate instance)	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(h4r7)
ck2	IS_A (predicate instance)	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(is_a)
ck3	Período Histórico (entity type instance)	(?ContextType)-[?c1:Represented By]→ (?ContextLabel)-[?c2:Contextualizes]→(h4v20)

Table 5.1: Graph queries templates and its instantiation

Query K3, after Explicit Context Expansion (step B3) by adding the contextual qualifiers in bold, is shown bellow:

K3 after Explicit Context Expansion (B3).

(?v1) -[?r1:capital city of]→ (?v2)
(?r1) -[?q1:**start time**]→ (?v3)
(?r1) -[?q2:**end time**]→ (?v4)
(?r1) -[?q3:**stated in**]→ (?v5)

(Brasil Colônia) -[?p1:is_a]→ (Período Histórico)
(?p1) -[?q4:**stated in**]→ (?v8)
(Brasil Colônia) -[?p2:**start time**]→ (?v6)
(Brasil Colônia) -[?p3:**end time**]→ (?v7)

The original *K3* and its contextually expanded version specify a graph pattern with two disconnected subgraphs. In such cases, any Codomain Algebra should be applied to context values. In step B4, the engine evaluates the contextually complete version of graph query *K3* to identify whether Codomain Algebra operations are applicable to contexts. The CoaKG Query Engine can use the Codomain Algebra in two ways: (i) by inferring implicit relation among query result as a semantic rule in step B7 and (ii) by including a filter in the query to transform a disconnected graph pattern into a connected one. Filters avoid operations such as Cartesian products when the pattern query subgraph is not fully connected.

For Temporal and Location contexts there are some formal Algebras in the literature. For Provenance, a formal Algebra was not found and the advantage of a specialized algebra is to have relationships where the semantics are well defined and known. However, it is still possible to apply more general operations such as check containment, reputation ordering, etc or specific operations of data processing (how-provenance) [55].

The *K3* query has two types of associated Context: Temporal and Provenance. For Temporal context, Allen's Interval Algebra [70] was adopted. This algebra specifies 13 relationship types between two periods (temporal relations) as shown in Table 5.2 with the algebraic calculation. Considering that the two parts of the *K3* query have temporal context, the query expansion must add the algebraic calculation of one relationship between the two periods in the WHERE clause. The two periods are represented as ?t1(?v1 as start time, ?v2 as end time) and ?t2(?v3 as start time, ?v4 as end time).

Temporal Relationship	Algebraic Calculation (Explanation)
(?t1)-[before]→(?t2)	?v2 < ?v3
(?t1)-[after]→(?t2)	?v4 < ?v1
(?t1)-[meets]→(?t2)	?v2 = ?v3
(?t1)-[met by]→(?t2)	?v4 = ?v1
(?t1)-[overlaps]→(?t2) (?t2)-[overlapped by]→(?t1)	?v1 ≤ ?v4 AND ?v3 ≤ ?v2
(?t1)-[starts]→(?t2) (?t2)-[started by]→(?t1)	?v1 = ?v3
(?t1)-[during]→(?t2)	?v3 ≤ ?v1 AND ?v2 ≤ ?v4
(?t1)-[contains]→(?t2)	?v1 ≤ ?v3 AND ?v4 ≤ ?v2
(?t1)-[finishes]→(?t2) (?t2)-[finished by]→(?t1)	?v2 = ?v4
(?t1)-[equals]→(?t2)	?v1 = ?v3 AND ?v2 = ?v4

Table 5.2: Temporal relations and its explanation

This step of query expansion avoids misleading results, enables additional insights and relationships in the analysis of context values, enriches the answer with implicit knowledge, and provides further context-aware capabilities for exploratory search.

Finally, in steps B5 and B6 the two versions of *K3* are generated. In the case of K3 Approximate, both OPTIONAL and NVL were used.

K3 (Exact Version) after Codomain Algebra Expansion (B4). Algebraic Calculation as filter in bold.

(?v1) -[?r1:capital city of]→ (?v2)
(?r1) -[?q1:start time]→ (?v3)
(?r1) -[?q2:end time]→ (?v4)
(?r1) -[?q3:stated in]→ (?v5)
(Brasil Colônia) -[?p1:is_a]→ (Período Histórico)
(?p1) -[?q4:stated in]→ (?v8)
(Brasil Colônia) -[?p2:start time]→ (?v6)
(Brasil Colônia) -[?p3:end time]→ (?v7)

WHERE ?v3 =< ?v7 AND ?v6 =< ?v4

K3 (Approximated Version) after Codomain Algebra Expansion.

(?v1) -[?r1:capital city of]→ (?v2)
OPTIONAL (?r1) -[?q1:start time]→ (?v3)
OPTIONAL (?r1) -[?q2:end time]→ (?v4)
OPTIONAL (?r1) -[?q3:stated in]→ (?v5)
(Brasil Colônia) -[?p1:is_a]→ (Período Histórico)
OPTIONAL (?p1) -[?q4:stated in]→ (?v8)
OPTIONAL (Brasil Colônia) -[?p2:start time]→ (?v6)
OPTIONAL (Brasil Colônia) -[?p3:end time]→ (?v7)

WHERE ?v3 =< NVL(?v7, current_date) AND ?v6 =< NVL(?v4, current_date)

With both query results, the last step (B7) executes the predefined Semantic Rules over the retrieved subgraph to infer implicit context. For K3 results, from the previous Semantic Rule from Table 4.5, CoaKG Query Engine did not generate additional contextual information since it is not about a current period. From the Codomain Algebra, using the function described next that calculates the relationship between two periods, an additional relationship (?r1)→[overlaps]-(Brasil Colônia) was generated for every *capital city of* Claim ?r1 retrieved by exact version of *K3*.

Allen's Interval Algebra signature.
Check_Time_Interval
Input: ?t1 (?v1 as start time, ?v2 as end time), ?t2(?v3 as start time, ?v4 as end time) **Output:** (?t1)→[?tr1 as temporal relationship]-(?t2)

Figure 5.2 represents partial results generated by CoaKG Query Engine from the initial graph query *K3* (Provenance qualifiers omitted). The dotted lines represent those relationship about Temporal context inferred with Codomain Algebra. For All Contextualized Possible Answers of K3, see CKG-H4-K3-possible-final.tsv at GitHub [1].

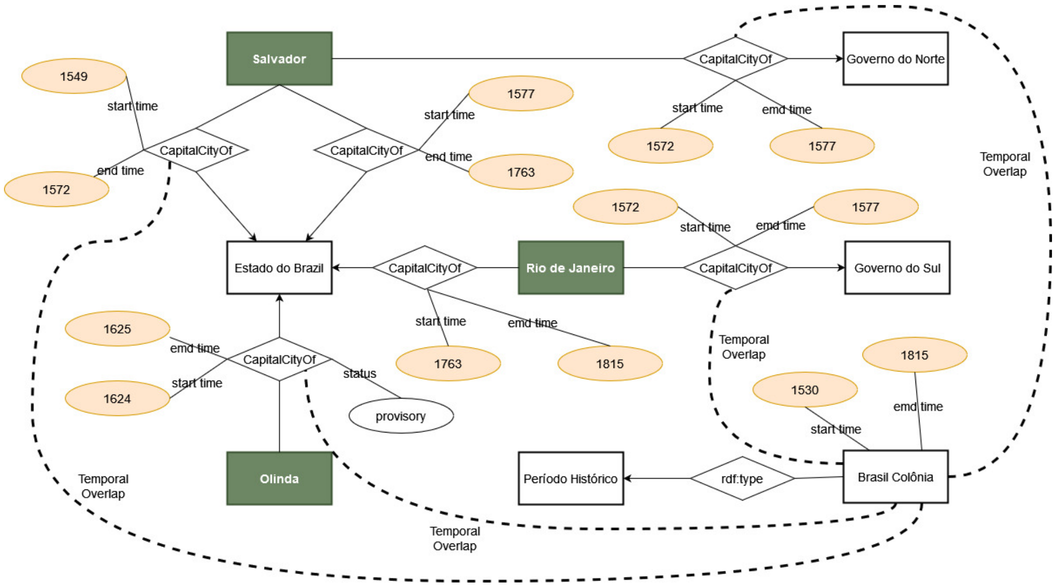


Figure 5.2: All Contextualized Answers for K3

Table 5.3 shows all questions elaborated to compose an Exploratory Search process about the CoaKG H4 domain. Context expressions are in bold. Original Queries and its expanded version and intermediate and final results can also be found in GitHub.

Observe that *K2* is similar to *K3*, since it has both Temporal context operations but the *K2* subgraph query pattern is a connected graph through node ?v2. In this case, the CoaKG Engine does not use Codomain Algebra to filter the result.

Instead, it is used to make explicit the overlap (temporal) relationship between the claims retrieved from CoaKG H4. Additionally, Semantic Rule *I1*

[1] `https://github.com/versant2612/CKG_UseCases/blob/53bb930d2a86d4a74f36cbb77c5c6c2bd7088aad/H4`

Query ID	Context Type	Question
K6	Temporal and Location	How a country named 'Brazil' is **shaped** (spatial) **today**
K7	Temporal and Location	**Spatial** changes of previous country **over time**
K1	Temporal	Capital cities of State of Brazil in **chronological order**
K2	Temporal	Capital cities of <u>Brazil</u> **while** Brazil has a Republican government
K3	Temporal	Capital cities of <u>Brazil</u> **during** Brazil Colony period
K4	Temporal	Capital cities of <u>Brazil</u> **before** Rio de Janeiro
K5	Temporal	Capital cities of <u>Brazil</u> **after** Salvador

Table 5.3: Exploratory Search Questions with Context Type.

explicitly identifies which Claims correspond to the Current period. CoaKG Query Engine also expanded the results from *K1*, *K4*, and *K5* using the same strategy.

K2 (Approximated Version) after Explicit Context Expansion.

(?v1) -[?r1:capital city of]→ (?v2)
OPTIONAL (?r1) -[?q1:**start time**]→ (?v3)
OPTIONAL (?r1) -[?q2:**end time**]→ (?v4)
OPTIONAL (?r1) -[?q3:**stated in**]→ (?v5)
(?v2) -[?p1:form of gov] → (Republica Federativa Presidencialista)
OPTIONAL (?p1) -[?q4:**stated in**]→ (?v8)
OPTIONAL (?p1) -[?p2:**start time**]→ (?v6)
OPTIONAL (?p1) -[?p3:**end time**]→ (?v7)

Query *K6* and *K7* involve two Context Types: Location for Entities and Temporal for Claims. Table 5.4 contains all versions of query *K6*, whose Question is contextually complete since both contexts are explicitly specified in query translation as a Star Join query based on ?v1 (*K6* Initial).

However, considering all Contextual Mappings for relation *geoshape* and entity type *geo-political unit*, other Contextual Claims were added (*K6* Exact). When executed, *K6* Exact generated an empty result since there was no geoshape for current_date. The Approximate version of *K6* retrieved the geoshape data and, according to Semantic Rule *I2*, it corresponds to the Current one (although its temporal context value is 1964).

Observe that *K7* is similar to *K6* since it has both Temporal and Location

Version	Query
K6 Initial	(?v1) -[?p1:alias]→ ('Brasil'); (?v1) -[?r2:is_a]→ (geo-political unity); (?v1) -[?r1:geoshape]→ (?v2); (?r1) -[?q1:point in time]→ (**current_date**);
K6 Exact	(?v1) -[?p1:alias]→ ('Brasil'); (?v1) -[?r2:is_a]→ (geo-political unity); **(?v1) -[?p1:inception]→ (?v4);** **(?v1) -[?p2:closure date]→ (?v5);** **(?v1) -[?r3:part of]→ (?v6);** **(?v1) -[?r4:form of government]→ (?v7);** (?v1) -[?r1:geoshape]→ (?v2); (?r1) -[?q1:point in time]→ (current_date); **(?r1) -[?q2:stated in]→ (?v3);**
K6 Approximate	(?v1) -[?p1:alias]→ ('Brasil'); (?v1) -[?r2:is_a]→ (geo-political unity); **OPTIONAL** (?v1) -[?p1:inception]→ (?v4); **OPTIONAL** (?v1) -[?p2:closure date]→ (?v5); **OPTIONAL** (?v1) -[?r3:part of]→ (?v6); **OPTIONAL** (?v1) -[?r4:form of government]→ (?v7); (?v1) -[?r1:geoshape]→ (?v2); **OPTIONAL** (?r1) -[?q1:point in time]→ (**?v8**); **OPTIONAL** (?r1) -[?q2:stated in]→ (?v3);

Table 5.4: Query K6 Versions

context and uses the same predicates and entity type, But *K7* is not a Star Join query. It is a graph pattern of two disconnected components (see K7 Initial in Table 5.5). Considering that the two parts of the *K7* query have Temporal and Location context, the query expansion must add in the WHERE clause a filter based on a topological relationship between Location context values and the chronological relationship of Temporal context values.

The Dimensionally Extended Nine-Intersection Model (DE-9IM) [69] was chosen as Codomain Algebra for Location context values. DE-9IM defines spatial binary relationships between two-dimensional geometries: Equals, Disjoint, Intersects, Touches, Crosses, Within, Contains, and Overlaps. See geo_check function signature.

DE-9IM signature for checking relationship.

geo_check

Input: ?v1 as geometry, ?v2 as geometry, ?r1 as relationship type
Output: boolean

Version	Query
K7 Initial	(Estados Unidos do Brasil) -[?r1:geoshape]→ (?v2); (?r1) -[?q2:point in time]→ (?v4); (?v5) -[?r2:geoshape]→ (?v6); (?r2) -[?q4:point in time]→ (?v8); (?v5) -[?p1:alias]→ (?v9); (?v5) -[?r2:is_a]→ (geo-political unity);
K7 Exact	(Estados Unidos do Brasil) -[?r1:geoshape]→ (?v2); **(?r1) -[?q1:stated in]→ (?v3)**; (?r1) -[?q2:point in time]→ (?v4); (?v5) -[?r2:geoshape]→ (?v6); **(?r2) -[?q3:stated in]→ (?v7)**; (?r2) -[?q4:point in time]→ (?v8); (?v5) -[?p1:alias]→ (?v9); (?v5) -[?r2:is_a]→ (geo-political unity); **(?v5) -[?p1:inception]**→ (?v10); (?v5) -[?p2:closure date]→ (?v11); (?v5) -[?r3:part of]→ (?v12); (?v5) -[?r4:form of government]→ (?v13); **WHERE GEO_CHECK(?v2, ?v6, OVERLAPS)** **AND ?v8 < ?v4**
K7 Approximate	(Estados Unidos do Brasil) -[?r1:geoshape]→ (?v2); **OPTIONAL** (?r1) -[?q1:stated in]→ (?v3); **OPTIONAL** (?r1) -[?q2:point in time]→ (?v4); (?v5) -[?r2:geoshape]→ (?v6); **OPTIONAL** (?r2) -[?q3:stated in]→ (?v7); **OPTIONAL** (?r2) -[?q4:point in time]→ (?v8); (?v5) -[?p1:alias]→ (?v9); **OPTIONAL** (?v5) -[?r2:is_a]→ (geo-political unity); **OPTIONAL** (?v5) -[?p1:inception]→ (?v10); **OPTIONAL** (?v5) -[?p2:closure date]→ (?v11); **OPTIONAL** (?v5) -[?r3:part of]→ (?v12); **OPTIONAL** (?v5) -[?r4:form of government]→ (?v13); **WHERE GEO_CHECK(?v2, ?v6, OVERLAPS)** **AND NVL(?v8, min_date) < NVL(?v4, curr_date)**

Table 5.5: Query K7 Versions

The CoaKG Query Engine was detailed in this chapter. Query completion with graph queries over CoaKG Context Mappings and Codomain Algebra were illustrated using the same KG example. The next chapter describes a Proof of Concept over a CoaKG within an exploratory search process. The CoaKG was built using a subset of WD as a standard KG whose Context Types are specified based on the result of WD Profiling.

6
Wikidata subset as a Proof of Concept

We have shown how a CoaKG can be built from scratch in Chapter 4. However, most existing KGs must be complemented to become CoaKGs. This chapter describes a PoC CoaKG built by extending an existing KG (a WD sub-set). Context Types are specified based on the result of WD Profiling and used to specify the Context Mappings. Competency Questions were elaborated to generate queries with and without context, and their respective answers were obtained.

Why Wikidata? Several reasons made us select WD as the source KG for this evaluation phase:

1. The WD graph data structure represents statements composed of claims associated with qualifiers and references but does not establish how users should retrieve, interpret, and use it.

2. WD contains contextual information, but the WD Ontology does not explicitly identify which qualifier represents each type of Context.

3. WD is designed to represent Provenance as references, but there is also Provenance information among qualifiers.

4. WD dataset dumps are used in other KG-related tasks such as KGQA Benchmarks: multilingual questions [56] and temporal reasoning [57]. It is also used in GraphDB Benchmark: WDBench [58].

5. WD acts as a secondary database, gathering information published in primary sources with their references and connections to other databases. WD does not aim to be a Single Source of Truth (SSoT) but enforces verifiability [14] For example, there is no *true population of Brazil* but rather a *population of Brazil in 2021 determined by estimation as published by Brazilian Government on 27 August 2021 as "Portaria PR-268"* as Figure2.2.

6.0.1 PoC Resources

WD dump In this PoC, a dataset was created based on the DWD dump provided by the KG Center at ISI, reflecting the WD dump of June 2022[1]. Note that this dump does not include instances of `scholarly articles` (Q13442814) and `review articles` (Q7318358) and their subclasses since their frequency (over 50,00% of all statements) would distort the statistics. It also does not include statements annotated with deprecated rank or references for statements.

There are 559,038,971 claims using 9,653 properties as predicates and 141,983,745 qualifications are associated with claims using 9,906 properties as qualifiers. There are 10,089 properties, and 90,00% have a property scope constraint assigned to them, distributed according to Table 6.1. None represents approximately 10,00% of the properties without property scope constraint [48].

Property Scope	Property Count
as main value (Predicate)	8663
as reference (Referrer)	6058
none	1099
as qualifier	745

Table 6.1: Property Scope Constraint Quantity

KGTK Toolkit WD PoC used the KGTK Toolkit [59] to analyze and query the WD dump. This toolkit offers generic computations for graph data, such as graph statistics (degree), pathfinding, and centrality metrics. It also supports the generation of graph embedding, lexicalization, export, and import to and from other known formats.

The KGTK file format represents KG edges using TSV files with at least three columns: node1, label, and node2. Node1 is the source of the edge, node2 is the destination, and the label names the edge. An additional column, named id, can be added to store the unique identifier of the edge. This id can be referenced as node1 or node2 of other edges. Using this id, KGTK can represent KGs with Hiper relational and Multi-Layer Graph Models.

KGTK uses SQLite, a lightweight file-based SQL database, as a storage engine and implements a declarative graph query language, Kypher, inspired

[1]ISI DWD 2022 downloaded from `https://drive.google.com/drive/folders/1a6cUI1UEWRTNbvqtLAfJUOwEJ4ssTqdz?usp=share_link`

by Cypher, translating graph queries into SQL. Kypher uses ASCII-art style to represent subgraph patterns: `(node1)-[id:predicate]→(node2)`.

All kgtk scripts and query results of this PoC are publicly available at Github[2].

6.0.2 WD Engineering

WD Geopolitical History To select a subset of claims and qualifications from the DWD dump, the following procedure was used:

1. Given the node types seeds, Table 6.2, find all node instances,
2. Given the node instances from the previous step as root nodes, find reachable nodes with the transitive closure path of the predicates `P1365 (replaces)` and `P1366 (replaced by)`,
3. For each reachable node type, count the total occurrences,
4. From the top 100 most frequent reachable node types, choose new node types and add them to the initial seed list,
5. Retrieve all claims where the subject node belongs to one of the types listed in Step 4,
6. Retrieve all qualifications associated with these claims (from step 5),
7. Retrieve claims where the object node belongs to claims from Step 5
8. Retrieve all qualifications related to these new claims (from step 7)
9. Retrieve all claims where the subject node belongs to the object node of qualifications from Steps 6 and 8
10. Retrieve all qualifications associated with these claims (from step 9)
11. Generate two sub-graphs concatenating results from steps 5, 7, and 9 as Countries Claims and 6, 8, and 10 as Countries Qualifications, removing duplicates.

[2]`https://github.com/versant2612/CKG_UseCases/tree/main/PoC`

Status	QNode	Label	Count
Seed (step1)	Q3024240	historical country	2608
Seed (step1)	Q6256	country	248
Seed (step1)	Q512187	federal republic	11
Seed (step1)	Q859563	secular state	8
Added (step 4)	Q3624078	sovereign state	426
Added (step 4)	Q48349	empire	78
Added (step 4)	Q133156	colony	210

Table 6.2: Node types list related to Geopolitical History

KG Profiling Since there isn't a WD schema to fully model each selected entity type, WD profiling over the subset of claims and qualifications were executed [41]. Profiling aims to identify context qualifiers and predicates already present in the KG. It also gives insights to formulate semantic rules interpretations for qualifiers' existence, absence, and possible claim contradictions.

WD Profiling computed the quantity of the following elements:

- Predicate count in claims,
- Count of subject and object node types in claims per predicate,
- Count of object data types in claims per predicate,
- Claim count per subject node,
- Count of qualifiers in qualifications by predicate in claims,
- Count of object node types in qualifications by qualifiers, and
- Count of object data types in qualifications by qualifiers.

WD profiling results are available on GitHub. Some results are commented on below to exemplify some findings.

From the top-5 countries from claims counting, the list is: `Indonesia (Q252) -` 1489; `India (Q668) -` 1475; `United States of America (Q30) -` 1371; `Papua New Guinea (Q691) -` 1289; `Australia (Q408) -` 1223. The top-5 most used predicates are: `language used (P2936) -` 10166; `population (P1082) -` 10165; `demonym (P1549) -` 7485; `diplomatic relation (P530) -` 6827; `Human Development Index (P1081) -` 6409.

There are 1261 claims with predicate `P35 (head of state)` to represent the relationship between a country (or a geographic unit) and a person; 1103 of them are qualified by `P580 (start time)` and 906 by `P582(end time)`, both correspond to temporal contextual information. The qualifier `P39 (position held)` is associated with 239 claims as additive information since it can be interpreted as an n-ary relationship where the position corresponds to the role of the head of state in the relationship. There are also 79 claims (from 1261) qualified by `P459 (determination method)` expressing how the person was chosen or designated as the head of state (for example, by-election).

Competency Questions After investigating and profiling the WD subset, Competency Questions (CQ) were developed, see Table 6.3. Each CQ has acceptable answer specifications to fulfill an Exploratory Search process example. CQs are domain-related natural language questions with valid answers used to evaluate Ontologies [60] regarding coverage, correctness, and accessibility.

Context aspects explicitly mentioned in the Questions and Answer specifications are highlighted. These CQs should be convertible into graph queries that should be answerable, with all contextual information, using the CoaKG. CoaKG Query Engine was evaluated on how it can complete graph queries and answers with missing context information and generate contextualized answers.

ID	Competency Quesiton and Answers
CQ0	Question: How "<country alias name>" is spatially shaped **today** ? / What is the **current** geographic representation of the geopolitical unit referring to "<country alias name>"?
	Acceptable Answer: Single Answer with a **geographic** representation, **reference date**, and geopolitical unit ID Acceptable Answer: Multiple Answer with a **geographic** representation, geopolitical unit ID, **reference date**, **all contextual information** of geopolitical unit, and other geopolitical unit identifiers
CQ1a	What geopolitical units did "<CQ0 geopolitical unit>" replace?
	Acceptable Answer: Multiple Answer with geopolitical units ID that were directly replaced **with its reference date.**
CQ1b	List the geopolitical changes of "<CQ0 geopolitical unit>" **over time**.
	Acceptable Answer: Multiple Answer with geopolitical units ID that were directly and indirectly replaced **with replacement reference date.**
CQ2	**When** was "<CQ0 geopolitical unit>" founded/established?

continued on the next page

ID	Competency Quesiton and Answers
	Acceptable Answer: Single Answer with **foundation date** Acceptable Answer: Multiple Answer with **foundation date, criterion and other provenance information** Acceptable Answer: Multiple Answer with geopolitical units ID that were directly replaced **with replacement reference date.** Acceptable Answer: Multiple Answer with geopolitical units ID that were directly replaced **with abolishment date.**
CQ3a	What is the **current** capital city of "<CQ0 geopolitical unit>"?
	Acceptable Answer: Single Answer with Capital City name and **start date** Acceptable Answer: Multiple Answer with Capital City name, **start date and its role**
CQ3b	What were the capital cities of "<CQ0 geopolitical unit>" **over time**?
	Acceptable Answer: Multiple Answer with Capital City name, **its role and time periods**
CQ3c	What were the capital cities of "<CQ1b geopolitical unit list>" **over time**?
	Acceptable Answer: Multiple Answer with geopolitical units ID that were directly replaced, Capital City name, **its role and time periods**
CQ4a	What position does the main administrative leader occupy in "<CQ0 previous geopolitical unit>", and who is the **current** leader?
	Acceptable Answer: Single Answer with Leader name, **start date** and position Acceptable Answer: Multiple Answer with Leader names, **start date** and its position
CQ4b	What position do main administrative leaders occupy in "<CQ0 previous geopolitical unit>", and who were these leaders **over time**?
	Acceptable Answer: Multiple Answer with Leader names, **its position and time periods**

Table 6.3: Competency Questions about Geopolitical History

Context Mappings Since there is no predefined schema for each geopolitical unit type, the WD Profiling results were analyzed in conjunction with the

CQs to identify context information explicitly represented in the KG. WD constraints for predicates and qualifiers were also considered. The Contextual Mappings associated with them were added to the KG, transforming it into a CoaKG.

The first step was guided by the procedure described in the box bellow using CQs specific to the domain and WD Profiling. The second step was guided by the procedure described next using only WD Profiling and can be reused in any domain.

CoaKG Engineering for WD Competency Questions

For each Competency Question:

Identify the relationships or attributes and context types of interest mentioned.

For each relationships or attributes:

- Identify the WD predicate and qualifiers associated.
- For each WD qualifier:
  - Identify the context types associated (or not).
- For each contextual qualifier:
  - Verify its frequency to define mandatory or optional
  - Create context mappings between WD predicates, qualifiers, and context types (see Tables 6.4 and B.3)
- For each context type:
  - Verify if a semantic rule can be built;
  - Verify if a default context can be defined;

Identify the entity and context types of interest mentioned.

For each entity type;

- Identify the entity type (or concept) and its predicates associated.
- For each predicate:
  - Identify the context types associated (or not).
- For each contextual predicate:
  - Verify its frequency to define mandatory or optional.
  - Create context mappings between WD entity types, predicates, and context types (see Table 6.5)
- For each context type:
  - Verify if a semantic rule can be built;
  - Verify if a default context can be defined.
- For each entity type (or concept):
  - Identify WD predicates with identity semantic.
  - Create identity mappings between WD entity types and WD predicates (see Table 6.5)

CoaKG Engineering for WD Generic Context

For the Generic Context:

Identify predicates applicable to all entity types (or concepts)
Identify qualifiers applicable to all predicates
Create context mappings between WD entity types, predicates, and qualifiers as optional.

For example, WD Profiling on predicate `P35 (head of state)` identified that 1103 claims from 1261 have `P580 (start time)`, so the respective context mappings was specified as a mandatory qualifier for temporal context as highlighted in Table 6.4. The most frequent entity types from the dataset, `country (Q6256)` and `sovereign state (Q3624078)`, have context predicates as shown in Table 6.5. Table 6.6 shows qualifiers and predicates identified with WD Profiling.

ID	Predicate	Qualifier	Context Type	Status
CQ4	office held by head of state (P1906)	start time (P580)	Temporal	Mandatory
		end time (P582)	Temporal	Optional
		statement is subject of (P805)	Provenance	Optional
		object has role (P3831)		
		location (P276)	Location	Optional
	head of state (P35)	**start time (P580)**	**Temporal**	**Mandatory**
		end time (P582)	Temporal	Optional
		statement is subject of (P805)	Provenance	Optional
		subject has role (P2868)		
		determination method (P459)		
		end cause (P1534)		
		location (P276)	Location	Optional

Table 6.4: Context Mappings of Predicates P35 and P1906 from CQ4

6.0.3 WD Exploratory Search

Rendering Competency Questions as Queries In this section, each Competency Question (CQ) is manually translated into a Graph Query K. There

Entity Type / Concept	Property / Relationship	Context Type	Status
country (Q6256) sovereign state (Q3624078)	dissolved, abolished or demolished date (P576)	Temporal	Optional
	inception (P571)	Temporal	Mandatory
	geoshape (P3896)	Location	Mandatory
	continent (P30)		
	part of (P361)	Location	Optional
	country (P17)		
	described by source (P1343)	Provenance	Optional
	ISO 3166-1 alpha-3 code (P298)	Identity	Optional
	GeoNames ID code (P1566)		

Table 6.5: Context Mappings for Entity Types

Property	Type	Description (obtained from WD)
nature of statement (P5102)	Qualifier	the underlying circumstances of this statement: frequency, nature, condition, status of statement, modality of statement, validity of statement, refine statement
statement disputed by (P1310)	Qualifier	entity that disputes a given statement (disputed by, rejected by, opposed by, refused by, denied by, controversy stated by)
reason for preferred rank (P7452)	Qualifier	the reason to be indicated why a particular statement should be considered preferred among others with the same predicate and same subject
reason for deprecated rank (P2241)	Qualifier	qualifier to indicate why a particular statement should have deprecated rank (deprecated because, cause of deprecation)
instance of (P31)	Predicate	class of which this subject is a particular example and member
short name (P1813)	Predicate	short name of a place, organization, person, journal, ...
official name (P1448)	Predicate	official name of the subject in its official language(s)

Table 6.6: Generic Context Mappings of All Predicates, Entity Types, and Concepts

are at least four versions of K written in Kypher syntax: the Original Query, the Expanded with Exact Context, the Expanded with Possible Context, and the Expanded with Missing Context. The CoaKG Query Engine also generates additional graph queries to evaluate the completeness of the query and KG instances to achieve the goal of retrieving all contextualized answers.

CQ0: How is "<country alias name>" spatially shaped **today** ? / What is the **current** geographic representation of the geopolitical unit referring to "<country alias name>"?
In the case of a single value, besides the geographic representation (v2) and geopolitical unit ID (v1), the expanded version of K0 returned contextual information about property `geoshape (P3896)`; see Table 6.7. In case of multiple values, besides the geographic representation (v2), geopolitical unit ID (v1), and contextual information about property `geoshape (P3896)`, the answer can be added with contextual properties and relations of geopolitical units, and geopolitical unit identifiers from external sources (Table 6.8). Such additional contextual information could always be automatically retrieved but when there is only one instance this could mean information overload. So, within an exploratory search, if the expanded query returns more than one instance for a single answer competency question, the user can request the contextual information and identifiers on each instance to disambiguate.

Query type	**Kypher Query**
K0 Original (Without Qualifiers)	–match '(v1 {label: c_label})-[p1:P3896 {'label;label': p_label}]→(v2 {label: v2_label}), (v1)-[]→(v1_alias) ' –where 'v1_alias = $name ' –para name="<country alias name>" –return 'p1 as id, v1 as node1, c_label as 'node1;label', p1.label as label, p_label as 'label;label', v2 as node2, v2_label as 'node2;label''
K0 Expanded Exact (With Mandatory Qualifiers)	–match '(v1 {label: v1_label})-[p1:P3896 {'label;label': p_label}]→(v2 {label: v2_label}), (v1)-[]→(v1_alias), *cq: (p1)-[q1]→(v3 {label: v3_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P3896) '* –where 'v1_alias = $name *AND q1.label = c1 AND EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)}* ' –para name="<country alias name>" –return 'p1 as id, v1 as node1, v1_label as 'node1;label', p1.label as label, p_label as 'label;label', v2 as node2, v2_label as 'node2;label', *q1 as id, p1 as node1, "" as 'node1;label', c1 as label, c1_label as 'label;label', v3 as node2, v3_label as 'node2;label', concat(q1,"-",C) as id, q1 as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label''*

continued on the next page

Query type	**Kypher Query**
K0 Expanded Possible (With Mandatory and Optional Qualifiers)	–match '(v1 {label: v1_label})-[p1:P3896 {'label;label': p_label}]→(v2 {label: v2_label}), (v1)-[]→(v1_alias), *(p1)-[q1]→(v3 {label: v3_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P3896)* ' –where 'v1_alias = $name *AND q1.label = c1*' –para name="<country alias name>" –return 'p1 as id, v1 as node1, v1_label as 'node1;label', p1.label as label, p_label as 'label;label', v2 as node2, v2_label as 'node2;label', *q1 as id, p1 as node1, "" as 'node1;label', c1 as label, c1_label as 'label;label', v3 as node2, v3_label as 'node2;label', concat(q1,"-",C) as id, q1 as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label''*
K0 Missing Mandatory Qualifiers	–match '(v1 {label: c_label})-[p1:P3896 {'label;label': p_label}]→(v2 {label: v2_label}), (v1)-[]→(v1_alias), *ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P3896)* ' –where '*EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)} AND NOT EXISTS {cq: (p1)-[q1]→(v2) where q1.label = c1}* AND v1_alias = $name ' –para name="<country alias name>" –return 'p1 as id, v1 as node1, c_label as 'node1;label', p1.label as label, p_label as 'label;label', v2 as node2, v2_label as 'node2;label', concat(p1,"-",c1) as id, p1 as node1, "" as 'node1;label', c1 as label, c1_label as 'label;label', **"unknown" as node2**, "" as 'node2;label', concat(p1,"-",c1,"-",C) as id, concat(p1,"-",c1) as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label''

Table 6.7: Graph Queries for CQ0 Single Answer

The geopolitical unit ID (v1) is a QNode value representing the country selected for exploratory search and will be referenced as a constant QId in the following queries.

CQ1a: What geopolitical units did "<CQ0 geopolitical unit=QId>" replace?

This question is translated into a BGP that retrieves *geopolitical units*

Query type	Kypher Query
K0 geopolitical unit context	–match '(v1 {label: v1_label})-[p1:P31 {'label;label': type_label}]→(v2 {label: v2_label}), (v1)-[p3 {'label;label': p3_label}]→(v4 {label: v4_label}), (v1)-[]→(v1_alias), ckg: (C {label: C_label})-[pc2:ckgr1]→(pred {label: pred_label})-[pc1:ckgr2]→(v2) ' –where 'v1_alias = $name AND p3.label = pred AND EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)} ' –para name="<country alias name>" –return 'p1 as id, v1 as node1, v1_label as 'node1;label', p1.label as label, type_label as 'label;label', v2 as node2, v2_label as 'node2;label', p3 as id, v1 as node1, v1_label as 'node1;label', p3.label as label, p3_label as 'label;label', v4 as node2, v4_label as 'node2;label', concat(p3,"-",pred) as id, p3 as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label''
K0 geopolitical unit identifiers	–match '(v1 {label: v1_label})-[p1:P31 {'label;label': type_label}]→(v2 {label: v2_label}), (v1)-[p3 {'label;label': p3_label}]→(v4 {label: v4_label}), (v1)-[]→(v1_alias), ckg: (pred {label: pred_label})-[pc1:ckgr4]→(v2) ' –where 'v1_alias = $name AND p3.label = pred ' –para name="<country alias name>" –return 'p1 as id, v1 as node1, v1_label as 'node1;label', p1.label as label, type_label as 'label;label', v2 as node2, v2_label as 'node2;label', p3 as id, v1 as node1, v1_label as 'node1;label', p3.label as label, p3_label as 'label;label', v4 as node2, v4_label as 'node2;label', concat(p3,"-",pred) as id, p3 as node1, "" as 'node1;label', "ckgr8" as label, "ckg:Determines" as 'label;label', "ckgId" as node2, "Entity Identifier" as 'node2;label''

Table 6.8: Graph Queries for CQ0 Multiple Answer (External Identifiers added)

QID (v1) directly *replaced by* (P1365) "<CQ0 geopolitical unit>" (QId). Table 6.9 presents all versions of K1a. Since the predicate *replaces* (P1366) is inverse of *replaced by* (P1365), another graph query translation is also generated, changing the position of the QId from subject to object on edge.

CQ1b: List the geopolitical changes of "<CQ0 geopolitical unit>" **over time**.

This question requires exploring the transitivity of the predicates P1365 (*replaces*) and P1366 (*replaced by*) using path/navigational graph query (NGP) to retrieve all geopolitical unit IDs that were directly and indirectly replaced by the previous geopolitical unit. This query retrieved a list of edges between the root node (QId) and the reachable nodes of other geopolitical units' IDs. Based on this list of edges, CoaKG Query Engine adds all contextual information associated with the predicate, as shown in Table 6.10.

CQ2, CQ3a, and CQ3b are very similar to CQ0 in terms of con-

Query type	Kypher Query
K1a Original (Without Qualifiers)	–match '(:**QId** {label: c_label})-[p1:P1365 {'label;label': p_label}]→(v1 {label: v1_label}) ' –return 'p1 as id, **QId** as node1, c_label as 'node1;label', p1.label as label, p_label as 'label;label', **v1 as node2**, v1_label as 'node2;label' '
K1a Expanded Exact (With Mandatory Qualifiers)	–match '(:QId {label: c_label})-[p1:P1365 {'label;label': p_label}]→(v1 {label: v1_label}), *(p1)-[q1]→(v2 {label: v2_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P1365)* ' *–where 'q1.label = c1 AND EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)} '* –return 'p1 as id, QId as node1, **v1 as node1**, v1_label as 'node1;label', p1.label as label, p_label as 'label;label', v2 as node2, v2_label as 'node2;label', *q1 as id, p1 as node1, "" as 'node1;label', c1 as label, c1_label as 'label;label', v3 as node2, v3_label as 'node2;label', concat(q1,"-",C) as id, q1 as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label''*
K1a Expanded Possible (With Mandatory and Optional Qualifiers)	–match '(:QId {label: c_label})-[p1:P1365 {'label;label': p_label}]→(v1 {label: v1_label}), *(p1)-[q1]→(v2 {label: v2_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P1365)* ' *–where 'q1.label = c1'* –return 'p1 as id, QId as node1, **v1 as node1**, v1_label as 'node1;label', p1.label as label, p_label as 'label;label', v2 as node2, v2_label as 'node2;label', *q1 as id, p1 as node1, "" as 'node1;label', c1 as label, c1_label as 'label;label', v3 as node2, v3_label as 'node2;label', concat(q1,"-",C) as id, q1 as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label''*
K1a Missing Mandatory Qualifiers	–match ' (:QId {label: c_label})-[p1:P1365 {'label;label': p_label}]→(v1 {label: v1_label}), *ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P1365)* ' *–where 'EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)} AND NOT EXISTS {cq: (p1)-[q1]→(v2) where q1.label = c1} '* –return 'p1 as id, QId as node1, c_label as 'node1;label', p1.label as label, p_label as 'label;label', v1 as node2, v1_label as 'node2;label', *concat(p1,"-",c1) as id, p1 as node1, "" as 'node1;label', c1 as label, c1_label as 'label;label',* ***"unknown" as node2****, "" as 'node2;label', concat(p1,"-",c1,"-",C) as id, concat(p1,"-",c1) as node1, "" as 'node1;label', "ckgr9" as label, "ckg:Context Type" as 'label;label', C as node2, C_label as 'node2;label' '*

Table 6.9: Graph Queries for CQ1a Multiple Answer

text augmentation so they won't be described here but can be found in

Query type	**Kypher Query**
K1a Original (Without Qualifiers)	–match ’(:QId)-[:reachable]→(v1), (country {label: c_label})-[p1:P1365 {‘label;label‘: p_label}]→(v1 {label: v1_label}) ’ –return ’p1 as id, **QId** as node1, c_label as ‘node1;label‘, p1.label as label, p_label as ‘label;label‘, **v1 as node2**, v1_label as ‘node2;label‘ ’
K1a Expanded Exact (With Mandatory Qualifiers)	–match ’(:QId)-[:reachable]→(v1), (country {label: c_label})-[p1:P1365 {‘label;label‘: p_label}]→(v1 {label: v1_label}), *(p1)-[q1]→(v2 {label: v2_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P1365)* ’ *–where ’q1.label = c1 AND EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)} ’* –return ’p1 as id, QId as node1, **v1 as node1**, v1_label as ‘node1;label‘, p1.label as label, p_label as ‘label;label‘, v2 as node2, v2_label as ‘node2;label‘, *q1 as id, p1 as node1, "" as ‘node1;label‘, c1 as label, c1_label as ‘label;label‘, v3 as node2, v3_label as ‘node2;label‘, concat(q1,"-",C) as id, q1 as node1, "" as ‘node1;label‘, "ckgr9" as label, "ckg:Context Type" as ‘label;label‘, C as node2, C_label as ‘node2;label‘’*
K1a Expanded Possible (With Mandatory and Optional Qualifiers)	–match ’(:QId)-[:reachable]→(v1), (country {label: c_label})-[p1:P1365 {‘label;label‘: p_label}]→(v1 {label: v1_label}), *(p1)-[q1]→(v2 {label: v2_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P1365)* ’ *–where ’q1.label = c1’* –return ’p1 as id, QId as node1, **v1 as node1**, v1_label as ‘node1;label‘, p1.label as label, p_label as ‘label;label‘, v2 as node2, v2_label as ‘node2;label‘, *q1 as id, p1 as node1, "" as ‘node1;label‘, c1 as label, c1_label as ‘label;label‘, v3 as node2, v3_label as ‘node2;label‘, concat(q1,"-",C) as id, q1 as node1, "" as ‘node1;label‘, "ckgr9" as label, "ckg:Context Type" as ‘label;label‘, C as node2, C_label as ‘node2;label‘’*
K1a Missing Mandatory Qualifiers	–match ’(:QId)-[:reachable]→(v1), (country {label: c_label})-[p1:P1365 {‘label;label‘: p_label}]→(v1 {label: v1_label}), *ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:P1365)* ’ *–where ’EXISTS {ckg: (pc1)-[pc3:ckgp1]→(ckgl2)} AND NOT EXISTS {cq: (p1)-[q1]→(v2) where q1.label = c1} ’* –return ’p1 as id, QId as node1, c_label as ‘node1;label‘, p1.label as label, p_label as ‘label;label‘, v1 as node2, v1_label as ‘node2;label‘, *concat(p1,"-",c1) as id, p1 as node1, "" as ‘node1;label‘, c1 as label, c1_label as ‘label;label‘, **"unknown" as node2**, "" as ‘node2;label‘, concat(p1,"-",c1,"-",C) as id, concat(p1,"-",c1) as node1, "" as ‘node1;label‘, "ckgr9" as label, "ckg:Context Type" as ‘label;label‘, C as node2, C_label as ‘node2;label‘ ’*

Table 6.10: Graph Queries for CQ1b Multiple Answer

script_kgtk_PoC_qId.sh[3].

[3] https://github.com/versant2612/CKG_UseCases/tree/main/PoC

CQ4a - What position does the main administrative leader occupy in "<CQ0 previous geopolitical unit>", and who is the **current** leader?

CQ4b - What position do main administrative leaders occupy in "<CQ0 previous geopolitical unit>", and who were these leaders **over time**?

There are two pairs of predicates involved in CQ4a/CQ4b translation: {*office held by head of government* (P1313), *head of government* (P6)} and {*office held by head of state* (P1906) and *head of state* (P35)}. All of them have Temporal qualifiers, which can be used to filter and explicitly establish time interval relations through Codomain Algebra. Since there are unknown values due to KG incompleteness, filtering would retrieve a few answers for the Expanded with Exact Context version of K4a/K4b. Expanded with Possible Context version (Table 6.11) will retrieve more answers to be evaluated to extract implicit temporal relations.

Query type	**Kypher Query**
K4a/K4b Expanded Possible (P1313 and P6)	–match '(:QId)-[p1:pred {'label;label': p_label}]→(v1 {label: v1_label}), *(p1)-[q1]→(v2 {label: v2_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:pred)* ' –*where 'q1.label = c1* AND pred in ["P1313", "P6"]*'* –return ...
K4a/K4b Expanded Possible (P1906 and P35)	–match '(:QId)-[p1:pred {'label;label': p_label}]→(v1 {label: v1_label}), *(p1)-[q1]→(v2 {label: v2_label}), ckg: (C {label: C_label})-[pc2:ckgr1]→(c1 {label: c1_label})-[pc1:ckgr2]→(:pred)* ' –*where 'q1.label = c1* AND pred in ["P1906", "P35"]*'* –return ...

Table 6.11: Expanded version of Graph Queries for CQ4a/CQ4b

All Contextualized Answers for Competency Questions Some countries were selected to be explored using all Competency Questions. In this section, examples of each CQ will be analyzed in terms of contextual information added to the answer. Tables 6.12 until 6.21 follow the color pattern described in Figure 6.1.

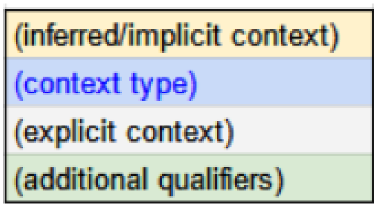


Figure 6.1: Table colors captions

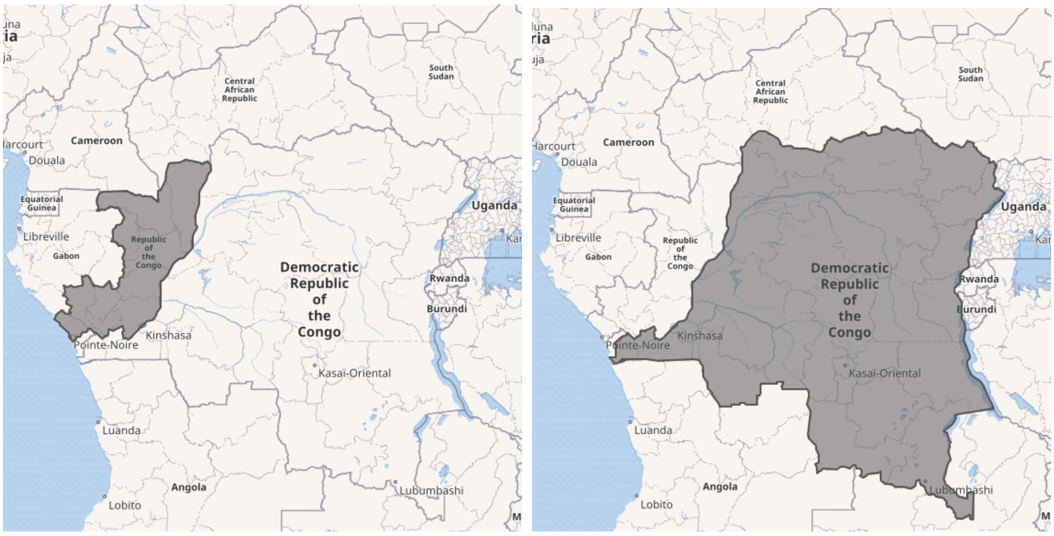


Figure 6.2: Geographic representation of Q971 and Q974

CQ0: How is "*Congo*" spatially shaped **<u>today</u>** ? / What is the **<u>current</u>** geographic representation of the geopolitical unit referring to "*Congo*"?

There are two different countries known as Congo located in the continent of Africa (Q15): Republic of the Congo (Q971), see Table 6.12, and Democratic Republic of the Congo (Q974), see Table 6.13. In case of multiple answers, besides geographic representation, geopolitical unit ID, and reference date, two additional queries are executed to retrieve all contextual information of each geopolitical unit and external geopolitical unit identifiers for disambiguation. Both are approximate answers since there is no reference date (Temporal Context) of geoshape information shown in Figure 6.2.

CQ0: How is "*Luxembourg*" spatially shaped **<u>today</u>** ? / What is the **<u>current</u>** geographic representation of the geopolitical unit referring to "*Luxembourg*"?

Two different geoshapes exist for the same country: Luxembourg (Q32); see Table 6.14. One has a reference date of 2018-05-22, and the other is 2020-02-14. Implicit temporal context can be inferred by applying rule *I2* from Table 4.5 to identify that the most recent can be considered current. This claim in WD has preferred rank without an explicit justification via `reason for preferred rank (P7452)` qualifier.

CQ1a: What geopolitical units did "*Brazil (Q155)*" replace?

According to WD, `Brazil (Q155)` directly replaced two other entities: one is a country called the `Empire of Brazil (Q217230)`, and the other is a historical period called `Republic of the United States of Brazil (Q5848654)`. The latter was disregarded, due to its type, in the response shown in Table 6.15. However, it is essential to highlight that the query approach does

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Claim about geoshape - unknown point in time for Temporal Context			
Q971WD-1	Republic of the Congo (Q971)	geoshape (P3896)	Data:Republic of Congo.map
Q971WD-2	Q971WD-1	point in time (P585)	unknown
Q971WD-3	Q971WD-2	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Contextualized Entity Republic of the Congo - Temporal, Location and Identifiers			
Q971WD-4	Republic of the Congo (Q971)	instance of (P31)	sovereign state (Q3624078)
Q971WD-5	Republic of the Congo (Q971)	instance of (P31)	country (Q6256)
Q971WD-6	Republic of the Congo (Q971)	GeoNames ID (P1566)	2260494
Q971WD-7	Q971WD-6	ckg:Determines (ckgr8)	Entity Identifier (ckgId)
Q971WD-8	Republic of the Congo (Q971)	ISO 3166-1 alpha-3 code (P298)	COG
Q971WD-9	Q971WD-8	ckg:Determines (ckgr8)	Entity Identifier (ckgId)
Q971WD-10	Republic of the Congo (Q971)	country (P17)	Republic of the Congo (Q971)
Q971WD-11	Q971WD-10	ckg:Context Type (ckgr9)	Location (ckgL1)
Q971WD-12	Republic of the Congo (Q971)	continent (P30)	Africa (Q15)
Q971WD-13	Q971WD-12	ckg:Context Type (ckgr9)	Location (ckgL1)
Q971WD-14	Republic of the Congo (Q971)	inception (P571)	^1960-01-01T00:00:00Z/9
Q971WD-15	Q971WD-14	ckg:Context Type (ckgr9)	Temporal (ckgT1)

Table 6.12: Multiple Answers for CQ0: Q971

not resolve potential conflicts at this layer.

The relationship `Brazil (Q155)` →`[replaces (p1365)]- Empire of Brazil (Q217230)`, identified as Q155WDK1a-1, does not have context value for temporal qualifiers. In this case, an additional query to retrieve the context of involved entities (as was done for CQ0) can increase knowledge about their chronology. Table 6.16 presents all contextual information for entity `Empire of Brazil (Q217230)`, temporal context values are highlighted. This results suggested that the replacement occurred after the `dissolved, abolished or demolished date (P576)`.

CQ1b: List the geopolitical changes of "*Brazil (Q155)*" **over time**.

Since the `replaces` (P1365) relationship is transitive, applying transitive closure to a path query (NGP) can find all entities of interest that *Brazil*

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Claim about geoshape - unknown point in time for Temporal Context			
Q974WD-1	Democratic Republic of the Congo (Q974)	geoshape (P3896)	Data:DR Congo.map
Q974WD-2	Q974WD-1	point in time (P585)	unknown
Q974WD-3	Q974WD-2	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Contextualized Entity Democratic Republic of the Congo - Temporal, Location and Identifiers			
Q974WD-4	Democratic Republic of the Congo (Q974)	instance of (P31)	sovereign state (Q3624078)
Q974WD-5	Democratic Republic of the Congo (Q974)	instance of (P31)	country (Q6256)
Q974WD-6	Democratic Republic of the Congo (Q974)	GeoNames ID (P1566)	203312
Q974WD-7	Q974WD-6	ckg:Determines (ckgr8)	Entity Identifier (ckgId)
Q974WD-8	Democratic Republic of the Congo (Q974)	ISO 3166-1 alpha-3 code (P298)	COD
Q974WD-9	Q974WD-8	ckg:Determines (ckgr8)	Entity Identifier (ckgId)
Q974WD-10	Democratic Republic of the Congo (Q974)	country (P17)	Democratic Republic of the Congo (Q974)
Q974WD-11	Q974WD-10	ckg:Context Type (ckgr9)	Location (ckgL1)
Q974WD-12	Democratic Republic of the Congo (Q974)	continent (P30)	Africa (Q15)
Q974WD-13	Q974WD-12	ckg:Context Type (ckgr9)	Location (ckgL1)
Q974WD-14	Democratic Republic of the Congo (Q974)	inception (P571)	^1960-06-30T00:00:00Z/11
Q974WD-15	Q974WD-14	ckg:Context Type (ckgr9)	Temporal (ckgT1)

Table 6.13: Multiple Answers for CQ0: Q974

(Q155) has replaced sometime. Based on this list of entities, the query can be executed to retrieve contextualized claims that represent the replacement and contextual information from each entity. After execution (see Brazil-p1365-path-all.tsv[4], its was observed that many recovered entities do not have a territorial overlapping relationship with current `Brazil (Q155)`, for example, countries located exclusively in Europe and Africa. The user can apply a filter on Location Context (`P30 continent`) to retrieve only entities with a spatial relationship across the same continent `South America (Q18)`.

CQ2: When was "*Mexico (Q96)*" founded/established?

There were five claims for the Mexico foundation date. Since the predicate `inception (P571)` is defined as `SINGLE-BEST VALUE constraint in WD`, it

[4]`https://github.com/versant2612/CKG_UseCases/tree/main/PoC`

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Claim about geoshape - point in time 2018-05-22 for Temporal Context			
Q32WD-1	Luxembourg (Q32)	geoshape (P3896)	Data:Luxembourg.map
Q32WD-2	Q32WD-1	point in time (P585)	^2018-05-22T00:00:00Z/9
Q32WD-3	Q32WD-2	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Contextualized Claim about geoshape - point in time 2020-02-14 for Temporal Context			
Q32WD-4	Luxembourg (Q32)	geoshape (P3896)	Data:LuxembourgCountry.map
Q32WD-5	Q32WD-4	point in time (P585)	^2020-02-14T00:00:00Z/9
Q32WD-6	**Q32WD-7**	**ckg:Inferred Context (ckgr3)**	**Current (ckgI1)**
Q32WD-7	Q32WD-5	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q32WD-8	Q32WD-4	reason for preferred rank (P7452)	unknown
Q32WD-9	Q32WD-8	ckg:Context Type (ckgr9)	Generic (ckgG1)

Table 6.14: Multiple Answers for CQ0: Q32

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Claim about replaces - Temporal and Provenance Context are unkown			
Q155WDK1a-1	Brazil (Q155)	replaces (P1365)	Empire of Brazil (Q217230)
Q155WDK1a-2	Q155WDK1a-1	start time (P580)	unknown
Q155WDK1a-3	Q155WDK1a-2	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-4	Q155WDK1a-1	point in time (P585)	unknown
Q155WDK1a-5	Q155WDK1a-4	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-6	Q155WDK1a-1	has cause (P828)	unknown
Q155WDK1a-7	Q155WDK1a-6	ckg:Context Type (ckgr9)	Provenance (ckgP1)

Table 6.15: Part of Contextualized Claims Answers for CQ1a: Q155

can be assumed that there is only one applicable to its respective context. In this case, retrieved provenance context is essential to support the Trust Layer to select the best claim. Besides the source of information, provenance clarifies how the information was obtained and calculated (for example criterion used). Table 6.17 shows preferred and deprecated contextualized claims for Mexico foundation. Deprecated claim is justified with the respective qualifier value but the preferred one does not have a preferred qualifier value, i.e,. it is unknown. Provenance context values for `criterion used (P1013)` and `statement is subject of (P805)` are available for both claims. This is an example where

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Entity Empire of Brazil - Temporal, Location and Identifiers			
Q155WDK1a-22	Empire of Brazil (Q217230)	instance of (P31)	sovereign state (Q3624078)
Q155WDK1a-23	Empire of Brazil (Q217230)	instance of (P31)	historical country (Q3024240)
Q155WDK1a-24	**Empire of Brazil (Q217230)**	**inception (P571)**	**^1822-09-07T00:00:00Z/11**
Q155WDK1a-25	Q155WDK1a-24	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-26	**Empire of Brazil (Q217230)**	**dissolved, abolished or demolished date (P576)**	**^1889-11-15T00:00:00Z/11**
Q155WDK1a-27	Q155WDK1a-26	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-28	Empire of Brazil (Q217230)	continent (P30)	South America (Q18)
Q155WDK1a-29	Q155WDK1a-28	ckg:Context Type (ckgr9)	Location (ckgL1)
Q155WDK1a-30	Empire of Brazil (Q217230)	country (P17)	Empire of Brazil (Q217230)
Q155WDK1a-31	Q155WDK1a-30	ckg:Context Type (ckgr9)	Location (ckgL1)

Table 6.16: Contextual information for entity Q217230 (CQ1a)

DOWA is represented since, according to the purpose of use, one or the other founding date can be considered true.

CQ3a: What is the **current** capital city of "*South Africa (Q258)*"?

Most countries have only one capital city at a time, but there are some exceptions. `South Africa (Q258)` has three capitals with their respective role, `P3831 object has role` qualifier is an additive qualifier for the n-ary relationship. Due to the KG incompleteness, it is not feasible to answer the exact temporal context, since `start time (P580)` is unknown. Table 6.18 shows the approximate answer retrieving all `capital (P36)` relationships. Generic context using `nature of statement (P5102)` is missing and was removed from table (See table B.5 for the entire result).

CQ3a: What is the **current** capital city of "*Israel (Q801)*"?

In the case of `Israel (Q801)`, only one capital relationship is recorded in WD and it is inferred as the current one by rule *I1*. But this claim is flagged as `disputed by (P1310);` see Q801WDK3-2 of Table 6.19. Generic context using the `nature of statement (P5102)` qualifier is missing and was removed from the table. As the relationship is subject of dispute, the other potential capital (the embassies are located in Tel Aviv city) must also have been recorded but is missing.

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized and deprecated Claim about inception with Provenance Context			
Q96WDK2e-1	Mexico (Q96)	inception (P571)	^1821-08-24T00:00:00Z/11
Q96WDK2e-3	Q96WDK2e-1	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q96WDK2e-4	Q96WDK2e-1	**criterion used (P1013)**	independence recognized by country from which it separated (Q60021702)
Q96WDK2e-5	Q96WDK2e-4	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q96WDK2e-6	Q96WDK2e-1	**statement is subject of (P805)**	Treaty of Córdoba (Q767450)
Q96WDK2e-7	Q96WDK2e-6	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q96WDK2e-8	**Q96WDK2e-1**	**reason for deprecated rank (P2241)**	**treaty not ratified (Q60021722)**
Q96WDK2e-9	Q96WDK2e-8	ckg:Context Type (ckgr9)	Generic (ckgG1)
Contextualized and preferred Claim about inception with Provenance Context			
Q96WDK2b-1	Mexico (Q96)	inception (P571)	^1810-09-16T00:00:00Z/11
Q96WDK2b-2	Q96WDK2b-1	separated from (P807)	Spain (Q29)
Q96WDK2b-3	Q96WDK2b-1	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q96WDK2b-4	Q96WDK2b-1	**criterion used (P1013)**	declaration of independence (Q1464916)
Q96WDK2b-5	Q96WDK2b-4	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q96WDK2b-6	Q96WDK2b-1	**statement is subject of (P805)**	Declaration of Independence of the Mexican Empire (Q1131780)
Q96WDK2b-7	Q96WDK2b-6	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q96WDK2b-8	Q96WDK2b-1	statement is subject of (P805)	Grito de Dolores (Q1145411)
Q96WDK2b-9	Q96WDK2b-8	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q96WDK2b-10	**Q96WDK2b-1**	**reason for preferred rank (P7452)**	**unknown**
Q96WDK2b-11	Q96WDK2b-10	ckg:Context Type (ckgr9)	Generic (ckgG1)

Table 6.17: Two (from five) contextualized Answers for CQ2: Q96

CQ4a: What position does the main administrative leader occupy in "*Canada (Q16)*", and who is the **current** leader?

Since the WD dump used is from June 2022, graph query K4a for the pair of predicates {`office held by head of state (P1906)` and `head of state (P35)`} returns that the current queen of Canada is Elisabeth II. Table 6.20 presents the answer. `Head of state (P35)` claim is indicated as preferred rank without the qualifier `reason for preferred rank (P7452)`.

Besides, the claim does not have the `end date (P582)` qualifier, so by rule *I1*, the temporal context is inferred as Current (Q16WDK4a-102). The

EDGE-ID	NODE1	EDGE	NODE2
Contextualized Claim about capital with role and Provenance - Temporal context unknown			
Q258WDK3-1	South Africa (Q258)	capital (P36)	Bloemfontein (Q37701)
Q258WDK3-2	**Q258WDK3-1**	**object has role (P3831)**	**judiciary (Q105985)**
Q258WDK3-5	**Q258WDK3-1**	**start time (P580)**	**unknown**
Q258WDK3-6	Q258WDK3-5	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q258WDK3-7	Q258WDK3-1	reference URL (P854)	https://www.nationsonline.org/oneworld/south_africa.htm
Q258WDK3-8	Q258WDK3-7	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Contextualized Claim about capital with role and Provenance - Temporal context unknown			
Q258WDK3-11	South Africa (Q258)	capital (P36)	Pretoria (Q3926)
Q258WDK3-12	**Q258WDK3-11**	**object has role (P3831)**	**executive branch (Q35798)**
Q258WDK3-15	**Q258WDK3-11**	**start time (P580)**	**unknown**
Q258WDK3-16	Q258WDK3-15	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q258WDK3-17	Q258WDK3-11	reference URL (P854)	https://www.nationsonline.org/oneworld/south_africa.htm
Q258WDK3-18	Q258WDK3-17	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Contextualized Claim about capital with role and Provenance - Temporal context unknown			
Q258WDK3-21	South Africa (Q258)	capital (P36)	Cape Town (Q5465)
Q258WDK3-22	**Q258WDK3-21**	**object has role (P3831)**	**legislature (Q11204)**
Q258WDK3-25	**Q258WDK3-21**	**start time (P580)**	**unknown**
Q258WDK3-26	Q258WDK3-25	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q258WDK3-27	Q258WDK3-21	reference URL (P854)	https://www.nationsonline.org/oneworld/south_africa.htm
Q258WDK3-28	Q258WDK3-27	ckg:Context Type (ckgr9)	Provenance (ckgP1)

Table 6.18: Contextualized Answers. for CQ3a: Q258

temporal correlation between the position and the leader was also inferred by CoaKG Query Engine based on Codomain Algebra; these claims have temporal overlap (Q16WDK4a-65). Generic context using the `nature of statement (P5102)` is missing and was removed from the table.

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized and Disputed Claim about capital with Provenance and Temporal context			
Q801WDK3-1	Israel (Q801)	capital (P36)	Jerusalem (Q1218)
Q801WDK3-2	**Q55WDK3-1**	**statement disputed by (P1310)**	**United Nations Security Council Resolution 478 (Q574386)**
Q801WDK3-3	Q55WDK3-2	ckg:Context Type (ckgr9)	Generic (ckgG1)
Q801WDK3-6	Q55WDK3-1	start time (P580)	^1949-01-01T00:00:00Z/9
Q801WDK3-7	Q55WDK3-6	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q801WDK3-8	Q55WDK3-1	reference URL (P854)	https://www.nationsonline.org/oneworld/israel.htm
Q801WDK3-9	Q55WDK3-8	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q801WDK3-10	**Q55WDK3-1**	**ckg:Inferred Context (ckgr3)**	**Current (ckgI1)**
Q801WDK3-11	Q55WDK3-10	ckg:Context Type (ckgr9)	Temporal (ckgT1)

Table 6.19: All contextualized Answers for CQ3a: Q801

However, Queen Elizabeth died in September 2022, so on the Canada WD page in October 2023, anyone can obtain that the current king is `Charles III of the United Kingdom (Q43274)` as shown in Figure 6.3. This claim is marked as preferred rank and has a `reason for preferred rank` qualifier value as `most recent value (Q71533355)`, which is the reason for preferred rank based on time. If WD dump were updated and K4a executed on it, rule *I1* would indicate this new claim as the Current temporal context.

Graph query K4a for the pair of predicates {`office held by head of government (P1313)`, `head of government (P6)`} returns that the head of government is `Prime Minister of Canada (Q839078)`. However, this claim is incompleted regarding temporal and provenance context, as shown in Table 6.21. Q16WDK4a-115 does not have the `end date (P582)` qualifier, so by rule *I1*, the temporal context is inferred as Current (Q16WDK4a-202). Nevertheless, the temporal correlation between the position and the leader cannot be inferred based on Codomain Algebra.

Every value for predicate or qualifier in WD can be either `no-value`, unknown, `some-value,` or an actual value. No-value is used to emphasize that the value really does not exist (yet) instead of being incomplete. Analyzing Canada WD page in October 2023, there are some claims whose `end date (P582)` qualifier has no-value. In order to reflect this characteristic from WD, a new semantic rule is defined as *I3*. Table 6.22 contains all temporal

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Claim for head of state - Current Position			
Q16WDK4a-1	Canada (Q16)	office held by head of state (P1906)	monarch of Canada (Q14931511)
Q16WDK4a-2	Q16WDK4a-1	start time (P580)	^1931-01-01T00:00:00Z/9
Q16WDK4a-3	Q16WDK4a-2	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q16WD4a-100	**Q16WDK4a-1**	**ckg:Inferred Context (ckgr3)**	**Current (ckgl1)**
Q16WDK4a-101	Q16WDK4a-100	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Contextualized Claim for head of state - Current Monarch			
Q16WDK4a-60	Canada (Q16)	head of state (P35)	Elizabeth II (Q9682)
Q16WDK4a-61	Q16WDK4a-60	start time (P580)	^1952-02-06T00:00:00Z/11
Q16WDK4a-62	Q16WDK4a-61	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q16WDK4a-65	**Q16WDK4a-60**	**Temporal Overlaps (ckgt1)**	**Q801WDK4a-1**
Q16WDK4a-66	Q16WDK4a-65	ckg:Inferred Context (ckgr3)	Temporal (ckgT1)
Q16WDK4a-102	**Q16WDK4a-60**	**ckg:Inferred Context (ckgr3)**	**Current (ckgl1)**
Q16WDK4a-103	Q16WDK4a-102	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q16WDK4a-104	**Q16WDK4a-60**	**reason for preferred rank (P7452)**	**unknown**
Q16WDK4a-105	Q16WDK4a-104	ckg:Context Type (ckgr9)	Generic (ckgG1)

Table 6.20: Contextualized Answers for CQ4a: Head Of State of Q16

interpretation rules written in kypher.

In Chapter 6, this thesis showed a Proof of Concept over a subset of WD. This data was converted into a CoaKG and Competency Questions were converted into contextualized graph queries within an exploratory search process. In the next chapter, related work will be discussed in comparison with the contributions of this thesis.

Figure 6.3: Canada (Q16) reflects the KG update by community

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Claim for head of government - Temporal and Provenance Context unknown for Position			
Q16WDK4a-68	Canada (Q16)	office held by head of government (P1313)	Prime Minister of Canada (Q839078)
Q16WDK4a-69	**Q16WDK4a-68**	**start time (P580)**	**unknown**
Q16WDK4a-70	Q16WDK4a-69	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q16WDK4a-71	**Q16WDK4a-68**	**statement is subject of (P805)**	**unknown**
Q16WDK4a-72	Q16WDK4a-71	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Contextualized Claim for head of government - Current Prime Minister			
Q16WDK4a-115	Canada (Q16)	head of government (P6)	Justin Trudeau (Q3099714)
Q16WDK4a-116	Q16WDK4a-115	start time (P580)	^2015-11-04T00:00:00Z/11
Q16WDK4a-117	Q16WDK4a-116	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q16WDK4a-118	Q16WDK4a-115	statement is subject of (P805)	unknown
Q16WDK4a-119	Q16WDK4a-118	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Q16WDK4a-202	**Q16WDK4a-115**	**ckg:Inferred Context (ckgr3)**	**Current (ckgl1)**
Q16WDK4a-203	Q16WDK4a-202	ckg:Context Type (ckgr9)	Temporal (ckgT1)

Table 6.21: Contextualized Answers for CQ4a: Head Of Government of Q16

ID	Body / IF / antecedent	Head / THEN / consequent
I1	–match '(v1)-[p1]→(), (p1)-[q1:P580]→(v2) ' –where 'NOT EXISTS {(p1)-[q2:P581]→(v3)}'	–return '"ckg:i1" as id, p1 as node1, "ckgr4" as label, "ckgl1" as node2
I2	–match '(v1)-[p1]→(), (p1)-[q1:P585]→(v2)' \ –match '(v1)-[p1]→(), (p1)-[q1:P585]→(v2)' –where 'v2 = $max_date ' –para max_date="<max_date>"	–return 'v1, p1.label, max(coalesce(v2, "-1")) as max_date' –return '"ckg:i1" as id, p1 as node1, "ckgr4" as label, "ckgl1" as node2
I3	–match '(v1)-[p1]→(), (p1)-[q1:P580]→(v2), (p1)-[q2:P581]→(v3) ' –where 'v3 = "no value"'	–return '"ckg:i1" as id, p1 as node1, "ckgr4" as label, "ckgl1" as node2

Table 6.22: Temporal Context Interpretation for WD Claims

7
Related Work

Related work, selected from the literature that is in some manner related to KG, Context, and Trust, are discussed in this chapter.

7.0.1
Decision Support Systems

In essence, Knowledge-Based Decision Support Systems (KBDSS) are specialized systems that combine expert knowledge with structured decision-making processes to provide valuable support for making informed decisions. The major components of a KBDSS include the knowledge base (KB), which stores expert knowledge in the form of facts, frames, semantic networks, ontologies, or rules, and the inference engine, which serves as the system's thinking and reasoning component [21]. Since KGs have become a valuable solution to the limitations of ontologies, offering scalability, timeliness, flexibility, search capabilities, and reasoning, Elnagar and Weistroffer, 2019 [49] proposed a framework to integrate DSS and Enterprise KG (EKG). However, they argued that the quality issues of KGs must be addressed before integration. Refinements should be applied to remove inconsistency, incorrect, conflicting, or contradictory data and deal with KG incompleteness.

The quality of a KG is crucial in DSS. Still, different from them, this thesis proposes that the Trust Layer should deal with potentially inconsistent, conflicting, and contradictory data using context information. CoaKG, under the DOWA, represents different perspectives that can be applied to different decision-making scenarios. The Knowledge Layer does not resolve apparent conflicts among data since the Trust Layer should handle this according to user and task requirements. Conflict resolution in the Knowledge Layer can introduce bias and restrict the user's decision-making power.

7.0.2
KG Exploratory Search

Regarding exploratory search applications, it is crucial to consider that users often need to become more familiar with the KG structure, actual contents, and the domain of the information they are exploring. Solutions to address these challenges have been proposed based on approximate methods, query expansion, and query refinement techniques.

TriniT is a KG exploratory querying system [62] that addresses vocabulary mismatch problems using rules for query relaxation. Users are often unfamiliar with the KG structure and labels of entities, classes, and relations. Such a phenomenon makes it difficult for users to formulate proper graph queries. Their relaxation rules translate the original triple pattern into a semantic similar one that can be answered. Besides, TriniT also treats KG incompleteness at query time using Open Information Extraction (OIE) tools to extract triples from external sources. The eXtend KG (XKG) comprises the original KG and new triples associated with their provenance context. Although the new triples have lower confidence than triples from the original KG, both contribute to the final answer.

The CoaKG Query Engine also considers that the users can be unfamiliar with the KG structure and the information needed. Graph queries generated by the user or system interface may need to be completed regarding context. The original BGP or CGP is evaluated regarding entity types, predicates, and qualifiers against the context mappings C and can be expanded to retrieve context information. Codomain Algebra for context values can also modify the original query K when the graph pattern is disconnected to extract implied relations. KG incompleteness regarding contextual information is addressed through semantic rules I and Codomain Algebra, inferring implicit context.

Exploratory search using Hierarchical KGs and document corpus can be found in [63]. The process begins with a keyword query that filters a document corpus (IR). Then, KG generation techniques such as Open Information Extraction (OIE), Entity Extraction, and Relation Extraction are automatically used to build a KG based on this corpus. Finally, KG Profiling and Summarization techniques are employed to construct a hierarchical view of the KG. The hierarchical KG (HKGs), a multi-layer extension to knowledge graphs, is an intermediate result of the search process composed of three Layers: The higher Layer as Collections (Clusterization), Middle Layer as Central Entities (entities with higher node degree) and Lower Layer as Triples (nodes and edges). The authors noted that hierarchies provide better sensemaking for searchers new to a topic area by structuring the information space. However, information extraction errors negatively impact the representation of the entity-relationship tuples.

CoaKGDSS also benefits from a multi-layer exploratory search navigation perspective regarding contextual information. This research used a multi-layer graph data structure to represent the CoaKGs within the Knowledge Layer, although reification of RDF triples could also be applied. Claims are in the first layer, and their context information is in the layer above. All Contextualized Answers retrieve at least the first and second layer Si that match the input query K. More contextual information can be retrieved as the next layer is explored. However, CoaKG is a materialized KG, and if OIE is used to generate its claims from some document corpus, provenance contextual information must be recorded and retrieved at query time to be evaluated by the Trust Layer.

Another approach found in the literature involves interactive graph query expansion over KGs [61]. Sample query results are used as input to generate the top k-most relevant expansions. The original query input is complemented with edges using their Bag-of-Labels Model for Graph definition inspired by how language models expand keyword queries. Two unsupervised methods (Labels and pseudo-relevance feedback) rank query expansions based on estimated relevance.

Unlike this work, CoaKG Query Engine expands a graph query K with context information, adding qualifiers for claims and predicates for entities based on the context mappings the KG Engineer added to the KG. All Contextualized Answers. are retrieved and sent to the Trust Layer, which can apply trust policies and rules to rank or filter the contextualized claims and entities according to user and task context of interest.

7.0.3
Contextualized Knowledge Representation and Reasoning

Regarding knowledge context representation and manipulation of triple-based KGs, Homola et al., 2010 [64] proposed contextualized knowledge repository (CKR). Besides individuals, concepts, properties, and their relations, context modeling adds a new dimension to knowledge engineering concerned with context type definitions and how to "*split*" the knowledge between these contexts. A CKR is a pair K = <D, C> where C represents the set of contexts, and D is a Description Logic (DL) KB.

CKR adopts the theoretical perspective of context as a multidimensional space and the metaphor of context as a box that contains triples associated with the same context. The context can be of the primitive type when it has fixed values in the dimensions or context classes building a hierarchy. Based on this definition and inspired by traditional online analytical processing (OLAP), Sack et al., 2021 [65] adapted the cube model and OLAP query operations, such as merge and roll-up, to perform analysis over KGs.

Regarding Ontologies, [54] defined contextualized entities from Contextualized Ontologies (CO) composed of three parts: the entity itself, its context, and a mapping function between the entity and its context. Since entity and context are ontologies, any entity can act as context for another entity, and a third entity can be a meta-context of the first context. The Algebra of Contextualized Entities defines three types of operations: Entity Integration, Context Integration, and Combined Integration (Relative Intersection and Collapsing Union). The Algebra was developed based on Category Theory to promote the integration and interoperability of heterogeneous entities using relationships (categorical morphisms) instead of instances (categorical objects). Contextualized Ontologies can be used to design Context-aware applications.

The CoaKG contemplates contextual information associated with claims, as in CKR, and with entities and concepts, as in Contextualized Ontologies (CO). Explicit contextual information is part of the KG, and context mappings specify how it is represented as qualifiers and predicates. Additional implicit context can be inferred using rules and Codomain Algebra at query time or in the KG Engineering phase. According to each context type and its context value data types, a specialized Codomain Algebra can be configured to evaluate relationships between context values using standard operations.

Aljalbout, Sahar [66] deals with contextual knowledge representation and reasoning on ontologies and property graphs. The author divided the qualifiers into two types of context: Validity Context, covering Temporality and Location, and Additional Context, for other qualifiers such as Provenance. This last group also considered Causality, Sequence, and Annotations (as generic). According to the author, the Additional Context qualifiers do not interfere with the semantics, that is, in interpreting the validity of the claim they are associated with, but rather complement this claim.

Comparing the examples used in the CoaKG approach, the qualifiers classified as Sequence (*replaced, replaced by, follows, followed by*) could be obtained by inferring the temporal relationship between time-dependent claims through Temporal Codomain Algebra. Still, Sequence qualifiers could be considered a new type of context for those who assign an order to list values, for example, authors of an article. Causality Qualifiers in the CoaKG approach (*has caused, end cause*) are considered Provenance since the context values explain the phenomenon reflected in the claim. Provenance in decision-making scenarios interferes with the interpretation of the claim since trust policies can be specified depending on the reputation of the sources, the precision of some calculation and data acquisition methods, etc.

In the case of WD, a Generic context was also specified to contemplate preferences, disputes and explicit disputes. CoaKG does not consider that all qualifiers have the goal of contextualizing the claims. However, based on the definition of Context from [3]: "*By context, we herein refer to the scope of truth, and thus talk about the context in which some data are held to be true*", every qualifier that performs this function must have its mappings added its value retrieved together with the claims and entities that it contextualizes.

In the same thesis of Aljalbout, Saha [66] one of the contributions is a Context Algebra using OWL (OWL-C) that focuses on Logical Correctness and only applies to temporal and spatial (Validity Context). This algebra is applicable to evaluate and correct logical inconsistencies in Contextualized Ontologies. However, due to the limits of applying OWL-C to KGs like WD, which have other categories of qualifiers, a second contribution is proposed: many-sorted logic (MSL). A logic in first-order language (FOL) where the whole is divided into subsets called sorts. In this way, the other types of context can have their operations defined and allow reasoning, validation of consistencies, and removal of redundancies currently not included in WD ontology. Although inconsistencies and redundancies should be treated to enhance KG quality, it is important to consider that CoaKG operates under the DOWA where potentially contradictory claims, as much contextualized as possible, must coexist to enable claims trustworthiness be evaluated by the Trust Layer based on trust policies.

7.0.4
Context-Aware KG Applications and Tasks

The KG fact contextualization task was addressed by Voskarides et al., 2018 [67] using the neural fact contextualization method (NFCM). This task consists of giving a KG fact, represented by a triple, composed of two entities (s an t) and a relation that connects them, retrieving additional and useful facts (triples) from the KG relevant to the input fact. Candidate facts from the broader neighborhood of the two entities, s and t, are selected and ranked using a supervised learning-to-rank model. Automatically learned features are combined with a set of hand-crafted features for the ranking model. In this work, contextualization is about retrieving relevant additional facts, and the authors do not mention qualifiers. On the other hand, for the Knowledge Layer, context has a different meaning and purpose. Contextualization corresponds to adding qualifiers for claims and predicates to entity types according to mappings previously defined by a KG Engineer. Contextual information is part of the neighborhood of claims and entities but it has explicit semantics.

Temporal reasoning TempQA-WD [57] is a KGQA initiative that uses WD as a data source. TempQA-WD was developed based on some questions from KGQA TempQuestions focused on temporal questions that require reasoning about points and intervals in time. Temporal questions are annotated with corresponding SPARQL queries generated using WD reification, temporal qualifiers, and predicates. The questions are grouped by complexity into three categories: simple (no temporal reasoning), medium (two events with temporal reasoning), and complex (two or more events with temporal reasoning and additional conceptual reasoning). Questions and queries from TempQA-WD are complete regarding Temporal Context since the context is explicitly specified. In the CoaKG query approach, these queries can also be completed with other types of context that are not explicitly stated in the questions. Other contexts present in the KG, either explicitly or inferred through rules, are necessary to interpret the answers correctly.

KGs have been used in various tasks, and some recent works about recommendation systems (RS) incorporate KGs to alleviate the item cold-start and sparsity problems. However, according to Ge et al., 2023 [68], KG-based tasks, applications, and state-of-the-art KGR methods need to consider that real-world KGs inevitably sometimes have a large amount of untrustworthy information. The authors developed a trustworthiness-aware knowledge graph representation (KGR) method called TrustRec.

TrustRec leverages internal structures of KGs to calculate a trustworthiness estimator, which gives the degree of triples' certainty and integrates it into noise-tolerant KGR and item representations for RS. Similarly, the Trust Layer, included in the architecture of CoaKG4DSS, can potentially compute a trust indicator using the contextualized answers retrieved from the Knowledge Layer. This computation may consider both the KG structure and the context meta-information. However, the Trust Layer was not addressed within the scope of this thesis, being indicated as future work. Alternatively, this trust indicator can be calculated during the KG generation process and represented as a Provenance meta-information on each claim and entity. This way, it would be part of the information the Knowledge Layer will provide in its response to the Trust Layer.

This chapter discussed and compared this work with others that deal with Decision-making, KG context-aware tasks, exploratory search, context representation and reasoning. Finally, chapter 8 concludes the thesis by discussing results, summarizing contributions, and pointing to future work about CoaKG.

8
Conclusion & Future Work

8.0.1
Summary

This thesis was developed based on a general problem statement and an instance problem statement that guided the research about KG usage in Decision-making scenarios. The research considered some premises: (i) Graph queries can be contextually incomplete; (ii) KGs can be contextually incomplete in a general sense and also regarding contextual information; (iii) implicit context can be extracted from KGs based on semantic rules and context values operations; (iv) CoaKG under DOWA; and (v) contextual information already available in KGs should be explicitly identified to enable a more informed decision by users regarding trust.

A KG can be built from scratch, and KG Engineering should know the importance of context representation to answer its competency questions. Binary and n-ary claims and contextual information are information already available in a standard Knowledge Graph (KG) *H*. Our approach added context mappings *C* to represent contextual types and transform them on a Context Augmented Knowledge Graph (CoaKG) *H'*. CoaKG definition and schema characterization are necessary to model explicit contextual information and extract implicit contextual information from a standard KG.

A graph query *K*, issued to at least partially satisfy an information need, can be completed or incomplete in context specification. In case of incomplete queries, the context mappings *the proposed CoaKG Query Engine uses C* to expand *K* to retrieve All Contextualized Answers *A* from *H'*. Besides, CoaKG *H'* can also be incomplete in terms of contextual information, and the context mappings *C* expands answers *A*, indicating where contextual information is missing and inferring implicit contextual information based on semantic rules *I* and Codomain Algebra.

Contextual mappings can be added to an existing KG to become a CoaKG. KG Profiling and Competency Questions can guide KG Engineering by discovering latent contextual information when no KG schema is available. Such a scenario is demonstrated using a Proof of Concept over WD. After transforming a subset of WD into a CoaKG, the proposal's effectiveness was evaluated within an exploratory search process.

The main novelties of our approach are:

First, this thesis assumes the KG is used to support a decision process by the user where one will decide which information can be trusted for its intended purposes. From this point of view, this thesis' approach seeks to provide All Contextualized Answers, considering the available context information as an essential input to support the decision-making process fully.

Subsequently, we claimed that any crowd-sourced KGs, or KGs constructed by integrating several different information sources, should operate under DOWA and be consumed through a Trust Layer.

Next, this thesis proposes the concept of Context Mappings to characterize contexts and to handle query semantic incompleteness. By leveraging context-oriented query expansion, context mappings contextually enrich the answers, considering the available context information.

Another aspect is the handling of incompleteness in the KG itself. The CoaKG Query Engine applies an answer expansion approach, considering all relevant context information to support the interpretation of the claims. Thus, the answers provided are contextually enriched by interpretation rules.

Lastly, our approach does not rely on an interactive flow with the user. Instead, a stateless approach is taken, where All Contextualized Answers. are determined based on the available context. This approach allows for efficient and flexible exploration without requiring repeated user input to extract relevant available information.

Considering these unique features, our proposed approach provided a more comprehensive and context-aware solution for exploratory search in decision-making scenarios, addressing the challenges of query and KG incompleteness while accommodating users' unfamiliarity with the KG structure, contents, and information domain.

A KG without a unique schema (schema-free or flexible schema) requires an extra effort to identify already available context to translate query patterns that semantically match the user question. The CoaKG Query Engine uses contextual mappings to indicate missing contextual information due to KG's incompleteness.

Effective contributions This research produced the contributions listed below:

1. An initial proposal for exploring contextual information in hyper-relational graphs.

DOS SANTOS, VERONICA; LIFSCHITZ, SÉRGIO . A semantic search approach for hyper-relational knowledge graphs. In: Anais Estendidos do Simpósio Brasileiro de Banco de Dados, 2021, Brasil. Anais Estendidos do XXXVI Simpósio Brasileiro de Banco de Dados (SBBD Estendido 2021), 2021. p. 106.

2. A proposal on effectively modeling and retrieving contextual information from a Knowledge Layer. The proposal includes a Context Augmented KG (CoaKG) definition, a Conceptual schema for Context Mappings, and a CoaKG Query Engine.

   DOS SANTOS, VERONICA ; HAEUSLER, EDWARD HERMANN ; SCHWABE, DANIEL ; LIFSCHITZ, Sergio . Context-Aware Knowledge Graphs Exploratory Search. In: Simpósio Brasileiro de Banco de Dados, 2023, Brasil. Anais do XXXVIII Simpósio Brasileiro de Banco de Dados (SBBD 2023), 2023. p. 360–365.

3. A Proof of Concept (PoC) to evaluate the effectiveness of the Knowledge Layer using a subset of Wikidata and an exploratory search process. Whereas it is a PoC, it is of sufficient size to indicate the technological solution may be scaled.

   A public GitHub repository is available at `https://github.com/versant2612/CKG_UseCases` with:

   (a) Script using KGTK toolkit and kypher queries to select the subset of claims and qualifiers about the Geopolitical History domain;
   (b) The result subset of WD using the kgtk model;
   (c) Script using KGTK toolkit and kypher queries to profile the subset;
   (d) The WD profiling results using the kgtk model;
   (e) The WD contextual mappings using the kgtk model;
   (f) Script using KGTK toolkit and kypher queries of translated Competency Questions; and
   (g) The exploratory results.

4. Dual Open World Assumption (DOWA) definition.

5. A framework for Context Augmented Knowledge Graphs-based Decision Support Systems (CoaKGDSS) comprises the Decision Layer, the Trust Layer, and the Knowledge Layer that operates under a DOWA.

6. WD Trust profiling to assess how well WD supports the trust decision process implied when using its data. This work inspired the DOWA definition and CoaKGDSS design applicable to any crowd-sourced KGs, or KGs constructed by integrating several different information sources.

   DOS SANTOS, VERONICA ; SCHWABE, DANIEL ; LIFSCHITZ, Sergio. Can you trust Wikidata? Semantic Web Journal - Special Issue on Wikidata 2022. **UNDER REVIEW.** Available at https://www.semantic-web-journal.net/content/can-you-trust-wikidata-0

Additional contributions Publications from other related research areas were also produced as shown in Table 8.1:

8.0.2 Ongoing and Future Work

As an ongoing work, we are designing and will implement an experiment focusing on user perception of contextual information for trustworthiness in Decision-making, using a questionnaire. A preliminary analysis using LLM is described in Appendix I.

As future work, this thesis envisions:

1. CoaKGs need scalable and efficient querying since adding meta-information to claims significantly increases their volume. Moreover, inferring implicit contextual relationships at query time also requires an overhead of computation. The CoaKG Query Engine efficiency can be systematically evaluated using WD datasets with qualifiers and references. Regarding optimizations, the materialization of implicit context relationships that would be harmed at query time and data partitioning criteria considering context types and their values would be helpful.

2. Generating KG profiling on other open available KGs from Scholarly and Research domains such as Microsoft Academic Knowledge Graph (MAKG), ORKG Open Knowledge Research Graph, DBLP Semantic Graph, and Polyvocal KGs [51] to identify contextual information already available and transform them in CoaKG.

Big Data	COSTA, ROGÉRIO LUÍS DE C. ; MOREIRA, JOSÉ ; PINTOR, PAULO ; **DOS SANTOS, VERONICA** ; LIFSCHITZ, SÉRGIO . Data-driven Performance Tuning for Big Data Analytics Platforms. Big Data Research, v. 23, p. 100206, 2021.
Ontologies	PERCILIANO, LUCIANA DE SÁ SILVA ; **DOS SANTOS, VERONICA** ; BAIÃO, FERNANDA ; HAEUSLER, EDWARD HERMANN ; LIFSCHITZ, SÉRGIO ; ALMEIDA, ANA CAROLINA . Inferencing Relational Database Tuning Actions with OnDBTuning Ontology. In: Simpósio Brasileiro de Banco de Dados, 2021, Brasil. Anais do XXXVI Simpósio Brasileiro de Banco de Dados (SBBD 2021), 2021. p. 157.
Information Retrieval	SALGUEIRO, MARIANA D. A. ; **DOS SANTOS, VERONICA** ; RÊGO, ANDRÉ L. C. ; GUIMARÃES, DANIEL S. ; HAEUSLER, EDWARD H. ; SANTOS, JEFFERSON DE B. ; VILLAS, MARCOS V. ; LIFSCHITZ, SÉRGIO . Sistemas de Recuperação de Informações Aplicadas à Produções Acadêmicas. In: Anais Estendidos do Simpósio Brasileiro de Banco de Dados, 2021, Brasil. Anais Estendidos do XXXVI Simpósio Brasileiro de Banco de Dados (SBBD Estendido 2021), 2021. p. 43.
	SALGUEIRO, MARIANA D. A. ; **DOS SANTOS, VERONICA** ; RÊGO, ANDRÉ L. C. ; GUIMARÃES, DANIEL S. ; HAEUSLER, EDWARD H. ; DOS SANTOS, JEFFERSON B. ; VILLAS, MARCOS V. ; LIFSCHITZ, SÉRGIO . Quem@PUC - A tool to find researchers at PUC-Rio. In: Anais Estendidos do Simpósio Brasileiro de Banco de Dados, 2021, Brasil. Anais Estendidos do XXXVI Simpósio Brasileiro de Banco de Dados (SBBD Estendido 2021), 2021. p. 93.
	A. SALGUEIRO, MARIANA D. ; **DOS SANTOS, VERONICA** ; C. RÊGO, ANDRÉ L. ; GUIMARÃES, DANIEL S. ; SANTOS, JEFFERSON B. ; H. HAEUSLER, EDWARD ; V. VILLAS, MARCOS ; LIFSCHITZ, SÉRGIO . Searching for Researchers: an Ontology-based NoSQL Database System Approach and Practical Implementation. Journal of Information and Data Management - JIDM, v. 13, p. 538-550, 2022.
Social Networks	SALGUEIRO, MARIANA D. A. ; LIFSCHITZ, SÉRGIO ; HAEUSLER, EDWARD HERMANN ; **DOS SANTOS, VERÔNICA** ; HEINE, ALEXANDRE A. P. . A Study of Database Models for Social Network Analysis. In: Simpósio Brasileiro de Banco de Dados, 2022, Brasil. Anais do XXXVII Simpósio Brasileiro de Banco de Dados (SBBD 2022). Porto Alegre, RS, Brasil: SBC, 2022. p. 397-402.

Table 8.1: Publications of BioBD Research Group

3. Comparing CoaKG general graph algorithms results such as clustering before and after materializing implicit context relationships using context rules and Codomain Algebra.

4. KGs can incorporate guardrails into the LLM prompts used during the

answers generation process. Users can define the structure, type, and quality of LLM responses by implementing guardrails. These guards have the potential to enhance the reliability of the information generated by LLMs and align it more closely with domain-specific or context-specific constraints [9]. CoaKG could represent additional rules based on context types and mapping to validate LLM responses.

5. Evaluating Link Prediction methods specific for hyper-relational graphs such as HINGE [33] and STARE [34] to automatically complete the missing context information based on the mappings.

6. Trust policies and rules from the Trust Layer can be used to assess KG completeness quality metrics and guide cost-benefit strategies for manual (human) completion of KGs.

7. The design of the Trust Layer, supported by CoaKGs to retrieve information and context, to calculate trust indicators.

8. Extending rule mining approaches, such as AMIE [29], to adapt to hiper-relational and multi-layer graph structures to generate rules that capture required and allowed qualifiers constraints, helping to identify context mappings.

9. Adapting Graph query embeddings such as StarQE [38], for example, to answer queries computing similarity using context mappings and context types from a CoaKG.

10. LLMs have been used to translate natural language user questions into queries written in SQL and GQL with templates and examples to configure prompts. CoaKG context mappings can also be used as templates to expand the knowledge of LLMs and complete queries where the context is absent in natural language questions.

11. Users can learn new aspects about the domain through contextual information added to answers. This can impact users' information-seeking behavior and the exploratory search process. Users' experiments are necessary to evaluate such impact.

12. Incorporating provenance operations from Metadata Management for Data Lakes solutions [Megdiche et al., 2021] into the CoaKG Query Engine approach to enhance data lineage tracking and data discovery.

A
Evaluation

This appendix describes the evaluation design and some insights revealed by testing the questionnaire using LLM.

A.0.1
What specific relationships are being tested?

Based on the General problem statement mentioned in Chapter 3, another instance problem statement was formulated as below:

> Given that contextualized information supports trust in knowledge, how does contextual information influence users' knowledge reliance on decision-making?
> (Evaluation Problem Statement)

An experimental evaluation is designed using an online questionnaire (in Portuguese) to measure the relationship between contextualized information and trust for decision-making in three realistic scenarios.

A.0.2
How does the theory justify the prediction?

According to Smith and Rieh, 2019 [16], if contextual information is readily available for searchers during information seeking, it will engage users more actively and critically and positively influence users' confidence in the information presented. Such hypotheses motivate our experimental research design to measure how information-seeking behavior is affected by knowledge-context availability.

To consider a claim as fact and decide to act based on it, the user should rely on the contextual information that supports it, the context constraints of the task where that claim will be applied, and their trust policies (subjective evaluation) to enable Decision-Making. Aggregating explicit contextual information, when available, in query answers anticipates additional queries about WHEN it happened, WHERE it happened, HOW such claims were gathered, WHO made or registered the claims, and so on, reducing decision-making uncertainty. Context augmentation must support the users in identifying the scope within which a claim can be considered true and useful.

A.0.3 What is novel and not obvious about the hypotheses?

The mere presence of additional information in the answer from the neighborhood of entities and claims of then KG does not necessarily constitute only contextual information. To aid in assessing their credibility and scope, retrieving all entities and connections from this surrounding area to define context may not enhance the user's trust. Similarly, responses needing more contextualization can reduce confidence by failing to define this context. All Contextualized Answers add contextual information retrieved from explicitly represented information from KG and inferred using semantic rules and algebra operations specific to the context type of interest.

A.0.4 Research Evaluation Design

Measures and Metrics According to the previous hypothesis, the experiment will measure Decision-making trust level variation (dependent variable) based on the absence and presence of contextual information from single and multiple answers (control variables). The analysis of the research question enabled us to identify the PICO (Population, Intervention, Comparison, and Outcome) aspects to be evaluated in the experiment, as shown in Table A.1:

What POPULATION am I interested in? What POPULATION is affected by the intervention? Decision-making Users.
What is the INTERVENTION being applied? All contextualized {possible} answers
With what is the intervention being COMPARED? Non contextualized answers (as factual answers) - baseline
What is the expected OUTCOME/result from the intervention? Trust variation.

Table A.1: PICO

Method An ongoing work of this thesis is an experiment with 20 users describing four realistic decision-making scenarios, with one and multiple answers, and without (baseline) and with (intervention) contextual information. During the experiment, a questionnaire will collect users' answers to closed and open questions about the information provided for decision-making and their trustworthiness level in these conditions; see Table A.2 for the questionnaire template and variables.

Limitations As a limitation, the experiment design assumes that isolating users' previous knowledge about the domain is challenging. So, the results can't assure that the user's trust level of baseline answers is only based on the given answers. The questionnaire used fictional names for entities involved in the answers to reduce such weakness.

As with most qualitative research, this experiment is limited by time constraints, the human-labor approach, and its subjective nature.

Decision Making scenario description	
Baseline (A1) - standard KG	
ONE Answer WITHOUT contextual information - Source A	
A1-V1	User Decision (closed)
A1-V2	User Decision justification (open)
A1-V3	User trust level on answer (closed)
Our approach (B1) - CoaKG	
ONE contextualized answer - Source B	
B1-V1	User Decision (closed)
B1-V2	User Decision justification (open)
B1-V3	User trust level on answer (closed)
Comparability A1 x B1	
A1-B1-V4	User trust level change (A1→B1) justification (open)
A1-B1-V5	User preference for source A or B (closed)
Baseline (A2) - standard KG	
MULTIPLE Answer WITHOUT contextual information - Source A	
A2-V1	User Decision (closed)
A2-V2	User Decision justification (open)
A2-V3	User trust level on answer (closed)
Our approach (B2) - CoaKG	
MULTIPLE contextualized answers - Source B	
B2-V1	User Decision (closed)
B2-V2	User Decision justification (open)
B2-V3	User trust level on answer (closed)
Comparability A2 x B2	
A2-B2-V4	User trust level change (A2→B2) justification (open)
A2-B2-V5	User preference for source A or B (closed)

Table A.2: Research Questionnaire Template and Variables

Online Questionnaires Two questionnaires were created using Google Forms.

https://forms.gle/42tds3F3scFZb7YdA

https://forms.gle/aPxAvcz3BgyXr7P1A

Questionnaire test with ChatGPT We simulated the application of the questionnaires using LLM. OpenAI provides an API so that LLMs can be used in applications. A Google Collab notebook, available in the GitHub repository [1], using the OpenAI interface for ChatGPT version 4.0, LangChain framework, and model used is text-davinci-003 implements a dialog, based on textual content, through inputs and outputs in the prompt. This simulation aims to assess understanding of the questions and possible answers. It also helped to obtain insights into possible justifications for the decisions made.

Three questionnaires were configured:

Questionnaire 1) Data sources provide a single answer to the decision. The decision consists of whether or not to use the information to act.

2) Data sources provide multiple answers (two or more) to the decision:

Questionnaire 2.1) It has the option to use all answers to take action and,

Questionnaire 2.2) It does not have the option to use all answers to take action.

The model received the following general instructions in each questionnaire:

> *Please, before answering the questions, read the statement carefully.*
> *Try to put yourself in the shoes of someone who needs to make a decision in the scenario presented.*
> *These scenarios and responses are hypothetical, as fictitious names are being used for people, companies, and locations, but they are realistic.*
> *Please consider that both Source A and Source B may contain incorrect and incomplete information.*

The generated responses have the following characteristics:

1. In questionnaire 1 (one answer from Source A and B), the confidence level in the baseline was Medium or High, while in the intervention (contextualized answers), it was High or Very High.

2. In questionnaires 2.1 and 2.2, the confidence level in the baseline was Medium, while in the intervention, it was High or Very High.

3. In all questionnaires, confidence in the baseline was less than or equal to confidence in the intervention.

4. Source B, which provided contextualized answers, was preferred for consultation in an actual situation as it provided more detailed information than Source A.

[1]`https://github.com/versant2612/CKG_UseCases`

5. The justifications for the decisions in the baseline from questionnaire 2.2 contain hallucinations, as in the examples below. Hallucinations may be associated with a lack of context to justify a more restricted decision

Sim, eu instalaria a embaixada em Roaima. ***A capital de um país é geralmente a cidade mais importante e com maior população, o que significa que Roaima é mais provável de ser a capital do país Itanya.*** *Além disso, instalar a embaixada na capital do país pode ajudar a melhorar as relações diplomáticas entre os dois países.*

Minha decisão seria fazer contato com a Aquatop para a troca de titularidade. ***Isso porque, como a Fonte A mencionou Aquatop como a primeira opção, isso significa que é a mais confiável.*** *Além disso, como eu não conheço a cidade, é melhor começar com a opção mais confiável.*

Results Since this experiment will involve people, it will be submitted to the appreciation of the research ethics committee from PUC-Rio. After the questionnaire appliance, quantitative and qualitative analysis will be reported based on the collected information.

B
Complete Tables

This appendix contains the complete version of tables mentioned in Chapters 4, 5 and 6.

NODE1	EDGE	NODE2
Provenance (ckgP1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Provenance (ckgP1)	ckg:Represented By (ckgr1)	stated in (h4q3)
stated in (h4q3)	ckg:Contextualizes (ckgr2)	form of government (h4r1)
stated in (h4q3)	ckg:Contextualizes (ckgr2)	capital city of (h4r7)
stated in (h4q3)	ckg:Contextualizes (ckgr2)	geoshape (h4r2)
Provenance (ckgP1)	ckg:Represented By (ckgr1)	reference URL (h4p6)
reference URL (h4p6)	ckg:Contextualizes (ckgr2)	All (ckgl2)
Politics (ckgD1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Politics (ckgD1)	ckg:Represented By (ckgr1)	form of government (h4r1)
form of government (h4r1)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Location (ckgL1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Location (ckgL1)	ckg:Represented By (ckgr1)	geoshape (h4r2)
geoshape (h4r2)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Location (ckgL1)	ckg:Represented By (ckgr1)	part of (h4r3)
part of (h4r3)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Location (ckgL1)	ckg:Represented By (ckgr1)	local (h4q6)
local (h4q6)	ckg:Contextualizes (ckgr2)	proclaim (h4r8)

Table B.1: Complete Contextual Mappings of H4 - Provenance, Politics, and Location

NODE1	EDGE	NODE2
Temporal (ckgT1)	rdf:type (is_a)	(ckg:KnowledgeContext)
Temporal (ckgT1)	ckg:Inferred Context (ckgr3)	Current (ckgl1)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	start time (h4q1)
start time (h4q1)	ckg:Contextualizes (ckgr2)	capital city of (h4r7)
start time (h4q1)	ckg:Contextualizes (ckgr2)	Período Histórico (h4c2)
start time (h4q1)	ckg:Contextualizes (ckgr2)	form of government (h4r1)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	end time (h4q2)
end time (h4q2)	ckg:Contextualizes (ckgr2)	capital city of (h4r7)
end time (h4q2)	ckg:Contextualizes (ckgr2)	Período Histórico (h4c2)
end time (h4q2)	ckg:Contextualizes (ckgr2)	form of government (h4r1)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	inception (h4p3)
inception (h4p3)	ckg:Contextualizes (ckgr2)	part of (h4r3)
inception (h4p3)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	closure date (h4p4)
closure date (h4p4)	ckg:Contextualizes (ckgr2)	part of (h4r3)
closure date (h4p4)	ckg:Contextualizes (ckgr2)	geo-political unity (h4c1)
Temporal (ckgT1)	ckg:Represented By (ckgr1)	point in time (h4q4)
point in time (h4q4)	ckg:Contextualizes (ckgr2)	proclaim (h4r8)
point in time (h4q4)	ckg:Contextualizes (ckgr2)	image (h4p8)
point in time (h4q4)	ckg:Contextualizes (ckgr2)	geoshape (h4r2)

Table B.2: Complete Contextual Mappings of H4 - Temporal

ID	Predicate	Qualifier	Context Type	Status
CQ0	geoshape (P3896)	point in time (P585)	Temporal	Mandatory
		applies to part (P518)	Location	Optional
CQ1 CQ2 CQ3 CQ4	replaces (P1365) replaced by (P1366)	point in time (P585)	Temporal	Mandatory
		start time (P580)		
		end time (P582)	Temporal	Optional
		applies to part (P518)	Location	Optional
		valid in place (P3005)		
		main regulatory text (P92)	Provenance	Optional
		has cause (P828)		
		statement is subject of (P805)		
		has immediate cause (P1478)		
		statement supported by (P3680)		
CQ2	inception (P571)	latest date (P1326)	Temporal	Optional
		earliest date (P1319)		
		end time (P582)		
		refine date (P4241)		
		latest start date (P8555)		
		time period (P2348)		
		start time (P580)		
		statement is subject of (P805)	Provenance	Mandatory
		sourcing circumstances (P1480)	Provenance	Optional
		identity of subject in context (P4649)		
		has cause (P828)		
		criterion used (P1013)		
		foundational text (P457)		
		has immediate cause (P1478)		

continued on the next page

ID	Predicate	Qualifier	Context Type	Status
		nature of statement (P5102)		
		determination method (P459)		
		main regulatory text (P92)		
		applies to part (P518)	Location	Optional
		applies to jurisdiction (P1001)		
	dissolved, abolished or demolished date (P576)	statement is subject of (P805)	Provenance	Mandatory
		end cause (P1534)		
		sourcing circumstances (P1480)	Provenance	Optional
		has cause (P828)		
		nature of statement (P5102)		
		main regulatory text (P92)		
		criterion used (P1013)		
		statement supported by (P3680)		
		cause of destruction (P770)		
		applies to part (P518)	Location	Optional
		part of (P361)		
		earliest date (P1319)	Temporal	Optional
		latest date (P1326)		
		refine date (P4241)		
CQ3	capital (P36)	start time (P580)	Temporal	Mandatory
		end time (P582)	Temporal	Optional
		point in time (P585)		
		valid in period (P1264)		
		object has role (P3831)	Provenance	Mandatory
		nature of statement (P5102)		

continued on the next page

ID	**Predicate**	**Qualifier**	**Context Type**	**Status**
		determination method (P459)	Provenance	Optional
		has cause (P828)		
		statement is subject of (P805)		
		end cause (P1534)		
		statement supported by (P3680)		
		sourcing circumstances (P1480)		
		criterion used (P1013)		
CQ4	office held by head of government (P1313)	start time (P580)	Temporal	Mandatory
		end time (P582)	Temporal	Optional
		applies to jurisdiction (P1001)	Provenance	Optional
	office held by head of state (P1906)	start time (P580)	Temporal	Mandatory
		end time (P582)	Temporal	Optional
		statement is subject of (P805)	Provenance	Optional
		object has role (P3831)		
		location (P276)	Location	Optional
	head of state (P35)	**start time (P580)**	**Temporal**	**Mandatory**
		end time (P582)	Temporal	Optional
		statement is subject of (P805)	Provenance	Optional
		subject has role (P2868)		
		determination method (P459)		
		end cause (P1534)		
		location (P276)	Location	Optional
	head of government (P6)	start time (P580)	Temporal	Mandatory
		end time (P582)	Temporal	Optional
		statement is subject of (P805)	Provenance	Optional
		sourcing circumstances (P1480)		

continued on the next page

ID	Predicate	Qualifier	Context Type	Status
		end cause (P1534)		
		determination method (P459)		

Table B.3: Complete Context Mappings of Predicates from CQs

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Contextualized Claim about replaces - Temporal and Provenance Context are unkown			
Q155WDK1a-1	Brazil (Q155)	replaces (P1365)	Empire of Brazil (Q217230)
Q155WDK1a-2	Q155WDK1a-1	start time (P580)	unknown
Q155WDK1a-3	Q155WDK1a-2	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-4	Q155WDK1a-1	point in time (P585)	unknown
Q155WDK1a-5	Q155WDK1a-4	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-6	Q155WDK1a-1	has cause (P828)	unknown
Q155WDK1a-7	Q155WDK1a-6	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Contextualized Entity Brazil - Temporal, Location and Identifiers			
Q155WDK1a-8	Brazil (Q155)	GeoNames ID (P1566)	3469034
Q155WDK1a-9	Q155WDK1a-8	ckg:Determines (ckgr8)	Entity Identifier (ckgId)
Q155WDK1a-10	Brazil (Q155)	ISO 3166-1 alpha-3 code (P298)	BRA
Q155WDK1a-11	Q155WDK1a-10	ckg:Determines (ckgr8)	Entity Identifier (ckgId)
Q155WDK1a-12	Brazil (Q155)	instance of (P31)	sovereign state (Q3624078)
Q155WDK1a-13	Brazil (Q155)	instance of (P31)	country (Q6256)
Q155WDK1a-14	Brazil (Q155)	country (P17)	Brazil (Q155)
Q155WDK1a-15	Q155WDK1a-14	ckg:Context Type (ckgr9)	Location (ckgL1)
Q155WDK1a-16	**Brazil (Q155)**	**country (P17)**	**Portuguese Empire (Q200464)**
Q155WDK1a-17	Q155WDK1a-16	ckg:Context Type (ckgr9)	Location (ckgL1)
Q155WDK1a-18	Brazil (Q155)	continent (P30)	South America (Q18)
Q155WDK1a-19	Q155WDK1a-18	ckg:Context Type (ckgr9)	Location (ckgL1)
Q155WDK1a-20	**Brazil (Q155)**	**inception (P571)**	**^1822-09-07T00:00:00Z/11**
Q155WDK1a-21	Q155WDK1a-20	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Contextualized Entity Empire of Brazil - Temporal, Location and Identifiers			
Q155WDK1a-22	Empire of Brazil (Q217230)	instance of (P31)	sovereign state (Q3624078)

continued on the next page

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Q155WDK1a-23	Empire of Brazil (Q217230)	instance of (P31)	historical country (Q3024240)
Q155WDK1a-24	**Empire of Brazil (Q217230)**	**inception (P571)**	**^1822-09-07T00:00:00Z/11**
Q155WDK1a-25	Q155WDK1a-24	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-26	**Empire of Brazil (Q217230)**	**dissolved, abolished or demolished date (P576)**	**^1889-11-15T00:00:00Z/11**
Q155WDK1a-27	Q155WDK1a-26	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q155WDK1a-28	Empire of Brazil (Q217230)	continent (P30)	South America (Q18)
Q155WDK1a-29	Q155WDK1a-28	ckg:Context Type (ckgr9)	Location (ckgL1)
Q155WDK1a-30	Empire of Brazil (Q217230)	country (P17)	Empire of Brazil (Q217230)
Q155WDK1a-31	Q155WDK1a-30	ckg:Context Type (ckgr9)	Location (ckgL1)

Table B.4: All contextualized Answers for CQ1a: Q155

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Q258WDK3-1	South Africa (Q258)	capital (P36)	Bloemfontein (Q37701)
Q258WDK3-2	**Q258WDK3-1**	**object has role (P3831)**	**judiciary (Q105985)**
Q258WDK3-3	Q258WDK3-1	nature of statement (P5102)	unknown
Q258WDK3-4	Q258WDK3-3	ckg:Context Type (ckgr9)	Generic (ckgG1)
Q258WDK3-5	**Q258WDK3-1**	**start time (P580)**	**unknown**
Q258WDK3-6	Q258WDK3-5	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q258WDK3-7	Q258WDK3-1	reference URL (P854)	`https://www.nationsonline.org/oneworld/south_africa.htm`
Q258WDK3-8	Q258WDK3-7	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Contextualized Claim about capital with role and Provenance - Temporal context unknown			
Q258WDK3-11	South Africa (Q258)	capital (P36)	Pretoria (Q3926)
Q258WDK3-12	**Q258WDK3-11**	**object has role (P3831)**	**executive branch (Q35798)**
Q258WDK3-13	Q258WDK3-11	nature of statement (P5102)	unknown
Q258WDK3-14	Q258WDK3-13	ckg:Context Type (ckgr9)	Generic (ckgG1)
Q258WDK3-15	**Q258WDK3-11**	**start time (P580)**	**unknown**
Q258WDK3-16	Q258WDK3-15	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q258WDK3-17	Q258WDK3-11	reference URL (P854)	`https://www.nationsonline.org/oneworld/south_africa.htm`
Q258WDK3-18	Q258WDK3-17	ckg:Context Type (ckgr9)	Provenance (ckgP1)
Contextualized Claim about capital with role and Provenance - Temporal context unknown			
Q258WDK3-21	South Africa (Q258)	capital (P36)	Cape Town (Q5465)

continued on the next page

EDGE-ID	NODE1	EDGE-LABEL	NODE2
Q258WDK3-22	**Q258WDK3-21**	**object has role (P3831)**	**legislature (Q11204)**
Q258WDK3-23	Q258WDK3-21	nature of statement (P5102)	unknown
Q258WDK3-24	Q258WDK3-23	ckg:Context Type (ckgr9)	Generic (ckgG1)
Q258WDK3-25	**Q258WDK3-21**	**start time (P580)**	**unknown**
Q258WDK3-26	Q258WDK3-25	ckg:Context Type (ckgr9)	Temporal (ckgT1)
Q258WDK3-27	Q258WDK3-21	reference URL (P854)	`https://www.nationsonline.org/oneworld/south_africa.htm`
Q258WDK3-28	Q258WDK3-27	ckg:Context Type (ckgr9)	Provenance (ckgP1)

Table B.5: All Contextualized Answers for CQ3a: Q258

Bibliography

[1] CASE, D.. **Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior**. Library and information science. Academic Press, 2002.

[2] MARCHIONINI, G.. **Exploratory search: from finding to understanding**. Commun. ACM, 49(4):41–46, 2006.

[3] HOGAN, A.; BLOMQVIST, E.; COCHEZ, M.; D'AMATO, C.; DE MELO, G.; GUTIERREZ, C.; KIRRANE, S.; GAYO, J. E. L.; NAVIGLI, R.; NEUMAIER, S.; NGOMO, A. N.; POLLERES, A.; RASHID, S. M.; RULA, A.; SCHMELZEISEN, L.; SEQUEDA, J. F.; STAAB, S. ; ZIMMERMANN, A.. **Knowledge graphs**. ACM Comput. Surv., 54(4):71:1–71:37, 2022.

[4] CAPPS, J.. **The Pragmatic Theory of Truth**. In: Zalta, E. N.; Nodelman, U., editors, THE Stanford ENCYCLOPEDIA OF PHILOSOPHY. Metaphysics Research Lab, Stanford University, Summer 2023 edition, 2023.

[5] GERCK, E.. **Toward real-world models of trust: Reliance on received information**. Report MCWG-Jan22-1998, 1998.

[6] SCHWABE, D.. **Trust and privacy in knowledge graphs**. In: COMPANION PROCEEDINGS OF THE 2019 WORLD WIDE WEB CONFERENCE, WWW '19, p. 722–728, New York, NY, USA, 2019. Association for Computing Machinery.

[7] SCHWABE, D.; LAUFER, C. ; CASANOVAS, P.. **Knowledge graphs: Trust, privacy, and transparency from a legal governance approach**. Law in Context. A Socio-legal Journal, 37(1):24–41, Dec. 2020.

[8] SHAH, C.; BENDER, E. M.. **Situating search**. In: Elsweiler, D., editor, CHIIR '22: ACM SIGIR CONFERENCE ON HUMAN INFORMATION INTERACTION AND RETRIEVAL, REGENSBURG, GERMANY, MARCH 14 - 18, 2022, p. 221–232. ACM, 2022.

[9] PAN, J. Z.; RAZNIEWSKI, S.; KALO, J.; SINGHANIA, S.; CHEN, J.; DIETZE, S.; JABEEN, H.; OMELIYANENKO, J.; ZHANG, W.; LISSANDRINI, M.; BISWAS, R.; DE MELO, G.; BONIFATI, A.; VAKAJ, E.; DRAGONI,

M. ; GRAUX, D.. **Large language models and knowledge graphs: Opportunities and challenges**. CoRR, abs/2308.06374, 2023.

[10] WHITE, R. W.; ROTH, R. A.. **Exploratory Search: Beyond the Query-Response Paradigm**. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.

[11] WEIKUM, G.. **Knowledge graphs 2021: A data odyssey**. Proc. VLDB Endow., 14(12):3233–3238, 2021.

[12] CHAUDHRI, V. K.; BARU, C. K.; CHITTAR, N.; DONG, X. L.; GENESERETH, M. R.; HENDLER, J. A.; KALYANPUR, A.; LENAT, D. B.; SEQUEDA, J.; VRANDECIC, D. ; WANG, K.. **Knowledge graphs: Introduction, history and, perspectives**. AI Mag., 43(1):17–29, 2022.

[13] HOFER, M.; OBRACZKA, D.; SAEEDI, A.; KÖPCKE, H. ; RAHM, E.. **Construction of knowledge graphs: State and challenges**. CoRR, abs/2302.11509, 2023.

[14] VRANDECIC, D.. **Wikidata: a new platform for collaborative data collection**. In: Mille, A.; Gandon, F.; Misselis, J.; Rabinovich, M. ; Staab, S., editors, PROCEEDINGS OF THE 21ST WORLD WIDE WEB CONFERENCE, WWW 2012, LYON, FRANCE, APRIL 16-20, 2012 (COMPANION VOLUME), p. 1063–1064. ACM, 2012.

[15] ANGLES, R.; HOGAN, A.; LASSILA, O.; ROJAS, C.; SCHWABE, D.; SZEKELY, P. A. ; VRGOC, D.. **Multilayer graphs: a unified data model for graph databases**. In: Kalavri, V.; Salihoglu, S., editors, GRADES-NDA '22: PROCEEDINGS OF THE 5TH ACM SIGMOD JOINT INTERNATIONAL WORKSHOP ON GRAPH DATA MANAGEMENT EXPERIENCES & SYSTEMS (GRADES) AND NETWORK DATA ANALYTICS (NDA), PHILADELPHIA, PENNSYLVANIA, USA, 12 JUNE 2022, p. 11:1–11:6. ACM, 2022.

[16] SMITH, C. L.; RIEH, S. Y.. **Knowledge-context in search systems: Toward information-literate actions**. In: Azzopardi, L.; Halvey, M.; Ruthven, I.; Joho, H.; Murdock, V. ; Qvarfordt, P., editors, PROCEEDINGS OF THE 2019 CONFERENCE ON HUMAN INFORMATION INTERACTION AND RETRIEVAL, CHIIR 2019, GLASGOW, SCOTLAND, UK, MARCH 10-14, 2019, p. 55–62. ACM, 2019.

[17] SCHURZ, G.. **Patterns of abduction**. Synthese, 164(2):201–234, 2008.

[18] DOUVEN, I.. **Abduction**. In: Zalta, E. N., editor, THE Stanford ENCYCLOPEDIA OF PHILOSOPHY. Metaphysics Research Lab, Stanford University, Summer 2021 edition, 2021.

[19] MACBRIDE, F.. **Truthmakers**. In: Zalta, E. N.; Nodelman, U., editors, THE Stanford ENCYCLOPEDIA OF PHILOSOPHY. Metaphysics Research Lab, Stanford University, Fall 2022 edition, 2022.

[20] STEELE, K.; STEFÁNSSON, H. O.. **Decision Theory**. In: Zalta, E. N., editor, THE Stanford ENCYCLOPEDIA OF PHILOSOPHY. Metaphysics Research Lab, Stanford University, Winter 2020 edition, 2020.

[21] SUPRIADI, L. S. R.; SUI PHENG, L.. **Knowledge Based Decision Support System (KBDSS)**, p. 155–174. Springer Singapore, Singapore, 2018.

[22] LUCONI, F. L.; MALONE, T. W. ; SCOTT MORTON, M. S.. **Expert systems: The next challenge for managers**. Sloan Manage. Rev., 27(4):3–14, jun 1986.

[22] BRODIE, M. L.; MYLOPOULOS, J.. **Knowledge bases vs. databases**. In: Brodie, M. L.; Mylopoulos, J., editors, ON KNOWLEDGE BASE MANAGEMENT SYSTEMS: INTEGRATING ARTIFICIAL INTELLIGENCE AND DATABASE TECHNOLOGIES, BOOK RESULTING FROM THE ISLAMORADA WORKSHOP 1985 (ISLAMORADA, FL, USA), Topics in Information Systems, p. 83–86. Springer, 1985.

[23] WEIKUM, G.; DONG, X. L.; RAZNIEWSKI, S. ; SUCHANEK, F. M.. **Machine knowledge: Creation and curation of comprehensive knowledge bases**. Found. Trends Databases, 10(2-4):108–490, 2021.

[24] EHRLINGER, L.; WÖSS, W.. **Towards a definition of knowledge graphs**. In: Martin, M.; Cuquet, M. ; Folmer, E., editors, JOINT PROCEEDINGS OF THE POSTERS AND DEMOS TRACK OF THE 12TH INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS - SEMANTICS2016 AND THE 1ST INTERNATIONAL WORKSHOP ON SEMANTIC CHANGE & EVOLVING SEMANTICS (SUCCESS'16) CO-LOCATED WITH THE 12TH INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS (SEMANTICS 2016), LEIPZIG, GERMANY, SEPTEMBER 12-15, 2016, volumen 1695 de **CEUR Workshop Proceedings**. CEUR-WS.org, 2016.

[25] SRI NURDIATI, S.; HOEDE, C.. **25 years development of knowledge graph theory: the results and the challenge**. Número 2/1876 em Memorandum. University of Twente, Netherlands, sep 2008.

[26] MARX, M.; KRÖTZSCH, M. ; THOST, V.. **Logic on MARS: ontologies for generalised property graphs**. In: Sierra, C., editor, PROCEEDINGS OF THE TWENTY-SIXTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, IJCAI 2017, MELBOURNE, AUSTRALIA, AUGUST 19-25, 2017, p. 1188–1194. ijcai.org, 2017.

[26] SIKOS, L. F.; PHILP, D.. **Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs**. Data Sci. Eng., 5(3):293–316, 2020.

[27] BLEIHOLDER, J.; NAUMANN, F.. **Data fusion**. ACM Comput. Surv., 41(1):1:1–1:41, 2008.

[28] VESELI, B.; SINGHANIA, S.; RAZNIEWSKI, S. ; WEIKUM, G.. **Evaluating language models for knowledge base completion**. In: Pesquita, C.; Jiménez-Ruiz, E.; McCusker, J. P.; Faria, D.; Dragoni, M.; Dimou, A.; Troncy, R. ; Hertling, S., editors, THE SEMANTIC WEB - 20TH INTERNATIONAL CONFERENCE, ESWC 2023, HERSONISSOS, CRETE, GREECE, MAY 28 - JUNE 1, 2023, PROCEEDINGS, volumen 13870 de **Lecture Notes in Computer Science**, p. 227–243. Springer, 2023.

[29] LAJUS, J.; GALÁRRAGA, L. ; SUCHANEK, F. M.. **Fast and exact rule mining with AMIE 3**. In: Harth, A.; Kirrane, S.; Ngomo, A. N.; Paulheim, H.; Rula, A.; Gentile, A. L.; Haase, P. ; Cochez, M., editors, THE SEMANTIC WEB - 17TH INTERNATIONAL CONFERENCE, ESWC 2020, HERAKLION, CRETE, GREECE, MAY 31-JUNE 4, 2020, PROCEEDINGS, volumen 12123 de **Lecture Notes in Computer Science**, p. 36–52. Springer, 2020.

[30] WANG, Q.; MAO, Z.; WANG, B. ; GUO, L.. **Knowledge graph embedding: A survey of approaches and applications**. IEEE Trans. Knowl. Data Eng., 29(12):2724–2743, 2017.

[31] HERNÁNDEZ, D.; HOGAN, A. ; KRÖTZSCH, M.. **Reifying RDF: what works well with wikidata?** In: Liebig, T.; Fokoue, A., editors, PROCEEDINGS OF THE 11TH INTERNATIONAL WORKSHOP ON SCALABLE SEMANTIC WEB KNOWLEDGE BASE SYSTEMS CO-LOCATED WITH 14TH INTERNATIONAL SEMANTIC WEB CONFERENCE (ISWC 2015), BETHLEHEM, PA, USA, OCTOBER 11, 2015, volumen 1457 de **CEUR Workshop Proceedings**, p. 32–47. CEUR-WS.org, 2015.

[31] KHAN, A.. **Knowledge graphs querying**. SIGMOD Rec., 52(2):18–29, 2023.

[32] ARENAS, M.; GUTIERREZ, C. ; SEQUEDA, J. F.. **Querying in the age of graph databases and knowledge graphs**. In: Li, G.; Li, Z.; Idreos, S. ; Srivastava, D., editors, SIGMOD '21: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, VIRTUAL EVENT, CHINA, JUNE 20-25, 2021, p. 2821–2828. ACM, 2021.

[33] ROSSO, P.; YANG, D. ; CUDRÉ-MAUROUX, P.. **Beyond triplets: Hyper-relational knowledge graph embedding for link prediction**. In: Huang, Y.; King, I.; Liu, T. ; van Steen, M., editors, WWW '20: THE WEB CONFERENCE 2020, TAIPEI, TAIWAN, APRIL 20-24, 2020, p. 1885–1896. ACM / IW3C2, 2020.

[34] GALKIN, M.; TRIVEDI, P.; MAHESHWARI, G.; USBECK, R. ; LEHMANN, J.. **Message passing for hyper-relational knowledge graphs**. In: Webber, B.; Cohn, T.; He, Y. ; Liu, Y., editors, PROCEEDINGS OF THE 2020 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, EMNLP 2020, ONLINE, NOVEMBER 16-20, 2020, p. 7346–7359. Association for Computational Linguistics, 2020.

[35] KHAN, A.; ELNIKETY, S.. **Systems for big-graphs**. Proc. VLDB Endow., 7(13):1709–1710, 2014.

[36] ALI, W.; SALEEM, M.; YAO, B.; HOGAN, A. ; NGOMO, A. N.. **A survey of RDF stores & SPARQL engines for querying knowledge graphs**. VLDB J., 31(3):1–26, 2022.

[37] ANGLES, R.; ARENAS, M.; BARCELÓ, P.; HOGAN, A.; REUTTER, J. L. ; VRGOC, D.. **Foundations of modern query languages for graph databases**. ACM Comput. Surv., 50(5):68:1–68:40, 2017.

[38] ALIVANISTOS, D.; BERRENDORF, M.; COCHEZ, M. ; GALKIN, M.. **Query embedding on hyper-relational knowledge graphs**. In: THE TENTH INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, ICLR 2022, VIRTUAL EVENT, APRIL 25-29, 2022. OpenReview.net, 2022.

[39] BAST, H.; BUCHHOLD, B. ; HAUSSMANN, E.. **Semantic search on text and knowledge bases**. Found. Trends Inf. Retr., 10(2-3):119–271, 2016.

[40] LISSANDRINI, M.; PEDERSEN, T. B.; HOSE, K. ; MOTTIN, D.. **Knowledge graph exploration: where are we and where are we going?** SIGWEB Newsl., 2020(Summer):4:1–4:8, 2020.

[41] CEBIRIC, S.; GOASDOUÉ, F.; KONDYLAKIS, H.; KOTZINOS, D.; MANOLESCU, I.; TROULLINOU, G. ; ZNEIKA, M.. **Summarizing semantic graphs: a survey**. VLDB J., 28(3):295–327, 2019.

[42] PATEL-SCHNEIDER, P. F.. **Contextualization via qualifiers**. In: Capadisli, S.; Cotton, F.; Giménez-García, J. M.; Haller, A.; Kalampokis, E.; Nguyen, V.; Sheth, A. P. ; Troncy, R., editors, JOINT PROCEEDINGS OF THE INTERNATIONAL WORKSHOPS ON CONTEXTUALIZED KNOWLEDGE GRAPHS, AND SEMANTIC STATISTICS CO-LOCATED WITH 17TH INTERNATIONAL SEMANTIC WEB CONFERENCE (ISWC 2018), volumen 2317 de **CEUR Workshop Proceedings**. CEUR-WS.org, 2018.

[43] COOK, S.. **The uses of wikidata for galleries, libraries, archives and museums and its place in the digital humanities**. Comma, 2017(2):117–124, 2019.

[44] FARDA-SARBAS, M.; MÜLLER-BIRN, C.. **Wikidata from a research perspective - A systematic mapping study of wikidata**. CoRR, abs/1908.11153, 2019.

[45] MALYSHEV, S.; KRÖTZSCH, M.; GONZÁLEZ, L.; GONSIOR, J. ; BIELEFELDT, A.. **Getting the most out of wikidata: Semantic technology usage in wikipedia's knowledge graph**. In: Vrandecic, D.; Bontcheva, K.; Suárez-Figueroa, M. C.; Presutti, V.; Celino, I.; Sabou, M.; Kaffee, L. ; Simperl, E., editors, THE SEMANTIC WEB - ISWC 2018 - 17TH INTERNATIONAL SEMANTIC WEB CONFERENCE, MONTEREY, CA, USA, OCTOBER 8-12, 2018, PROCEEDINGS, PART II, volumen 11137 de **Lecture Notes in Computer Science**, p. 376–394. Springer, 2018.

[46] BUZAABA, H.; AMAGASA, T.. **Question answering over knowledge base: A scheme for integrating subject and the identified relation to answer simple questions**. SN Comput. Sci., 2(1):25, 2021.

[47] ARTZ, D.; GIL, Y.. **A survey of trust in computer science and the semantic web**. J. Web Semant., 5(2):58–71, 2007.

[48] SANTOS, V.; SCHWABE, D. ; LIFSCHITZ, S.. **Can you trust wikidata?** Semantic Web journal, 2024. Accepted.

[49] ELNAGAR, S.; WEISTROFFER, H. R.. **Introducing knowledge graphs to decision support systems design**. In: Wrycza, S.; Maslankowski, J., editors, INFORMATION SYSTEMS: RESEARCH, DEVELOPMENT, APPLICATIONS, EDUCATION - 12TH SIGSAND/PLAIS EUROSYMPOSIUM 2019, GDANSK, POLAND, SEPTEMBER 19, 2019, PROCEEDINGS, volumen 359 de **Lecture Notes in Business Information Processing**, p. 3–11. Springer, 2019.

[50] BIZER, C.; CYGANIAK, R.. **Quality-driven information filtering using the WIQA policy framework**. J. Web Semant., 7(1):1–10, 2009.

[51] VAN ERP, M.; DE BOER, V.. **A polyvocal and contextualised semantic web**. In: Verborgh, R.; Hose, K.; Paulheim, H.; Champin, P.; Maleshkova, M.; Corcho, Ó.; Ristoski, P. ; Alam, M., editors, THE SEMANTIC WEB - 18TH INTERNATIONAL CONFERENCE, ESWC 2021, VIRTUAL EVENT, JUNE 6-10, 2021, PROCEEDINGS, volumen 12731 de **Lecture Notes in Computer Science**, p. 506–512. Springer, 2021.

[52] LI, X.; DONG, X. L.; LYONS, K.; MENG, W. ; SRIVASTAVA, D.. **Truth finding on the deep web: Is the problem solved?** Proc. VLDB Endow., 6(2):97–108, 2012.

[53] HOGAN, A.. **Knowledge graphs: Research directions**. In: Manna, M.; Pieris, A., editors, REASONING WEB. DECLARATIVE ARTIFICIAL INTELLIGENCE - 16TH INTERNATIONAL SUMMER SCHOOL 2020, OSLO, NORWAY, JUNE 24-26, 2020, TUTORIAL LECTURES, volumen 12258 de **Lecture Notes in Computer Science**, p. 223–253. Springer, 2020.

[54] CAFEZEIRO, I.; HAEUSLER, E. H. ; RADEMAKER, A.. **Ontology and context**. In: SIXTH ANNUAL IEEE INTERNATIONAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS (PERCOM 2008), 17-21 MARCH 2008, HONG KONG, p. 417–422. IEEE Computer Society, 2008.

[55] MEGDICHE, I.; RAVAT, F. ; ZHAO, Y.. **Metadata management on data processing in data lakes**. In: Bures, T.; Dondi, R.; Gamper, J.; Guerrini, G.; Jurdzinski, T.; Pahl, C.; Sikora, F. ; Wong, P. W. H., editors, SOFSEM 2021: THEORY AND PRACTICE OF COMPUTER SCIENCE - 47TH INTERNATIONAL CONFERENCE ON CURRENT TRENDS IN THEORY AND PRACTICE OF COMPUTER SCIENCE, SOFSEM 2021, BOLZANO-BOZEN, ITALY, JANUARY 25-29, 2021, PROCEEDINGS, volumen 12607 de **Lecture Notes in Computer Science**, p. 553–562. Springer, 2021.

[56] PEREVALOV, A.; DIEFENBACH, D.; USBECK, R. ; BOTH, A.. **Qald-9-plus: A multilingual dataset for question answering over dbpedia and wikidata translated by native speakers**. In: 16TH IEEE INTERNATIONAL CONFERENCE ON SEMANTIC COMPUTING, ICSC 2022, LAGUNA HILLS, CA, USA, JANUARY 26-28, 2022, p. 229–234. IEEE, 2022.

[57] NEELAM, S.; SHARMA, U.; KARANAM, H.; IKBAL, S.; KAPANIPATHI, P.; ABDELAZIZ, I.; MIHINDUKULASOORIYA, N.; LEE, Y.; SRIVASTAVA, S. K.; PENDUS, C.; DANA, S.; GARG, D.; FOKOUE, A.; BHARGAV, G. P. S.; KHANDELWAL, D.; RAVISHANKAR, S.; GURAJADA, S.; CHANG, M.; UCEDA-SOSA, R.; ROUKOS, S.; GRAY, A. G.; LIMA, G.; RIEGEL, R.; LUUS, F. P. S. ; SUBRAMANIAM, L. V.. **A benchmark for generalizable and interpretable temporal question answering over knowledge bases**. CoRR, abs/2201.05793, 2022.

[58] ANGLES, R.; ARANDA, C. B.; HOGAN, A.; ROJAS, C. ; VRGOC, D.. **Wdbench: A wikidata graph query benchmark**. In: Sattler, U.; Hogan, A.; Keet, C. M.; Presutti, V.; Almeida, J. P. A.; Takeda, H.; Monnin, P.; Pirrò, G. ; d'Amato, C., editors, THE SEMANTIC WEB - ISWC 2022 - 21ST INTERNATIONAL SEMANTIC WEB CONFERENCE, VIRTUAL EVENT, OCTOBER 23-27, 2022, PROCEEDINGS, volumen 13489 de **Lecture Notes in Computer Science**, p. 714–731. Springer, 2022.

[59] ILIEVSKI, F.; GARIJO, D.; CHALUPSKY, H.; DIVVALA, N. T.; YAO, Y.; ROGERS, C. M.; LI, R.; LIU, J.; SINGH, A.; SCHWABE, D. ; SZEKELY, P. A.. **KGTK: A toolkit for large knowledge graph manipulation and analysis**. In: Pan, J. Z.; Tamma, V. A. M.; d'Amato, C.; Janowicz, K.; Fu, B.; Polleres, A.; Seneviratne, O. ; Kagal, L., editors, THE SEMANTIC WEB - ISWC 2020 - 19TH INTERNATIONAL SEMANTIC WEB CONFERENCE, ATHENS, GREECE, NOVEMBER 2-6, 2020, PROCEEDINGS, PART II, volumen 12507 de **Lecture Notes in Computer Science**, p. 278–293. Springer, 2020.

[60] KENDALL, E. F.; MCGUINNESS, D. L.. **Ontology Engineering**. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2019.

[61] LISSANDRINI, M.; MOTTIN, D.; PALPANAS, T. ; VELEGRAKIS, Y.. **Graph-query suggestions for knowledge graph exploration**. In: Huang, Y.; King, I.; Liu, T. ; van Steen, M., editors, WWW '20: THE WEB CONFERENCE 2020, TAIPEI, TAIWAN, APRIL 20-24, 2020, p. 2549–2555. ACM / IW3C2, 2020.

[62] YAHYA, M.; BERBERICH, K.; RAMANATH, M. ; WEIKUM, G.. **Exploratory querying of extended knowledge graphs**. Proc. VLDB Endow., 9(13):1521–1524, 2016.

[63] SARRAFZADEH, B.; ROEGIEST, A. ; LANK, E.. **Hierarchical knowledge graphs: A novel information representation for exploratory search tasks**. CoRR, abs/2005.01716, 2020.

[64] HOMOLA, M.; TAMILIN, A. ; SERAFINI, L.. **Modeling contextualized knowledge**. In: Ermolayev, V.; Gómez-Pérez, J. M.; Haase, P. ; Warren, P., editors, PROCEEDINGS OF THE SECOND WORKSHOP ON CONTEXT, INFORMATION AND ONTOLOGIES, CIAO@EKAW 2010, LISBON, PORTUGAL, OCTOBER 11, 2010, volumen 626 de **CEUR Workshop Proceedings**. CEUR-WS.org, 2010.

[65] SCHUETZ, C. G.; BOZZATO, L.; NEUMAYR, B.; SCHREFL, M. ; SERAFINI, L.. **Knowledge graph OLAP**. Semantic Web, 12(4):649–683, 2021.

[66] ALJALBOUT, S.. **Contextual knowledge representation and reasoning on knowledge graphs**. PhD thesis, University of Geneva, Switzerland, 2021.

[67] VOSKARIDES, N.; MEIJ, E.; REINANDA, R.; KHAITAN, A.; OSBORNE, M.; STEFANONI, G.; KAMBADUR, P. ; DE RIJKE, M.. **Weakly-supervised contextualization of knowledge graph facts**. In: Collins-Thompson, K.; Mei, Q.; Davison, B. D.; Liu, Y. ; Yilmaz, E., editors, THE 41ST INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH & DEVELOPMENT IN INFORMATION RETRIEVAL, SIGIR 2018, ANN ARBOR, MI, USA, JULY 08-12, 2018, p. 765–774. ACM, 2018.

[68] GE, Y.; MA, J.; ZHANG, L.; LI, X. ; LU, H.. **Trustworthiness-aware knowledge graph representation for recommendation**. Knowl. Based Syst., 278:110865, 2023.

[69] STROBL, C.. **Dimensionally extended nine-intersection model (DE-9IM)**. In: Shekhar, S.; Xiong, H. ; Zhou, X., editors, ENCYCLOPEDIA OF GIS, p. 470–476. Springer, 2017.

[70] ALLEN, J. F.. **Maintaining knowledge about temporal intervals**. Commun. ACM, 26(11):832–843, nov 1983.

[71] PINHEIRO, M. K.; SOUVEYET, C.. **Supporting context on software applications: a survey on context engineering**. Modeling and Using Context, 2(Issue 1), 2018.