The DARPA Wikidata Overlay: Wikidata as an ontology for natural language processing

Elizabeth Spaulding[1], Kathryn Conger[1], Anatole Gershman[2], Rosario Uceda-Sosa[3], Susan Windisch Brown[1], James Pustejovsky[4], Peter Anick[4] and Martha Palmer[1]
[1]University of Colorado Boulder, [2]Language Technologies Institute, Carnegie Mellon University, [3]IBM Research, T.J.Watson, [4]Brandeis University
{elizabeth.spaulding,kathryn.conger,susan.brown,martha.palmer}@colorado.edu, anatoleg@andrew.cmu.edu, rosariou@us.ibm.com, {pustejovsky,panick}@brandeis.edu

Abstract

With 102,530,067 items currently in its crowd-sourced knowledge base, Wikidata provides NLP practitioners a unique and powerful resource for inference and reasoning over real-world entities. However, because Wikidata is very entity focused, *events* and *actions* are often labeled with eventive nouns (e.g., the process of diagnosing a person's illness is labeled "diagnosis"), and the typical participants in an event are not described or linked to that event concept (e.g., the medical professional or patient). Motivated by a need for an adaptable, comprehensive, domain-flexible ontology for information extraction, including identifying the roles entities are playing in an event, we present a curated subset of Wikidata in which events have been enriched with PropBank roles. To enable richer narrative understanding between events from Wikidata concepts, we have also provided a comprehensive mapping from temporal Qnodes and Pnodes to the Allen Interval Temporal Logic relations.

1 Introduction

An ontology is a necessary framework for practical system representation of domain-specific world knowledge. However, since we each perceive the world somewhat differently, and different domains require access to varying levels of granularity, a unique, universal ontology is not a realistic goal. The biomedical portal[1] alone lists 1213 different biomedical ontologies. The largest structured repository of knowledge about world entities is Wikidata (Vrandečić and Krötzsch, 2014), but, as should be expected from any collective knowledge repository, it has many inconsistencies, circular subclass links, and partially overlapping concepts and gaps. Since every practical project designed for a specific task needs a consistent ontology, our challenge is to provide a common, mutually agreed upon vocabulary that speeds up the process of incorporating new domains or evolving existing ones by automatically leveraging existing knowledge bases, such as Wikidata.

We introduce the DARPA Wikidata Overlay (DWD Overlay), a curated subset of Wikidata enriched with PropBank (Kingsbury and Palmer, 2002) roles. We chose Wikidata as our primary resource for ontological concepts because of its extensive coverage of concepts, its linking of those concepts to each other, and the ability to contribute additional concepts to the database as needed. Importantly, Wikidata concepts are linked to relevant Wikipedia entries, a distinct advantage for NLP applications concerned with current events. We turned to PropBank as our source of participant roles because of its wide coverage of verbs and eventive nouns that we could easily match to Wikidata concepts and because its roles could be represented as both broad, general roles (e.g., ARG0, ARG1) and as more event-specific (e.g., attacker, victim). Further, to enable richer narrative understanding between events, we have provided a comprehensive mapping from temporal Qnodes and Pnodes to the Allen Interval Temporal Logic relations (Allen, 1983, 1984). While Wikidata is a multilingual project, and PropBank is a language-independent semantic description, both resources use a mainly English interface and semantic roles from resources like PropBank can be, to some degree, language-specific (Burchardt et al., 2006).

The overlay is currently hosted in a JSON format[2], designed to give users the ability to browse concepts and easily ingest the ontology structure into their computing applications. We hope that by establishing a robust and accurate mapping be-

[1]https://bioportal.bioontology.org/ontologies

[2]https://github.com/e-spaulding/xpo

Proceedings of the 19th Joint ACL - ISO Workshop on Interoperable Semantic Annotation, pages 1–10
June 20, 2023.

tween PropBank and Wikidata, we enable the usage of well-established NLP methods for event extraction using PropBank combined with the inference power and massive coverage of Wikidata.

2 Background and Related Work

Wikidata[3] is a large, crowd-sourced knowledge base. Each item in Wikidata refers to a either a concept ("president") or a real-world instantiation of a concept ("Joe Biden"), and is called a *Qnode*. Qnodes are connected to one another via *Pnodes*, which represent the relation in <*subject, relation, object*> triples in Wikidata.

The original impetus for using Wikidata to support DARPA programs came from the DARPA AIDA and KAIROS programs, closely followed by DARPA MAA. Each of the aforementioned DARPA programs is described in more detail below, as well the motivation for coming to a consensus on an approach to ontology development that could be quickly adapted to new domains.

Active Interpretation of Disparate Alternatives (AIDA), now completed, aimed at the organization of natural language news and social network information into competing, alternative hypotheses (narratives) about events and situations. AIDA systems integrated multi-modal knowledge elements into a common semantic representation suitable for hypothesis generation. Task Area 1 (TA1) performers applied cutting-edge NLP techniques to extract and co-refer knowledge elements from streaming text, images and videos, producing entries in a knowledge base (KB) for each input document. The program goals originally included the challenging restriction of TA2's and TA3's performing their tasks without access to raw input data. Without consulting the original TA1 inputs, TA2 performers link together entity and event entries from individual documents into a unified KB using cross-document co-reference techniques. The TA1's could provide TA2's the names of named entities but could not originally provide TA2's with the lexical items or images for other types of entities. Instead, TA2's passed along type information for entity and event entries from an Ontology. This made the development of a program-wide ontology that all performers could utilize a high priority. TA3 performers then mined the unified TA2 KB for competing hypotheses.

At the outset of the program, AIDA worked on developing an expressive, semantic Program Ontology that could be used by performers to encode and exchange KBs and hypotheses. Midway through the program, the Program Ontology contained hundreds of entity, relation, and event types developed using a data-driven approach inspired by pre-existing knowledge resources. These included previous Linguistic Data Consortium (LDC) annotation efforts, such as ACE (Strassel and Mitchell, 2003; Doddington et al., 2004; Song and Strassel, 2008) and ERE (Aguilar et al., 2014; Song et al., 2015) and their extensions for other programs such as DEFT [4], as well as publicly available resources such as YAGO (Suchanek et al., 2007; Hoffart et al., 2013; Mahdisoltani et al., 2015; Pellissier Tanon et al., 2020), FrameNet (Fillmore and Baker, 2009), PropBank (Kingsbury and Palmer, 2002), VerbNet (Kipper-Schuler, 2005) and the Reference Event Ontology (Brown et al., 2017).

An ongoing tension existed between performers' desires for expressive, expansive ontology models and LDC's need for a manageable ontology supporting cost-effective corpus construction and annotation that can assist in evaluation of resulting system output (Tracey et al., 2022). AIDA's solution was to charge LDC with selecting from previous programs those elements of the Program Ontology that would support salient entities and events in current program evaluation scenarios. This approach resulted in a patchwork AIDA Annotation Ontology where the connections between different ontological elements were sometimes obscured. For instance, Geographical Areas and Geographical Points were suggested in the Program Ontology, but without clear definitions. LDC chose Geographical Points to define a Location subtype that could be used for Addresses. They chose Geographical Areas to define a subtype of Facility that could be used for installations covering a significant area, larger than a point, such as Borders and Checkpoints. This choice of labels helped annotators to distinguish Addresses and Borders from other types of Locations and Facilities, but can be confusing to ontologists more familiar with Geographical Point and Area as two subtypes of Spatial Region.

The MAA and KAIROS programs described below both relied heavily on the AIDA Annotation Ontology, reflecting ongoing program needs for an expressive ontological data model that can be easily

[3] https://www.wikidata.org/

[4] https://www.darpa.mil/program/deep-exploration-and-filtering-of-text

extended to new domains and evaluation scenarios.

Modeling Adversarial Activity (MAA), now completed, was directed towards mathematical and computational methods for graph alignment and merging as well as subgraph detection and subgraph matching. MAA used the AIDA ontology, and MAA graphs were direct projections of AIDA RDF graphs into a property graph format that supported efficient and scalable graph analytics developed by MAA performers. MAA predominantly used the LDC Annotation Ontology as encoded in the AIDA Interchange Format. MAA also focused on the transactional aspects of interactions in addition to entity- and event-based knowledge graphs. The MAA evaluation phase included data sources related to financial topics, e.g. scientific publications and social media, and required modeling temporal events and entities with both physical and abstract attributes. In addition to the modeling of such data sources, the AIDA Annotation Ontology was also used by MAA performers to develop approximate entity alignment and subgraph matching algorithms.

Knowledge-directed Artificial Intelligence Reasoning Over Schemas (KAIROS) is ongoing and shifts the focus from the alternative hypotheses in AIDA to extracting sequences of events with temporal structure, such as narrative schemas. The goal is an AI system that can identify, link and temporally sequence complex events and their subsidiary elements and participants. For KAIROS the TA1's induce new schemas to create a library of schemas, and the TA2's are supposed to detect instances of these schemas in data. Since the schemas are intended to abstract away from the specific words and phrases that initially indicate them, there is a similar reliance on ontological types. A major focus of the first phase of KAIROS was the identification and definition of a set of Event Primitives that can comprise the schema elements. Most of these are recycled from AIDA, although sometimes at a more coarse-grained level. New ones have also been defined. During this effort, additional argument slots were added to many of the AIDA events.

The suggestion of shifting focus to Wikidata was made during the attempts to merge AIDA entity and event types into a nascent KAIROS ontology. Trying to quickly expand an existing although partial AIDA ontology to cover new domains highlighted its gaps as well as the difficulty of finding rational locations for new types without recourse to an overarching upper level ontology. Wikidata was not originally expected or intended to follow good principles of ontology development (Noy and Mcguinness, 2001), but a lot of effort on the part of many conscientious contributors had resulted in a reasonable approximation. After a few successful experiments with mapping the existing AIDA and KAIROS Entity and Event types to Wikdata Qnodes, a Cross-Program Ontology subcommittee was formed. DARPA approved the subcommittee's proposal to adopt Wikidata as a shared, general resource for entity and event identification, and the DARPA Wikidata Overlay was born.

The DWD Overlay should be contrasted with DARPA Wikidata (DWD), which is a large Wikidata dump adhering to the AIDA "Time Machine" constraint: due to the program's strict evaluation schedule, to properly track the inferences the systems are making automatically, it is sometimes necessary to ensure that program performers do not have access to vital information that they are supposed to detect or induce automatically. Thus, the DWD takes a large portion of Wikidata restricted to information before 2010. The DWD itself has enabled research on knowledge graphs (Wang et al., 2022). The overlay started by pulling only from DWD during the programs, but has expanded into the full Wikidata catalog.

3 Methodology

Node type	Total	Top level	PropBank
Entities	276	68	0
Events	5,167	479	5,164
Relations	216	152	144
Temporal relations	8	8	1

Table 1: Current coverage of the overlay, version 5.4.5.

The first task in the shift from the comparatively small, domain-specific LDC annotation tagset to Wikidata involved manually mapping the 200+ AIDA/KAIROS LDC Entity, Relation and Event types, subtypes and sub-subtypes to Wikidata Qnodes. Every such mapping was subject to at least two passes from human curators, sometimes with conflicts generating extensive discussion with a larger group. In cases of dispute between a Qnode and its superclass, the superclass was selected to ensure wider coverage. Existing Entities, Relations and Events were carefully examined in turn, and an upper-middle level ontology for each category was manually extracted from Wikidata, and subjected to careful vetting, to simplify down stream inference tasks.

Because Wikidata Qnodes are especially oriented towards entities rather than events, entities were relatively straightforward. Events are difficult to delineate and place in hierarchies, making their representation inconsistent across ontologies. Mapping AIDA/KAIROS events to Wikidata Qnodes was therefore unsurprisingly more difficult than mapping the entities. Several AIDA/KAIROS event types were found to have no plausible Wikidata Qnode. Mapping the AIDA/KAIROS relations to Wikidata also required careful manual effort. As many relations as possible were mapped to Pnodes. However, in cases where no Pnode could be found for an AIDA/KAIROS relation, it was mapped to a Qnode.

3.1 Enriching Wikidata events with PropBank roles

Step 1: a semi-automatic mapping. Because Wikidata provides no information about the participants or arguments of an event, we added this information semi-automatically to an expansion of the original 132 DWD events that were based on LDC event types. Around 5,000 additional Qnodes were identified as event classes (e.g., Q7944 "earthquake" – a class vs. Q211386 "1906 San Francisco Earthquake" – an instance) and linked to PropBank rolesets using rules. In Wikidata, class Qnodes have a "subclass of" property that points to one or more class Qnodes (e.g., Q7944 "earthquake" is a subclass of Q8065 "natural disaster"). A Qnode was considered an "event candidate" if it was a descendant (a direct or indirect subclass) of the Wikidata event Q1190554 "occurrence". This filtering produced close to 30,000 "event candidates" many of which we would not consider events. For example, in Wikidata, Q18534 "metaphor" is a descendant from "occurrence". We created an exclusion list of 11 high-level non-events such as Q223557 "physical object" and eliminated all event candidates that descended only from the Qnodes on the exclusion list. After some manual editing, we ended up with about 4,500 Qnode events.

Next, we used PropBank rolesets to create argument frames for the event Qnodes. We used lexical matching of the node and roleset labels and aliases to obtain a rough mapping of Qnodes to rolesets. When a Qnode did not lexically match any rolesets, we ascended the class hierarchy to find the nearest ancestor with a roleset mapping. This resulted in many events mapped to the same roleset, e.g., many specific diseases mapped to ill.01. While this produced reasonable argument frames for most of the event Qnodes, it was a noisy mapping from the PropBank rolesets to Qnodes with many rolesets mapped to multiple Qnodes. The number of Qnodes per roleset was somewhat reduced by excluding the subclasses of the mapped nodes, e.g., if roleset R was mapped to Q1 and Q2 and Q2 was a descendant of Q1, Q2 was deleted from the mapping. But that still left many one-to-many mappings and quite a few one-to-one mappings were not optimal.

Step 2: comprehensive annotation. Because these one-to-many mappings presented a problem for performers, a manual review was initiated for the PropBank-Wikidata mappings, starting with those PropBank rolesets that map to more than 10 Wikidata nodes. The review expanded into an ongoing comprehensive annotation project. Annotators evaluate the degree to which existing Wikidata Qnodes match each PropBank roleset (there are 11,277 rolesets total, the 5000 from above are being reviewed first). Existing Qnodes that closely match the general meaning and granularity of a PropBank roleset are preserved. In cases where no suitable Qnode can be found for a roleset, annotators recommend adding a new Qnode to Wikidata itself to match the sense of the roleset exactly. Additionally, when a Wikidata Qnode and a PropBank roleset are related but differ in scope, annotators document the cause of a mismatch for use in creating more fine-grained mapping relationships. Table 2 summarizes the progress of the mapping. Finally, annotators check for incorrect semi-automatic mappings, ensuring that only high quality mapping are retained.

Event templates. Finally, every event in the overlay is enriched with event templates based on their PropBank mapping, which provide a way to induce past-tense natural language sentences from extracted events with slots filled. For example, Q11398090 "creation" has this template:

```
<A0_pag_creator> created <A1_ppt_thing_
created> using <A2_vsp_materials_used>
at <AM_loc>
```

Just like with the PropBank roles, templates were first automatically generated and then a slower manual curation process was initiated to vet the templates and ensure quality. Some automatically generated templates are not grammatical but are still included for broader, albeit noisier, coverage. The templates can be used in encoder-decoder models for argument extraction (Li et al., 2021; Du et al., 2022) as well as for easier human browsing and analysis in both the overlay itself and after event extraction.

3.2 Enriching Wikidata relations with PropBank roles

LDC began with a small amount of relations geared towards specific domains, spanning topics such as affiliations, locations, personal relationships, measurements, and part-whole relations. The relations worked for the domains they were built for, but were not comprehensive enough for open domain text, and the hierarchical structure was geared more towards ease of annotation rather than robust, principled ontological representation.

Relations are represented in Wikidata as Pnodes (P for "property") which allow for relational $<$*subject, Pnode, object*$>$ triples to act as the main expressive component (triples are called "statements") in Wikidata.

Manually mapping Pnodes to PropBank. Often, Pnodes lend themselves to mapping to PropBank roles (e.g. P50 *author* maps cleanly to two roles in the author.01 roleset in PropBank). To offer more support for relation extraction and inference using the overlay, we began mapping these relation Pnodes to PropBank rolesets, as well. Out of the 216 relations currently listed in the overlay, 144 have PropBank mappings. Other Pnodes do not map easily to PropBank; for example, P1120 "number of deaths" implies a more complex event causing multiple deaths that doesn't correspond to a single verb, and the roleset "die.01" doesn't necessarily imply multiple deaths and thus does not have a specific slot for quantity. Future work may include additional event decompositions of such Pnodes, taking causality into account, but is not included in the current version of the overlay.

3.3 Event-event relations

In addition to the mapping of the original LDC Relations and the additional Pnode relations, special attention was paid to temporal relations. The overlay identifies a handful of Wikidata Qnodes and Pnodes as temporal relations based on Allen's Interval Temporal Logic (Allen, 1983, 1984), exemplified in Figure 1.

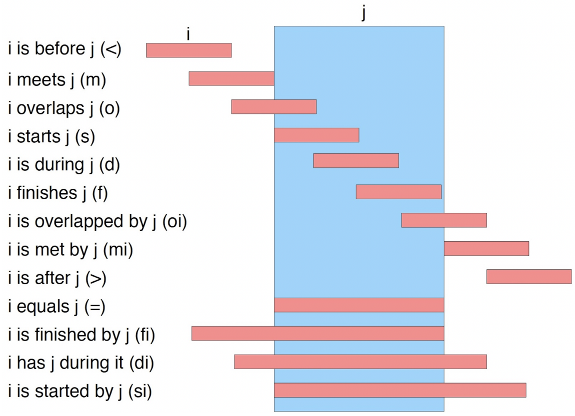

Figure 1: The full 13 relations from Allen's Interval Temporal Logic (Allen, 1983).

Event reasoning requires temporal reasoning, which is concerned with representing and reasoning about both anchoring and ordering relationships between temporal intervals and events. Temporally situating events in a narrative involves two strategies: establishing a *relative ordering* of the events to each other, and a *temporal anchoring* of each event relative to a fixed time, such as an overt temporal expression, *yesterday*, or Reichenbach's speech time (Reichenbach, 1947). To this end, we adopt Allen's interval temporal logic (Allen, 1983, 1984), which is an attempt to model events directly in a temporal relation calculus. In this system, temporal intervals are considered primitives, while constraints (e.g., on actions) are expressed as relations between intervals. There are 13 basic (binary) interval relations, where six are inverses of the other, excluding equality.

Allen's interval-based notion of events also forms the interpretive core of TimeML (Pustejovsky et al., 2003), ISO-TimeML (Pustejovsky, 2017), the multilingual resources built on ISO-TimeML community (Im et al., 2009; Bittar et al., 2011; Caselli et al., 2011), as well as the shared tasks based on ISO-TimeML (Verhagen et al., 2007, 2010; UzZaman et al., 2012). The representation of events as reified intervals with constraints can be mapped to formal calculi used in temporal reasoning, e.g., DAML-Time (Hobbs and Pustejovsky, 2003), as well as Interval Temporal Logic (Pratt-Hartmann, 2007). This strategy also allows one to interpret the ordering of events in discourse and narratives as an interval constraint satisfaction problem, which has had a significant influence on recog-

	Semi-automatic	Human	Total mappings
Original (v5.3.0)	4,567	136	4,703
Mapping changed	*- 121*	*+ 121*	
Mapping retained	*- 406*	*+ 406*	*= 989 unique Qnodes*
Mapping added	*- 0*	*+ 462*	*covering 1,089 rolesets*
Current (v5.4.5)	4,040	1,125	5,165
New Qnodes recommended		2,792	

Table 2: A summary of the progress on PropBank-Wikidata annotation integration into the DWD overlay. *Italics* show human-curated PropBank-Qnode mappings: either retained from the original semi-automatic mapping, changed from the original, or added from outside the original 4.5k). The bottom row shows that around 2,800 rolesets were found the have no plausible Qnode by humans, and a Qnode addition to Wikidata was recommended.

nizing narrative event chains and identifying event schemas (Chambers and Jurafsky, 2008, 2009), as well as more recent work on script learning and frame induction (Cheung et al., 2013; Pichotta and Mooney, 2014).

Q/Pnode	Label	Allen interval
Q79030196	before	i is before j
P156	followed by	i meets j
P155	follows	i is met by j
P1382	partially coincident with	i and j partially overlap
Q6014822	inclusion	i occurs within j
Q79030284	after	i is after j
Q842346	equality	i equals j

Table 3: Wikidata nodes that represent temporal relations based on Allen intervals

4 Discussion

Use cases and limitations. The DWD overlay has mainly been used (Zhan et al., 2023) as the primary resource for general-purpose event extraction. The PropBank mappings in the overlay enabled Zhan et al. (2023) to create a large dataset starting from PropBank annotations and ending with Wikidata event Qnodes chosen from our mappings by Mechanical Turk workers.

Although the overlay has facilitated these advances in event extraction, many limitations have been identified. The many-to-one mappings (many Qnodes per PropBank roleset) proved to be the biggest limiting factor, as well as inaccuracy of automatic mappings, and the inclusion of low frequency nodes. The ongoing annotation project described in Section 3.1 should ameliorate these limitations.

Tension between resources. The same advantages we gain from combining Wikidata with PropBank—extensive coverage of real-world entities and concepts, plus a large, rich set of participant roles for events—create the largest problems. Both resources are powerful as NLP tools in their own right, each created for a slightly different purpose under a slightly different ethos. As discussed in Section 3, PropBank is action and event oriented, while Wikidata is oriented towards entities, more often containing nominalizations and nominal forms of events, if an event is represented at *all*.

When determining how entity-denoting nodes in Wikidata are to be mapped to PropBank event predicates, it is useful to examine how events are lexicalized in language. For English, the most frequent lexical realization of an event is a predicative verb, e.g., *eat*, *sink*, *write*, *sign*. This is followed by event nominalizations (e.g., *arrival*, *explosion*, *decay*) and activity nominalizations (e.g., *eating*, *sinking*, *writing*), and finally event nominals, e.g., *meal*, *war*, *accident*. As mentioned above, since Wikidata is largely organized around reified (entity-centric) conceptual nodes, it is not surprising that both nominalizations and nominal forms are more commonly represented as Qnodes for event denotations. For example, event predicates denoting activities that have clearly unambiguous nominalizations can be found represented as Qnodes in Wikidata:

- **Activity Nominalization**:
  eat.01 - eating_Q213449
  sink.01 - sinking_Q30880545
  write.01 - writing_Q86647781
  sign.01 - * "signing" is a specialized sense

In these cases, the PropBank mapping is easy. Many reifications of event nominalizations in Wikidata, however, tend to denote the result of the event, rather than the activity or event itself:

- **Result Nominalization**:

sign.01 - sign_Q3695082/signature_Q188675
dream.01 - dream_Q36348

Hence, mapping from Wikidata concepts to PropBank participant roles requires discernment between concepts that map to events themselves and concepts that should fill participant slots of events that are not represented in Wikidata at all.

Deciding when to allow imperfect mappings for the sake of coverage, yet at the expense of semantic integrity, has been a constant tension in the annotation project. For this reason, our annotators have recommended adding thousands of Qnodes to Wikidata itself to match the sense of PropBank rolesets exactly.

5 Future Work

The PropBank-Wikidata annotation project is still ongoing. The overlay is expected to become higher quality and less noisy as the project progresses. However, we hope to eventually *retire* the overlay by integrating it into Wikidata itself. Integrating our mappings into Wikidata itself will allow maintenance to be handled by the crowd-sourcers that already maintain Wikidata. In the meantime, we anticipate that our unique resource provides opportunities for further advancements in the field of semantic annotation and ontologies for natural language processing.

5.1 Adding event structures to Wikidata itself

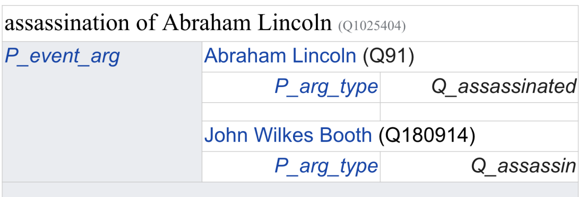

Figure 2: Sample of a Wikidata statement including proposed Pnodes and Qnodes for event arguments based on PropBank. Proposed nodes in *italics*.

We plan to incorporate our mapped PropBank roles into Wikidata itself. By moving these roles into Wikidata, researchers will eventually be able to use Wikidata directly and repeated updating of the DWD would not be necessary. In discussions with Wikidata, it was suggested we hire a Wikidata consultant—someone who is already a frequent contributor to Wikidata—to assist in adding information to Wikidata itself. It was also decided to first release this addition to Wikidata as an appendix. This would allow users to try the enhancement before fully altering the main Wikidata structure.

Specifically, we propose using new special-purpose Qnodes to represent event arguments in Wikidata. For example, the 'killer' in Q844482 (killing) will be Q_Q844482_killer (with the appropriate number replacing the '_Q844482_killer' part). These "event role" Qnodes will include the following proposed Pnodes: P_role_index, P_role_function, P_role_description, P_role_in, and P_selectional_preference, which are shown in Table 4 exemplifying their usage for the proposed *killer* Qnode.

Multiple statements with P_selectional _preference should be interpreted as an "OR", i.e., the filler of the role slot should descend from at least one of the selectional preference Qnodes. The meaning of "descend" could be application-specific, but, generally, we mean a combination of "subclass of", "parent taxon" and "instance of" properties.

Once we complete the mapping of the PropBank rolesets to Wikidata Qnodes, we can create the event role Qnodes automatically. Since there are about 11,400 PropBank rolesets with 2-4 roles each, we can expect about 25,000-40,000 new event role Qnodes. It might also be possible to cluster the event role Qnodes and create a "subclass of" hierarchy. We want to stress that the proposed event role Qnodes are not lexical or grammatical constructs. The existence of a killer in a killing event is not tied to any language or grammar. It is a part of the "killing" concept.

Wikidata contains many Qnodes representing event instances. For example, Q1025404 (assassination of Abraham Lincoln) is an instance of Q3882219 (assassination). Our proposal will create Q_assassin and Q_assassinated event role Qnodes. We propose to create one new property P_event_arg with a qualifier P_arg_type to represent the roles in an event instance, which we show in Figure 2.

In the process of mapping PropBank to Wikidata, we have identified hundreds of gaps in coverage in the Wikidata event hierarchy. Therefore, we additionally plan to add event Qnodes where our annotators noted they could find no matching Qnodes for a particular PropBank roleset.

5.2 Evaluation of ontologies

Ontologies can be *formally* evaluated via principles (Oltramari et al., 2010). These modes of evalu-

Proposed Pnode	Possible values	Value for Q_killer
role index	0, 1, 2, ... or "M"	0
role function	a Qnode representing PropBank role functions	Q392648 (agent)
role description	a string	"killer"
role in	the event class Qnode	Q844482 (killing)
selectional preference	a Qnode which stipulates the ancestor of the potential role filler	Q5 (human)

Table 4: Our proposed Pnode additions to Wikidata that give information for the event arguments of event role Qnodes.

ation are time- and resource-consuming, requiring philosophical training and manual human effort. Another mode of evaluating ontologies is application-based: one can rank ontologies based on metrics used for applications of the ontologies themselves. From this observation, a few different research questions that we could answer with our overlay emerge: do ontologies that receive a positive formal evaluation also perform well on NLP tasks? Has our manual curation work resulted in better downstream performance, or a better evaluation using formal principles? We would like to address these questions in future work, along with exploring different ways of incorporating this resource into downstream applications.

6 Conclusion

We introduced the DWD Overlay, a curated subset of Wikidata enriched with PropBank roles for use as an ontology for natural language processing. Our mapping combines the extensive coverage of ontological concepts and the inference power of Wikidata with participant roles in PropBank, providing a comprehensive, open domain resource for information extraction especially geared toward natural language newstext. While the DWD Overlay already includes 1,125 manually curated Qnode-PropBank mappings and 4,040 semi-automatically induced Qnode-PropBank mappings, event templates for every event, as well as mappings to Allen Interval Temporal Logic relations, the human annotation is still a work in progress and the overlay is expected to continue to increase in quality.

Acknowledgments

We gratefully acknowledge the support of DARPA FA8750-18-2-0016-AIDA – RAMFIS: Representations of vectors and Abstract Meanings for Information Synthesis and a sub award from RPI on DARPA KAIROS Program No. FA8750-19-2-1004. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA or the U.S. government.

References

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 45–53, Baltimore, Maryland, USA. Association for Computational Linguistics.

James F. Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

James F Allen. 1984. Towards a general theory of action and time. *Artificial intelligence*, 23(2):123–154.

André Bittar, Pascal Amsili, Pascal Denis, and Laurence Danlos. 2011. French timebank: an iso-timeml annotated reference corpus. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 130–134. Association for Computational Linguistics.

Susan Windisch Brown, Claire Bonial, Leo Obrst, and Martha Palmer. 2017. The rich event ontology. In *Proceedings of the Events and Stories in the News Workshop*, pages 87–97.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado, and Manfred Pinkal. 2006. The salsa corpus: a german corpus resource for lexical semantics. In *Proceedings of LREC*, Genoa, Italy.

Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. 2011. Annotating events, temporal expressions and relations in Italian: the It-TimeML experience for the Ita-TimeBank. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 143–151. Association for Computational Linguistics.

Nathanael Chambers and Dan Jurafsky. 2008. Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the Conference on Empiri-*

cal Methods in Natural Language Processing, pages 698–706. Association for Computational Linguistics.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 602–610. Association for Computational Linguistics.

Jackie Chi Kit Cheung, Hoifung Poon, and Lucy Vanderwende. 2013. Probabilistic frame induction. *arXiv preprint arXiv:1302.4813*.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Xinya Du, Zixuan Zhang, Sha Li, Pengfei Yu, Hongwei Wang, Tuan Lai, Xudong Lin, Ziqi Wang, Iris Liu, Ben Zhou, Haoyang Wen, Manling Li, Darryl Hannan, Jie Lei, Hyounghun Kim, Rotem Dror, Haoyu Wang, Michael Regan, Qi Zeng, Qing Lyu, Charles Yu, Carl Edwards, Xiaomeng Jin, Yizhu Jiao, Ghazaleh Kazeminejad, Zhenhailong Wang, Chris Callison-Burch, Mohit Bansal, Carl Vondrick, Jiawei Han, Dan Roth, Shih-Fu Chang, Martha Palmer, and Heng Ji. 2022. RESIN-11: Schema-guided event prediction for 11 newsworthy scenarios. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations*, pages 54–63, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Charles J. Fillmore and Collin Baker. 2009. 313 A Frames Approach to Semantic Analysis. In *The Oxford Handbook of Linguistic Analysis*. Oxford University Press.

Jerry Hobbs and James Pustejovsky. 2003. Annotating and reasoning about time and events. In *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, volume 3.

Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from wikipedia. *Artificial Intelligence*, 194:28–61.

Seohyun Im, Hyunjo You, Hayun Jang, Seungho Nam, and Hyopil Shin. 2009. Ktimeml: specification of temporal and event expressions in korean text. In *Proceedings of the 7th Workshop on Asian Language Resources*, pages 115–122. Association for Computational Linguistics.

Paul R Kingsbury and Martha Palmer. 2002. From TreeBank to PropBank. In *LREC*, pages 1989–1993.

Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

Sha Li, Heng Ji, and Jiawei Han. 2021. Document-level event argument extraction by conditional generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 894–908, Online. Association for Computational Linguistics.

Farzaneh Mahdisoltani, Joanna Asia Biega, and Fabian M. Suchanek. 2015. YAGO3: A knowledge base from multilingual Wikipedias. In *Conference on Innovative Data Systems Research*.

N. Noy and Deborah Mcguinness. 2001. Ontology development 101: A guide to creating your first ontology. *Knowledge Systems Laboratory*, 32.

Alessandro Oltramari, Aldo Gangemi, Chu-Ren Huang, Nicoletta Calzolari, Alessandro Lenci, and Laurent Prévot. 2010. *Synergizing ontologies and the lexicon: a roadmap*, 1 edition, page 72–78. Cambridge University Press.

Thomas Pellissier Tanon, Gerhard Weikum, and Fabian Suchanek. 2020. YAGO 4: A reason-able knowledge base. In *The Semantic Web*, page 583–596, Cham. Springer International Publishing.

Karl Pichotta and Raymond Mooney. 2014. Statistical script learning with multi-argument events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 220–229.

I Pratt-Hartmann. 2007. From timeml to interval temporal logic. In *Proceedings of the Seventh International Workshop on Computational Semantics (IWCS-7)*, pages 166–180.

James Pustejovsky. 2017. Iso-timeml and the annotation of temporal information. In *Handbook of Linguistic Annotation*, pages 941–968. Springer.

James Pustejovsky, José M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3:28–34.

Hans Reichenbach. 1947. *Elements of Symbolic Logic*. The Free Press; London: Collier-Macmillan.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015. From light to rich ERE: Annotation of entities, relations, and events. In *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 89–98, Denver, Colorado. Association for Computational Linguistics.

Zhiyi Song and Stephanie Strassel. 2008. Entity translation and alignment in the ACE-07 ET task. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Stephanie Strassel and Alexis Mitchell. 2003. Multilingual resources for entity extraction. In *Proceedings of the ACL 2003 Workshop on Multilingual and Mixed-language Named Entity Recognition*, pages 49–56, Sapporo, Japan. Association for Computational Linguistics.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, page 697–706, New York, NY, USA. Association for Computing Machinery.

Jennifer Tracey, Ann Bies, Jeremy Getman, Kira Griffitt, and Stephanie Strassel. 2022. A study in contradiction: Data and annotation for aida focusing on informational conflict in russia-ukraine relations. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1831–1838.

Naushad UzZaman, Hector Llorens, James Allen, Leon Derczynski, Marc Verhagen, and James Pustejovsky. 2012. Tempeval-3: Evaluating events, time expressions, and temporal relations. *arXiv preprint arXiv:1206.5333*.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007)*, pages 75–80.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th international workshop on semantic evaluation*, pages 57–62.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Jiang Wang, Filip Ilievski, Pedro Szekely, and Ke-Thia Yao. 2022. Augmenting knowledge graphs for better link prediction. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, page 2277–2283, Vienna, Austria. International Joint Conferences on Artificial Intelligence Organization.

Qiusi Zhan, Sha Li, Kathryn Conger, Martha Palmer, Heng Ji, and Jiawei Han. 2023. GLEN: General-purpose event detection for thousands of types. Preprint on webpage at `https://arxiv.org/abs/2303.09093`.