DiDO: An Ontology for Dialogue and Discourse Representation

Kelsey Rook[1,*], Deborah L. McGuinness[1]

[1]Rensselaer Polytechnic Institute, 110 8th St Troy NY, 12180

Abstract

We introduce the Discourse In Dialogue Ontology (DiDO), a modular framework designed to represent structural, semantic, and pragmatic features generically present in conversation. The core of DiDO represents the structure of conversation, and supports the representation of dialogue from datasets and dialogue systems in multi-party, human, agent, and mixed human-agent settings. DiDO's discourse module abstractly models phenomena such as dialogue acts and discourse relations, acting as a tagset and dimension neutral backbone between annotation schemas. DiDO also uses RDF 1.2 to apply metadata such as discourse phenomena to in-dialogue assertions, facilitating context-aware querying over conversational context with respect to temporality and perspective. DiDO's generic approach facilitates the annotation, representation, and analysis of dialogue data across different domains and applications. We detail the methodology used to develop DiDO, and demonstrate its usefulness in dataset alignment and dialogue understanding through a use case evaluation on multi-party meeting transcripts. This modeling with DiDO enables complex querying over assertions, interlocutor perspective, and the temporal flow of dialogue.

Keywords

ontology, ontology engineering, dialogue, dialogue representation

1. Introduction

Conversations are the primary vehicle for human information exchange. With the recent ubiquity of dialogue agents in everyday life, dialogue has also become an interface between humans and software systems. This shift makes the representation of knowledge towards and from dialogue even more important when considering machine understanding and automated reasoning, resulting in a large volume of research on grounding dialogue systems in semi-structured or structured knowledge. Additionally, the hybridization of professional environments and improvements to transcription technology have increased both the volume of machine-processed dialogue and the feasibility of building dialogue datasets from diverse settings.

Studies in linguistics and discourse analysis identify several unique attributes of dialogue that necessitate specialized representation. Unlike formal documents, dialogue is characterized by cross-turn assertions, where facts are constructed incrementally across multiple speakers [1]. Furthermore, high pronoun density requires robust coreference resolution to disambiguate entities.

Dialogue functions as an information exchange system in which interlocutors choose their information contributions and evolve their internal state based on other contributions. Each dialogue partner maintains their own context, rather than all interlocutors relying on a single, shared context, and evidence for particular conclusions is frequently sparse and distributed non-contiguously across multiple turns [1]. The general consensus in linguistics is that pragmatics, the study of how context contributes to meaning, is as vital to understanding dialogue contributions as the semantic content itself [2]. Utterances with the same semantic content frequently serve different communicative functions that can usually be determined by human interlocutors through context, but that may be more difficult for software agents such as LLMs to distinguish [3]. While logically rigorous hierarchies and ontologies

Proceedings of FOIS 2026 Satellite events co-located with the 16th International Conference on Formal Ontology in Information Systems (FOIS 2026), September 21-25 , Vitória, Brazil

*Corresponding author.

rookk@rpi.edu (K. Rook); dlm@cs.rpi.edu (D. L. McGuinness)

0000-0003-0471-7159 (K. Rook); 0000-0001-7037-4567 (D. L. McGuinness)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-27

CEUR-WS.org/Vol-4342/paper5.pdf

exist to represent specific elements like dialogue acts, illocutionary force, or argumentation, there is a notable absence of a mid-level ontology that concurrently represents dialogue structure, pragmatic theory, and assertions from dialogue. To address this gap, we propose the Discourse in Dialogue Ontology (DiDO), a generic, unifying representation that prioritizes the structural skeleton of dialogue and its associated provenance at the top level.

1.1. Contribution and Scope

DiDO's contribution is the unified modeling of four properties that no currently available dialogue ontology provides: (1) explicit dialogue structure (turns, interlocutors, temporal ordering), (2) discourse and dialogue act modeling that is agnostic to any specific tagset or theoretical commitment, (3) first-class representation of assertions extracted from dialogue, and (4) provenance tracing back to source datasets. Table 1 shows that existing frameworks each support a subset of these properties, typically within a single domain (meetings, task-oriented dialogue, or scientific discourse).

The tag-set agnosticism of DiDO is a structural decision: DiDO does not model the content of any dialogue act or discourse relation taxonomy (e.g. which categories exist, what they mean), but rather the "slot" that any such category fills with respect to an utterance. The `dido:DialogueActAnnotation` class is a place to attach an arbitrary external label, not a competing classification scheme. DiDO specifies that an utterance has a function, and that this function can be sourced from typed nodes carrying an external tagset's labels. We do not claim this resolves theoretical disagreement about discourse categorization; we claim only that competing theories can be annotated within the same structural container, which is what permits cross-dataset querying, as demonstrated in Section 8.

DiDO's discourse module makes explicit the communicative function (the dialogue act) underlying an utterance, which is one component of pragmatic meaning. It does not formalize broader extra-textual contextual factors — socio-cultural norms, shared history between interlocutors, and similar context that some pragmatic theories treat as central to meaning. DiDO allows the attachment of this information if a user's annotation schema captures it, but does not model it natively.

The landscape of dialogue data spans many settings such as business meetings, task-oriented human-agent sessions, and casual chit-chat. Because these datasets originate from diverse academic communities, they may be terminologically inconsistent. For instance, the atomic dialogue unit that serves a specific function is variably called an "utterance", "turn", or "functional segment" across schemas [4]. DiDO addresses this disparity through a normalized ontology that formally defines these concepts in relation to each other, ensuring the equivalent ideas in datasets and dialogue systems can be mapped and queried through a shared conceptual vocabulary.

By connecting signals over larger conversational units and providing human-readable class labels, DiDO enables sophisticated corpus querying and data integration. This perspective-aware, multi-dimensional approach ensures that the inherent subjectivity of human conversation is preserved and leveraged, rather than lost in translation to a formal knowledge base. While most ontology-aligned knowledge bases represent factual knowledge of the world, DiDO supports the tracking of separate interlocutor perspectives, allowing the tracking of individual user beliefs and group argumentation over time. The establishment and maintenance of common ground is crucial in dialogue systems for successful interaction [5]; instead of dialogue history being compressed and flattened, or growing too large to be useful, DiDO has the capability to support query-based information retrieval of the dialogue state through time. Our contributions are as follows:

- We review the background to our work, which includes seminal discourse and dialogue theory, and define discourse terminology used throughout the paper. Additionally, we survey existing ontologies and models of dialogue and abstract discourse structures pertaining to dialogue (Section 2).
- We describe the development process of DiDO including use cases and competency questions (Section 3), and the resulting ontology modules: DiDO-Core (Section 4), DiDO-Discourse (Section 5), and DiDO-Assertions (Section 6).

- We conduct a competency question-based evaluation, demonstrating aligning across two distinct dialogue corpora and retrieval of complex dialogue state information (Section 8).

2. Related Work

2.1. Discourse and Dialogue Theory

The interpretation of dialogue requires understanding of both the literal text and its meaning in the larger context of a conversation. While semantics concerns the inherent meaning of language entities, pragmatics concerns the meaning of language in context [2]. Within dialogue, pragmatics can be viewed as the relation between utterances and context [6], and is crucial to understanding the meaning of dialogue beyond what is seen in individual utterances. Two components of dialogue acts—the semantic content and communicative function [7]—are sometimes defined more precisely as actions corresponding with intended information state updates. In the classic example, "Can you pass me the salt?" [8], the intent of the speaker must be inferred from context rather than the literal content of the utterance. The meaning of utterances is contingent on its position in a conversation, as well as the "common ground" established between interlocutors.

A foundational framework in the domain of discourse for dialogue is Speech Act Theory [9], which posits that utterances function as actions in the world; that is, they perform social acts such as requesting, promising, or informing. Austin's original formulation [10] named "performatives" as explicit first-person declarative utterances (e.g. "I hereby promise"). Searle's later systematization replaced this distinction with a general taxonomy of illocutionary acts. Speech Act Theory is further formalized and refined in the framework of dialogue acts [11], the understood communicative function of which specifies how interlocutors update their information states. Further, certain dialogue acts inherently depend on acts that have taken place earlier in dialogue; these relations have been formalized in a wide range of proposed classifications.

Current ontologies and taxonomies of dialogue acts largely build on Searle's set of categories: Assertives (committing to the truth of a proposition), directives (requesting action from the listener), commissives (committing to future action), expressives (conveying psychological states), and declarations (altering reality through speech). These categories provide a philosophical foundation from which many dialogue act taxonomies have been expanded into more granular categories for modeling dialogic interactions. These include MRDA (Meeting Recorder Dialogue Act) corpus [12], which introduces tags specifically for multi-party interactions such as meetings and group discussions. The DIT++ dialogue act taxonomy separates dialogue acts into several dimensions [13], accounting for dialogue-related tasks, how speakers process previous turns and perceive other interlocutors' internal states, social obligations, and turn management; that is, the mechanics of how speakers hold or release the floor. DiAML, the Dialogue Act Markup Language, presents an abstract syntax for dialogue act structure in triple form (S, A, f_d) consisting of a sender S, a set of addressees A, and the dialogue function f_d [7]. This framework eventually informed the ISO 24617-2 standard for dialogue act annotation [14]. Additionally, diverse sets of rhetorical relations have been proposed to define relations and represent semantic dependence between dialogue acts, as well as between various tag sets.

Dialogue acts and other theories of discourse have been used in NLP research as features to better capture conversational dynamics [15, 16, 17]. Automatic speech and dialogue act classification are well-explored tasks [18, 19] with state-of-the-art methodologies achieving from 75-90% accuracy depending on taxonomy granularity [3]. As a result, the consideration of dialogue acts and other discourse features in real-time dialogue systems and other situations in which hand-annotated data is unavailable is increasingly tenable.

2.2. Discourse and Dialogue Ontologies

Efforts to provide structured representations of dialogue have largely resulted in specialized, domain-dependent ontologies. Early work in this domain serves scientific collaboration and personalized

Framework	Dialogue structure?	Discourse structure?	Pragm. & Intent?	Assert.?	Setting agnostic?	Domain agnostic?	Provenance?
MMDO [20]	Partial†	Yes; not tag-set agnostic	Yes	No	No (Meetings)	Yes	Yes
Silva et al. 2021 [21]	No	Yes; not tag-set agnostic	Yes	Yes	No (TOD)	Yes	No
CoAKTinG [22]	Partial†	No	No	No	No (Meetings)	Yes	Yes
WEBDial [23]	No	Yes; not tag-set agnostic	Yes	Yes	No (TOD)	No	No
OLiA	No	Yes	No	No	Yes	Yes	No
SWAN	No	Yes	No	Yes	No (Sci. Lit.)	No (Neuro.)	Yes
Discourse Elements Ontology	No	Yes	No	No	No (Sci. Lit.)	Yes	No
DiDO (Ours)	**Yes**	**Yes; tag-set agnostic**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Table 1
A comparison of existing ontologies for dialogue and discourse modeling, and the features they support the modeling of. Pragm.=Pragmatics; Assert.=Assertions. †Partial: models meeting-level structure (sessions, participants) but not fine-grained turn-level temporal structure

assistant understanding for the meeting setting. The CoAKTinG project [22] resulted in the Compendium tool for meeting understanding, supported by a meeting ontology representing artifacts such as meeting instances, presentations, and information transfer events. Similarly, the Multimodal Discourse Ontology (MMDO) supported the Cognitive Assistant that Learns and Organizes (CALO) [20]. The MMDO models communicative acts in three layers: domain-level information which is encoded into a symbolic message, and then into a physical signal. The MMDO also includes communicative roles, and a static high-level tagset for dialogue acts from the MALTUS (Multidimensional Abstract Layered Tagset for Utterances) tagset [24]. These ontologies are not publicly available.

In task-oriented dialogue (TOD), ontologies are frequently used to map user requests to domain-specific actions. For example, Silva et al. [21] demonstrate how linking ontologies to discourse act modules clarify user intents. Utterances are parsed and aligned with an ontology of dialogue acts and entities, from which an agenda is constructed to provide a solution to the most plausible user meaning. The discourse acts included in the ontology were chosen due to their frequency in the TOD setting, including “Show”, “DisplayValue”, and “Filter”. Santamaria et al. [25] present a representation schema towards the construction of personalized dialogue agents. They argue that a suitable dialogue memory representation should capture accumulation of knowledge over time, and dual representation of knowledge and speaker persona to facilitate reasoning over different worldviews. This schema uses dialogue acts to provide context to the personal, factual, and social assertions expressed.

Several ontologies exist that model certain aspects of discourse separately from the notion of conversation. For example, the Discourse Elements Ontology (DEO) [26] describes rhetorical elements within documents, and the SWAN (Semantic Web Applications in Neuromedicine) ontology [27] describes discussions and collaborations within the biomedical community through scientific discourse networks. The OLiA (Ontologies of Linguistic Annotation) [28] suite of ontologies provides a robust background for various linguistic categories, mapping concrete annotations from various corpora to reference terms. This includes discourse extensions to facilitate interoperability of annotations of various discourse phenomena including anaphora, topical structure, discourse structure, and discourse relations, restricted to the annotation of monologic text.

While frameworks exist to represent discourse and dialogue, they remain fragmented by domain specificity or narrow structural scope, as can be seen in Table 1. We position DiDO as a "backbone" to abstractly represent the connection between these aspects of dialogue, supporting provenance, multidimensional dialogue act modeling, and description of propositions made in utterances.

3. Ontology Development

To develop DiDO, we follow best practices from the guide Ontology Development 101 [29, 30]. This includes determining scope using competency questions, reuse of established ontologies, and keeping a term list from which we build a class hierarchy and the remainder of the ontology. We selected this methodology because dialogue, as established in Section 2, spans heterogeneous theoretical traditions and annotation practices. A CQ-driven process lets us scope DiDO around the queries a backbone ontology must support, rather than around any single tagset or task's structure. Additionally, elements of Modular Ontology Modeling (MOMo) [31], such as diagram-first modeling and a focus on modularity, suit DiDO's use case well, allowing users to adopt only the structures they need without importing the full ontology. We use Protégé [32] as a starting point for drafting DiDO, as well as tools from the Ontology Development Kit (ODK) [33] to manage imports and build releases. We maintain these releases, as well as artifacts such as term lists, at the following GitHub repository: https://github.com/shmewtep/DiDO

3.1. Use Case

DiDO is designed with a focus on interoperability between existing language and dialogue domain resources, as well as specialized annotations of discourse and other dialogue phenomena. We maintain an abstract representation that remains neutral toward conflicting linguistic theories. DiDO does not adjudicate between competing theories of discourse, but provides a shared structural container into which annotations can be placed and queried jointly. This allows users to map their preferred vocabularies to a common structural backbone, providing a container for annotations rather than dictating "correct" attributions.

DiDO's initial use case focuses on the textual transcripts of spoken dialogue. While auditory and visual cues are fundamental to human communication, DiDO prioritizes the information extraction pipeline, which typically begins with signals converted to text, or outputs from text-based online dialogue systems. Dialogue is inherently multi-speaker and temporal, and individual contributions are likely to reflect distinct, sometimes conflicting viewpoints. This contrasts with authored documents, which typically present a single integrated perspective rather than attributing claims to distinct, temporally distinguishable speakers. As a result, knowledge derived from dialogue will frequently reflect differing viewpoints, evolving beliefs, and varying levels of uncertainty. Without the added context of who made a statement, and when, this knowledge is useless. For example, if two speakers assert contradictory claims, or if one speaker expresses conflicting statements over time, a flat representation of these assertions in a knowledge graph would introduce ambiguity. To address this, we propose the development of an ontology designed to represent dialogue structure and related assertions, with respect to speaker identity and temporal context.

DiDO is composed of the following modules: DiDO-Core, DiDO-Discourse, and DiDO-Assertions. DiDO-Core encompasses high-level dialogue structure and provenance, DiDO-Discourse includes utterances and information regarding their semantic meaning, pragmatic meaning, and intent, and finally, DiDO-Assertions manages the reification of dialogue-level assertions. The separation of these modules is intended to enhance reusability by allowing users to adapt DiDO to their use cases only as needed, and extensibility, by partially decoupling DiDO-core from its extensions and thus allowing future iterations to introduce more modules without requiring a major refactoring of the base dialogue structure. The goals of DiDO-Core are reflected in the set of competency questions in Table 2.

3.2. Ontology Reuse

To provide a foundation in a well-established project, DiDO adopts the Semanticscience Integrated Ontology (SIO) mid-level ontology [34] as its structural core. In particular, DiDO uses SIO's upper-level modeling of objects, processes, and their attributes (qualities, roles, measurement values) to model the dialogue process. Its mid-level modeling of information content additionally supports the modeling of granular dialogue artifacts. Spatial and temporal qualification of SIO entities is captured through a set

	Competency question
Dialogue Structure	
CQ1	How many **participants** are involved in *dialogue*?
CQ2	What is the **textual content** of *turn*?
Interlocutors and Roles	
CQ3	Which **participant** had the floor for *turn*?
CQ4	Is **participant** a *human* or a *software agent*?
Temporality	
CQ5	Which **turn** most closely follows *turn*?
CQ6	Which **turns** occurred between *time* and *time*?
CQ7	Which **turns** overlap with *turn*?
CQ8	What is the **duration** of *dialogue*?
Provenance	
CQ9	Which **dataset** is *dialogue* from?
CQ10	What is the **ID** of *dialogue* within its dataset?

Table 2
Competency Questions related to DiDO's core modeling

of relations, and information in the form of literals, such as the output audio and transcribed text from an utterance, are captured as instances of information content entities associated with specific objects or processes.

The temporal dimension of dialogue in DiDO is modeled using the Time Ontology in OWL (OWL-Time), which provides a vocabulary for expressing facts about ordering relations among instants, intervals, and durations with various reference systems. OWL-Time provides the semantic flexibility needed to reconcile heterogeneous date and time formats, and analyze events based on temporal reasoning and topological relations such as "overlaps" and "contains". We use the PROV Ontology (PROV-O) [35] to model the provenance of dialogue artifacts such as transcripts, particularly pertaining to those associated with datasets. Further, we utilize POWLA [36], an OWL reconstruction of PAULA [37], a generic XML formalism for representing linguistic annotation. The namespaces and IRIs of the ontologies used can be seen in Table 3.

prefix	ontology	IRI
dido	Discourse in Dialogue Ontology	http://purl.org/dido/dido.owl
sio	Semanticscience Integrated Ontology	https://semanticscience.org/ontology/sio.owl
time	Time Ontology in OWL	http://www.w3.org/2006/time
prov	PROV Ontology	http://www.w3.org/ns/prov
dcat	Data Catalog Vocabulary	http://www.w3.org/ns/dcat
powla	POWLA	http://purl.org/powla/powla.owl

Table 3
Ontologies used in DiDO

4. DiDO-Core

The core structure of dialogue involves the dialogue entity itself, interlocutors, and turns, which are the entities most easily and commonly captured in dialogue systems and dialogue datasets with transcripts. We align the primary entities of dialogue with SIO's process-role-entity modeling, treating *Dialogue* as a process that unfolds over time. The participant or *Interlocutor* role is realized in *Dialogue*, and may be fulfilled by humans or software agents. The inclusion of software agents as valid interlocutors reflects the increasing prevalence of human-agent dialogues in deployed systems; we acknowledge that artificial agents may not function identically to human-human communication at the level of pragmatics and intentionality. The next crucial concept is the dialogue *Turn*, which represents a continuous speech segment from a single *Interlocutor*, bounded by periods of that *Interlocutor's* inactivity. We use OWL-

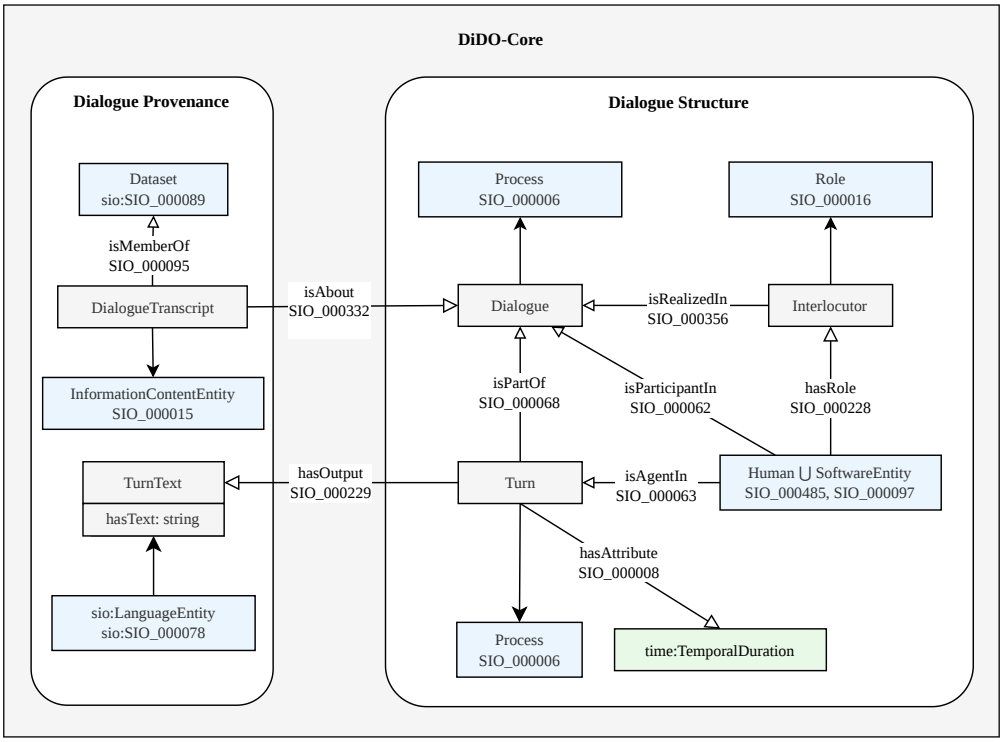

Figure 1: Conceptual diagram representing DiDO-Core. Gray boxes represent classes created in DiDO. Solid arrows indicate subclass relationships.

	Competency question
CQ11	What **dialogue acts** are associated with the utterances supporting *assertion*?
CQ12	What **utterances** were made as part of a *dialogue act type*?
CQ13	Did any interlocutors agree with **utterance**?

Table 4
Competency Questions related to DiDO's discourse modeling

Time to model the durations of *Turns*, due to its adaptability to a multitude of temporal reference systems. Additionally, OWL-Time supports Allen Algebra [38], which allows temporal reasoners to determine the types of overlap that occur between turns. This modeling, including simple provenance tracking extensible through PROV-O, can be seen in Figure 1.

5. DiDO-Discourse

A core challenge in information extraction (IE) from dialogue lies in modeling the dynamic nature of belief and perspective. Conversation is not a static collection of facts; it is a collaborative process where beliefs are proposed, challenged, and iteratively revised over time. DiDO-Discourse integrates discourse theory to contextualize literal text of utterances, which moves DiDO's modeling from a record of who said what to a dynamic model of discourse that preserves underlying meaning.

Discourse refers broadly to language use beyond an isolated utterance, encompassing both multi-utterance organization (e.g., rhetorical or coherence relations between segments) and language in its social and communicative context. Pragmatics concerns specifically how context contributes to meaning, i.e., how identical semantic content yields different communicative effects depending on who utters it, when, and under what shared assumptions [2]. A dialogue act is the theoretical construct describing the communicative function an utterance performs (e.g., requesting, asserting, agreeing), grounded in

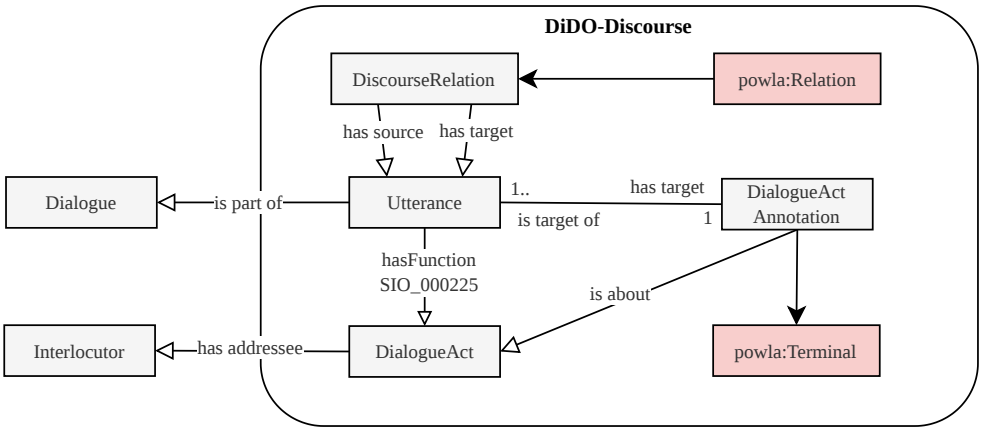

Figure 2: Dialogue act modeling with DiDO-discourse. *powla:Terminal* nodes correspond to atomic units in terms of the annotation schema.

```
:DALayer a powla:Layer ;
    rdfs:label "Dialogue Act Layer" .
:DiscourseLayer a powla:Layer ;
    rdfs:label "Discourse Relation Layer" .

:node_s1 a powla:Terminal ;
    powla:hasLayer :DALayer ;
    powla:hasLabel "Suggest" ;
    powla:hasStringValue "If we remove the LCD display, we could save ourselves a fair
    ↪  amount." .

:node_s2 a powla:Terminal ;
    powla:hasLayer :DALayer ;
    powla:hasLabel "Assess" ;
    powla:hasStringValue "I think if we remove the LCD display then there was absolutely no
    ↪  point to these meetings." .

:rel_objection a powla:Relation ;
    powla:hasLayer :DiscourseLayer ;
    powla:hasLabel "Objection" ;
    powla:hasSource :node_s2 ;
    powla:hasTarget :node_s1 .
```

Figure 3: Turtle representation of a dialogue snippet in the AMI corpus using DiDO and POWLA

Speech Act Theory [9]. We use "discourse act" synonymously with "dialogue act" throughout this paper to refer to this same construct; where prior work we cite uses "discourse act" to mean something more general (e.g., any labeled segment of discourse, not specifically tied to speech act theoretic function), we note this explicitly. A discourse relation is distinct from a dialogue act: it describes a rhetorical or semantic dependency between two dialogue acts or utterances (e.g., one utterance being raised as an objection to another), rather than the function of a single utterance in isolation.

We additionally distinguish "Turn" from "Utterance", since these two terms are a particularly clear instance of the cross-dataset terminological inconsistency motivating this paper. We define a "Turn" as a continuous period of speech by a single interlocutor, bounded by surrounding periods of that interlocutor's inactivity; a Turn may contain multiple sentences. We define an "Utterance" as a single functional unit serving one or more dialogue acts. The relationship between these two units is not fixed across datasets, and this is not merely a labeling difference but a difference in what each dataset treats as the atomic unit of analysis. The AMI Meeting Corpus, discussed further in Section 8, distinguishes

the two directly: a single speaker turn is routinely segmented into several shorter utterances, each independently bearing its own dialogue act. DailyDialog [39], by contrast, conflates them: each turn in its native representation is annotated with exactly one dialogue act regardless of how many sentences it contains, making Turn and Utterance effectively 1-to-1 in that dataset. A representation that hard-codes either assumption, that a turn is always exactly one utterance, or that an utterance is always a sub-span of a longer turn, cannot represent both corpora without misrepresenting AMI's turns or inventing turn boundaries that DailyDialog does not specify. We treat Turn and Utterance as distinct classes related by, but not equated with, one another for exactly this reason, and revisit this distinction in Section 8.

While DiDO defines *Turns* in a dialogue by their surrounding periods of inactivity, *Utterances* are defined by the function or functions they serve in dialogue, referred to as dialogue acts. Each utterance may be associated with one or more dialogue acts which provide insight into the communicative intent and information state of the interlocutor. For example, an utterance tagged as a statement opinion (e.g. "I think the browser should be a viewer") represents a different truth value and commitment level within the speaker's mental model than a non-opinion statement (e.g., "The browser is a viewer"). Competency questions related to these issues can be seen in Table 4.

When modeling dialogue acts, we do not choose a particular tagset to encode or make any statements on how dialogue segments should be divided, but rather how dialogue elements relate to the concept of a dialogue act. DiDO defines structural elements and valid attributes, but maintains an abstract representation that allows users to "plug in" their preferred vocabularies. This design is critical for subjective domains like sentiment analysis: coming to a consensus over which set of descriptors encompass the whole of human emotion as represented in communication is doomed to fail, but certain tagsets are designed to be more beneficial in specific application settings.

The complexity of dialogue act annotation varies from flat tagsets for which annotation involves assigning a single tag to each utterance in the dataset, to more complex models that involve categories on multiple dimensions. Several theories also specify rhetorical relations that are possible between specific dialogue acts, such as question-answer structures. Despite this variance, dialogue act modeling can be abstracted to a set of annotations on an utterance for a set of dimensions, rhetorical relations between dialogue acts, and relations between utterances and interlocutors (source, addressee). This complexity is largely captured by the metamodeling of the ISO 24617-2 standard [14], which organizes communicative functions by semantic dimension, and further specifies qualifiers and dependence relations. However, this metamodel does not generalize to annotation schemes used in many widely-used dialogue corpora, including the AMI Meeting Corpus [40]. Such taxonomies have great practical relevance, enabling large-scale corpora annotations with reasonable inter-annotator agreement, with categories that act as approximations of the dialogue function of utterances [41]. DiDO's modeling abstracts over both types of dialogue act taxonomy, using the generic *Dialogue Act Annotation* and *Discourse Relation* classes for which POWLA accommodates the instantiation of specific dimensions and tagsets, as seen in Figure 2. This allows a corpus annotated with AMI's flat taxonomy to be queried alongside one annotated using ISO 24617-2, using the same SPARQL queries against DiDO's shared structure.

6. DiDO-Assertions

To effectively model the nuances of what is expressed in dialogue, DiDO explicitly represents assertions that are expressed within the dialogue as distinct informational entities. This is achieved through reification, in which an RDF triple is treated as a resource, modeled as an instance of an *Assertion* linked to its conversational context. DiDO leverages RDF 1.2 to represent knowledge triples as first-class objects (instances of an *Assertion* class). RDF-1.2 introduces quoted triple terms that can then be referenced in the subject/object of another triple. This approach allows essential contextual metadata—-speaker identity, temporal attribution, and dialogue act context (the latter serving as a proxy for the speaker's intent)—to be associated directly with the claim itself, as in Figure 4.

Alignment of assertions with speakers and utterances, along with the integration of domain ontologies, can allow for complex reasoning and querying that facilitates many downstream tasks such as dialogue

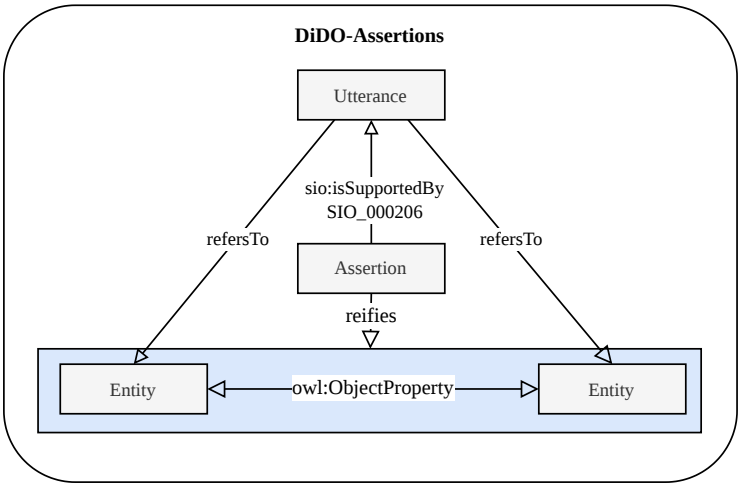

Figure 4: DiDO-Assertion modeling

	Competency question
CQ14	Which **interlocutor** made *assertion*?
CQ15	What **utterances** support *assertion*?
CQ16	What **utterances** did *interlocutor* make regarding *entity*?
CQ17	What **assertions** were made regarding *entity*?
CQ18	Did **speaker** revise their initial stance on *assertion*?
CQ19	Which **assertions** did **interlocutor** support?

Table 5
Competency questions related to DiDO-Assertion

summarization and question answering. Dialogue acts, which are decided by both semantic and pragmatic attributes of an utterance, are also represented as important context for utterances and the associated triples. Related competency questions can be seen in Table 5.

Flat dialogue representations can capture who said what, but do not typically capture how the dialogue's state of asserted claims evolves over its course. For instance, if *Speaker S1* makes an assertion in *Turn 1* and then makes a contradictory assertion in *Turn 10*, a simple provenance model would result in a knowledge graph containing two coexisting, logically contradictory triples with no means of expressing that the second supersedes or revises the first. Using quoted triples alongside discourse modeling, DiDO represents each assertion over time, enabling inference of the speaker's intent to revise their assertion. Competency questions related to these capabilities can be seen in Table 5.

7. Publishing and Resource Availability

DiDO is available in Turtle, RDF, and JSON-LD formats and is published under an MIT license. Artifacts including Competency Questions, SPARQL queries, terms, and other documentation are available at https://github.com/shmewtep/DiDO. The ontology itself is available in the GitHub repository, and at the URI: http://purl.org/dido/dido.owl. The DiDO repository is maintained by the authors as part of ongoing research into dialogue knowledge representation. Versioned releases are managed using the Ontology Development Kit (ODK) [33] and hosted on GitHub.

8. Evaluation

We conducted basic error detection using the OOPS! ontology pitfall scanner [42]. The only important errors detected were missing domains and ranges from well-established reused ontologies, and the majority of warnings stemmed from the lack of standardized labels and definitions in the POWLA ontology. With respect to FAIR principles, DiDO is Accessible (MIT license, multiple serialization formats), Interoperable (extensively reuses established ontologies), and Reusable (19 CQs, SPARQL queries, modular design). Findability is partially met by the PURL, GitHub repository with documentation, and website, and will be fulfilled by community registry upon the next release.

8.1. Competency Question-Based Evaluation

To validate the utility and expressivity of the DiDO ontology, we conducted a comprehensive competency-question (CQ) based evaluation. We evaluate the ontology's coverage by executing a suite of 19 SPARQL queries, each corresponding to a competency question defined to test one or more DiDO modules. The complete set of executable SPARQL queries are openly available in our project repository, as well as RDF files containing a subset of the AMI Meeting Corpus aligned to DiDO. An interactive UI and SPARQL endpoint pre-loaded with these aligned datasets is also available, allowing direct execution of the CQs' corresponding queries: https://triplydb.com/KelseyRook/AMIES2002bDiDO/sparql.

The AMI Meeting Corpus was selected to test DiDO's capacity to handle standard annotations. AMI utilizes a flat custom schema for its dialogue act and discourse relation labels. By mapping the AMI structure, we demonstrate DiDO's ability to preserve discourse and dialogue properties. While standard dialogue acts can be mapped programmatically, higher-order semantics such as belief tracking and stance attribution require a more complex structure. To assess the efficacy of DiDO's Assertion module––specifically its ability to track the provenance of claims––we manually annotated AMI dialogues with RDF triples representing core domain assertions. By linking these logical triples back to the source utterances using the `sio:isSupportedBy` predicate, we were able to successfully execute temporal queries. For example, our evaluation verified that the ontology can accurately query which specific interlocutor introduced an assertion, who subsequently agreed or disagreed with it, and track instances where an interlocutor revised their stance over the course of the interaction.

8.2. Cross-Corpus Alignment

A central claim of this paper is that DiDO supports alignment across datasets that differ not only in format, but in structure and terminology. We selected AMI and DailyDialog [39] for this demonstration. AMI is distributed as XML and drawn from multi-party (3–4 participant) face-to-face meetings, with precise utterance-level timestamps and a granular custom dialogue act taxonomy (e.g., `Stall`, `Assess`, `Elicit-Inform`). DailyDialog is available via Hugging Face, with no timestamps, and four standardized speech act categories (`Inform`, `Question`, `Directive`, `Commissive`). Critically, AMI contains specific text segments attributed with dialogue acts alongside dialogue turns, while DailyDialog conflates them; this is a structural difference in what DailyDialog considers to be a functional unit of dialogue.

Table 6 summarizes the mapping used for both corpora. Both are aligned to the same set of DiDO classes and SIO properties: a dialogue session, an interlocutor (meeting participants for AMI; alternating speaker roles for DailyDialog), an utterance and associated turn, and a dialogue act drawn from each corpus's native taxonomy. For AMI, each annotated speaker segment is instantiated as both a `dido:Turn` and a `dido:Utterance`, with multiple utterances linked to a single longer turn wherever AMI's source segmentation places several functionally distinct units within one continuous speaker contribution. For DailyDialog, whose native representation provides no sub-turn segmentation, we instantiate each annotated unit as both a `dido:Utterance` and its implicit `dido:Turn` in a one-to-one relationship. This shared structure lets us execute a single SPARQL query retrieving utterances with a particular function across both corpora simultaneously.

We note that RDF 1.2 is currently a candidate recommendation and thus not citable as a W3C standard,

Concept	AMI mapping	DailyDialog mapping
`dido:Dialogue`	Meeting session (e.g. `ex:dialogueES2002a`)	Individual conversation (e.g. `ex:dialogue_dd_0`)
`dido:Utterance`	Individual speaker segment (e.g. `ex:utteranceES2002a0`)	Individual text turn (e.g. `ex:utterance_dd_0_0`)
`dido:Turn`	Individual speaker turn (e.g. `ex:turnES2002a0`); may contain several Utterances	Same text span as the Utterance (e.g. `ex:utterance_dd_0_0`); 1-to-1 with Utterance
`dido:Interlocutor`	Meeting participants (e.g. `ex:humanB`, `ex:humanC`)	Alternating speaker roles (e.g. `ex:human_dd_0_A`)
`dido:DialogueAct`	Custom AMI subclasses (e.g. `Stall`, `Offer`)	Standard speech act subclasses (e.g. `Directive`, `Question`)

Table 6
Mapping of AMI and DailyDialog native structures to a shared set of DiDO classes and properties.

and is not supported by many tools that are operable on RDF 1.1 artifacts; however, a procedure is available to convert RDF 1.2 quoted triples to a standard RDF 1.1 reification structure [43].

9. Conclusion

While DiDO establishes a foundation for modeling textual dialogue, we acknowledge the limitations of supporting text-based dialogue only. Multimodal features such as gaze, gesture, and prosody are indicative of intent and meaning. Additionally, DiDO-Discourse is in need of further consideration: while we adopt a general modeling that does not align with any particular theory of dialogue acts or fixed set of annotations, we acknowledge that there is diverse thought in the areas of pragmatics and discourse, and that the discourse elements represented in DiDO are a simplified subset of pragmatics. For example, DiDO does not currently model extra-contextual factors such as socio-cultural norms. Our current scope is simplified, which we plan to examine in future work through justification of the subset we choose to model, or the expansion of the DiDO-Discourse module.

In this paper, we introduced the Discourse in Dialogue Ontology (DiDO) to address the lack of a generalized, domain-agnostic ontology for dialogue modeling. By abstracting the basic structure of dialogue and utilizing a flexible annotation framework, DiDO supports interoperability across highly heterogeneous dialogue corpora. We intend to explore the usage of DiDO not only for static dialogue annotation, but in real-time dialogue systems. Future research will explore the deployment of DiDO for the on-line management of dialogue histories in conversational systems, as well as the continued refinement of the ontology itself based on the complexities of discourse theory.

Acknowledgments

Thanks to Dr. Tomek Strzalkowski for his assistance regarding dialogue and discourse theory.

This paper is based upon work supported in part by the United States Air Force under Contracts No. FA8750-21-C-0075 and FA8750-23-F-0080.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of USAF or the U.S. Government.

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude Sonnet in order to: grammar and spelling check, formatting assistance, paraphrase and reword. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Yu, K. Sun, C. Cardie, D. Yu, Dialogue-based relation extraction, arXiv preprint arXiv:2004.08056 (2020).

[2] S. C. Levinson, Pragmatics, Cambridge university press, 1983.

[3] A. Qamar, J. Tong, R. Huang, Do llms understand dialogues? a case study on dialogue acts, in: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025, pp. 26219–26237.

[4] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. V. Ess-Dykema, M. Meteer, Dialogue act modeling for automatic tagging and recognition of conversational speech, Computational linguistics 26 (2000) 339–373.

[5] L. Varonina, S. Kopp, Knowledge modelling for establishment of common ground in dialogue systems 7 (2021) 9–26. URL: http://journals.openedition.org/ijcol/797. doi:10.4000/ijcol.797.

[6] D. Jurafsky, Pragmatics and computational linguistics, The handbook of pragmatics (2006) 578–604.

[7] H. Bunt, J. Alexandersson, J. Carletta, J.-W. Choe, A. C. Fang, K. Lee, V. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, et al., Towards an iso standard for dialogue act annotation, in: Seventh conference on international language resources and evaluation (LREC'10), 2010.

[8] H. Bunt, B. Black, Computational pragmatics (2013).

[9] J. R. Searle, F. Kiefer, M. Bierwisch, et al., Speech act theory and pragmatics, volume 10, Springer, 1980.

[10] J. Austin, How to do things with words. (1962).

[11] H. Bunt, Context and dialogue control, Think Quarterly 3 (1994) 19–31.

[12] E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, H. Carvey, The icsi meeting recorder dialog act (mrda) corpus, in: Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004, 2004, pp. 97–100.

[13] H. Bunt, The dit++ taxonomy for functional dialogue markup, in: AAMAS 2009 workshop, towards a standard markup language for embodied dialogue acts, 2009, pp. 13–24.

[14] H. Bunt, J. Alexandersson, J.-W. Choe, A. C. Fang, K. Hasida, V. Petukhova, A. Popescu-Belis, D. Traum, Iso 24617-2: A semantically-based standard for dialogue annotation (2012).

[15] R. Higashinaka, M. Nakano, K. Aikawa, Corpus-based discourse understanding in spoken dialogue systems, in: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, 2003, pp. 240–247.

[16] F. Dettori, M. Forasassi, L. Veronese, L. Lestingi, V. Scotti, M. G. Rossi, Do you understand how i feel?: Towards verified empathy in therapy chatbots, arXiv preprint arXiv:2601.08477 (2026).

[17] C. Bothe, S. Magg, C. Weber, S. Wermter, Discourse-wizard: discovering deep discourse structure in your conversation with rnns, arXiv preprint arXiv:1806.11420 (2018).

[18] V. Raheja, J. Tetreault, Dialogue act classification with context-aware self-attention, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 3727–3733.

[19] Y. Liu, K. Han, Z. Tan, Y. Lei, Using context information for dialog act classification in dnn framework, in: Proceedings of the 2017 conference on empirical methods in natural language processing, 2017, pp. 2170–2178.

[20] J. Niekrasz, M. Purver, A multimodal discourse ontology for meeting understanding, in: S. Renals, S. Bengio (Eds.), Machine Learning for Multimodal Interaction, volume 3869, Springer Berlin Heidelberg, 2006, pp. 162–173. URL: http://link.springer.com/10.1007/11677482_14. doi:10.1007/11677482_14, series Title: Lecture Notes in Computer Science.

[21] J. Silva, D. Melo, I. Rodrigues, J. Seco, C. Ferreira, J. Parreira, An ontology based task oriented dialogue:, in: Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, SCITEPRESS - Science and Technology Publications, 2021, pp. 96–107. URL: https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0010711900003064. doi:10.5220/0010711900003064.

[22] M. S. Bachler, S. J. B. Shum, D. C. D. Roure, D. T. Michaelides, K. R. Page, Ontological Mediation of Meeting Structure: Argumentation, Annotation, and Navigation (2003).
[23] M. Veyret, J.-B. Duchene, K. Afonouvi, Q. Brabant, G. Lecorve, L. M. Rojas-Barahona, WEBDial, a multi-domain, multitask statistical dialogue framework with RDF, 2024. URL: http://arxiv.org/abs/2401.03905. doi:10.48550/arXiv.2401.03905. arXiv:2401.03905 [cs].
[24] A. Clark, A. Popescu-Belis, Multi-level dialogue act tags, in: Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004, 2004, pp. 163–170.
[25] S. B. Santamarıa, L. Krause, L. Donatelli, P. Vossen, The role of personal perspectives in open-domain dialogue (2023).
[26] A. Constantin, S. Peroni, S. Pettifer, D. Shotton, F. Vitali, The Document Components Ontology (DoCO), Semantic Web 7 (2016) 167–181. URL: https://journals.sagepub.com/doi/full/10.3233/SW-150177. doi:10.3233/SW-150177.
[27] P. Ciccarese, E. Wu, G. Wong, M. Ocana, J. Kinoshita, A. Ruttenberg, T. Clark, The SWAN biomedical discourse ontology, Journal of Biomedical Informatics 41 (2008) 739–751. URL: https://linkinghub.elsevier.com/retrieve/pii/S1532046408000580. doi:10.1016/j.jbi.2008.04.010.
[28] C. Chiarcos, M. Sukhareva, Olia–ontologies of linguistic annotation, Semantic Web 6 (2015) 379–386.
[29] N. F. Noy, D. L. McGuinness, et al., Ontology development 101: A guide to creating your first ontology, 2001.
[30] E. F. Kendall, D. L. McGuinness, Ontology engineering, Morgan & Claypool Publishers, 2019.
[31] C. Shimizu, K. Hammar, P. Hitzler, Modular ontology modeling, Semantic Web 14 (2023) 459–489.
[32] M. A. Musen, The protégé project: a look back and a look forward, AI Matters 1 (2015) 4–12. URL: https://doi.org/10.1145/2757001.2757003. doi:10.1145/2757001.2757003.
[33] N. Matentzoglu, D. Goutte-Gattat, S. Z. K. Tan, J. P. Balhoff, S. Carbon, A. R. Caron, W. D. Duncan, J. E. Flack, M. Haendel, N. L. Harris, et al., Ontology development kit: a toolkit for building, maintaining and standardizing biomedical ontologies, Database 2022 (2022) baac087.
[34] M. Dumontier, C. J. Baker, J. Baran, A. Callahan, L. Chepelev, J. Cruz-Toledo, N. R. Del Rio, G. Duck, L. I. Furlong, N. Keath, et al., The semanticscience integrated ontology (sio) for biomedical research and knowledge discovery, Journal of biomedical semantics 5 (2014) 14.
[35] P. Groth, L. Moreau, Prov-overview, W3C Working Group Note 1135 (2013) 881–906.
[36] C. Chiarcos, A generic formalism to represent linguistic corpora in rdf and owl/dl., in: LREC, 2012, pp. 3205–3212.
[37] S. Dipper, M. Götze, U. Küssner, M. Stede, Representing and querying standoff xml, Data Structures for Linguistic Resources and Applications (2007) 337–346.
[38] J. F. Allen, Towards a general theory of action and time, Artificial intelligence 23 (1984) 123–154.
[39] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, S. Niu, Dailydialog: A manually labelled multi-turn dialogue dataset, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2017, pp. 986–995.
[40] W. Kraaij, T. Hain, M. Lincoln, W. Post, The ami meeting corpus (2005) 1–4.
[41] A. Popescu-Belis, Dimensionality of dialogue act tagsets: An empirical analysis of large corpora, Language Resources and Evaluation 42 (2008) 99–107.
[42] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, International Journal on Semantic Web and Information Systems (IJSWIS) 10 (2014) 7–34.
[43] P.-A. Champin, et al., RDF 1.2 Interoperability, W3C Working Draft, W3C, 2026. URL: https://www.w3.org/TR/rdf12-interop/.