Large-Scale Knowledge Graphs as a Tool for Enhanced Robotic Perception

Mark Adamik, Ilaria Tiddi, Romana Pernisch, Stefan Schlobach

[1]Vrije Universiteit Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam The Netherlands
m.adamik@vu.nl, r.pernisch@vu.nl, i.tiddi@vu.nl, k.s.schlobach@vu.nl

Abstract

Autonomous robotic systems depend on their perception and understanding of their environment for informed decision-making. One of the goals of the Semantic Web is to make knowledge on the Web machine-readable, which can significantly aid robots by providing background knowledge, and thereby support their understanding. In this paper, we present a reasoning system that uses the Ontology for Robotic Knowledge Acquisition (ORKA) to integrate the sensory data and perception algorithms of the robot, thereby enhancing its autonomous capabilities. This reasoning system is subsequently employed to retrieve and integrate information from the Semantic Web, thereby improving the robot's comprehension of its environment. To achieve this, the system employs a Perceived-Entity Linking (PEL) pipeline that associates regions in the sensory data of the robotic agent with concepts in a target knowledge graph. As a use-case for the linking process, the Perceived-Entity Typing task is used to determine the more fine-grained subclass of the perceived entities. Specifically, we provide an analysis of the performance of different knowledge graph embedding methods on the task using a annotated observations and WikiData as a target knowledge graph. The experiments indicate that relying on pre-trained embedding methods results in an increased performance when using TransE as the embedding method for the observations of the robot. This contribution advances the field by demonstrating the potential of integrating Semantic Web technologies with robotic perception, thereby enabling more nuanced and context-aware decision-making in autonomous systems.

Introduction

Although there have been many advances in Robotics & AI in the past decades, we are still far from having robotic agents that possess the ability to efficiently perform a wide range of tasks autonomously in the dynamic environment of the real world. This ultimately limits their utility in terms of task performance, the range of situations in which they can be deployed, and in some cases also by compromising the safety of the environment (Sciutti et al. 2018). Compared to virtual AI agents, where the risk of physical damage and injury is usually limited, the need for safe and reliable interaction with their surroundings is much greater in the case of embodied agents.

One of the reasons why the vision of capable and safe robotic agents is still out of reach is the lack of common-sense reasoning, and insufficient understanding of contextual information. A robot can only perceive[1] what it has been programmed to, either explicitly by creating some form of world model (Besl and Jain 1985) using some perception pipeline (e.g. computer vision algorithms), or implicitly by reacting to the stimulus coming from the sensors (Brooks 1986). In both cases, the range of possible interpretation of the sensory information is limited by the implicit and explicit knowledge that the agent is equipped with.

Despite the extensive availability of common-sense information in machine-readable formats via the Semantic Web (Antoniou et al. 2012), its application in the context of embodied agents has been limited. Semantic Web databases (i.e. large-scale knowledge graphs) such as WikiData (Vrandecic and Krötzsch 2014), DBpedia (Auer et al. 2007) or YAGO (Suchanek, Kasneci, and Weikum 2007) are intended to be used by developers to provide background information for various tasks, such as movie recommendation, retrieving information on historical figures, etc.. These databases also contain general, common-sense concepts such as physical objects and their properties, which could potentially be utilized by robotic agents (Adamik and Schlobach 2023).

To bridge this gap, the Perceived-Entity Linking task has been proposed, as the task of recognizing entities in the sensory data of the robot, and linking the recognized entities to a target knowledge graph (Adamik et al. 2024a). In this paper, we focus on creating a system to support various solutions that solve the above task. We address the following research questions:

RQ1: *How can a robotic agent that is equipped with a set of sensors associate the incoming sensory data with a corresponding unique identifier in a target knowledge graph?*
RQ2: *How can the resources in large-scale knowledge graphs be used to improve robotic perception?*

Concerning **RQ1**, we first introduce a robotic perception system that uses the Ontology for Robotic Knowledge Ac-

[1]Although what exactly constitutes perception is not a trivial matter, for the purposes of this work, by perception we mean the creation of some kind of symbolic data structure.

quisition (ORKA)[2] to organise the perception pipeline of robotic agents and link the perceived entities to resources in knowledge graphs on the Semantic Web. Then, the system is used to examine **RQ2** by evaluating how state-of-the-art graph embedding techniques perform on one type of Perceived-Entity Linking task called the Perceived-Entity Typing, which aims at determining the subclass of the entity label provided by the perception system of the robotic agent.

The contribution of the paper is two-fold, as we 1) present a modular, open-source implementation of a system[3] that links robotic perception to the Semantic Web, and 2) provide an evaluation of our use-case on how state-of-the-art knowledge graph embedding methods could be used to improve robotic perception.

Related Work

Although the Semantic Web has been around for over two decades, there is only a limited work that focuses on harnessing its capabilities to enhance robotic perception.

Stanton et al. (Stanton and Williams 2003) focuses on describing the RoboCup domain by using Resource Description Framework (RDF) to define the entities perceived by the robot. Fischer et al. (Fischer et al. 2018) combines WikiData, WordNet, and an object detection algorithm for action and tool recommendation. The closest related example to our efforts is Young et al. (Young et al. 2016), which introduces an object learning system that relies on DBpedia to suggest labels for previously unseen objects. An extension of this effort (Young et al. 2017) is a deep-learning-based vision system where the context-based suggestions for candidate labels for the objects were filtered using the results of a web-mining process. Although not relying on the Semantic Web resources, notable approaches, such as RoboEarth (Waibel et al. 2011; Tenorth et al. 2013) aimed at establishing a unified knowledge-sharing platform for robotic agents. Unfortunately since the project has been discontinued, no implementation is available online.

Examples outside of the field of robotics where Semantic Web is used to gather background information on physical objects can be found in the domain of Internet of Things (IoT). In (Rossetto et al. 2020), entities in peoples' lifelogs are linked to WikiData to enable semantic search. In (Wu et al. 2017), the Semantic Web of Things are proposed to link sensory metadata to entities in DBpedia. While this approach is similar in principle to the PEL task performed by robotic agents, it does not take the aspect of embodiment into account.

Lastly, some approaches have examined the concept of multimodal entity linking, where the entities to be linked to a target knowledge graph are contained in either images (Zhang, Li, and Yang 2021) or a combination of text and images (Zheng et al. 2022). Although freely available multimodal entity linking databases (e.g. (Gan et al. 2021)) make the development of multimodal entity-linking algorithms possible, all the mentioned resources focus on celebrities, movies or tweets, and do not include commonsense concepts that robotic agents could use.

While some of the approaches described above use Semantic Web resources to improve robotic perception in some way, our proposed method differs in that 1) we provide a general system that facilitates the linking process and allow for interoperability with large-scale knowledge graphs, and 2) we utilize knowledge graph embedding methods used to refine the labels provided by the perception pipeline.

Robotic Perception Linking System

The system we propose is built on the Robotic Operating System (ROS)[4] middle-ware, and comprises of three main layers: (i) the *Sensory Layer*, which serves as an abstraction layer from the physical devices to make the sensory data accessible to (ii) the *Observation Layer*, which in turn enriches the sensory data with semantics and results in an observation graph describing the perceived entities. The contents of the observation graphs are then provided to the (iii) *Linking Layer*, which facilitates the matching between the observed entities and the ones contained in the external knowledge sources (we refer to this as Perceived-Entity Linking). The main components of the system are outlined in Figure 1.

Sensory Layer

As a first step, the sensory information of the robot is made available by the use of ROS publishers, which serve as an abstraction over the low-level implementation of the sensory device, and allows for the accessing of the sensory data. The data is sent over predefined ROS topics using a specific message type, both of which need to be defined by the engineers of the system for the publishers. Both the physical robot and the simulation environment follow the same operational structure, which allows interoperability of the different physical and virtual systems and the Observation Layer, independently of the implementation details.

Observation Layer

The Observation Layer enriches the sensory data with the necessary semantics to enhance the robot's cognitive capabilities, and transforms the data into a format that could be linked to external knowledge sources.

The main component of the layer is ORKA (Adamik et al. 2024b), the ontology built based on the Semantic Sensory Network (SSN) and the Extensible Observation Ontology (OBOE) alignment modules.[5]. As the output of sensors and the corresponding perception algorithms that process the sensory data do not carry any explicit meaning, the main purpose of ORKA is to provide the necessary background knowledge on the sensors and algorithms to allow for the autonomous interpretation of the sensory stream. The primary classes of the ontology represent the `Observations` as the result of the information processing coming from the `Sensors` and their associated `Procedures`, which concern `Entities` and their respective `Observable`

[2]https://github.com/Dorteel/orka

[3]https://github.com/Dorteel/pel_ros

[4]https://ros.org/

[5]https://www.w3.org/2015/spatial/wiki/Alignment_to_OBOE.html

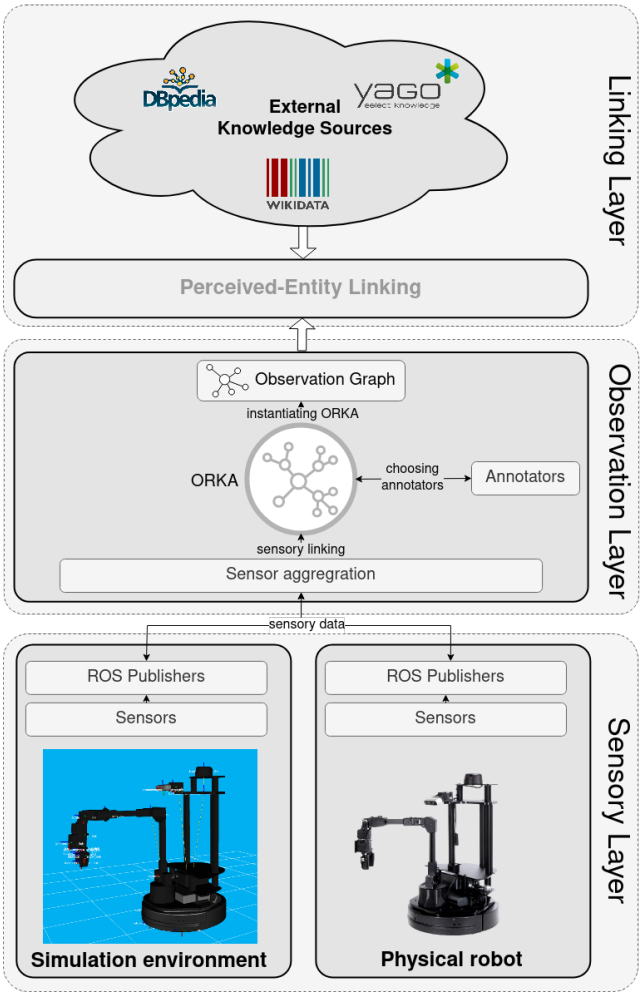

Figure 1: An overview of the proposed system with a simulated or physical robot, where first the sensory data is collected and matched to the ontology, then, once enriched with annotations from perception algorithms (annotators), the resulting observation graph is linked to entities contained in large-scale knowledge graphs.

`Properties` (e.g. color, shape, location). The ontology allows for an automatic reasoning over the sensory and algorithmic capabilities and the properties of the observed entities.

Sensor aggregation The sensory information published on the different ROS topics are aggregated and matched to the sensor classes contained in ORKA. To facilitate the matching, the required information in the ontology can be accessed using the `publishedOnTopic` predicate that stores the corresponding topic of the given sensor, whereas the `hasMessageType` predicate ensures the parsing of the correct message types.

Annotators Once the sensory data has been aggregated, the ontology is used to overview the perception algorithms (also called annotators) that are available to the agent. All annotators are present in the ontology as instances of the `Procedures` class. The reasoning step over the annotators is performed using a Pellet (Sirin et al. 2007) reasoner and Semantic Web Rule Language (SWRL) rules (Horrocks et al. 2004) to infer new knowledge about the sensory capabilities and the properties of objects[6]. An example of the choosing process could be selecting the right annotator that is capable of detecting the class required by the task at hand.

Since the target of our linking process is large-scale knowledge graphs, ORKA also serves as an intermediary knowledge graph that restricts the scale of the target knowledge graph by focusing on the most relevant properties of the observed entities, which are contained in ORKA.

Observation Graphs Once the sensory data of the robot has been linked to ORKA, and enriched by annotations of the appropriate perception algorithms, these processes result in an instantiation of the ORKA called *observation graph*, that contains the sensory data, information about the sensors that acquired the data, as well as the algorithms, their outputs and the characteristics of the entities about which the sensors and algorithms provide information. Entities contained in the observation graph are then selected for the Perceived-Entity Linking process.

Linking Layer

The Linking Layer performs the Perceived-Entity Linking between the entities in the observation graph and the resources in the external knowledge bases. To aid the linking process, ORKA specifies the SPARQL endpoints for the different knowledge graphs, as well as the URI's to the main concepts contained in the ontology (e.g. properties of objects and labels provided by the annotators). The current implementation of ORKA contains links to DBpedia, WikiData and YAGO, using the `hasDBpediaURI`, `hasWikiDataURI` and `hasYagoURI` predicates respectively. The Perceived-Entity Linking process could be performed either by SPARQL-based[7] queries (Adamik et al. 2024a) over the entities in the knowledge graph, or by employing different link prediction methods to establish the connection.

In the approach described in the following section, we focus on establishing the links to the subclasses of the perceived entities contained in the observation graph, by employing knowledge graph embedding techniques.

Use-Case

Perceived-Entity Typing

To illustrate how the Semantic Web resources would ideally extend the perception capabilities of robotic agents, we describe the Perceived-Entity Typing (PET) task, which is inspired by the Entity Typing problem of Natural Language Processing (Dai and Zeng 2023). The task is to provide a more fine-grained label based on the context of the entity, which in the case of the system described in the previous

[6]A complete list of the SWRL rules used is available in the ontology repository.

[7]https://www.w3.org/TR/sparql11-query/

Figure 2: An example of the simulation environment (top left), the annotated view of the robot (bottom left), and the resulting observation graph when aggregated with the additional sensory data (right).

section is provided by the observation graph. The PET task is performed following a traditional entity linking pipeline, where first candidate entities are selected from the target knowledge graph, then a sorting process ranks the candidates to select the most appropriate candidate.

For example, let us consider a robot that is tasked to fetch a mug, while its perception systems are only able to recognize cups. As mugs are considered a special type of cups, that are usually made of ceramic material or glass, and have a handle part, the robot would need to reason about the properties of the subclasses of cups to determine which of the observed entities belong to the more specific type of cup named mug. To achieve this, knowledge graph emebdding methods are deployed to translate the properties of the entities into an embedding space. Afterwards, a SPARQL query is used to gather candidate entities by considering the sibling classes of the determined label. The candidates are then filtered based on the similarity of each candidate to the given entity, considering the properties of the entity present in the observation graph.

Implementation

To instantiate the system, and as a test-bed for the linking process, we use ROS Noetic and the WeBots[8] robot simulation environment. Although the system is designed to work with both physical and simulated agents, we present here the simulation environment, and leave the implementation of the complete system of the physical robot to future work. The annotators implemented are an object detector algorithm based on YOLO (Redmon et al. 2016), a color detector, material detector based on a Vision Transformer model (Dosovitskiy et al. 2020), and a structure detector capable of detecting part-of relationships. An example of the sensory data and the resulting observation graph can be seen in Figure 2. In the current implementation, observation graphs are created upon request, as setting the system to provide the graphs every t time-period results in great computational overhead. Additionally, we use ORKA's pre-defined links (hasWikidataURI, hasDBpediaURIs etc.) to link entities identified by the annotators to entities of the target Knowledge Graph.

This ensures that entities are correctly linked to the target knowledge graph and thereby eliminates confounding factors for performing the PET task. Investigating how to automatically establish a link between entities in ORKA and an existing KG is out of the scope of the current paper, but an existing method can be found in (Adamik et al. 2024a). To perform the matching of the observed entities to the candidate entities, a new embedding is formed from the part of the observation graph concerning the properties of the entity. The new embedding $\mathbf{E}_o$ for the entity is calculated by a weighted sum of the embeddings $\mathbf{e}_i^p$ of the properties that

[8]https://cyberbotics.com/

Concept Name	Wikidata ID	TransE	SimplE	RotatE	QuatE	DistMult	ComplEx
office chair	Q1021686	**0.518**	0.410	0.370	0.457	0.374	0.253
wheelchair	Q191931	0.515	0.440	0.417	0.485	0.415	0.306
fauteuil	Q981009	0.420	**0.485**	**0.438**	**0.599**	0.394	**0.369**
wing chair	Q1847455	0.354	0.395	0.316	0.387	0.347	0.260
litter	Q476850	0.351	0.480	0.398	0.532	**0.431**	0.309
rocking chair	Q14963	0.365	0.479	0.333	0.525	0.371	0.266
Bergère	Q820539	0.241	0.259	0.038	0.430	0.306	0.225
electric chair	Q185639	0.328	0.322	0.092	0.369	0.241	0.268
high chair	Q1622390	0.349	0.454	0.319	0.577	0.385	0.317
folding chair	Q1744391	0.315	0.483	0.413	0.554	0.404	0.319

Table 1: Similarity scores for the top 10 related concepts to an observation regarding an office chair (ground truth), recognised as a chair. In this example, the observation embeddings of TransE resulted in being closest to the correct entity.

the entity possesses (e.g. label, color, shape):

$$\mathbf{E}_o = \sum_{i=1}^{n} \mathbf{w}_i \mathbf{e}_i^p$$

To achieve this step, we rely on the pre-trained knowledge graph embedding models using GraphVite (Zhu et al. 2019), to combine the nodes representing entities and their properties as represented in the observation graph and match to the closest candidates selected from the target knowledge graph. The graph embedding mdoels are trained on the WikiData5m (Wang et al. 2021) dataset.

Once the embeddings of the entities in the observation graph have been created, the matching process starts with a candidate generation step, where the target knowledge graph is queried for the available subclasses corresponding to the entity label to be identified. The SPARQL query we used to select candidates from WikiData can be seen in Listing 1.

Listing 1: SPARQL Query used to select the top k subclass candidates for the PET task from WikiData given an *entityId*, ordered based on the number of incoming links to the subclass.

```
SELECT
    ?subclass ?subclassLabel (COUNT(?link)
        AS ?linkcount)
WHERE {
    ?subclass wdt:P279* wd:{entityId}.
    ?link ?anylink ?subclass.
    SERVICE wikibase:label {bd:serviceParam
         wikibase:language "en".}
GROUP BY ?subclass ?subclassLabel
ORDER BY DESC(?linkcount) LIMIT {k}
```

The observation embedding $\mathbf{E}_o$ is then used to calculate the similarity scores to the embeddings of each of the each candidate, and rank them accordingly..

Experiment

In this section, we focus on assessing the performance of pre-trained knowledge graph embedding models using an annotated dataset of visual sensory inputs. Relying on the distributional hypothesis as the guiding principle, the aim of the experiment is to test whether the pre-trained embedding techniques encode sufficient information to perform the Perceived-Entity Typing task.

Experiment Setup We used the WikiData5m dataset and manually annotated images, both from a simulation setting and real world examples, to form a small dataset with 36 examples as ground truth with fine-grained subclass labels for a given entity. Using a household robot as an example, the chosen entities contained two sublasses of cups (measuring cups and mugs), two subclasses of chairs (office chairs and wheelchairs), and three subclasses of apples. The annotations for the entities contained the labels, materials, colors, shapes, and part-of relationships, in line with the available annotators described earlier. The shape annotation was only used for the apples, while the part-of annotation was used for chairs with wheels and mugs with handles. To create different variations of each entity type, different colors and detected materials were used.

Following a traditional entity linking-based pipeline, first the candidate subclasses were acquired by querying WikiData. Each candidate subclass, as well as the observation graph, was then embedded using one of the pre-trained knowledge graph embedding methods. In this experiment, we used the top 15 candidate labels returned from WikiData, and filtered the results to the entities contained in Wikidata5m. We compared TransE (Bordes et al. 2013), DistMult (Yang et al. 2015), ComplEx (Trouillon et al. 2016), SimplE (Kazemi and Poole 2018), RotatE (Sun et al. 2019) and QuatE (Zhang et al. 2019) knowledge graph embedding methods, by calculating the cosine similarity between the observation graph embedding and the candidates. An example of the resulting similarity scores with an observation embedding containing the annotations of an office chair, as well as their selected candidates, can be seen in Table 1.

Results The results showed that out of the 36 examples, TransE could determine the correct subclass in 14 of these cases, where mugs, wheelchairs and office chairs were assigned the correct label. The next best performing embed-

Concept Name	wd id	Sample#	TransE			DistMult		RotatE		QuatE		SimplE	ComplEx
			hits@1	hits@3	hits@5	hits@3	hits@5	hits@3	hits@5	hits@3	hits@5	hits@5	hits@5
Mug	Q386215	6	1	1	1	1	1	0.333	1	1	1	1	0.666
Measuring cup	Q907099	3	0	0.666	1	0	0	0	0	0	0	0	0.666
Office chair	Q1021686	9	0.222	1	1	0	0.333	0	1	0	0	0	0
Wheelchair	Q191931	8	0.75	1	1	1	1	1	1	0	0	0.375	0.625
Golden Delicious	Q201996	4	0	0	1	0	0	1	1	0	0	0	0
Granny Smith	Q506040	3	0	0	0	0	0	0	1	0	0	0	1
Gala apple	Q494259	3	0	1	1	0	0	1	1	0.666	1	0	1
Summary:		36	0.389	0.778	0.917	0.389	0.472	0.472	0.917	0.222	0.25	0.25	0.472

Table 2: A summary of the results of the PET task on determining the correct subclass for the cups, chairs, and apples detected in the 36 different observation graphs embedded using several different embedding approaches. The results show that TransE has the highest hits@1 performance on the examples containing the mug, wheelchair, and office chairs. For the sake of brevity, columns without a single hit are omitted.

ding model was RotatE, with 0.472 overall accuracy of hits@3 and 0.917 overall accuracy of hits@5. Interestingly, RotatE performs better with observations containing apples than TransE, although the limited sample size for each of the examples limits our ability to draw strong conclusions. A summary of the results can be seen in Table 2.

Discussion

The results presented in the previous section indicated that using TransE as embedding model yielded promising results in our dataset. The correctly linked entities indicate that certain properties, such as the part-of relation, which both the mugs and the office and wheelchairs contained, could be one of the reasons for the increased performance. Another factor contributing to the superior performance compared to other embedding models could be the manner in which the observational embeddings were generated, as summing over the property- and relation-embeddings align with the specific characteristics of the TransE embedding method.

Limitations

One of the main limitations of the experiment stems from the limited symbolic descriptions of entities in the target knowledge graph, in this case WikiData5m. Although the dataset contains millions of entities, its inclusion of WikiData resources describing general concepts, such as qualities of objects or common physical objects that the robot would need to detect, is limited. For example, the general concepts related to size (`wdt:Q322481`) are not contained in the dataset. Furthermore, the embedding models are trained using WikiData and Wikipedia textual corpus. However, a multimodal embedding model such as (Wang et al. 2019) could exploit the sensory data and lift the limitation of relying on purely symbolic description of the observed entities, and potentially improve the matching results.

Open Challenges

On a system level, the question of how to address the error and noise in the sensory data and the perception algorithms, as well as the scalability of the system, needs to be investigated. Related to PEL, unsolved challenges include the integration of environmental context involving other objects into the embeddings, and the temporal linking of observation graphs with the entities they contain. Furthermore, some properties of objects are irrelevant, or instance-specific, instead of class-specific (i.e. most objects can have multiple colors). Using context vectors as suggested by the use of conceptual spaces could provide a way to prioritize certain properties for different concepts (Adams and Raubal 2009). The availability, completeness and safety related to the reliability of the contents of Semantic Web resources is not yet investigated. While some of these issues have been pointed out in (Adamik and Schlobach 2023), no comprehensive analysis of the resources is given so far.

Conclusion

In this work, we presented a reasoning system that organizes the sensory data of robotic agents based on the ORKA ontology, and improves robotic perception with matching the perceived entities to the Semantic Web. We described the Perceived-Entity Typing task, which aims at improving robotic perception, by refining the label of the entity contained in the observation graph. Furthermore, we explored the performance of pre-trained embedding models from WikiData on these tasks. The results suggest that while some embedding methods provide promising initial results, the datasets should be enhanced by including more commonsense concepts, that could be used by the robots.

As future work, we plan to implement and embed our system in a physical robot's perception-action loop, enabling it to guide actions based on the observation graph. Instead of relying on pre-trained embedding models for the Perceived-Entity Typing task, we aim to build a database of physical object descriptions, expecting this to improve embedding performance.

This work presents the first attempt at introducing a system for solving the Perceived-Entity Typing task to utilize the Semantic Web more effectively. We invite the community to contribute and build upon these initial results.

References

Adamik, M.; Pernisch, R.; Tiddi, I.; and Schlobach, S. 2024a. Advancing Robotic Perception with Perceived-Entity Linking. In *The Semantic Web - ISWC 2024 - 23rd International Semantic Web Conference, Baltimore, The United States of America, November 11-15, 2024, Proceedings*, Lecture Notes in Computer Science. Springer.

Adamik, M.; Pernisch, R.; Tiddi, I.; and Schlobach, S. 2024b. ORKA: An Ontology for Robotic Knowledge Acquisition. In *Proceedings of the 24th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2024*, Lecture Notes in Computer Science. Springer.

Adamik, M.; and Schlobach, S. 2023. Towards a Definition and Conceptualisation of the Perceived-Entity Linking Problem. In Pomarlan, M.; Beßler, D.; Borgo, S.; and Diab, M., eds., *Proceedings of the Workshop on Ontologies for Autonomous Robotics co-located with The 32nd IEEE International Conference on Robot and Human Interactive Communication (IEEE RO-MAN 2023), Busan, S. Korea, August 28, 2023*, volume 3595 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Adams, B.; and Raubal, M. 2009. A Metric Conceptual Space Algebra. In Hornsby, K. S.; Claramunt, C.; Denis, M.; and Ligozat, G., eds., *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009, Proceedings*, volume 5756 of *Lecture Notes in Computer Science*, 51–68. Springer.

Antoniou, G.; Groth, P.; van Harmelen, F.; and Hoekstra, R. 2012. *A Semantic Web Primer, 3rd Edition*. MIT Press. ISBN 978-0-262-01828-9.

Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; and Ives, Z. G. 2007. DBpedia: A Nucleus for a Web of Open Data. In Aberer, K.; Choi, K.; Noy, N. F.; Allemang, D.; Lee, K.; Nixon, L. J. B.; Golbeck, J.; Mika, P.; Maynard, D.; Mizoguchi, R.; Schreiber, G.; and Cudré-Mauroux, P., eds., *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*, volume 4825 of *Lecture Notes in Computer Science*, 722–735. Springer.

Besl, P. J.; and Jain, R. C. 1985. Three-Dimensional Object Recognition. *ACM Comput. Surv.*, 17(1): 75–145.

Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; and Yakhnenko, O. 2013. Translating Embeddings for Modeling Multi-relational Data. In Burges, C. J. C.; Bottou, L.; Ghahramani, Z.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, 2787–2795.

Brooks, R. A. 1986. A robust layered control system for a mobile robot. *IEEE J. Robotics Autom.*, 2(1): 14–23.

Dai, H.; and Zeng, Z. 2023. From Ultra-Fine to Fine: Fine-tuning Ultra-Fine Entity Typing Models to Fine-grained. In Rogers, A.; Boyd-Graber, J. L.; and Okazaki, N., eds., *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, 2259–2270. Association for Computational Linguistics.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR*, abs/2010.11929.

Fischer, L.; Hasler, S.; Deigmoeller, J.; Schnuerer, T.; Redert, M.; Pluntke, U.; Nagel, K.; Senzel, C.; Ploennigs, J.; Richter, A.; and Eggert, J. 2018. Which tool to use? Grounded reasoning in everyday environments with assistant robots. In Steinbauer, G.; and Ferrein, A., eds., *Proceedings of the 11th Cognitive Robotics Workshop 2018, co-located with 16th International Conference on Principles of Knowledge Representation and Reasoning, CogRob@KR 2018, Tempe, AZ, USA, October 27th, 2018*, volume 2325 of *CEUR Workshop Proceedings*, 3–10. CEUR-WS.org.

Gan, J.; Luo, J.; Wang, H.; Wang, S.; He, W.; and Huang, Q. 2021. Multimodal Entity Linking: A New Dataset and A Baseline. In Shen, H. T.; Zhuang, Y.; Smith, J. R.; Yang, Y.; César, P.; Metze, F.; and Prabhakaran, B., eds., *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021*, 993–1001. ACM.

Horrocks, I.; Patel-Schneider, P. F.; Boley, H.; Tabet, S.; Grosof, B.; Dean, M.; et al. 2004. SWRL: A semantic web rule language combining OWL and RuleML. *W3C Member submission*, 21(79): 1–31.

Kazemi, S. M.; and Poole, D. 2018. SimplE Embedding for Link Prediction in Knowledge Graphs. In Bengio, S.; Wallach, H. M.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, 4289–4300.

Redmon, J.; Divvala, S. K.; Girshick, R. B.; and Farhadi, A. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 779–788. IEEE Computer Society.

Rossetto, L.; Baumgartner, M.; Ashena, N.; Ruosch, F.; Pernischová, R.; and Bernstein, A. 2020. LifeGraph: A Knowledge Graph for Lifelogs. In Gurrin, C.; Schoffmann, K.; Jonsson, B. T.; Dang-Nguyen, D.; Lokoc, J.; Tran, M.; and Hürst, W., eds., *Proceedings of the Third ACM Workshop on Lifelog Search Challenge, LSC@ICMR 2020, Dublin, Ireland, June 8-11, 2020*, 13–17. ACM.

Sciutti, A.; Mara, M.; Tagliasco, V.; and Sandini, G. 2018. Humanizing Human-Robot Interaction: On the Importance of Mutual Understanding. *IEEE Technol. Soc. Mag.*, 37(1): 22–29.

Sirin, E.; Parsia, B.; Grau, B. C.; Kalyanpur, A.; and Katz, Y. 2007. Pellet: A practical OWL-DL reasoner. *J. Web Semant.*, 5(2): 51–53.

Stanton, C. J.; and Williams, M. 2003. Grounding Robot Sensory and Symbolic Information Using the Semantic Web. In Polani, D.; Browning, B.; Bonarini, A.; and Yoshida, K., eds., *RoboCup 2003: Robot Soccer World Cup*

VII, volume 3020 of *Lecture Notes in Computer Science*, 757–764. Springer.

Suchanek, F. M.; Kasneci, G.; and Weikum, G. 2007. Yago: a core of semantic knowledge. In Williamson, C. L.; Zurko, M. E.; Patel-Schneider, P. F.; and Shenoy, P. J., eds., *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, 697–706. ACM.

Sun, Z.; Deng, Z.; Nie, J.; and Tang, J. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Tenorth, M.; Perzylo, A. C.; Lafrenz, R.; and Beetz, M. 2013. Representation and Exchange of Knowledge About Actions, Objects, and Environments in the RoboEarth Framework. *IEEE Trans Autom. Sci. Eng.*, 10(3): 643–651.

Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; and Bouchard, G. 2016. Complex Embeddings for Simple Link Prediction. In Balcan, M.; and Weinberger, K. Q., eds., *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, 2071–2080. JMLR.org.

Vrandecic, D.; and Krötzsch, M. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM*, 57(10): 78–85.

Waibel, M.; Beetz, M.; Civera, J.; D'Andrea, R.; Elfring, J.; Gálvez-López, D.; Häussermann, K.; Janssen, R.; Montiel, J. M. M.; Perzylo, A.; Schießle, B.; Tenorth, M.; Zweigle, O.; and van de Molengraft, R. 2011. RoboEarth. *IEEE Robotics Autom. Mag.*, 18(2): 69–82.

Wang, X.; Gao, T.; Zhu, Z.; Zhang, Z.; Liu, Z.; Li, J.; and Tang, J. 2021. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *Trans. Assoc. Comput. Linguistics*, 9: 176–194.

Wang, Z.; Li, L.; Li, Q.; and Zeng, D. 2019. Multimodal Data Enhanced Representation Learning for Knowledge Graphs. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14-19, 2019*, 1–8. IEEE.

Wu, Z.; Xu, Y.; Yang, Y.; Zhang, C.; Zhu, X.; and Ji, Y. 2017. Towards a Semantic Web of Things: A Hybrid Semantic Annotation, Extraction, and Reasoning Framework for Cyber-Physical System. *Sensors*, 17(2): 403.

Yang, B.; Yih, W.; He, X.; Gao, J.; and Deng, L. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Young, J.; Basile, V.; Kunze, L.; Cabrio, E.; and Hawes, N. 2016. Towards Lifelong Object Learning by Integrating Situated Robot Perception and Semantic Web Mining. In Kaminka, G. A.; Fox, M.; Bouquet, P.; Hüllermeier, E.; Dignum, V.; Dignum, F.; and van Harmelen, F., eds., *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, volume 285 of *Frontiers in Artificial Intelligence and Applications*, 1458–1466. IOS Press.

Young, J.; Kunze, L.; Basile, V.; Cabrio, E.; Hawes, N.; and Caputo, B. 2017. Semantic web-mining and deep vision for lifelong object discovery. In *2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017*, 2774–2779. IEEE.

Zhang, L.; Li, Z.; and Yang, Q. 2021. Attention-Based Multimodal Entity Linking with High-Quality Images. In Jensen, C. S.; Lim, E.; Yang, D.; Lee, W.; Tseng, V. S.; Kalogeraki, V.; Huang, J.; and Shen, C., eds., *Database Systems for Advanced Applications - 26th International Conference, DASFAA 2021, Taipei, Taiwan, April 11-14, 2021, Proceedings, Part II*, volume 12682 of *Lecture Notes in Computer Science*, 533–548. Springer.

Zhang, S.; Tay, Y.; Yao, L.; and Liu, Q. 2019. Quaternion Knowledge Graph Embeddings. In Wallach, H. M.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E. B.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, 2731–2741.

Zheng, Q.; Wen, H.; Wang, M.; and Qi, G. 2022. Visual Entity Linking via Multi-modal Learning. *Data Intell.*, 4(1): 1–19.

Zhu, Z.; Xu, S.; Tang, J.; and Qu, M. 2019. GraphVite: A High-Performance CPU-GPU Hybrid System for Node Embedding. In Liu, L.; White, R. W.; Mantrach, A.; Silvestri, F.; McAuley, J. J.; Baeza-Yates, R.; and Zia, L., eds., *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, 2494–2504. ACM.