



HAL
open science

Ontology based Event Knowledge Graph enrichment using case based reasoning

Rajesh Piryani, Nathalie Aussenac-Gilles, Nathalie Jane Hernandez, Cédric Lopez, Camille Pradel

► To cite this version:

Rajesh Piryani, Nathalie Aussenac-Gilles, Nathalie Jane Hernandez, Cédric Lopez, Camille Pradel. Ontology based Event Knowledge Graph enrichment using case based reasoning. 20th International Conference on Semantic Systems SEMANTICS 2024, VU University Amsterdam; Institute for Applied Informatics (InfAI); Semantic Web Company, Sep 2024, Amsterdam, Netherlands. pp.1-18. hal-04653598

HAL Id: hal-04653598

<https://hal.science/hal-04653598>

Submitted on 19 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

April 2024

Ontology based Event Knowledge Graph enrichment using case based reasoning

Rajesh PIRYANI^{a,b,1}, Nathalie AUSSENAC-GILLES^{a,b,c}, Nathalie HERNANDEZ^{a,d},
Cédric LOPEZ^e and Camille PRADEL^f

^aIRIT, Toulouse, France

^bUniversite Toulouse III Paul Sabatier (UT3), Toulouse, France

^cCNRS, Toulouse, France

^dUniversite Toulouse Jean Jaures (UT2J), Toulouse, France

^eEmvista, Jacau, France

^fChapsVision, Toulouse, France

ORCID ID: Rajesh Piryani <https://orcid.org/0000-0003-3374-0657>, Nathalie Aussenac-Gilles <https://orcid.org/0000-0003-3653-3223>, Nathalie Hernandez <https://orcid.org/0000-0003-3845-3243>, Camille Pradel <https://orcid.org/0000-0001-9944-563X>

Abstract. Purpose: The first objective of this research is to represent an event with 5W1H characteristics (who, what, where, when, why, and how) through ontologies. The second objective is to propose an approach for enriching an event knowledge graph (EKG) based on this ontology using EvCBR, an outperforming case-based reasoning algorithm found in the literature. Furthermore, we have studied the impact of each W (Who, Where, and When) on the performance of EvCBR on the Wikipedia Causal Event dataset. **Methodology:** We proposed the XPEventCore ontology to represent 5W1H characteristics of events by integrating multiple event ontologies (SEM and FARO) and introduced new object properties for representing Cause and Method to answer “Why” and “How” questions. We adopted this XPEventCore ontology for a specific use case (the MR4AP Wikipedia dataset), and populated the EKG. Furthermore, we adapted EvCBR, a case-based reasoning approach, to enrich this EKG. **Findings:** XPEventCore ontology provides a structured and adaptable foundation for capturing the essential facets of an event. It can be adapted to any domain (like MR4AP Wikipedia dataset) and populated to generate EKGs. Then, we applied EvCBR, and subsequent analysis revealed that reasoning had a significant impact. Notably, EvCBR outperformed with reasoning approaches on the dataset. **Originality:** XPEventCore ontology is the first ontology that represents an event with 5W1H characteristics.

Keywords. Event ontology, Knowledge graph enrichment, Case-based reasoning, Event knowledge graph, 5W1H

1. Introduction

Knowledge graphs (KG) are an important component of information systems that require structured knowledge, both domain-specific and domain-independent. They have

¹Corresponding Author: Rajesh Piryani, IRIT, UPS(UT3), Toulouse, France, E-mail:rajesh.piryani@irit.fr.

April 2024

greatly enhanced the ability to equip intelligent systems with comprehensive knowledge, potentially transforming the way we access and utilize information. Within the realm of knowledge representation, event-centric knowledge such as Event KGs (EKGs) are pivotal in communicating dynamic and procedural knowledge [1]. Such knowledge representation has paved the way for numerous downstream applications, ranging from search and question-answering to recommendation systems, financial quantitative investments, and text generation [2,3,4,5,6,7].

In the state of the art, an event is described as *anything that happens at a particular time and location* [8] such as meetings, cultural shows, phone calls and purchases, but also business buyouts, changes of management, and health crises. Knowledge of these events is essential for humans to make decisions that will have an impact on future events. Many innovative applications can benefit or even emerge from a technology capable of extracting events from various sources, representing them, aggregating them, and exploiting them to predict future events. We can for example cite: anticipating demand for sanitary products, the supervision of cultural, advertising or festive events; but also the study of competitions or commercial markets. In all these applications, events are represented with more and more detailed features. Features may characterize the actors involved (Who organizes or takes part in the event), the nature of the event (What happens), its location (Where), the time when it takes place and its duration (When), but also the reason for its happening (Why) or the way it happened (How) [25]. An event is considered fully comprehended when all 5W1H components (Who, what, where, when, why, and how) have been accounted for. This type of event is commonly referred to as an *explainable event* [9]. In our work, we define an event as *the specific occurrence of something (an action or series of actions) that happens at a certain time, a certain place, due to certain reasons, involving one or more participants (such as objects or humans) and which can frequently be described as a change of state.*

In recent years, several ontologies have been published to represent events, making their semantics more explicit and allowing to automatically reason and deduce additional information [10]. Among them are the Event Ontology [11], the Linked Open Description of Events (LODE) [12], the F-model [13], and the Simple Event Model (SEM) [14]. Examples of upper ontologies representing events are BFO [15], UFO [16], DOLCE [17], YAMATO [18], SUMO [19] and examples of domain-specific ontologies are CIDOC-CRM [20], ABC [21]. However, through comprehensive analysis of the existing literature [22], there is currently no ontology available that adequately represents an explainable event incorporating the 5W1H characteristics. Examples of explainable event are news events. There is a need for an ontology to represent such events from news articles with 5W1H characteristics in order to answer to the journalistic questions, namely: who did what, when, where, why, and how.

In this paper, we propose the XPEventCore ontology to represent 5W1H characteristics events by integrating event ontologies such as SEM [14] and FARO [23]. We introduce new object properties for representing Cause and Method to answer the questions “Why” and “How” respectively. Furthermore, we adopt this event ontology to represent the MR4AP Wikipedia dataset (a company dataset), which contains events with 5W characteristics and other discourse relations between events, such as *before or purpose*.

Populating these ontologies to build KGs is prone to errors and incompleteness in the initial stages due to human errors, algorithmic flaws, or missing information in the data source for automated construction [10]. The process of KG refinement involves

adding missing information, correcting errors, and simplifying the graph by eliminating redundancies. While some literature uses KG completion or KG identification to refer to KG refinement, these only address the addition of missing information or the identification of redundancies or errors [10]. In contrast, KG refinement encompasses all updates to the KG, including filling in missing information, detecting and correcting errors, and eliminating redundancies [10]. In this paper, we use EvCBR [24], an outperforming case-based reasoning approach, to predict properties of unseen effect events. For example, in Figure 1, there are two events $E1$ and $E2$ connected through *has_cause* relation; EvCBR aims to predict the properties of unseen event $E2$, that are $a1$ and $a2$. EvCBR involves retrieving similar past cases and determining the path from EKG to predict causal properties. We propose to adapt EvCBR to make predictions about new properties of events connected through other relations such as *purpose* and temporal relation *before*.

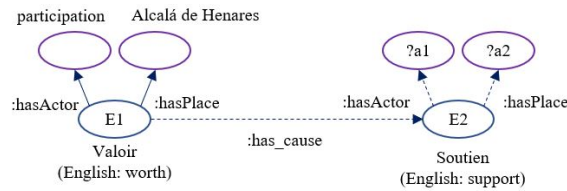


Figure 1. Example of predicting properties for event $E2$

The contributions in this paper can be summarized as follows:

- We propose the XPEventCore² ontology that incorporates various ontologies such as SEM and FARO to represent the 5W1H attributes (who, what, when, where, why, and how). We introduce novel object properties for Cause and Method to address the "Why" and "How" aspects, respectively.
- We use EvCBR, a case-based reasoning approach, for improving the refinement of EKGs. Our contribution is to focus on its faculty to propose purpose and temporal relations, taking into account the generalization nature of EvCBR and the knowledge represented in the ontology about event characteristics.
- We study the impact of considering 5W1H attributes for events on the refinement of two different EKGs, chosen for their different characteristics. One is used by a company [44] and one has been extracted from wikidata [24].

2. State of the art

2.1. Event Ontologies

Chen and Li [25] categorized events into four groups based on 5W1H: Explainable events (events with 5W1H non-empty values), completely formed events (events with What, Who, Where, When, and How values), well-formed events (events with What, When, Where, and Who values) and loosely formed events (events with less than four non-empty values in 5W1H representation). Further, Hamborg et al. [26] discussed that

²We will release our code publicly at <https://github.com/rpiryani/xpEventCore>

“why” answers the reason or cause of the event, and “how” answers the method by which an action was performed. Hamborg et al. [26] research work focused on extractions of event characteristics from News articles. Other research work [27] focused on social media event detection and defined events as an occurrence of interest in the real world that initiates a dialogue on the event-related topic among diverse social media users. Another common definition is a series of actions accomplished between more than one agent (referring to people or machines) [28]. Like Allan et al.[8], Dou et al.[29] described social media events as a change in the size of textual data that concerns the associated topic at a specific time and location. Wang et al. [30] designed a verb-based approach to extract news event information using 5W1H.

In recent years, several ontologies were proposed and published to represent events using semantic web technologies, such as Event Ontology [11], LODÉ[12], F-model [13], SEM [14]. Examples of upper ontologies representing events are BFO [15], UFO [16], DOLCE [17], YAMATO [18], SUMO [19] and examples of domain-specific ontologies are CIDOC-CRM [20], ABC [21]. Considering the 5W1H features, the focus of the ontologies listed in Table 1 differs. The majority of them represent well-formed events with 4Ws (What, When, Where, and Who) except CIDOC-CRM, F-model, and FARO. Chen and Li [25] call *explainable events* long-term events with a richer representation which are explained thanks to the 5WH properties, making explicit latent relationships (e.g., evolution and causality). We consider these 6 dimensions as the basis for competency questions that an event ontology should satisfy [22]. None of the ontologies in our review represents such type of events [22]. In this paper, we propose XPEventCore, a novel ontology to represent four categories of events with 5W1H characteristics.

Table 1. Event Ontologies representing components of 5W1H. For more details about how these dimensions are represented in each ontology, see Table 3 of [22]

Name of Ontology	What (change and action)	Where (possible place)	When (specific time)	Who (Participants)	Why (causes)	How (method)
CIDOC-CRM [20]	✓	✓	✓	✓	✓	✗
Event ontology [11]	✓	✓	✓	✓	✗	✗
DUL	✓	✓	✓	✓	✗	✗
LODE [12]	✓	✓	✓	✓	✗	✗
F-model[13]	✓	✓	✓	✓	✓	✗
SEM [14]	✓	✓	✓	✓	✗	✗
ABC[31]	✓	✓	✓	✓	✗	✗
FARO [23]	✓	✗	✗	✗	✓	✗

2.2. KG Enrichment

Although ontologies and rules are utilized during the initial stages of construction, Knowledge Graphs (KGs) are still susceptible to errors and incompleteness [10]. Even after construction, refinement is necessary to complete missing information and correct

April 2024

inaccuracies. Some literature considers KG refinement to be synonymous with KG completion [32] or KG identification [33]. KG completion is the identification or prediction and addition of missing information, such as link prediction (LP), relation prediction, and attribute completion, while excluding error correction or redundancy removal [10].

Recent surveys [34,35] have extensively detailed a significant number of embedding-based methods developed for LP in KG. However, these methods often exhibit performance trade-offs depending on the dataset and the effectiveness of the model hyperparameters tuning [36]. Although GNN-based models have shown great performance, there is an ongoing debate about the semantics captured in such models, and the evaluation process for LP models is being continually reevaluated [24]. Moreover, most embedding-based models currently operate in the closed-world setting. Only a few recent works have considered inductive link prediction by using additional knowledge sources such as text or hyper-relational facts [24]. Recently, Shirai et al. [24] have introduced a case-based reasoning model EvCBR to predict properties about new consequent events based on similar cause-effect events present in the KG. It utilizes statistical measures to find similar events and employs path-based predictions. It requires no training steps. EvCBR has better performances in comparison to three embedding-based models (TansE, ComplEx and RotatE), one GNN model (NoGE), and rule-based LP models (ProbCBR and AnyBURL) that adopt analogous scoring and sampling techniques as EvCBR.

In this paper, we adapted the EvCBR algorithm to predict the properties of the events connected through causal, temporal, and purpose relations. We consider two different EKGs to evaluate the impact of 5W1H attributes. Both of them are relatively small (86,693 triples for one and 758,857 for the other one). Existing embedding models and other models require a large amount of training data, whereas EvCBR does not require much training and focuses on the EKG.

3. Methodology

We first present XPEventCore, the Event Ontology representing 5W1H characteristics, which we then specialise for a specific use case and populate from a given dataset to get an EKG. Next, we explain how we predict properties of unseen events connected through causal, temporal, and purpose relations by adapting the EvCBR approach to this EKG.

3.1. Event Ontology

XPEventCore provides a structured and adaptable foundation for capturing the essential facets of an event. By leveraging XPEventCore, events can be represented consistently across contexts and interconnect diverse datasets. Its flexibility allows to customize its structure, extending or refining it to accommodate particular scenarios' unique characteristics and requirements.

3.1.1. Core part of ontology

From comprehensive literature analysis [22], it appears that currently, no ontology adequately represents an explainable event incorporating the 5W1H characteristics. SEM is a widely reused ontology, as evidenced by its application in numerous research works [37,38,39,40,41,42,43] where it has been extended to represent events that fulfill their

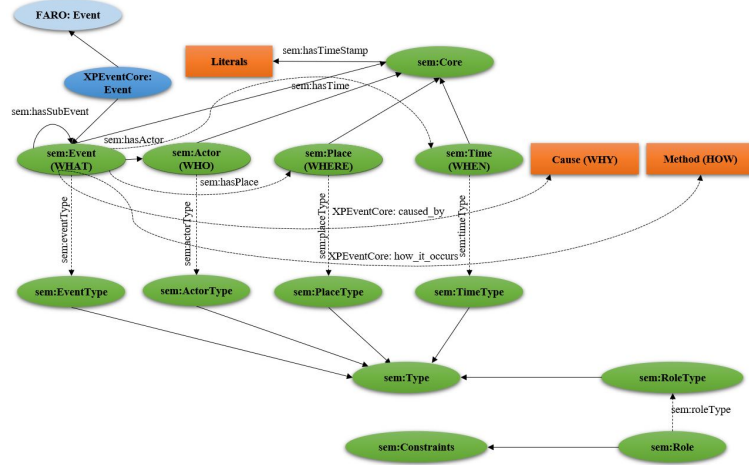


Figure 2. XPEventCore Ontology. Regular arrows are `rdfs:subClassOf` properties between the classes, and dash lines with arrows represent `rdfs:domain` and `rdfs:range` restrictions on properties.

requirements. Nevertheless, the SEM model only represents four of the 5W1H characteristics and does not allow linking different events or providing relations between classes of the same type. Whereas FARO does not represent “When” explicitly, it allows temporal relations, which means FARO has an implicit answer to “When” question. FARO neither explicitly defines classes or properties for “Where” and “Who”. LOD ontology aims for minimal modeling of events. CIDOC-CRM has many classes and properties, but only a subset defines the event. It does not represent the “How” element of 5W1H.

To overcome these limitations, we have developed the XPEventCore ontology by extending SEM. XPEventCore explicitly addresses the representation of explainable events incorporating the 5W1H characteristics. It also integrates other ontologies, including the FARO ontology, known for its rich set of relations. New classes and properties have been introduced within XPEventCore. As shown in Figure 2, to represent events, we have mitigated the limitation of SEM by creating `XPEventCore:Event` as a subclass of both `sem:Event` and `FARO:Event` that provides a rich set of relations between events. Other key classes of XPEventCore are reused from SEM, such as `sem:Actor` representing active/passive entities participating in the event, `sem:Place` representing the location, and `sem:Time` representing the time. We have also introduced new object properties as subproperties of `sem:eventProperty` to represent *Why* and *How* by `XPEventCore:caused_by`, and `XPEventCore:how_it_occurs` respectively. Events are connected to entities, places, time, cause, and method through the object properties `sem:hasActor`, `sem:hasPlace`, `sem:hasTime`, `XPEventCore:caused_by` and `XPEventCore:how_it_occurs` respectively. Figure 3 illustrates the example of XPEventCore with 5W1H. In this example, we cannot use `FARO:causes` because, according to the definition provided by FARO, `FARO:causes` can connect an event with its effect, but not with another (caused) event. Generally, “Why” constituent is the most difficult to obtain among other constituents since human resources need to be significantly involved [25]. Two meaningful relationships (i.e., the content reference relationship and the dependence relationship) among events can provide information about the Why dimension [25].

April 2024

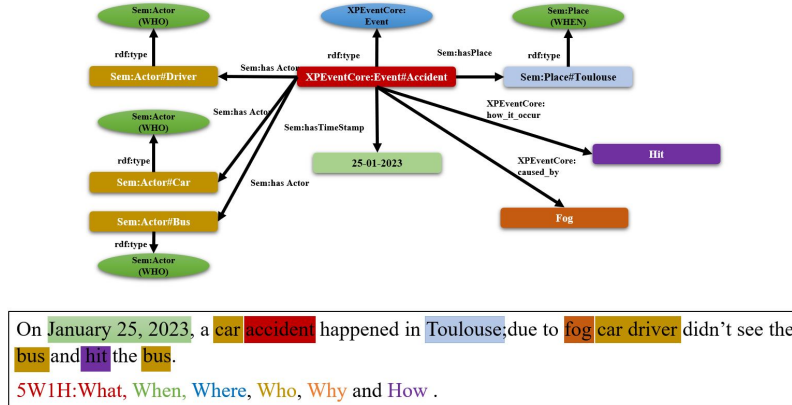


Figure 3. Representation of Example through XPEventCore to represent 5W1H.

3.1.2. Adaptation and Population of XPEventCore Ontology to Specific Use case

To effectively employ XPEventCore for a specific purpose, the primary step is to identify the event and its pertinent features in a comprehensive manner. If the established criteria are not entirely satisfied, it may be necessary to generate new classes and object properties to depict the event accurately. Subsequently, the EKG should be populated by systematically integrating the identified event and its relevant characteristics into the customized XPEventCore infrastructure. Figure 4 illustrates the steps for adapting XPEventCore to a specific use case and all of these steps are explained below:

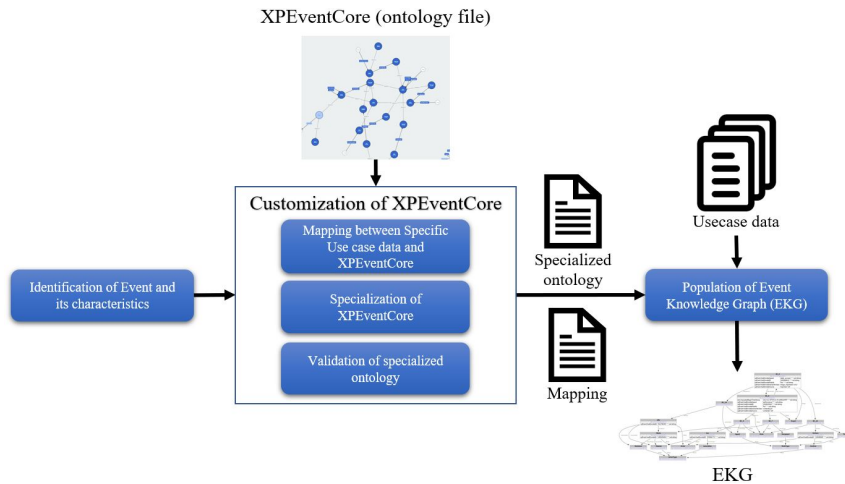


Figure 4. Adaptation steps

1. **Identification of event and its characteristics:** We start by defining what are the essential characteristics of an event (What, Who, Where, When, and How values) for a given use case and how to extract them from data or texts before they representation in the EKG. It is not compulsory for the event to have all 5W1H.

April 2024

2. **Customization of XPEventCore (Figure 4):** First, the XPEventCore classes and properties are mapped to the specific use case data, i.e. to event and their characteristics as assessed in the use case. Second, the ontology is tailored when predefined XPEventCore structures do not fully align with the use case requirements. Third, the specialized ontology is validated through Ooops. The customization of the XPEventCore step produces two outputs: a specialized ontology and mappings.
3. **Population of Event Knowledge Graph (EKG) (Figure 5):** First, we pre-process the data to identify events and their characteristics, using NLP techniques when data is text. It uses mappings to create event instances and their characteristics according to ontology, and builds the EKG.

For evaluation purposes, we have adapted the XPEventCore ontology for the MR4AP Wikipedia dataset. The proposed ontology and EKG are available online³.

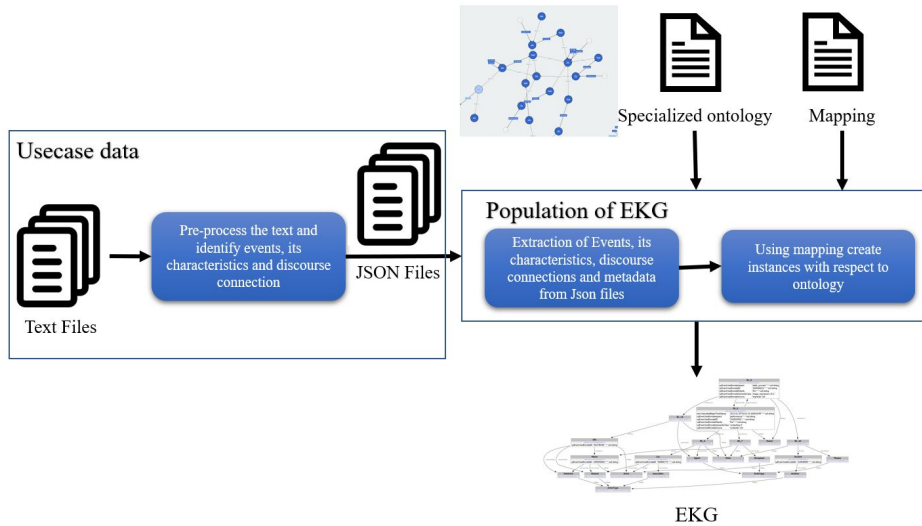


Figure 5. Pipeline for Event Knowledge Graph Population

3.2. Overview on the EvCBR approach

EvCBR [24] is a case-based reasoning model. It predicts the properties of new subsequent events based on similar cause-effect events present in the KG. EvCBR makes path-based predictions without training by identifying similar occurrences using statistical measurements like entity similarity, case head similarity, case tail similarity and case selection. Generally, this task is a 2-hop Link Prediction (LP) task as illustrated in Figure 6: the first hop is a causal relation connecting a cause event to a new effect event, and the second hop is a property of the new effect event that is desired to be predicted. We apply this approach to other relations, such as temporal and purpose relations. Section 4 shows how the approach performs on such relations and gives an overview on the main steps.

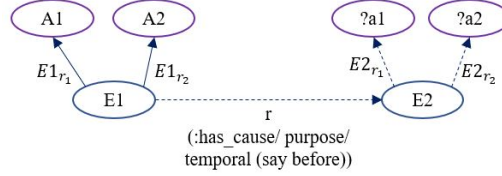


Figure 6. Problem statement

3.2.1. Problem Statement

Figure 6 illustrates the problem statement. The primary objective of the case-based reasoning approach is to predict properties of unseen events connected through causal/ temporal/ purpose relations. This task is represented as a query triple (E_1, r, E_2) where E_1 and E_2 are the events and r could be causal (*has_cause*)/ temporal/ purpose relation. The objective is to predict the properties of E_2 in the form of $(E_2, E_{2r}, ?a)$. The assumption is that $E_1 \in E$ while $E_2 \notin E$.

3.2.2. Methodology

The case-based reasoning approach is the one presented in [24] adapted to event concepts with their relations and 5W1H properties. We have proposed a generalisation of the approach in order to apply it to predict any relation (causal/ temporal/ purpose). The process has three steps illustrated in Figure 7: Finding similar event cases according to the selected relation, finding and applying a prediction path, and refining the prediction. It starts from a new causal event, say Valoir, then proceeds to predict actors and places of a new unseen event, say Soutien as depicted in Figure 7.

1. **Finding similar causal/temporal/purpose event cases:** For a given input event for which we want to make predictions, first, retrieve cases of $E_1 - E_2$ event pairs in EKG that are similar to the new input event. Then, cases are selected based on the similarity of their E_1 to the new query event as well as the similarity of their E_2 properties to the properties of our new unseen E_2 event for which we want to do the predictions. The similarity of individual entities in EKG is computed based on outgoing connection. For case similarity, a measure of importance has been computed for each relation based on how common the relation is; it gives more weight to rarer relations or entities when selecting which cases to retrieve [24].
2. **Retrieving and applying prediction paths:** This process is an important step in making predictions based on EKG data. Firstly, similar events are retrieved from the EKG. Then a useful path to make predictions is searched through the EKG. For each retrieved case with event pairs E_1 and E_2 , the objective is to find in the EKG a prediction path of relations that can be followed from the E_1 entity to the E_2 properties. The confidence of each prediction path is then scored, with higher confidence being assigned to the path that leads to the correct entity in a majority of the retrieved cases [24]. The score is calculated based on the precision of following the paths in similar $E_1 - E_2$ event pairs. An aggregated prediction score for each property of the E_2 event is computed using the confidence scores. This

³<https://github.com/rpiryani/xpEventCore>

April 2024

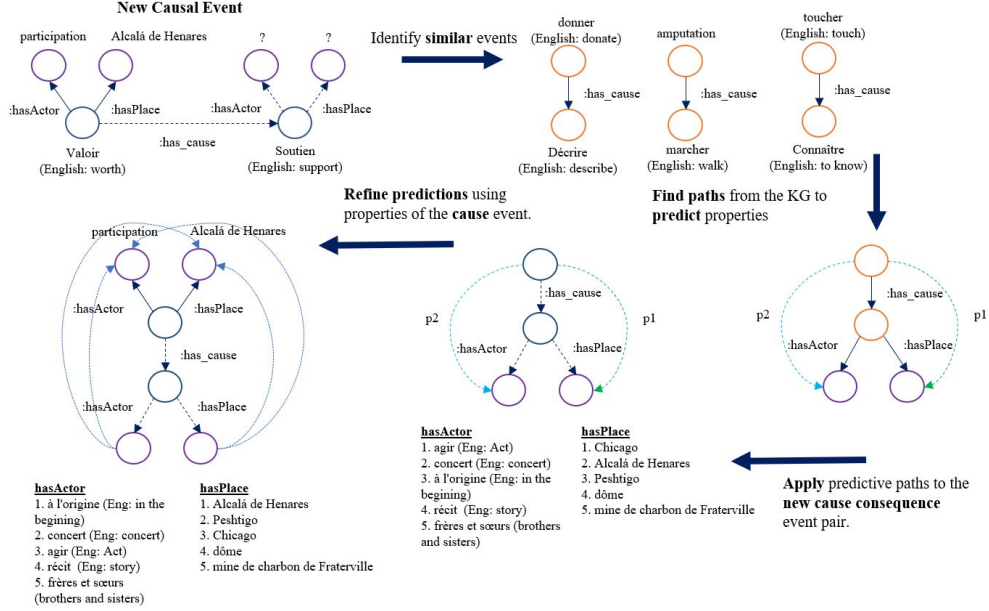


Figure 7. Example of Case-based reasoning approach

path-based prediction method allows EvCBR to provide more explainability than many embedding-based models [24]. Once the prediction path has been learned, starting from the query event, the path can be followed to make predictions about the unseen $E2$ event properties.

3. **Refining prediction:** The last step is to refine prediction results. The intuition behind this is that if EvCBR predictions about the $E2$ event are correct, then we should be able to follow the procedure in the reverse direction to accurately predict the input $E1$ event's properties starting from the $E2$ event. This step in reverse direction gives a way to ground the confidence of predictions in terms of the ground truth properties of the input $E1$ event.

4. Experiments

To evaluate the added value of considering 5W1H attributes of an event represented with our ontology, we have considered two EKGs, presented in section 4.1. We then assessed the interest of the attributes by carrying out various evaluations in section 4.2.

4.1. Considered datasets

For our experimental purpose, we used considered two EKGs constructed for the MR4AP Wikipedia dataset [44] and Wikidata Causal Event dataset[24]. The MR4AP dataset and Wikidata Causal Event dataset are different from each other; one represents the events identified in sentences using Emvista MR4AP technology, whereas the other one is curated from Wikidata and Wikipedia.

Table 2. MR4AP EKG Description

Without Reasoning					
S. No.	Name of Relation	No. of Pairs	No. of triples	No. of pairs selected randomly for evaluation	No. of two hops to evaluate
1	Causal	588	With Event Pairs: 86,693 (Overall: 157,905)	100	364
2	Temporal	1625	86,693	100	324
3	Purpose	790	86,993	100	355
With Reasoning					
1	Causal	814	With Event Pairs: 21,3926 (Overall: 299,310)	100	900
2	Temporal	1625	213,926	100	647
3	Purpose	790	213,926	100	823

MR4AP dataset The MR4AP dataset was constructed by the company Emvista using its meaning representation formalism (MR4AP) [44]. MR4AP employs a subset of VerbNet roles to which some labels are added to identify temporal, spatial, discourse, and coreference relations. We populated the EKG from the MR4AP Wikipedia dataset using Event ontology and the process described in Section 3.1.2. Then, we constructed the dataset for adapted EvCBR evaluation from this EKG. As adapted EvCBR is a 2-hop LP task for properties of unseen entities, the test dataset has been split based on entities rather than individual triples. The dataset has been split into training triples, validation connection, validation triples, test connection and test triples. In the dataset, any entity contained in the training triples as part of the training set, while head entities of triples in the test validation triples are part of the test and validation sets, respectively. The test connection denotes relations from an entity in the training set to an entity in the test set. The test triples then indicate outgoing triples from an entity in the test set to one in the training set. For causal, from $E1 - E2$ event pairs, we randomly selected 100 $E2$ events connected through *has_cause* relation with $E1$ events to serve as the test set. Similarly, we repeated the process for temporal and purpose relations.

Wikidata Causal Event dataset It was curated using Wikidata and Wikipedia. This dataset has been used for the evaluation of EvCBR [24]. It has 758,857 triples and 1,953 pairs of events, encompassing 157 unique event classes. These triples connect 284 unique “cause” events to 311 unique “effect” events. From a set of cause-effect event pairs, we randomly select 100 effect events to serve as the test set.

4.2. Results and Discussion

We apply adapted EvCBR to rank predictions for each tail entity in the test set of 2-hop connections and report the mean reciprocal rank (MRR) and Hits@K metrics. We evaluate adapted EvCBR in various ways, such as the impact of using reasoner on adapted EvCBR on a causal relation dataset. Similarly, we investigate the adapted EvCBR on

Table 3. EvCBR Results on MR4AP Causal Event Dataset

Without Reasoning				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.4264	0.3709	0.5192	0.5275
EvCBR _{refinedResult}	0.3883	0.3242	0.489	0.5027
EvCBR _{base + ref res}	0.4236	0.3626	0.5192	0.5302
With Reasoning				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.6653	0.6278	0.7689	0.8022
EvCBR _{refinedResult}	0.5948	0.5433	0.7067	0.7289
EvCBR _{base + ref res}	0.6617	0.6167	0.7756	0.8044

temporal and purpose relations. The EvCBR_{base} model performs step 1 and step 2 presented in section 3.2.2, EvCBR_{refinedResult} model steps 1 and 3 and EvCBR_{base+refres} model performs all 3 steps.

4.2.1. Eval1: impact of using a reasoner

First we evaluated the impact of exploiting the ontology axioms (such as *has_consequence* is the inverse of *has_cause*) to enrich the KG with new inferred triples thanks to a reasoner⁴. In the MR4AP EKG, 127,233 triplets are inferred. Table 2 provides the detailed information about the dataset. Table 3 shows the evaluation results of three different EvCBR models on MR4AP Causal Event dataset with and without reasoning.

Overall, the results show that the EvCBR_{base} model performed the best with reasoning, while the EvCBR_{refinedResult} model performed the worst. The combination of the two models in EvCBR_{base+refres} improved the performance slightly. The introduction of reasoning improved the performance of all models across all metrics thanks to the inference of *rdf:type* relations. In the majority of cases, the range of these relations falls between 1 to 20, sometimes under 50, which boosts EvCBR performance.

4.2.2. Eval2: EventCBR on other relations (temporal and purpose)

We have evaluated the impact of reasoning on other relations. Table 4 show the performance of the EvCBR model on two datasets - MR4AP Purpose Event and Temporal Event. In the Purpose Event dataset, the model's performance improved significantly when reasoning was applied, as indicated by the increase in MRR and H@1 scores. The EvCBR_{base} model had the highest MRR and H@1 scores among all the models. In the Temporal Event dataset, the model performance also improved with reasoning, but the improvement was not as significant as in the Purpose Event dataset. The EvCBR_{base} model had the highest MRR score, while the EvCBR_{base} and EvCBR_{base + ref res} models had the highest H@1 score. Overall, the results suggest that the EvCBR model can benefit from reasoning, particularly in the Purpose Event dataset. However, further analysis and experimentation may be necessary to understand the model limitations and to suggest improvements in the prediction algorithm: identifying which axioms generate noise and which ones generate relevant new triples. For example, some inferred *rdf:type* relations are redundant and relevant to find good properties.

⁴We used the RDFS-Plus (optimised) reasoner available on graphDB, the triplestore we have chosen

Table 4. EvCBR Results on MR4AP Purpose and temporal Event Datasets

MR4AP Purpose Event Dataset								
Model	Without Reasoning				With Reasoning			
	MRR	H@1	H@10	H@100	MRR	H@1	H@10	H@100
EvCBR _{base}	0.4867	0.4403	0.5682	0.5881	0.7226	0.6961	0.7819	0.8076
EvCBR _{refinedResult}	0.4679	0.4176	0.5597	0.5739	0.6932	0.6495	0.7659	0.7782
EvCBR _{base + ref res}	0.4852	0.4375	0.5682	0.5852	0.7214	0.6887	0.7831	0.8076
MR4AP Temporal Event Dataset								
Model	Without Reasoning				With Reasoning			
	MRR	H@1	H@10	H@100	MRR	H@1	H@10	H@100
EvCBR _{base}	0.3394	0.3333	0.3519	0.3611	0.6609	0.6528	0.6768	0.6864
EvCBR _{refinedResult}	0.3274	0.3179	0.3395	0.3457	0.6576	0.6544	0.664	0.6704
EvCBR _{base + ref res}	0.3398	0.3302	0.3519	0.3611	0.6609	0.6528	0.6768	0.6848

4.2.3. Evaluation 3: impact of Ws on Wikidata Causal Event Dataset

Our objective is to investigate the impact of Ws (Who, Where, and When) in the Wikidata Causal Event dataset as it has been used to evaluate EvCBR [24]. This study requires to identify the properties (connected to the event) in the dataset that represent the Ws. For this, we replaced the Wikidata properties mentioned in [38] like “P17”, “P276”, “P625”, “P131”, “P30” with Where, “P710”, “P664”, “P112” with Who and “P585”, “P580”, “P582”, “P571”, “P576”, “P577” with When. In Table 6, this dataset is referred to as Dataset 2, whereas Dataset 1 refers to the dataset without any replacement. In Table 6, Dataset 2 results indicate a slight improvement in MRR and H@1. However, we replaced only a limited number of properties with Ws in this experiment.

Prompt 1 Classification of Wikidata Properties

Property id: ...
Label: ...
Description: ...
Analyze the Property id and label using the Wikidata property description to determine the most accurate description of its objectives from the following options:
Where: possible places.
When: specific time.
Who: active/passive entities.
Why: possible reason of cause.
How: method by which an action was performed.
E: None of the above.
F: Not sure.
Mention only the selected option in one word

Therefore, to classify additional Wikidata properties in Ws, we used the chatGPT model. We collected all Wikidata properties metadata information from Wikidata endpoint⁵ using SPARQL and classified the properties into Ws using the GPT-3.5-turbo model with the provided Prompt 1. Table 5 presents the resulting Wikidata properties classification statistics. The majority of properties are classified as option E (around

⁵<https://query.wikidata.org/sparql> (retrieved on March 05, 2023)

Table 5. Wikidata properties classification statistics using chatgpt

Categories	No. of properties classified	Matched with Wikidata Causal Event Dataset
Who	2264	1125
Where	3339	1695
When	318	183
Why	118	82
How	339	152
E: None of the above	5331	2736
F: Not sure	8	4
Total	11717	5977

Table 6. EvCBR Results on Wikidata Causal Events Dataset

Dataset 1: Wikidata Causal Events dataset				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.2019	0.1239	0.3337	0.5202
EvCBR _{refinedResult}	0.2068	0.1460	0.3166	0.4982
EvCBR _{base + ref res}	0.2080	0.1362	0.3264	0.5252
Dataset 2: Wikidata Causal Events dataset replacing some relation with who, where, when				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.2023	0.1397	0.3118	0.4838
EvCBR _{refinedResult}	0.2134	0.1708	0.2962	0.4554
EvCBR _{base + ref res}	0.2084	0.1522	0.3118	0.4851
Dataset 3: Results related to replacing some relation with frequent W's				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.1973	0.1153	0.3362	0.5117
EvCBR _{refinedResult}	0.2008	0.1411	0.3031	0.4785
EvCBR _{base + ref res}	0.2028	0.1276	0.3252	0.5129

45.49%), Where (around 28.49%) and Who (around 19.32%) followed by How (around 2.89%), When (around 2.71%), Why (around 1.0%) and F (around 0.06%).

Next, we selected the 50 most frequent Wikidata properties from the Wikidata Causal Event dataset and associated labels from chatbot classification, then performed the manual annotations. Out of 50, 20 are classified as Where, 10 as Who, followed by 4 as When, 0 as Why, and 0 as How. Then, we replaced these 34 classified properties with ontological properties: `sem:hasPlace` (Where), `sem:hasActor` (Who), and `sem:hasTimeStamp` (When) and applied the EvCBR algorithm. The results of this experiment (Dataset 3 in Table 6), are more similar to the one of Dataset 1. Overall, the EvCBR_{refinedResult} model performed better than the base model in all three datasets. However, combining the base and refined result models led to even better results in Dataset 1 and Dataset 3. In Dataset 2, the EvCBR_{refinedResult} model achieved the best results. The reason behind the improvement in Dataset 2 and Dataset 3 is that, in some cases, the ranking of properties is improved compared to that without replacement. We can conclude that the replacement of properties can have a positive impact on the modeling process, and further exploration of this methodology may lead to even more favorable outcomes.

Table 7. EvCBR Results on Wikidata and MR4AP Causal Events Dataset

Dataset: Wikidata Causal Events dataset				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.2019	0.1239	0.3337	0.5202
EvCBR _{refinedResult}	0.2068	0.1460	0.3166	0.4982
EvCBR _{base + ref res}	0.2080	0.1362	0.3264	0.5252
Dataset: MR4AP Causal Events dataset				
Model	MRR	H@1	H@10	H@100
EvCBR _{base}	0.4264	0.3709	0.5192	0.5275
EvCBR _{refinedResult}	0.3883	0.3242	0.489	0.5027
EvCBR _{base + ref res}	0.4236	0.3626	0.5192	0.5302

4.2.4. Eval4: comparing of EventCBR on two different graphs with all the Ws

The table 7 shows the comparison of the performance of three different versions of the EvCBR model on two different datasets: Wikidata Causal Events dataset and MR4AP Causal Events dataset. The findings suggest that the EvCBR model exhibits superior performance on the MR4AP dataset than on the Wikidata dataset. Additionally, the combination algorithm EvCBR_{base + ref Res} surpasses the other two models. These outcomes showcase the potential of the EvCBR model in data analysis and its adaptability to enhance performance in diverse scenarios.

5. Conclusion

In this paper, we proposed the XPEventCore ontology to represent an event with 5W1H characteristics and used it to populate EKGs. XPEventCore integrates multiple existing event ontologies (SEM and FARO) and introduces new object properties representing “cause” and “method” to answer “Why” and “How” questions respectively. The XPEventCore ontology provides a structured and adaptable foundation for capturing the essential facets of an event and can be adapted to any domain. We further used EvCBR, a case-based reasoning approach, to enrich an EKG built using XPEventCore. We adapted it in various ways. Our analysis of the adapted EvCBR on a dataset created from our EKG revealed that reasoning significantly improved its performance, and that the base EvCBR outperformed the other variants with reasoning approaches. We also evaluated the adapted EvCBR methodology on a dataset that included other types of relationships between events, like causal, temporal, and purpose. The results showed that the performance of EvCBR varied depending on the type of relation being considered. Specifically, the performance of EvCBR was significantly better for causal relations compared to temporal and purpose relations. Further, we investigated the impact of each W (Who, Where, and When) on the performance of EvCBR on the Wikipedia Causal Event dataset. In the future, the primary objective is to expand the size of the dataset and investigate the effectiveness of generative neural networks (GNNs), embeddings, and large language models in enhancing an Event Knowledge Graph. This would enable us to develop more efficient and effective approaches to our work, thus elevating the quality of our research outcomes.

References

- [1] Guan S, Cheng X, Bai L, Zhang F, Li Z, Zeng Y, Jin X, Guo J. What is Event Knowledge Graph: A Survey. *IEEE Transactions on Knowledge and Data Engineering* [Internet]. 2022;1–20. Available from: <http://dx.doi.org/10.1109/tkde.2022.3180362>
- [2] Rudnik C, Ehrhart T, Ferret O, Teysou D, Troncy R, Tannier X. Searching News Articles Using an Event Knowledge Graph Leveraged by Wikidata. *Companion Proceedings of The 2019 World Wide Web Conference* [Internet]. 2019 May 13; Available from: <http://dx.doi.org/10.1145/3308560.3316761>
- [3] Yang C, Li W, Zhang X, Zhang R, Qi G. A Temporal Semantic Search System for Traditional Chinese Medicine Based on Temporal Knowledge Graphs. *Semantic Technology* [Internet]. 2020;13–20. Available from: http://dx.doi.org/10.1007/978-981-15-3412-6_2
- [4] Souza Costa T, Gottschalk S, Demidova E. Event-QA. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* [Internet]. 2020 Oct 19; Available from: <http://dx.doi.org/10.1145/3340531.3412760>
- [5] Wu J, Zhu X, Zhang C, Hu Z. Event-centric Tourism Knowledge Graph—A Case Study of Hainan. *Lecture Notes in Computer Science* [Internet]. 2020;3–15. Available from: http://dx.doi.org/10.1007/978-3-030-55130-8_1
- [6] Cheng D, Yang F, Wang X, Zhang Y, Zhang L. Knowledge Graph-based Event Embedding Framework for Financial Quantitative Investments. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* [Internet]. 2020 Jul 25; Available from: <http://dx.doi.org/10.1145/3397271.3401427>
- [7] Colas A, Sadeghian A, Wang Y, Wang DZ. Eventnarrative: A large-scale event-centric dataset for knowledge graph-to-text generation. *arXiv preprint arXiv:2111.00276*. 2021 Oct 30.
- [8] Allan J, Papka R, Lavrenko V. On-line new event detection and tracking. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* [Internet]. 1998 Aug; Available from: <http://dx.doi.org/10.1145/290941.290954>
- [9] Kolajo T, Daramola O. Human-centric and semantics-based explainable event detection: a survey. *Artificial Intelligence Review* [Internet]. 2023 Jun 22;56(S1):119–58. Available from: <http://dx.doi.org/10.1007/s10462-023-10525-0>
- [10] Subagdja B, Shanthoshigaa D, Wang Z, Tan AH. Machine Learning for Refining Knowledge Graphs: A Survey. *ACM Computing Surveys*. 2024 Feb 23;56(6):1–38.
- [11] Yves R, Samer A, Mark S, Frederick G. The music ontology. In *Proceedings of the International Conference on Music Information Retrieval, ISMIR 2007* (Vol. 2007, pp. 417–422).
- [12] Shaw R, Troncy R, Hardman L. LODÉ: Linking Open Descriptions of Events. *Lecture Notes in Computer Science* [Internet]. 2009;153–67. Available from: http://dx.doi.org/10.1007/978-3-642-10871-6_11
- [13] Scherp A, Franz T, Saathoff C, Staab S. F—a model of events based on the foundational ontology dolce+DnS ultralight. *Proceedings of the fifth international conference on Knowledge capture* [Internet]. 2009 Sep; Available from: <http://dx.doi.org/10.1145/1597735.1597760>
- [14] Van Hage WR, Malaisé V, Segers R, Hollink L, Schreiber G. Design and use of the Simple Event Model (SEM). *Journal of Web Semantics* [Internet]. 2011 Jul;9(2):128–36. Available from: <http://dx.doi.org/10.1016/j.websem.2011.03.003>
- [15] Grenon P, Smith B. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition & Computation* [Internet]. 2004 Mar;4(1):69–104. Available from: http://dx.doi.org/10.1207/s15427633scc0401_5
- [16] Guizzardi G, Halpin T. Ontological foundations for conceptual modelling. *Applied Ontology* [Internet]. 2008;3(1–2):1–12. Available from: <http://dx.doi.org/10.3233/ao-2008-0049>
- [17] Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L. Sweetening Ontologies with DOLCE. *Lecture Notes in Computer Science* [Internet]. 2002;166–81. Available from: http://dx.doi.org/10.1007/3-540-45810-7_18
- [18] Mizoguchi R, Borgo S. YAMATO: Yet-another more advanced top-level ontology. Borgo S, Galton A, Kutz O, editors. *Applied Ontology* [Internet]. 2022 Mar 15;17(1):211–32. Available from: <http://dx.doi.org/10.3233/ao-210257>
- [19] Niles I, Pease A. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001* 2001 Oct 17 (pp. 2–9).
- [20] Doerr M. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine*. 2003 Sep 15;24(3):75–.

April 2024

- [21] Poli R, Healy M, Kameas A. Theory and Applications of Ontology: Computer Applications [Internet]. Springer Netherlands; 2010. Available from: <http://dx.doi.org/10.1007/978-90-481-8847-5>
- [22] Piryani R, Aussenac-Gilles N, Hernandez N. Comprehensive Survey on Ontologies about Event. In-ESWC Workshops on Semantic Methods for Events and Stories (SEMMES @ ESWC 2023) 2023 May 28 (Vol. 3443, pp. 1-15). ceur-ws.org.
- [23] Rebboud Y, Lisená P, Troncy R. Beyond causality: Representing event relations in knowledge graphs. In International Conference on Knowledge Engineering and Knowledge Management 2022 Sep 20 (pp. 121-135). Cham: Springer International Publishing.
- [24] Shirai S, Bhattacharjya D, Hassanzadeh O. Event Prediction using Case-Based Reasoning over Knowledge Graphs. Proceedings of the ACM Web Conference 2023 [Internet]. 2023 Apr 30; Available from: <http://dx.doi.org/10.1145/3543507.3583201>
- [25] Chen X, Li Q. Event modeling and mining: a long journey toward explainable events. The VLDB Journal [Internet]. 2019 Jul 1;29(1):459–82. Available from: <http://dx.doi.org/10.1007/s00778-019-00545-0>
- [26] Hamborg F, Breiting C, Gipp B. Giveme5w1h: A universal system for extracting main events from news articles. arXiv preprint arXiv:1909.02766. 2019 Sep 6.
- [27] Hasan M, Orgun MA, Schwitter R. A survey on real-time event detection from the Twitter data stream. Journal of Information Science [Internet]. 2017 Mar 17;44(4):443–63. Available from: <http://dx.doi.org/10.1177/0165551517698564>
- [28] Hakeem A, Sheikh Y, Shah M. CASEE: a hierarchical event representation for the analysis of videos. In AAAI 2004 Jul 25 (pp. 263-268).
- [29] Dou W, Wang X, Ribarsky W, Zhou M. Event detection in social media data. In IEEE VisWeek workshop on interactive visual text analytics-task driven analytics of social media content 2012 Oct (pp. 971-980).
- [30] Wang W, Zhao D, Zou L, Wang D, Zheng W. Extracting 5W1H Event Semantic Elements from Chinese Online News. Lecture Notes in Computer Science [Internet]. 2010; 644–55. Available from: http://dx.doi.org/10.1007/978-3-642-14246-8_62
- [31] Lagoze C, Hunter J. The ABC Ontology and Model, Journal of Digital Information. Article No. 77, 2001-11-06. 2001;2(2).
- [32] Kejrival M, Knoblock CA, Szekely P. Knowledge graphs: Fundamentals, techniques, and applications. MIT Press; 2021 Mar 30.
- [33] Pujara J, Miao H, Getoor L, Cohen W. Knowledge Graph Identification. Lecture Notes in Computer Science [Internet]. 2013;542–57. Available from: http://dx.doi.org/10.1007/978-3-642-41335-3_34
- [34] Wang M, Qiu L, Wang X. A Survey on Knowledge Graph Embeddings for Link Prediction. Symmetry [Internet]. 2021 Mar 16;13(3):485. Available from: <http://dx.doi.org/10.3390/sym13030485>
- [35] Wang Q, Mao Z, Wang B, Guo L. Knowledge Graph Embedding: A Survey of Approaches and Applications. IEEE Transactions on Knowledge and Data Engineering [Internet]. 2017 Dec 1;29(12):2724–43. Available from: <http://dx.doi.org/10.1109/tkde.2017.2754499>
- [36] Ruffinelli D, Broscheit S, Gemulla R. You can teach an old dog new tricks! on training knowledge graph embeddings.
- [37] van Hage WR, Malaisé V, de Vries GKD, Schreiber G, van Someren MW. Abstracting and reasoning over ship trajectories and web data with the Simple Event Model (SEM). Multimedia Tools and Applications [Internet]. 2011 Jan 7;57(1):175–97. Available from: <http://dx.doi.org/10.1007/s11042-010-0680-2>
- [38] Blin I. Building Narrative Structures from Knowledge Graphs. Lecture Notes in Computer Science [Internet]. 2022;234–51. Available from: http://dx.doi.org/10.1007/978-3-031-11609-4_38
- [39] Gottschalk S. Creation, Enrichment and Application of Knowledge Graphs.
- [40] Chen L, Liu D, Yang J, Jiang M, Liu S, Wang Y. Construction and application of COVID-19 infectors activity information knowledge graph. Computers in Biology and Medicine [Internet]. 2022 Sep;148:105908. Available from: <http://dx.doi.org/10.1016/j.combiomed.2022.105908>
- [41] Ma Y, Wang Z, Li M, Cao Y, Chen M, Li X, Sun W, Deng K, Wang K, Sun A, Shao J. MMEKG: Multi-modal Event Knowledge Graph towards Universal Representation across Modalities. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations [Internet]. 2022; Available from: <http://dx.doi.org/10.18653/v1/2022.acl-demo.23>
- [42] Bosanska DC, Huptych M, Lhotská L. A Pipeline for Population and Analysis of Personal Health Knowledge Graphs (PHKGs). In SEMANTiCS (Posters & Demos) 2022.
- [43] Wu J, Zhu X, Zhang C, Hu Z. Event-centric Tourism Knowledge Graph—A Case Study of Hainan. Lecture Notes in Computer Science [Internet]. 2020;3–15. Available from: <http://dx.doi.org/10.1007/978-3->

April 2024

030-55130-8_1

- [44] Giordano B, Lopez C. MR4AP: Meaning representation for application purposes. In Proceedings of the Fourth International Workshop on Designing Meaning Representations 2023 Jun (pp. 110-121).