

CoaKG: A Contextualized Knowledge Graph Approach for Exploratory Search and Decision Making

Veronica dos Santos¹ ✉ 

Informatics Department, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil

Daniel Schwabe ✉ 

Informatics Department Associate Researcher, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil

Altigran Soares da Silva ✉ 

Institute of Computing (IComp), Federal University of Amazonas (UFAM), Manaus, Brazil

Sérgio Lifschitz ✉ 

Informatics Department, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil

— Abstract —

In decision-making scenarios, an information need arises due to a knowledge gap when a decision-maker needs more knowledge to make a decision. Users may take the initiative to acquire knowledge to fill this gap through exploratory search approaches using Knowledge Graphs (KGs) as information sources, but their queries can be incomplete, inaccurate, and ambiguous. Although KGs have great potential for exploratory search, they are incomplete by nature. Besides, for both Crowdsourced KGs and KGs constructed by integrating several different information sources of varying quality to be effectively consumed, there is a need for a Trust Layer. Our research aims to enrich and allow querying KGs to support context-aware exploration in decision-making scenarios. We propose a

layered architecture for Context Augmented Knowledge Graphs-based Decision Support Systems with a Knowledge Layer that operates under a Dual Open World Assumption (DOWA). Under DOWA, the evaluation of the truthfulness of the information obtained from KGs depends on the context of its claims and the tasks carried out or intended (purpose). The Knowledge Layer comprises a Context Augmented KG (CoaKG) and a CoaKG Query Engine. The CoaKG contains contextual mappings to identify explicit context and rules to infer implicit context. The CoaKG Query Engine is designed as a query-answering approach that retrieves all contextualized answers from the CoaKG. A Proof of Concept (PoC) based on Wikidata was developed to evaluate the effectiveness of the Knowledge Layer.

2012 ACM Subject Classification Information systems → Decision support systems; Information systems → Information integration; Information systems → Graph-based database models

Keywords and phrases Knowledge Graphs, Context Search, Decision Support

Digital Object Identifier 10.4230/TGDK.3.1.4

Category Research

Supplementary Material *Software (gtk Scripts, Dataset)*: https://github.com/versant2612/CKG_UseCases/tree/main/PoC; archived at `sswh:1:dir:488e76ffa33b3dc7c3b587ba55d5135218683253`

Funding *Altigran Soares da Silva*: This research is partially supported by FAPEAM under the Neural-Bond Project (UNIVERSAL 2023 Proc. 01.02.016301.04300/2023-04); and by CNPq under Projects IAIA (406417/2022-9) and TILD-IAR (408490/2024-1); and an individual grant from CNPq to Altigran da Silva (307248/2019-4).

Sérgio Lifschitz: This research is partially supported by CAPES (PUC-Rio's institutional funding).

Received 2024-10-08 **Accepted** 2025-06-07 **Published** 2025-07-04

¹ corresponding author



© Veronica dos Santos, Daniel Schwabe, Altigran Soares da Silva, and Sérgio Lifschitz; licensed under Creative Commons License CC-BY 4.0

Transactions on Graph Data and Knowledge, Vol. 3, Issue 1, Article No. 4, pp. 4:1–4:27



Transactions on Graph Data and Knowledge

TGDK Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

It is widely acknowledged that humans make decisions based on their prior knowledge about the domain, common sense, and often on past personal experience and beliefs. Such judgments allow people to act under uncertainty and manage the risk of negative consequences. However, when an agent has inadequate knowledge to achieve a goal due to a knowledge gap - a difference between an ideal state that allows informed decision-making and the actual state of knowledge - an information need situation arises [8]. This information need drives an exploratory process in order to close this gap.

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 the dollar exchange rate variation and gasoline prices. For instance, decisions involving health and well-being, such as getting vaccines or vacation trips, are also influenced by information-seeking behaviors. Such behavior is part of people's daily lives [8].

Different information needs require different search strategies that specialized computational tools can support [26]. More recent developments on conversational Artificial Intelligence (AI) have made Large Language Models (LLMs) part of information-seeking tools [20]. Although LLMs 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 improve the perception of trustworthiness, but this does not guarantee their reliability [34]. Indeed, users must be aware that the response provided by LLM and machine learning (ML) methods is in fact the most likely (probable) answer (based on its training corpus), regardless of its actual truthfulness.

The central problem addressed in this article is that traditional KGs used for decision-making and exploratory searches often lack the proper qualification of relations and attributes that affect the accuracy of the answers for the task at hand. Another issue is that users frequently submit ambiguous queries or omit critical contextual elements, such as temporal or spatial information, leading to responses that may not fully address their information needs. This gap creates a significant challenge in ensuring the trustworthiness and usefulness of the answers provided, particularly when users make critical decisions based on the retrieved information. Finally, when the KG contains contextual information, it is still necessary to differentiate this from other types of meta-information.

The primary contribution of this paper is to propose the concept of a *Concept Augmented Knowledge Graph* (CoaKG) that explicitly represents contexts, coupled with a novel framework that enhances decision-making by supporting context-aware exploratory search. This framework presupposes the existence of a Trust Layer, that filters trusted information according to the context, the purposes, and the policies of the agent making the decision. It does so by addressing ambiguity and incompleteness in queries submitted to the underlying CoaKG to retrieve relevant information for the task at hand. We briefly discuss the decision support architecture, but actual details are out of the scope of this paper.

We argue for the feasibility of the approach by presenting a realistic PoC based on a real-world KG, Wikidata. We demonstrate how the CoaKG Query Engine can infer implicit context and expand graph queries to retrieve all relevant contextualized answers. This approach significantly improves the relevance of information, making it more useful for complex decision-making scenarios.

The paper is organized as follows. This section has introduced the problem, motivation, and research goals, highlighting the challenges of traditional KGs in decision-making contexts. Section 2 provides background information, discussing key concepts such as uncertainty in information-seeking, trust in knowledge-based systems, and the limitations of current KG structures. Section 3

details the proposed framework for CoaKG, including the architecture of the Knowledge Layer, Trust Layer, and Decision Layer. Section 4 presents the PoC, demonstrating the effectiveness of the proposed approach through experiments using a subset of Wikidata. Next, Section 5 discusses related work compared with our approach. Finally, Section 6 concludes the paper, summarizing the contributions and outlining future research directions.

2 Background

Uncertainty is an important concept both in information-seeking and decision-making. Users are engaged in reducing the uncertainty associated with information to make a decision. 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 [36]. However, simply consuming more information is not guaranteed to reduce uncertainty [8]. Contextualized information can help reduce uncertainty.

2.1 Contextualized Answers

Since **Context** is an overloaded concept, we adopt a definition proposed by Hogan et al., 2022 [19]: “*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 outlines the limits of interpreting information in time, place, and according to its origin. Therefore, even if the original query does not specify the context information to be added, this context information must be retrieved to enable correct interpretation of the answers.

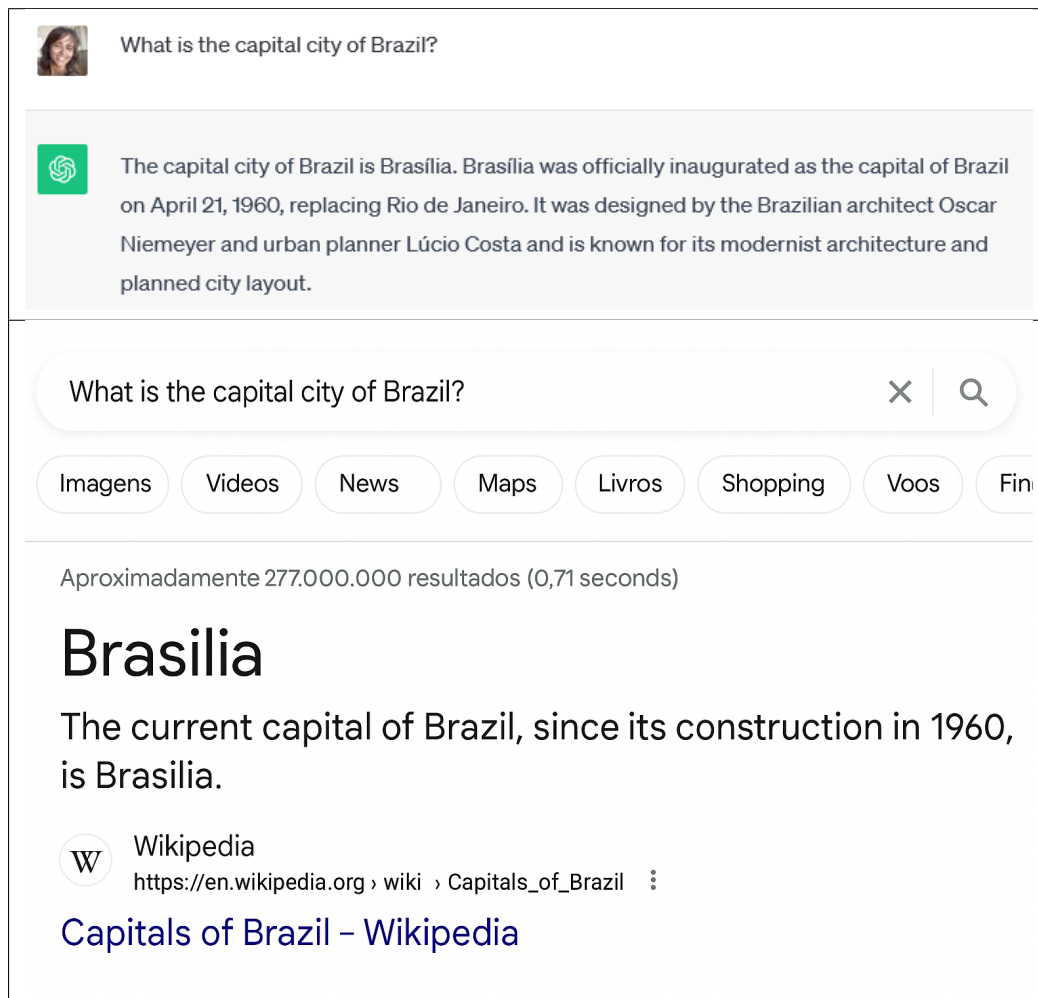
Information from multiple, distributed, and even contradictory sources, requires contextual information for trustworthiness evaluation. Accepting a claim as a fact depends on the contextual information that qualifies it, the context constraints of the task in which that claim will be applied, and the individual, organization, or community trust policies. Although this additional meta-information is necessary, it is possible that the user does not include it explicitly in the initial formulation of the query.

For example, if one asks a LLM, *e.g.*, ChatGPT, and an online search engine, *e.g.*, Google Search: “*What is the capital of Brazil?*” they would typically receive answers such those illustrated in Figure 1. Observe that both assume the default temporal context of the question as **current** even though the query did not mention this. It would satisfy an information need for actions such as deciding where his country should install an embassy in Brazil or to plan a trip to Brazil’s capital next month.

Anyone familiar with the domain knows that the “*capital of*” relationship has an intrinsic temporal context (it is a 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. However, for other more infrequent (or specialized) search tasks, the temporal context should be explicitly specified as part of the query. For example, suppose that the user is a historian and has found a document written during the Brazilian Colonial Period between 1530 and 1815 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, this would be a valid answer considering the temporal scope of the search. So, in all scenarios, the usefulness and validity of the answers depend on the circumstances of the task that motivated the information search.

As stated by Smith and Rieh [35], if a knowledge context is available to searchers during information seeking, they will engage with the information more actively and critically, and it can also positively influence users’ trust in the information provided. They defined “*Knowledge Context*” as meta-information individuals can use to interpret information on a search engine

4:4 Context Augmented Knowledge Graphs (CoaKG)



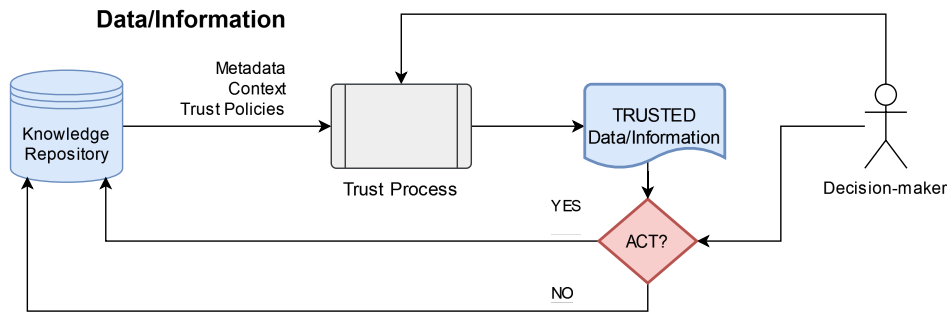
■ **Figure 1** Answers from ChatGPT and Google Search, based on tests conducted in November 2023.

results page. Such hypotheses motivate experimental research designs that reflect how information-seeking behavior is affected by knowledge context availability, as opposed to focusing on improving retrieval system performance in terms of precision and recall.

2.2 Trusted Decisions

Trust (in information) can be defined as “*knowledge-based reliance on received information*” [32, 17], that is, someone trusts (or not) the knowledge acquired if the truthfulness of this knowledge implies the decision to carry out some action [32]. In this article, we assume that the scenario illustrated in Figure 2 represents how an agent relies on trust to take action, as proposed by Schwabe et al., 2020 [33]

The input of the Trust Process are data/information composed by the data itself, general metadata, context metadata, and trust policies retrieved from a 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 and guides the process. The output of the Trust Process is Trusted Data/Information, and the user decides to act or not based on it. Actions triggered by the decisions generate additional information stored back in the Knowledge Repository to be used in upcoming decisions.



■ **Figure 2** Information Flow in Trusted Decisions. Data and metadata (general and contextual) flow from the Knowledge Repository into the Trust Process, where trust policies specific to the user and task are applied. These trust policies guide the transformation of raw data into trusted data/information. The resulting trusted data informs the Decision-maker, generating new information that is fed back into the Knowledge Repository for subsequent use.

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 a philosophical point of view [7]. The Pragmatic Theories of Truth redirect 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. Pragmatic Theories of Truth typically regard it as a result of individuals' procedures and commitments when solving problems, formulating claims, or conducting scientific investigations [7].

2.3 Knowledge-Based Decision Support Systems

An essential concern in Decision Support Systems (DSS) to reduce uncertainty is **Trust** since the decision-maker must trust the information's veracity to make an informed decision.

DSS were developed in the early 1970s for semistructured or unstructured decision-making activities. A DSS that supports strategic decision-makers using trustable expert knowledge is called a Knowledge-Based DSS (KBDSS). The main components of a KBDSS are (i) a knowledge base (with an inference engine) and (ii) a decision support shell [37].

A Knowledge Base (KB) represents facts some agent believes about the outside world, which are typically incomplete and operate under the Open World Assumption (OWA). The absence of KB information only means that this knowledge was not made explicit [5]; its truth value is unknown. Query answers are only those that can be logically proven; if it is not possible to prove a fact, it cannot be assumed that the negation of that fact is true (fact). Decision-makers should make decisions despite the lack of complete information, but should be aware of this. In the words of Carl Sagan [40]: “*the absence of evidence is not evidence of absence*”.

2.4 Knowledge Graphs

Knowledge can also be structured as a KG. Several KG definitions can be found in the literature; some 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*” [14]. According to Weikum [39], KGs are KBs modeled as a graph, since relationships are the focus of the analysis (semantic networks). However, they consider the term KG a misnomer, since KGs are not limited to binary relationships; they also encompass higher-arity tuples and intentional data through constraints and rules.

KGs have emerged as practical tools for knowledge sharing and discovery due to their ability to represent complex domain data and cover long-tail scenarios. A significant use case where KGs have become a key asset is web search engines [39]. KGs offer data modeling and integration flexibility since a pre-defined 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 [10, 18]. Every edge corresponds to a statement or claim in a KG. Other initiatives such as Wikidata² build a crowd-sourced, multilingual, large-scale KG [38].

According to Marx et al., 2017 [27], although there was no universally accepted definition at that time, what differentiates a KG from a dataset organized in a graph data structure is the need to enrich these data with information context. Context attributes such as temporal and spatial data added to statements result in higher-arity tuples. Provenance is another contextual information critical to KG consumption since a KG can represent information from different perspectives and purposes. Preserving the provenance of knowledge by recording its source, methods, and extraction time is essential to managing and curating the KG as its content evolves over extended periods [39]. Graph data structures for KGs address this using composite objects and qualification of statements.

There are several proposals for KG data structures [2], some more straightforward, like RDF (directed graphs with labeled edges) and LPG (labeled and directed property graphs), as well as some more complex and abstract, such as the multi-layer graph (graphs with higher-arity relationships and with identifiers on the edges). The Wikidata hyper-relational graph data model represents entities as *data items (nodes)* and *statements using a **subject-predicate-object** structure (edges)*. *The subject is always a data item, the predicate is a property, and the object is either a data item or a literal, with its type constrained by the definition of the property. Additionally, Wikidata allows statements to be further enriched through qualifiers as well as ranking and references. However, it does not allow claim qualifications or rankings to have qualification or references.*

A multi-layer graph data structure, as defined in [2], 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 as shown in Figure 3.

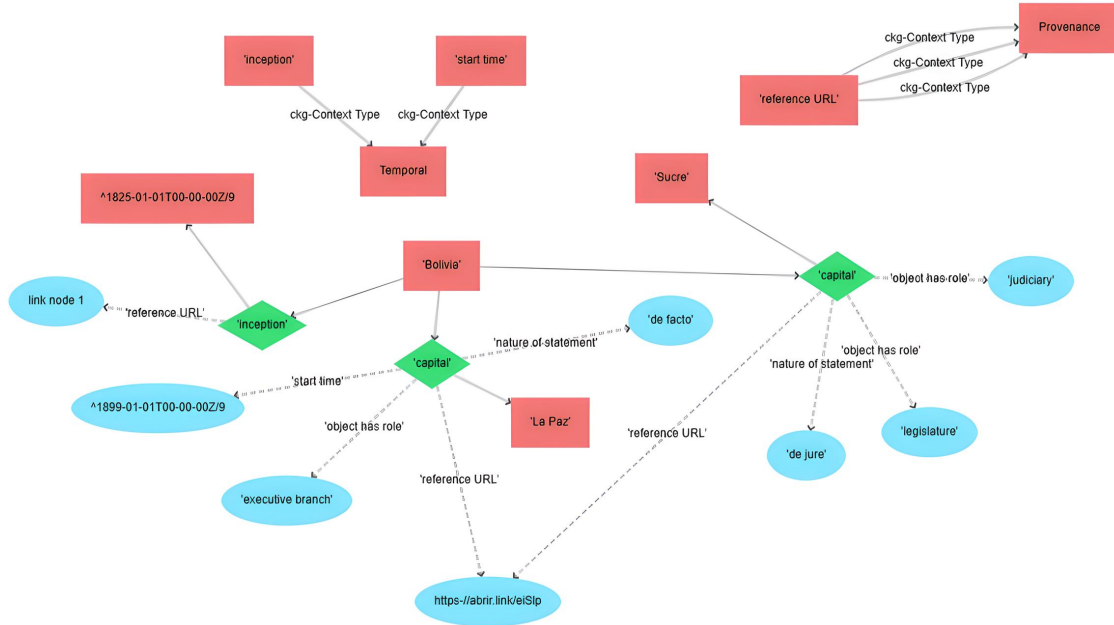
A multi-layer graph H is an n -layer graph, where each layer results from the nested use of edge IDs, and n represents the highest layer associated with an edge identifier within H . A multi-layer graph, $H = (O, \gamma)$, represents a graph data structure where edges on edges can exist. In any concrete 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.

Elements of O can be:

- V is a finite set of vertices (or nodes) that represents concepts, entity types, and instances;
- L is a finite set of vertices (or nodes) that represents literals corresponding to property and qualifier values;
- R is a finite set of binary, directed or undirected, relationship types;
- P is a finite set of property (or attribute) types, and
- Q is a finite set of qualifier types.

While γ contains:

² <http://www.wikidata.org>



■ **Figure 3** Wikidata snippet represented as multi-layer graph edges. Nodes in red rectangles and edges with filled arrows belong to the first layer. Nodes in blue ellipses and dashed edges belong to the second layer.

- E is a finite set of edges representing relationships based on a relationship type R between two vertices from V ;
- pV is a finite set of edges representing relationships, based on a property P , between one vertex from V and another from L ;
- qE is a finite set of edges representing relationships, based on a qualifier Q , between one edge from E or pV or qE or qP and a node from V , and
- qP is a finite set of edges representing relationships, based on a qualifier Q , between one edge from E or pV or qE or qP and a node from L .

From this definition, it is clear that qualifiers (edge properties) can be defined at arbitrary depths (i.e., qualifiers of qualifiers, etc...) For example, a provenance qualifier may provide information about the source of a statement, and a second level qualifier may provide information about who provided that source, and so on, This is trivially represented using Multi Layer Graphs, but is not allowed in the WD data model. In other graphs models, such as RDF or LPGs, it requires reification of statements, which add complexity to the resulting representation (and ultimately also affects the performance of queries). Such a syntactic alternative enables context representation and an in-depth exploratory search over contextual information throughout n-layers, but does not establish how context semantics should be interpreted.

3 Context Augmented Knowledge Graph-based Decision Support Systems

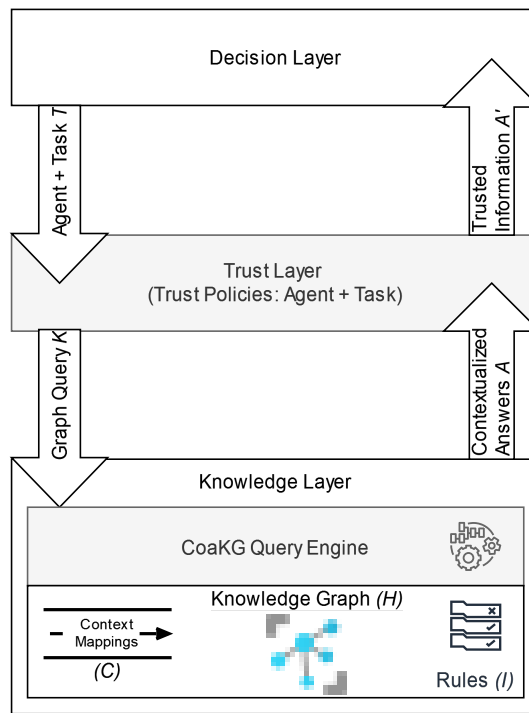
An analysis of the current status of Wikidata [31] concluded that given the characteristics of the Wikidata data structures (such as the lack of references for qualifiers, incorrect specification of property scope in addition to omissions and ambiguities of definitions in constraints) and in the data (violations of constraints that generate incompleteness and possible interpretation errors), it is necessary to have an additional and separate layer to adequately support the trust process.

4:8 Context Augmented Knowledge Graphs (CoaKG)

Given the premises that:

- Crowd-sourced KGs, such as Wikidata, or KGs generated through the automatic integration of multiple information sources with varying degrees of quality and trustworthiness, must be used via a Trust Layer [32, 33] in decision-making scenarios and
- Trust is typically assessed using policies based on hard evidence or reputation-based estimation and these mechanisms should also account for the contextual information, as well as the circumstances and associations related to the decision-making goal, to allow a well-founded trust assessment [3].

This research proposes a theoretical and abstract layered architecture of DSS supported by KG, named CoaKG4DSS, which incorporates a Trust Layer placed between Knowledge and Decision Layers, as shown in Figure 4.



■ **Figure 4** CoaKG4DSS. This figure describes the proposed multi-layer architecture for a DSS where each layer operates as an independent module, performing a specific set of functions and providing services to the upper layer. The Trust Layer is responsible for processing each user query submitted through an exploratory search approach and transforming it into a graph query, denoted as K . Additionally, the Trust Layer takes contextualized answers A from the Knowledge Layer and refines it to supply the Decision Layer with trusted information A' .

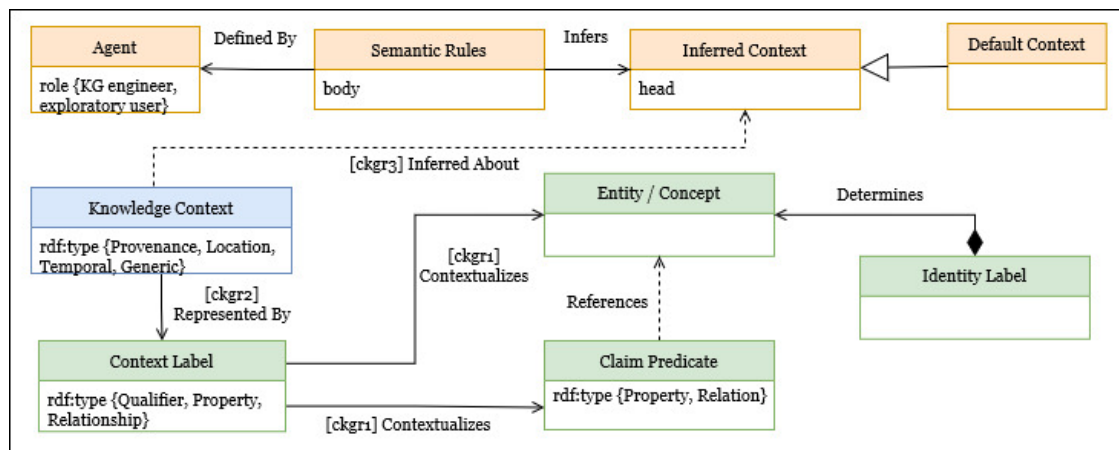
- **Decision Layer:** Given the goal of performing some action, decision-making occurs above the Trust Layer, which can operate under uncertain environments employing a variety of approaches: explicit programming (anticipate all scenarios), supervised learning (based on training examples), optimization (specify possible decision strategies and a performance measure to be maximized); planning (known models and deterministic problems); and reinforcement learning (strategy is learned according to a performance measure) [28]. The Decision Layer is already present in KBDSS architectures found in the literature [37].

- **Trust Layer:** 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. Trust policies take contextualized answers A as input and generate trusted information A' that satisfies the task and agent criteria. The Trust Layer transforms information into trusted information by applying task- and user-dependent trust policies using both information and contextual information.
- **Knowledge Layer:** Enables explicit context representation, retrieval, and interpretation enhancements to support decision-making since context is essential for identifying the scope of a claim that is considered valid and useful. Furthermore, to provide all the information the Trust Layer requires, the knowledge retrieved from the KG should be explicitly contextualized, at least in terms of Provenance, Temporal, and Location dimensions. KGs can incorporate claims representing different perspectives that are apparently contradictory. Therefore, decision-makers must rely on their trust policies to decide what is true and useful in the context of their information needs and tasks.

Due to these characteristics and differences from KBDSS, the Knowledge Layer operates under DOWA. **Contrary to OWA, under DOWA, the mere presence of a statement in a KG does not imply its truth. Therefore, DOWA posits that truthfulness is user-dependent trust and is determined by a trust evaluation that depends on the contexts of claims and tasks being carried out or intended.**

A CoaKG definition as shown in Definition 1 and the mapping schema as shown in Figure 5 characterize the elements necessary to model explicit contextual information and extract implicit contextual information from a standard KG. The mapping schema demonstrates how various components interact to enhance query responses with both explicit and inferred contexts. The knowledge context, represented by attributes such as provenance, location, and temporal information, is mapped to entity types and concepts through context labels and claim predicates.

► **Definition 1.** *CoaKG $H' = \langle H, C, I \rangle$ A multi-layer KG H with entities, concepts, claims and context information; the context mappings C between KG elements and context information; and interpretations I as rules to extract implicit context. A CoaKG extends a standard KG with resources that enable it to explicitly represent contextual information and infer implicit context.*



■ **Figure 5** CoaKG Schema for Context Mappings and Interpretation Rules, illustrating the schema for identifying contextual information from the CoaKG. These mappings allow for the contextualization of claims and identifying entities, while semantic rules infer additional context where needed.

4:10 Context Augmented Knowledge Graphs (CoaKG)

A common way to contextualize claims (statements) is by adding property-value pairs as qualifiers to statements. 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 [29].

For example, consider the Wikidata statement “*Born on the Fourth of July (Q471159) has cast member (P161) Tom Cruise (Q37079) with qualification character role (P453) Ron Kovic (Q530602)*”. The **character role** (P453) qualifier complements the **cast member** (P161) relationship with additional information but does not affect its truthfulness evaluation. In this case, reference qualifiers such as the **reference URL** (P854) can help verify the veracity of the information. An example is the link to the catalog of this film on the IMDB platform, which is a popular website about films. On this site, it is possible to consult the entire cast, including their roles. Another reference qualifier associated with **reference URL** (P854) is **retrieved** (P813) which specifies when the access to this site was made. Reference qualifiers add provenance context to the statements in Wikidata.

Since explicit contextual information is part of the KG, context mappings specify how it is represented as qualifiers and predicates among claims and metadata. According to the proposed Conceptual Schema, context mappings added to a standard KG transform it into a CoaKG. Similarly to Contextualized Ontologies [6], in CoaKGs, mappings establish the role of each element, *i.e.*, whether it functions as an entity in the standard KG or as context information. The context specification C works as a meta schema above the KG schema, adding semantics to its meta-information.

Three context types are usually found in the literature (Temporal, Location, and Provenance) and are common in several application domains.

- **Temporal** contexts encompass dates, periods, events, and so on, allowing 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. WHEN-type questions that may arise during the exploratory search can also be answered.
- **Location** contexts encompass places, coordinates, regions, spatial shapes, etc, enabling spatially-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 belongs. 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 the WHERE-type questions that may arise during the exploratory search.
- **Provenance** contexts encompass information sources, agents, processes, methods, and so on, allowing 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. In addition to 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.

Note that other domain-specific context types can be added to the schema. Contextual information can also help reveal temporal, spatial, and lineage relations among claims, entities, or concepts not explicitly represented in the KG.

The schema includes agents, such as KG engineers or users who carry out explorations, who define semantic rules that infer an implicit context. Rules can be represented as an implication between an antecedent (or IF clause or BODY rule) and a consequent (or THEN or HEAD rule).

The semantic rules consist of a set of claims, its explicit context, and operators (\wedge , v , $\#$) as antecedents that infer a relative context as consequent. The inferred and default context is used to refine the retrieved information, ensuring that responses are accurate and contextually relevant.

The contents of the Knowledge Layer can be queried through the Context Augmented Query Engine, which uses a Graph Query Language (GQL) to query the underlying KG to retrieve both trust policies and contextualized knowledge (*i.e.*, claims and metadata). Such an engine also includes an Inference Engine to execute Semantic Rules using Contextual mappings and Co-domain Algebra functions associated with each type of Context. These mechanisms are used to answer potentially incomplete graph queries concerning context information.

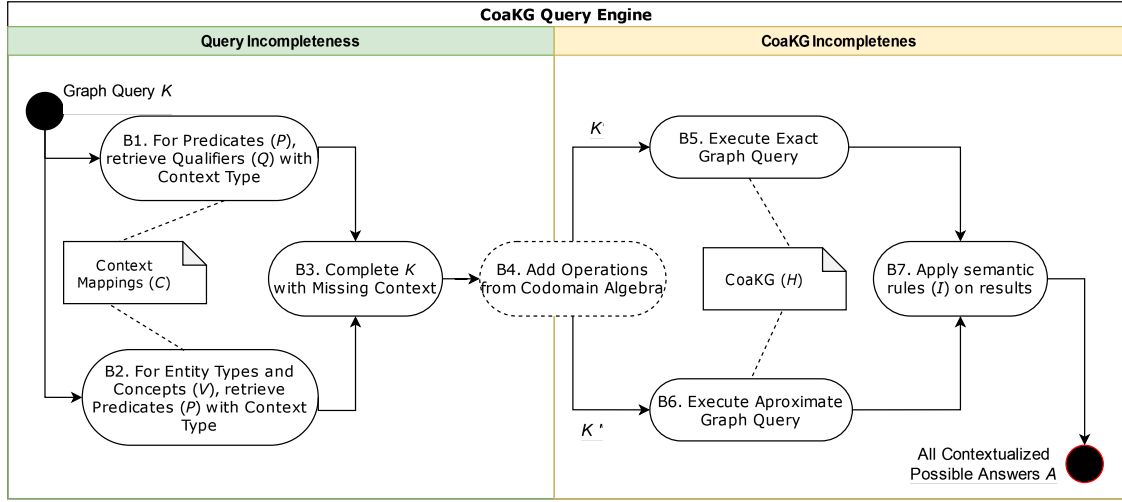
The *Trust Layer* receives the user query and converts it into a graph query K to be used as input to the CoaKG Query Engine as shown in figure 6. CoaKG Query Engine transforms the original graph query K , which is potentially contextually incomplete, into a contextually complete query. Any query K is considered contextually complete if all mappings created by the KG Engineer that reference elements mentioned in the original query are used to retrieve all context information represented in the KG. Furthermore, the query can also be modified to add hidden relationships using specific Co-domain Algebra operations generating an expanded graph query K .

A Co-domain Algebra defines a finite set of operations over co-domain values in various contexts, enabling the inference of relationships not directly materialized in the KG. These relationships between two elements of a KG (nodes or edges) can then be explicitly represented in the query response by adding new edges. An example of Co-domain Algebra for date values in the temporal context is *Allen's Interval Algebra* [1], which defines 13 basic relationships between time intervals, such as *before*, *after*, *during*, *starts*, *ends*, *among others*. Another example, pertaining to the location context, is the *Region Connection Calculus* [12], which specifies spatial relationships such as *intersection*, *inclusion*, *disconnection*, *touch*, and *more*. Specific algebraic operations enable the explicit representation of implicit relationships across contextual information. Examples of such implicit relations are claims co-occurrence in time or spatial overlap between entities.

The CoaKG query engine generates two versions of the expanded graph query K : exact and approximated, to be executed in steps B5 and B6. The exact graph query K' assumes that KG is complete and all mandatory context information are represented and can be retrieved by this query. The approximated graph query K'' deals with KG incompleteness and optional (not mandatory) context information. Three types of modification can be applied: (i) OPTIONAL clause for context qualifiers, properties, and relations, (ii) replace constant with variables for context values, and (iii) check context values for NULL to substitute by DEFAULT VALUES in WHERE clauses. After executing the queries, it is possible to apply semantic rules defined by the KG engineer or by the user performing the exploration. The rules are applied to infer the implicit context among the set of responses to produce all possible contextualized responses. We illustrate these steps in Section 4.

We present more details about the CoaKG schema and the CoaKG Query Engine in [13]. In that article we illustrated the transformation of a contextual incomplete query into a complete one through an example of exploratory search about capital cities of Brazil during the colonial period³.

³ The previous work focused on presenting the CoaKG schema and the internal mechanisms of the CoaKG Query Engine. In this paper, we substantially broaden the scope by introducing the complete architectural framework - including the Trust Layer - and by providing a PoC to assess its applicability in knowledge-based decision support scenarios.



■ **Figure 6** CoaKG Query Engine for Query and CoaKG Incompleteness. Considering K as input, it generates two versions of the contextually complete query using mappings and Co-domain Algebra. Semantic rules are applied to the query results to generate all possible contextualized responses.

Knowledge Engineering of CoaKGs

A CoaKG can be built from scratch when contextual modeling is a concern of the KG purpose. However, our proposed CoaKG schema can also be superimposed on a given existing KG (thus generating a corresponding CoaKG). In this situation, the fundamental point is to identify existing predicates that can function as contextual predicates and types and semantic rules to complete missing contextual information. Figure 7 shows the CoaKG engineering phases that must be executed to build a CoaKG.

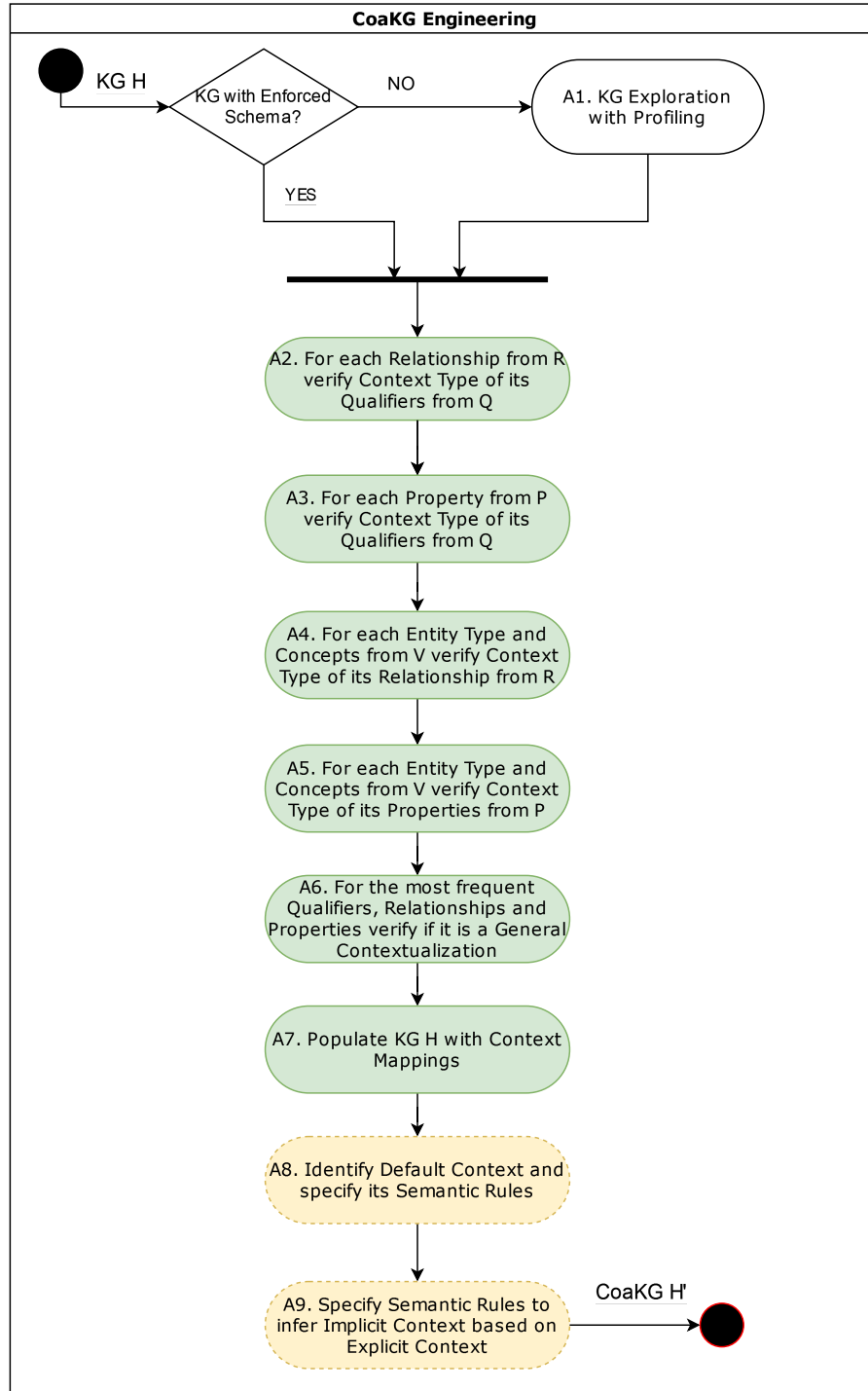
A standard KG without a formal schema should first be submitted to KG Summarization & Profiling [25] to identify its latent schema, as we have done with Wikidata (see Section 4). The process does not require specific domain knowledge since it uses general graph algorithms and returns a high-level overview of KG content. This result will assist knowledge engineers in evaluating context dimensions already present in the KG and help formulate semantic rules interpretations for the existence or absence of qualifiers and possible claims contradictions.

Context types and element identification are challenging since any information associated with an KG element (Entity, Concept, or Claim) can, in principle, be considered context. Knowledge engineers should consider the relevance of each component for the purpose of the context-aware application [30]. Another essential task is to identify gaps in existing KG instances concerning the context information necessary for a more informative response (degenerate CoaKG).

The following section describes a PoC over a CoaKG within an exploratory search process. The CoaKG was built using a subset of Wikidata as a standard KG whose Context Types are specified based on the result of Wikidata Profiling.

4 Proof of Concept

This section details our PoC for the CoaKG at the Knowledge Layer, showcasing its application in enhancing exploratory search and decision-making. We outline the methodology for building the CoaKG from a subset of Wikidata, including the profiling procedure to identify contextual predicates (properties or relationships). Competency questions were used to evaluate the Knowledge



■ **Figure 7** CoaKG Engineering steps. Given any standard KG during the engineering, it is necessary to identify if it has an enforced schema. KG Profiling must be applied to identify a latent schema if it doesn't. Based on the KG (latent) schema, the engineer can identify context types associated with the predicates, qualifiers, and entity types to specify the mappings. Finally, semantic rules can also be identified.

Layer operation and demonstrate how the CoaKG Query Engine retrieves and contextualizes answers, highlighting its effectiveness in providing context-aware information for informed decision-making.

Since anyone can build a CoaKG by extending an existing standard KG, we used a subset of Wikidata to demonstrate this pathway. We chose Wikidata because it serves as a rich knowledge repository, consolidating information from primary sources along with their references and connections to other databases. For example, Wikidata includes data about historical events, linking details from official records and academic publications to related entities and concepts, facilitating a comprehensive understanding of the topic.

Although Wikidata contains contextual information, its ontology does not explicitly identify which qualifier represents each type of context. Also, Wikidata is designed to represent provenance as a reference, but there is also Provenance information among qualifiers. For our PoC, context types were specified based on the result of a Wikidata profiling procedure identified relevant contextual predicates. These were then used to specify the context mappings.

The Wikidata graph data structure represents statements of claims associated with qualifiers and references, but does not establish how users should retrieve, interpret, and use them. For this PoC, a set of Competency Questions was elaborated to generate queries with and without context, and their respective answers were obtained.

4.1 Wikidata Selection

For this PoC, we created a KG about geopolitical entities based on a subset of Wikidata available in a dump provided by the KG Center at ISI⁴, reflecting the Wikidata version of June 2022⁵ and composed by a set of tab-separated column-based text format (TSV) files. We used the KGTK Toolkit [22] to analyze and query this Wikidata dump. This toolkit offers generic computations for graph data, such as graph statistics (degree), pathfinding, and centrality metrics, and a graph query language named Kypher, a variant of Cypher adapted to manipulate multi-layer graph structure since Cypher was designed to LPG. All KGTK scripts and query results of this PoC are publicly available in Github⁶.

To select a subset of claims and qualifications from this dump, we used the following procedure:

1. Given the node types seeds from the domain of interest, shown in Table 1, 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 number of occurrences,
4. From the top 100 most frequent reachable node types, select 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.

⁴ <https://www.isi.edu/centers-ckg/>

⁵ Downloaded from https://drive.google.com/drive/folders/1a6cUI1UEWRTNbvqtLAfJU0wEJ4ssTqdz?usp=share_link

⁶ https://github.com/versant2612/CKG_UseCases/tree/main/PoC

■ **Table 1** Node types list related to Geopolitical History.

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

After selecting the subset of Wikidata nodes and edges relevant to the domain of Geopolitical History, a profiling process was conducted to extract the latent schema of this subgraph. Since no Wikidata schema exists to fully model each selected entity type, profiling was carried out over the subset of claims and qualifications. [9]. Profiling aims to identify context qualifiers and predicates already present in the KG and their frequencies of occurrence. It also gives insights into formulating semantic rule interpretations for qualifiers' existence, absence, and possible claim contradictions. The profiling script⁷ and datasets result that revealed the latent WD schema are available on GitHub⁸. Some results are commented on below to illustrate some findings.

We retrieved 163.129 claims with 83.264 qualifications about Countries. The top-5 countries with most claims, are: **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 qualifiers 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).

4.2 Competency Questions

Next, Competency Questions (CQ) were formulated (Table 2). CQs are domain-related natural language questions with valid answers that are used to evaluate Ontologies [23] regarding coverage, correctness, and accessibility. Each CQ has acceptable response specifications to fulfill the CQ requirements. Commonly, CQ are elicited from interactions with domain experts but we manually investigated the profiling results to identify temporal- and spatial-related questions that could be answered. While conducting our research, we did not find an approach for the automatic generation of CQs based on Wikidata similar to the RevOnt proposal [11].

The contextual aspects explicitly mentioned in the question and answer specifications are highlighted. CQ0 and CQ1 are spatial-related while the others are temporal-related events. CQ1 has two versions to add the temporal variation aspect. CQ2 requires provenance context

⁷ https://github.com/versant2612/CoaKG_UseCases/blob/815c67732f5fd3c84fc2abb2cf77214d0f9f56e9/PoC/script_kgtk_WD_PoC.sh

⁸ Partial latent WD schema generated by profiling is available at https://github.com/versant2612/CoaKG_UseCases/blob/09fac1a926cba9d7b43f33f72da9c2b77c7a1858/PoC/WDschema.md

4:16 Context Augmented Knowledge Graphs (CoaKG)

in the response. CQ3 has three temporal variations and CQ4 has two. These CQs should be convertible into graph queries that should be answerable, with all contextual information, using the CoaKG. The CoaKG Query Engine was evaluated to determine how to complete graph queries and responses with missing context information and generate contextualized answers.

■ **Table 2** Competency Questions about Geopolitical History.

ID	Competency Question and Answers
CQ0	<u>Question:</u> 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 Answers with geopolitical units ID that were, directly and indirectly, replaced with replacement reference date .
CQ2	<u>When</u> was “<CQ0 geopolitical unit>” founded/established?
	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 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 geopolitical unit>”, and who were these leaders over time ?
	Acceptable Answer: Multiple Answer with Leader names, its position and time periods

4.3 Context Mappings and Rules

During CoaKG Engineering, since there is no predefined schema for each type of geopolitical unit, the profiling results were analyzed with the CQs to identify context information explicitly represented in the KG. Wikidata constraints for predicates and qualifiers were also considered.

The first step was guided by the procedure described in the first box of Appendix A using CQs specific to the domain and following the profiling. The second step was guided by the procedure described in Appendix A using only profiling, which can be reused for any domain. The contextual mappings⁹ associated with them were added to the KG, transforming it into a CoaKG.

To illustrate, Wikidata profiling on predicate P35 (**head of state**) identified that 1103 out of 1261 claims have a P580 (**start time**) qualifier, so in the respective context mappings it was specified as a mandatory qualifier for temporal context as highlighted in Table 3. The most frequent types of entity from the dataset, **country** (Q6256) and **sovereign state** (Q3624078), have context predicates as shown in Table 4.

■ **Table 3** Context Mappings of Predicates P35 and P1906 from CQ4.

Query 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

Wikidata's predicate or qualifier values can be either **no-value**, unknown, **some-value**, or an actual value. **No-value** emphasizes that the value does not exist (yet) instead of being incomplete. To reflect this characteristic of Wikidata and also deal with the absence of **end date** (P582) qualifier, two interpretation rules were defined and written in Kypher as shown in Table 5. The *Match* and *where* clauses correspond to the Body (IF) of rules while the *return* clause corresponds to their Head (THEN).

4.4 Wikidata Contextualized Answers

Each CQ and its variations were manually translated into a graph query *K*. There are at least four versions of *K* written in Kypher syntax: Original Query, Expanded with Exact Context, Expanded with Possible Context, and Expanded with Missing Context. We also generated

⁹ Contextual mappings generated by the procedures described in appendix A are available at https://github.com/versant2612/CoaKG_UseCases/blob/815c67732f5fd3c84fc2abb2cf77214d0f9f56e9/PoC/context_mappings.tsv

4:18 Context Augmented Knowledge Graphs (CoaKG)

■ **Table 4** Context Mappings for Entity Types.

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 5** Temporal Context Interpretation for Wikidata Claims.

ID	Interpretation Rule
I1	<code>-match '(v1)-[p1]→(), (p1)-[q1:P580]→(v2) '</code> <code>-where 'NOT EXISTS {(p1)-[q2:P581]→(v3)}'</code> <code>-return "ckg:i1" as id, p1 as node1, "ckgr3" as label, "Inferred Context" as label:label, "ckgl1" as node2, "Current" as node2:label'</code>
I3	<code>-match '(v1)-[p1]→(), (p1)-[q1:P580]→(v2), (p1)-[q2:P581]→(v3) '</code> <code>-where 'v3 = "no value"'</code> <code>-return "ckg:i1" as id, p1 as node1, "ckgr3" as label, "Inferred Context" as label:label, "ckgl1" as node2, "Current" as node2:label'</code>

additional graph queries to evaluate the completeness of the query and KG instances to achieve the goal of retrieving all contextualized answers. The CoaKG Query Engine, which has not yet been implemented, is the the component of the CoaKG4DSS architecture responsible for regenerating queries automatically¹⁰.

We selected a few countries to explore using all CQ. Next, we present some examples of CQ to illustrate the contextual information added to the answer.

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

There are five claims regarding the founding date of Mexico in Wikidata. The predicate `inception` (P571) in Wikidata is defined with a **SINGLE-BEST VALUE constraint**¹¹, meaning that this type of constraint specifies that a Wikidata property typically holds a single value per item. Therefore, it can be assumed that only one value is applicable in each respective context. Each claim includes a qualification for `criterion used` (P1013) provenance qualifier. In this case, the retrieved provenance context is essential to support the Trust Layer and evaluate each one’s trustworthiness. Besides the source of information, provenance clarifies how the information was obtained or calculated. This is an example where DOWA is present since, according to the purpose of use, the decision-maker can consider one or the other foundation date as true and adequate to the task.

¹⁰The script with the translation template for K queries referring to CQs is available at https://github.com/versant2612/CoaKG_UseCases/blob/815c67732f5fd3c84fc2abb2cf77214d0f9f56e9/PoC/script_kgtk_PoC_qId.sh

¹¹For more explanation about this constraint see https://www.wikidata.org/wiki/Help:Property_constraints_portal/Single_best_value

Table 6 shows preferred and deprecated contextualized claims for the foundation of Mexico. The deprecated claim is justified with the respective qualifier value, but the preferred one does not have a preferred qualifier value, that is, it is unknown. Provenance context values for **criterion used** (P1013) and **statement is subject of** (P805) are available for both claims.

■ **Table 6** Two (out of five) contextualized Answers for CQ2: Q96.

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)

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

Since the Wikidata 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 7 presents the answer. Note that head of state (P35) claim is indicated as preferred rank, although the value for reason for preferred rank (P7452) is unknown.

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 temporal relationship between the position and the leader was also inferred by the CoaKG Query Engine based on Co-domain Algebra; these claims have temporal overlap (Q16WDK4a-65).

4:20 Context Augmented Knowledge Graphs (CoaKG)

■ **Table 7** Contextualized Answers for CQ4a: Head Of State of Q16.



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)
Q16WDK4a-100	Q16WDK4a-1	ckg:Inferred Context (ckgr3)	Current (ckgI1)
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 (ckgI1)
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 8** Contextualized Answers for CQ4a: Head Of Government of Q16.

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 (ckgI1)
Q16WDK4a-203	Q16WDK4a-202	ckg:Context Type (ckgr9)	Temporal (ckgT1)

However, Queen Elizabeth died in September 2022, so on the Canada Wikidata page in October 2023, it is possible to retrieve that the current king is **Charles III of the United Kingdom** (Q43274) as shown in Figure 8. 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. Analyzing the Canada Wikidata page in October 2023, there are some claims whose **end date** (P582) qualifier has no value, but this preferred claim does not have **end date** (P582) qualifier.

If the Wikidata dump had been updated and *K4a* executed on it, rule *I1* would indicate this new claim as the Current temporal context. This is an example of how interpretation rules can infer implicit context information and generate a more contextualized answer.

head of state	 Elizabeth II	
	start time	6 February 1952
	proxy	Harold Alexander, 1st Earl Alexander of Tunis Vincent Massey Georges Vanier
	end time	Mary Simon 8 September 2022
head of state	 Charles III of the United Kingdom	
	start time	8 September 2022
	proxy	Mary Simon
	reason for preferred rank	most recent value

■ **Figure 8** Canada (Q16) reflects the KG update by community.

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 (Q16WDK4a-68) is incomplete regarding temporal and provenance context, **end date** (P580) and **statement is subject of** (P805) qualifier values are unknown, as shown in Table 8.

Regarding to **head of government** (P6), Q16WDK4a-115 claim has **start date** (P580) but does not have the **end date** (P582) qualifier. So, by rule *I1*, the temporal context is inferred as Current (Q16WDK4a-202). In such a scenario, the temporal relationship between the position (Q16WDK4a-68) and the leader (Q16WDK4a-115) cannot be inferred by the CoaKG Query Engine based on Co-domain Algebra. However, all missing required contextual information is explicitly indicated in the answer.

5 Related Work

In essence, 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 KB [37]. A framework to integrate DSS and Enterprise KG (EKG) as its knowledge repository was proposed by [15]. 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 to address KG incompleteness.

We agree that the quality of statements/claims is crucial in DSS. Still, different from them, in our proposed architecture for DSS 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 the data, as the Trust Layer should handle this according to the user and task requirements. Conflict resolution in the Knowledge Layer can introduce bias and restrict the user's decision-making power.

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, KG-based tasks, applications, and state-of-the-art KGR methods need to consider that real-world KGs inevitably end up having a large amount of untrustworthy information [16]. The authors developed a trustworthiness-aware KG representation 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 proposed architecture, can compute a trust indicator based on the trust policies of the decision-making task to rank the contextualized answers retrieved from the Knowledge Layer. This computation may consider both the KG instances and the context mappings.

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 [41] that addresses vocabulary mismatch problems using rules for query relaxation. Users are often unfamiliar with the KG structure and the labels of entities, classes, and relations. This 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 users may be unfamiliar with the KG structure and the information needed. Graph queries generated by the user or system interface or the Trust Layer may need to be completed regarding context. The original graph query K is evaluated in relation to entity types, concepts, predicates, and qualifiers against the context mappings C and can be expanded to retrieve context information. Co-domain Algebra for context values can also modify the graph query when the graph pattern is disconnected to extract implied relations. KG incompleteness regarding contextual information is addressed through semantic rules I and Co-domain Algebra, inferring implicit context.

Another approach found in the literature involves interactive graph query expansion over KGs [24]. 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 A are retrieved and sent to the Trust Layer, which can apply trust policies and rules using context information to rank or filter the contextualized claims and entities according to user and task context of interest.

Concerning knowledge context representation and manipulation of triple-based KGs, Homola et al., 2010 [21] proposed contextualized knowledge repository (CKR). A CKR is a pair $K = \langle D, C \rangle$ where C represents the set of contexts, and D is a Description Logic 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 CoaKG also contemplates explicit contextual information associated with claims, represented as triples, besides entities and concepts. Unlike CKR, CoaKG adopts a multi-layer graph data structure that enables in-depth exploration of context information navigating through n -layers. Additionally, implicit context can be inferred using rules and Co-domain Algebra at query time by CoaKG Query Engine.

6 Conclusion

In this section, we review our results and contributions, and point out future work.

Considering the flow of information for decision-making presented (Figure 2) and the opportunity to use KGs as knowledge repositories, some relevant issues related to the context needed to be addressed:

- How can each element type be differentiated if the knowledge repository consists of entities and claims with respective meta-information regarding the context?
- How to identify the different types of context represented in the knowledge graph?
- How to identify the lack of information about a relevant context for decision-making?
- If queries do not specify context, how can they be modified to retrieve the relevant context?
- If context information is implicit and not materialized in the repository, how can it be inferred using the retrieved information?

Our research answered these questions, preparing and querying KGs to support context-aware exploration in decision-making scenarios. We designed a framework for CoaKG4DSS composed of a Decision Layer, a Trust Layer, and a Knowledge Layer. Our observation of how KGs operate in practice led to the formulation of DOWA, hitherto implicit in crowd-sourced KGs such as Wikidata. DOWA asserts that the mere presence of a statement in a KG does not warrant its truthfulness.

Overall, 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 should operate under DOWA; and (v) contextual information, absent or already available in KGs, should be explicitly identified to enable a more informed decision by users regarding trust.

A graph query K , issued so that the information retrieved at least partially satisfies an information need, may be complete or incomplete with respect to context specification. In the case of incomplete queries, the proposed CoaKG Query Engine uses the context mappings C to expand the graph query K to retrieve all Contextualized Answers A from H .

This PoC is a small-scale approach used to demonstrate the feasibility of an idea. Our PoC aims to demonstrate that the Knowledge Layer is capable of providing Contextualized Answers to the Trust Layer, enabling the construction of DSSs based on CoaKGs. Since we cannot rely on a knowledge expert, we chose a domain that is relatively easy to understand, explore, and develop CQs. We selected a subset of Wikidata, restricted to the Geopolitical History domain, using a method of filtering nodes, claims, and qualifiers to generate a standard KG represented in the multilayer model. Then, we applied our KG Engineering steps using this standard KG as input to identify the explicit context, built the mappings using the proposed schema, and

added these mappings to the KG itself, transforming it into a CoaKG. We also developed two examples of rules to infer the implicit context. The CQs were manually converted into graph queries that were completed by simulating the operation of the proposed CoaKG Query Engine. The scripts and graph queries developed, as well as the datasets and contextualized responses generated throughout the PoC are publicly available on GitHub, allowing reproducibility.

We acknowledge that our PoC currently focuses on Wikidata and does not explicitly evaluate CoaKG against other existing KGs. Given the scope of this work, we frame this aspect as future work.

Although the POC only used a subset of Wikidata, we claimed that any KG, crowd-sourced or constructed by integrating several different information sources, should operate under DOWA and be consumed through a Trust Layer. However, the Trust Layer was not detailed for the PoC, being indicated as future work. We are currently evaluating the use of the Reference Ontology of Trust (ROT) [4] to specify trust policies. ROT can answer questions about the nature of trust and its influential factors. Trust policies can be used to identify trustworthiness evidence that the decision maker agent (trustor) would like to receive to trust Contextualized Answers A (trustee). It could also be used by the Trust Layer to generate trust calibration signals that transform Contextualized Answers A into Trusted Information A' .

References

- 1 James F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, November 1983. doi:10.1145/182.358434.
- 2 Renzo Angles, Aidan Hogan, Ora Lassila, Carlos Rojas, Daniel Schwabe, Pedro A. Szekely, and Domagoj Vrgoc. Multilayer graphs: a unified data model for graph databases. In Vasiliki Kalavri and Semih Salihoglu, 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, pages 11:1–11:6. ACM, 2022. doi:10.1145/3534540.3534696.
- 3 Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. *J. Web Semant.*, 5(2):58–71, 2007. doi:10.1016/j.websem.2007.03.002.
- 4 Riccardo Barattella, Glenda C. M. Amaral, Tiago Prince Sales, Renata S. S. Guizzardi, and Giancarlo Guizzardi. The many facets of trust. In Nathalie Aussenac-Gilles, Torsten Hahmann, Antony Galton, and Maria M. Hedblom, editors, *Formal Ontology in Information Systems - Proceedings of the 13th International Conference (FOIS 2023)*, Sherbrooke, Quebec, Canada, July 17-20, 2013 and Virtual Event, September 18-20, 2013, volume 377 of *Frontiers in Artificial Intelligence and Applications*, pages 17–31. IOS Press, 2023. doi:10.3233/FAIA231115.
- 5 Michael L. Brodie and John Mylopoulos. Knowledge bases vs. databases. In Michael L. Brodie and John Mylopoulos, 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, pages 83–86. Springer, 1985.
- 6 Isabel Cafezeiro, Edward Hermann Haeusler, and Alexandre Rademaker. Ontology and context. In *Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom 2008)*, 17-21 March 2008, Hong Kong, pages 417–422. IEEE Computer Society, 2008. doi:10.1109/PERCOM.2008.21.
- 7 John Capps. The Pragmatic Theory of Truth. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2023 edition, 2023.
- 8 Donald Case. *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Library and information science. Academic Press, 2002. URL: <https://books.google.com.br/books?id=unXgAAAAAAAJ>.
- 9 Sejla Cebiric, François Goasdoué, Haridimos Kondylakis, Dimitris Kotzinos, Ioana Manolescu, Georgia Troullinou, and Mussab Zneika. Summarizing semantic graphs: a survey. *VLDB J.*, 28(3):295–327, 2019. doi:10.1007/s00778-018-0528-3.
- 10 Vinay K. Chaudhri, Chaitanya K. Baru, Naren Chittar, Xin Luna Dong, Michael R. Genesereth, James A. Hendler, Aditya Kalyanpur, Douglas B. Lenat, Juan Sequeda, Denny Vrandečić, and Kuansan Wang. Knowledge graphs: Introduction, history and, perspectives. *AI Mag.*, 43(1):17–29, 2022. doi:10.1609/aimag.v43i1.19119.
- 11 Fiorela Ciroku, Jacopo de Berardinis, Jongmo Kim, Albert Merono Penuela, Valentina Presutti, and Elena Simperl. Revont: Reverse engineering of competency questions from knowledge graphs via language models. *Journal of Web Semantics*, 82, May 2024. Publisher Copyright: © 2024. doi:10.1016/j.websem.2024.100822.
- 12 Anthony Cohn, Brandon Bennett, John Gooday, and Mark Gotts. Qualitative spatial representation and reasoning with the region connection calculus.

- GeoInformatica*, 1, October 1997. doi:10.1023/A:1009712514511.
- 13 Veronica dos Santos, Edward Haeusler, Daniel Schwabe, and Sergio Lifschitz. Context-aware knowledge graphs exploratory search. In *Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 360–365, Porto Alegre, RS, Brasil, 2023. SBC. doi:10.5753/sbbd.2023.233377.
 - 14 Lisa Ehrlinger and Wolfram Wöß. Towards a definition of knowledge graphs. In Michael Martin, Martí Cuquet, and Erwin Folmer, 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 (SuCESS'16) co-located with the 12th International Conference on Semantic Systems (SEMANTiCS 2016)*, Leipzig, Germany, September 12–15, 2016, volume 1695 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016. URL: <https://ceur-ws.org/Vol-1695/paper4.pdf>.
 - 15 Samaa Elnagar and Heinz Roland Weistroffer. Introducing knowledge graphs to decision support systems design. In Stanislaw Wrycza and Jacek Maslankowski, editors, *Information Systems: Research, Development, Applications, Education - 12th SIGSAND/PLAIS EuroSymposium 2019, Gdansk, Poland, September 19, 2019, Proceedings*, volume 359 of *Lecture Notes in Business Information Processing*, pages 3–11. Springer, 2019. doi:10.1007/978-3-030-29608-7_1.
 - 16 Yan Ge, Jun Ma, Li Zhang, Xiang Li, and Haiping Lu. Trustworthiness-aware knowledge graph representation for recommendation. *Knowl. Based Syst.*, 278:110865, 2023. doi:10.1016/j.knosys.2023.110865.
 - 17 Ed Gerck. Toward real-world models of trust: Reliance on received information. *Report MCWG-Jan22-1998*, 1998. doi:10.13140/RG.2.1.2913.6727.
 - 18 Marvin Hofer, Daniel Obraczka, Alieh Saeedi, Hanna Köpcke, and Erhard Rahm. Construction of knowledge graphs: State and challenges. *CoRR*, abs/2302.11509, 2023. doi:10.48550/arXiv.2302.11509.
 - 19 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d’Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge graphs. *ACM Comput. Surv.*, 54(4):71:1–71:37, 2022. doi:10.1145/3447772.
 - 20 Aidan Hogan, Xin Luna Dong, Denny Vrandečić, and Gerhard Weikum. Large language models, knowledge graphs and search engines: A crossroads for answering users’ questions. *pre-print*, 2025. doi:10.48550/arXiv.2501.06699.
 - 21 Martin Homola, Andrei Taminlin, and Luciano Serafini. Modeling contextualized knowledge. In Vadim Ermolayev, José Manuel Gómez-Pérez, Peter Haase, and Paul Warren, editors, *Proceedings of the Second Workshop on Context, Information and Ontologies, CIAO@EKAW 2010, Lisbon, Portugal, October 11, 2010*, volume 626 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2010. URL: <https://ceur-ws.org/Vol-626/regular5.pdf>.
 - 22 Filip Ilievski, Daniel Garijo, Hans Chalupsky, Naren Teja Divvala, Yixiang Yao, Craig Milo Rogers, Rongpeng Li, Jun Liu, Amandeep Singh, Daniel Schwabe, and Pedro A. Szekely. KGTK: A toolkit for large knowledge graph manipulation and analysis. In Jeff Z. Pan, Valentina A. M. Tamma, Claudia d’Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal, editors, *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II*, volume 12507 of *Lecture Notes in Computer Science*, pages 278–293. Springer, 2020. doi:10.1007/978-3-030-62466-8_18.
 - 23 Elisa F. Kendall and Deborah L. McGuinness. *Ontology Engineering*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2019. doi:10.2200/S00834ED1V01Y201802WBE018.
 - 24 Matteo Lissandrini, Davide Mottin, Themis Palpanas, and Yannis Velegrakis. Graph-query suggestions for knowledge graph exploration. In Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen, editors, *WWW ’20: The Web Conference 2020, Taipei, Taiwan, April 20–24, 2020*, pages 2549–2555. ACM / IW3C2, 2020. doi:10.1145/3366423.3380005.
 - 25 Matteo Lissandrini, Torben Bach Pedersen, Katja Hose, and Davide Mottin. Knowledge graph exploration: where are we and where are we going? *SIGWEB Newsl.*, 2020(Summer):4:1–4:8, 2020. doi:10.1145/3409481.3409485.
 - 26 Gary Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46, 2006. doi:10.1145/1121949.1121979.
 - 27 Maximilian Marx, Markus Krötzsch, and Veronika Thost. Logic on MARS: ontologies for generalised property graphs. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19–25, 2017*, pages 1188–1194. ijcai.org, 2017. doi:10.24963/ijcai.2017/165.
 - 28 Tim A. Wheeler Mykel J. Kochenderfer and Kyle H. Wray. *Algorithms for Decision Making*. The MIT Press, August 2022. eBook.
 - 29 Peter F. Patel-Schneider. Contextualization via qualifiers. In Sarven Capadisli, Franck Cotton, José M. Giménez-García, Armin Haller, Evangelos Kalampokis, Vinh Nguyen, Amit P. Sheth, and Raphaël Troncy, editors, *Joint Proceedings of the International Workshops on Contextualized Knowledge Graphs, and Semantic Statistics co-located with 17th International Semantic Web Conference (ISWC 2018)*, volume 2317 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018. URL: <https://ceur-ws.org/Vol-2317/article-01.pdf>.
 - 30 Manuele Kirsch Pinheiro and Carine Souveyet. Supporting context on software applications: a survey on context engineering. *Modeling and Using Context*, 2(Issue 1), 2018. doi:10.21494/ISTE.OP.2018.0275.
 - 31 Veronica Santos, Daniel Schwabe, and Sérgio Lifschitz. Can you trust wikidata? *Semantic*

- Web journal*, 15(6):2271–2292, 2024. Accepted. doi:10.3233/SW-243577.
- 32 Daniel Schwabe. Trust and privacy in knowledge graphs. In Sihem Amer-Yahia, Mohammad Makhadmeh, Ashish Goel, Geert-Jan Houben, Kristina Lerman, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia, editors, *Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, WWW '19, pages 722–728, New York, NY, USA, 2019. ACM. doi:10.1145/3308560.3317705.
 - 33 Daniel Schwabe, Carlos Laufer, and Pompeu Casanovas. Knowledge graphs: Trust, privacy, and transparency from a legal governance approach. *Law in Context. A Socio-legal Journal*, 37(1):24–41, December 2020. doi:10.26826/law-in-context.v37i1.126.
 - 34 Chirag Shah and Emily M. Bender. Situating search. In David Elsweiler, editor, *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, Regensburg, Germany, March 14 - 18, 2022*, pages 221–232. ACM, 2022. doi:10.1145/3498366.3505816.
 - 35 Catherine L. Smith and Soo Young Rieh. Knowledge-context in search systems: Toward information-literate actions. In Leif Azzopardi, Martin Halvey, Ian Ruthven, Hideo Joho, Vanessa Murdock, and Pernilla Qvarfordt, editors, *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, CHIIR 2019, Glasgow, Scotland, UK, March 10-14, 2019*, pages 55–62. ACM, 2019. doi:10.1145/3295750.3298940.
 - 36 Katie Steele and H. Orri Stefánsson. Decision Theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2020 edition, 2020.
 - 37 Leni Sagita Riantini Supriadi and Low Sui Pheng. *Knowledge Based Decision Support System (KBDSS)*, pages 155–174. Springer Singapore, Singapore, 2018. doi:10.1007/978-981-10-5487-7_7.
 - 38 Denny Vrandečić. Wikidata: a new platform for collaborative data collection. In Alain Mille, Fabien Gandon, Jacques Misselis, Michael Rabinovich, and Steffen Staab, editors, *Proceedings of the 21st World Wide Web Conference, WWW 2012, Lyon, France, April 16-20, 2012 (Companion Volume)*, pages 1063–1064. ACM, 2012. doi:10.1145/2187980.2188242.
 - 39 Gerhard Weikum. Knowledge graphs 2021: A data odyssey. *Proc. VLDB Endow.*, 14(12):3233–3238, 2021. doi:10.14778/3476311.3476393.
 - 40 Gerhard Weikum, Xin Luna Dong, Simon Razniewski, and Fabian M. Suchanek. Machine knowledge: Creation and curation of comprehensive knowledge bases. *Found. Trends Databases*, 10(2-4):108–490, 2021. doi:10.1561/19000000064.
 - 41 Mohamed Yahya, Klaus Berberich, Maya Ramanath, and Gerhard Weikum. Exploratory querying of extended knowledge graphs. *Proc. VLDB Endow.*, 9(13):1521–1524, 2016. doi:10.14778/3007263.3007299.

A CoaKG Engineering

This appendix details the engineering process for creating the Context Augmented Knowledge Graph (CoaKG) using Wikidata (WD) as a foundational resource. It outlines the steps taken to develop context mappings for both competency questions and generic contexts, facilitating a structured approach to contextualizing information.

Listing 1 describes the procedure for addressing WD competency questions by identifying relevant relationships, attributes, and context types, and creating appropriate mappings.

Listing 2 focuses on the engineering of generic context by identifying widely applicable predicates and qualifiers across entity types, enabling the establishment of optional context mappings.

■ Listing 1 CoaKG Engineering for WD Competency Questions.

```

For each Competency Question:
  Identify relationships or attributes and context types of interest
  mentioned.
  For each relationship or attribute:
    Identify the WD predicate and qualifiers associated.
    For each WD predicate:
      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
    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 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
    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 semantics.
    Create identity mappings between WD entity types and
    WD predicates

```

■ Listing 2 CoaKG Engineering for WD Generic Context.

```

For the Generic Context, using entity types, predicates, and qualifiers
frequency:
  Identify predicates applicable to most entity types (or concepts);
  Create context mappings between WD entity types and predicates
  of interest as optional.
  Identify qualifiers applicable to most predicates;
  Create context mappings between WD predicates and qualifiers
  of interest as optional.

```