Towards a Definition and Conceptualisation of the Perceived-Entity Linking Problem

Mark Adamik[1,*], Stefan Schlobach[1]

[1]*Vrije Universiteit Amsterdam, Netherlands*

Abstract

In this paper we present an initial conceptualization of the problem of Perceived-Entity Linking (PEL), which is inspired by the entity linking task used in the Natural Language Processing domain. The task was adopted to represent a problem knowledge-driven embodied systems face, which concerns the linking of the representations of perceived entities to a target knowledge graph. We provide an initial description of the problem, demonstrated with a motivating example, and followed by a preliminary case study, where we identify some of the challenges and opportunities PEL presents both to the engineers of knowledge-driven autonomous agents and to the knowledge engineers of the Semantic Web community.

Keywords

knowledge representation, ontology, robotics, perception

1. Introduction

Autonomous agents, such as robotic systems, base their operation on the sensory information extracted from the environment. In order to be able to operate in an autonomous manner, these systems need to perceive different aspects of the often dynamically changing environment. Perception has been a fundamental problem of many fields of science, from philosophy through psychology and biology to computer vision and artificial intelligence. Consequently, many notions and definitions of perception exist, but for the purposes of this paper, perception is defined as the process of creating symbolic representations through sensory receptors from stimulus coming from the environment.

Within the field of robotics, one of the most popular methods to represent symbolic knowledge is in the form of ontologies, which can be defined as shared conceptualisations of a domain[1]. There are several different ontologies that are being used for representing knowledge in the field of robotics (for a recent survey of ontology-based robotic systems, consult [2, 3]). Most of these systems have some form of representation of perceptual data. These knowledge structures however are usually engineered by domain experts, and therefore could be considered as a top-down approach to creating conceptualizations. From a philosophical point of view however, it can be argued that all symbolic knowledge humanity ever acquired have come through some form of perceptual input, and therefore - in theory - knowledge could be coupled with some (set of) perceptual data. Harnad [4] defined the direction of the coupling process that points from the

---

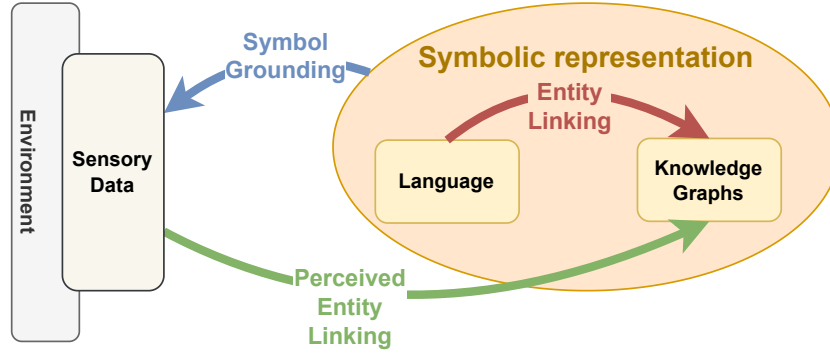CEUR Workshop Proceedings (CEUR-WS.org)

Figure 1: A simplified view of the Perceived Entity Linking problem, and its relation to the other, similar problems. Here the Sensory Data could mean both raw sensory data and processed data.

symbolic description to the direction of the sensory data as the symbol grounding problem, which has been an influential idea for many different fields, such as cognition, language and computer science.

In this paper, we present a first attempt to conceptualize the problem that we call Perceived-Entity Linking (PEL), which is the process of linking perceptual data to a corresponding entity in a target knowledge graph, where perceptual data could mean either the raw sensory output of the sensing devices, or the outputs of the perception pipelines implemented in the given system. The term of PEL was coined using a problem from the Natural Language Processing (NLP) domain called Entity Linking as an inspiration, which is the process of linking entities occurring in text to the corresponding entities in a target knowledge graphs. A simplified conceptualisation of these terms can be seen in Figure 1. The conceptualisation process follows an ontology-based approach. As this task opens up a vast array of problems and possible directions, we limit our initial investigations to the domain of physical objects and their properties.

The following section provides a definition and analysis of the PEL task, followed by Section 3, that provides a general overview of some of the ontology-based approaches to robotics, and their relation to PEL. Section 4 provides a preliminary case study using a single physical object and comparing some of the representations, while addressing the shortcoming and challenges posed to the task of PEL. Section 5 concludes the paper, outlining some possible directions of future work and open challenges.

2. Problem Analysis

2.1. Problem Description

Underlying the vision of the semantic web is the assumption that computers have access to structured collections of information and inference rules that allow automated reasoning to be done on the information[5]. This information is then stored openly, usually in the form of knowledge graphs. Although the knowledge shared in these graphs such as DBpedia [6], ConceptNet [7] and WikiData [8] are often considered common-sense knowledge, a robot cannot readily access the information contained within, and can't relate it to the environment without

human engineering. Accessing the knowledge encoded in these resources requires the autonomous agent to establish a connection between the context it's currently embedded in, and the target knowledge graph. In order to do so, the agent first needs to represent the context through its sensory capabilities. In an ontological approach, the Semantic Sensor Network (SSN) Ontology [9] is aimed at describing the sensory capabilities and the observations provided by these sensors. This framework however is not sufficient to model the PEL problem.

Among the examined knowledge-based robotic systems, this connection is commonly established by one of the two ways:

1. Hand-crafting the system to link to the correct identifier in an online knowledge-graph (e.g. [10, 11]).
2. Reducing the problem of PEL to a local[1] problem by not utilizing general, external knowledge graphs, but instead creating an separate knowledge framework, often with focus on a specific set of tasks (e.g. [12, 13, 14, 15]).

In both cases, first a local representation is created, which then - in the case of the first approach - is linked to an external knowledge graph. This suggests that the problem of PEL could be conceptualized with two levels, where the first level is concerned with using the sensory data and integrating different perceptual processes to create a representation using the knowledge representation format the knowledge-driven system is utilizing. The second level is concerned with aligning these representations with external knowledge graphs. The the two levels of PEL therefore try to answer the following questions:

Level one: Given a set of sensors and their outputs, how could the sensory data be linked to the corresponding entities in the knowledge base of the knowledge-driven system?

Level two: Given a representation of a perceived entity and a target knowledge graph, how could the representation be linked to a corresponding entity in a target knowledge graph?

There are two distinctive differences in the two levels. The first one lies in the input, where level one has a finite set of sensors and their corresponding outputs corresponding to physical stimulus the sensor is responding to, whereas the possible representations serving as input for level two are less constrained. The second difference lies in the target knowledge graph. The first level has a representation that is tailored to the system by the designers, whereas the target knowledge graph in level two is usually designed for representing general knowledge.

2.2. Motivating Example

Consider the following example: a robot equipped with an RGB camera captures an image of an orange. In order for the system to create a symbolic representation that could be associated with the concept of an orange, the object needs to be detected. Taking a popular object detection algorithm, such as Yolo [17] as an example, the output of the algorithm would be a label, a bounding box indicating the location of the object on the image, and a value between 0 and 1 indicating the prediction certainty. Taking only these values, as a starting point, if the robot would be required to access the entity based on the label 'orange' in a large knowledge base such as WikiData, additional knowledge would be required to differentiate between the several

---

[1]In this context local refers to a system crafted by the designers to solve the PEL problem, and not necessarily to physical locality.
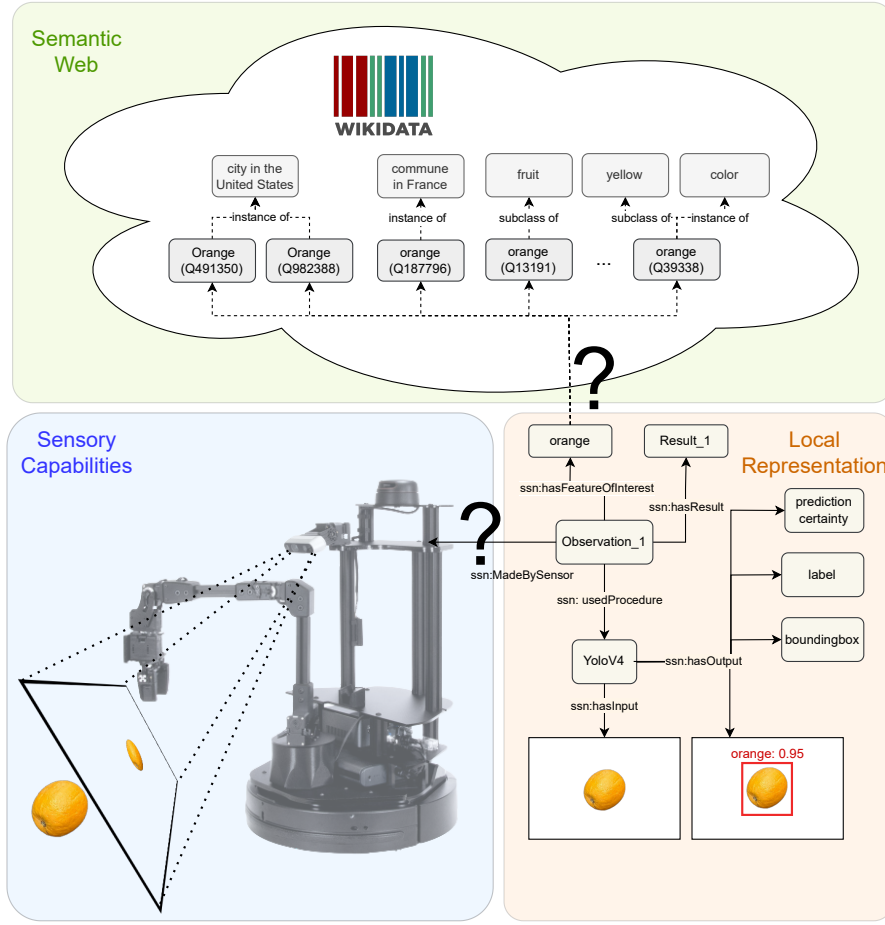
Figure 2: A motivating example of the PEL problem, formulated using the SSN ontology and WikiData [16] as the target knowledge graph. The pipeline consists of a single procedure, namely a YoloV4 object detection algorithm, and the detection instance is encompassed into an *Observation* instance. An instance of type *featureOfInterest* denotes the identity of the observed entity. The two levels of PEL consist of determining how features of the sensory data create a *featureOfInterest*, and determining which resource should the *featureOfInterest* be linked to in the target knowledge graph.

possibilities. In this specific scenario, the robot would need to have a (local) representation, from which it could be inferred that the label *orange* represents a class which is a subclass of the *fruit* class. Using a local knowledge graph to link to the correct entity would transform the problem of PEL to an ontology alignment problem [18]. We would argue however, that this solution would only solve the Level 2 problem, as it would only move the problem to a local representation (namely linking the label, bounding box and prediction certainty to class *Orange*). Furthermore, this solution assumes a label is produced by the perception system. However, in the case where an object detection algorithm is not able to provide a label (e.g. if the implemented perception system can only estimate the size, shape and color of the object), the name of the class and subclass couldn't be used, and the matching would need to be performed based on other resources

in the knowledge graph.

2.3. Similar Tasks

In [19], the authors describe the process of reification as the bidirectional mapping between symbolic systems and the sensory data. They propose a reification engine, based on biologically inspired cognitive architectures, that is capable of performing the reification task using *PerCepts*, which are abstracted symbolic properties of objects. The problem of Data Association is mainly concerned with tracking sensory measurements, e.g. linking the measurements together over time. While the techniques developed in this field could be considered relevant to the problem of PEL, the main concern of data-association seems to be mainly finding interframe correspondences between objects [20].

Another related term is *anchoring*, introduced by Chella et al. [21], which is considered a special case of the symbol grounding problem, where symbols represent individual physical objects. While this term is used in a very similar way to PEL, the main differences are that PEL is concerned with linking to knowledge graphs, and the linked entities are not limited to physical objects, but could also represent properties as well.

3. Related Work

This section briefly reviews how robotic systems that employ ontologies represent properties of the perceived objects, and the means of obtaining their measurements, when specified. The robotic systems considered are largely taken from recent surveys [2, 3, 22]. The main emphasis when investigating these systems are on their representations of physical objects and the processes that support the calculation of these properties.

Diab et al. developed a reasoning framework called Perception and Manipulation Knowledge (PMK) [15], where both object properties and algorithms are represented in the ontology. Their described method uses an AR tag tracking approach, and encodes the semantic knowledge manually, circumventing the problem of PEL.

The Ontology-based Unified Robot Knowledge (OUR-K) framework [23] has a promising structure, as it integrates low-level perceptual sensory data with the extracted perceptual features (i.e. color, texture and SIFT) and couples these features to perceptual concepts. For all three perceptual features the corresponding algorithm is linked using the *hasFeatureAlgorithm* predicate. Unfortunately, no online version of the system can be found, therefore the exact implementation of the perceived entity linking process is unknown.

Socio-physical Model of Activities (SOMA)[24] is an ontological framework designed for aiding robots to solve everyday manipulation activities, with a special focus on home environments. SOMA is built on the DUL ontology and also serves as a foundational ontology behind the KnowRob[13] framework. SOMA introduces a more nuanced object representation, and using DUL as the upper ontology, divides qualities into social (e.g. cleanliness) and physical qualities, where physical qualities are further divided into intrinsic and extrinsic qualities. In this model, extrinsic qualities depend on the environment, while intrinsic qualities are independent of context. Although SOMA is not concerned with the detailed sensory capabilities of the robot,

another extension of KnowRob, the Semantic Robot Description Language (SRDL) [25] includes a taxonomy of the different sensory categories and some software categories.

A related project, RoboEarth [12] was an ambitious international effort running between 2010 and 2014 with an overarching purpose of providing a platform for robots to store and share information in a scalable and reusable way. Much of the infrastructure around the project however seems to be unavailable.

RoboSherlock [14, 26] is a cognitive vision system that extends the previously mentioned KnowRob framework. It provides a perception pipeline that can accommodate different perceptual processes. It is built on a promising work by Nyga et al. [27] that proposes a partial solution to PEL, where a collection of image processing techniques called annotators are used to provide a detailed description of the objects perceived by the robot.

The following section presents a simple use-case that aims to investigate how some of the ontologies represented in this section could solve the two different levels of PEL described in the previous section.

4. Case Study

In this section we use a preliminary proof-of-concept case study to compare how some of the knowledge-based robotic systems, namely PMK[2] and RoboSherlock[3] could represent an orange on the two levels of the PEL task using their currently available ontologies to create an instance of an orange. In this manually created example an ideal object detection algorithm is assumed, that can provide all the information the ontologies afford. Since KnowRob, SOMA, RoboSherlock and SRDL described above are part of the same infrastructure, only RoboSherlock is considered for the case study as it is the most relevant to the PEL task. The other frameworks described in the previous section had no available implementations that were readily accessible online.

4.1. Linking to a Local Representation

The ontologies of both the PMK and RoboSherlock systems were examined using Protégé[4] (see Figure 3). In this use-case, both ontologies represent two key aspects of the object: *Material* and *Color*, with both ontologies lacking the exact color (orange). PMK considers three more classes: *ObjectName*, *Weight* and *SIFT features*. SIFT features however are only representative of the current viewpoint, therefore we would not consider it to be a good representative of a constant quality. RoboSherlock includes the shape and size of the object, both of which could be considered important features that represent the core identity of the concept. Although the relative nature of the size representation should be addressed.

It is important to note that RoboSherlock is focused solely on visual features, therefore weight and objectname are not considered in this example. These features however could also be included when considering the KnowRob framework as a whole.

---

[2]https://github.com/MohammedDiab1/PMK/tree/master/PMK/interfaceprologcpp/owl
[3]https://github.com/RoboSherlock/robosherlock/tree/master/robosherlock/owl
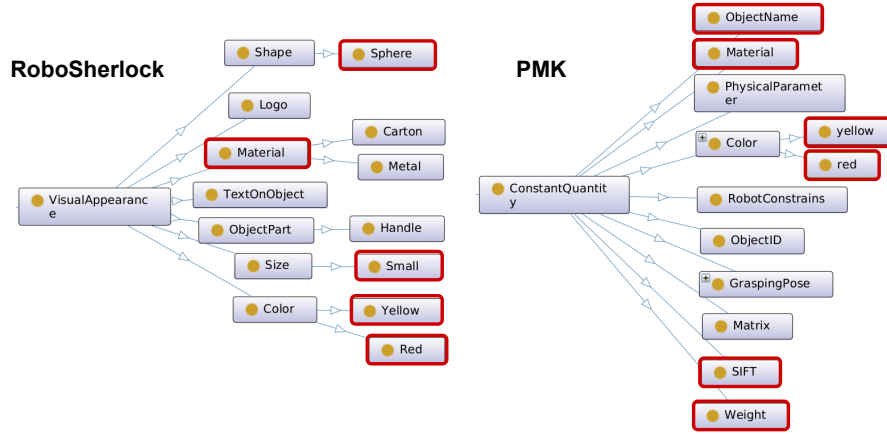[4]https://protege.stanford.edu/

Figure 3: Examining the object representation capabilities of the RoboSherlock and PMK ontologies using the OntoGraf plugin of Protégé. The classes that are possibly relevant to the perception of an orange are highlighted. Since neither of the ontologies contain the color *orange* therefore both red and yellow are marked.

4.2. Linking to an External Representation

For this task, three target knowledge graphs containing common-sense knowledge are considered, namely ConceptNet [5], DBPedia [6] and WikiData [7]. As the contents of Yago [28] is based on WikiData, it is excluded from the comparison. For all three resources, the relationships providing information about the physical or visual qualities of the fruit orange is considered.

Upon examining all the three resources a perceived instance of orange could be linked to, ConceptNet had the least structured information, mixing several meanings of orange. WikiData and DBpedia presented more structured information, and contained images depicting oranges that served as grounding. While several pieces of information that is characteristic of the concept *orange fruit* is represented, such as the color and the parent classes (DBpedia and WikiData), and nutritional information (DBpedia only) there is no information about the material, shape, or the average sizes or weights of the fruits that could be used for the PEL task.

This preliminary investigation has revealed some weaknesses and opportunities in both the Semantic Web resources and ontology-based approaches

5. Conclusion and Future Work

In this paper we introduced the task of Perceived-Entity Linking, which addresses the problem of how the different representations abstracted from the perceptual sensory data could be linked to a target knowledge graph. We performed a preliminary analysis of this problem, during which two layers of the linking task were identified, namely linking to a local representation that is engineered for the specific system, and linking to an external knowledge source. After briefly

---

[5]https://conceptnet.io/c/en/orange
[6]https://dbpedia.org/page/Orange_(fruit)
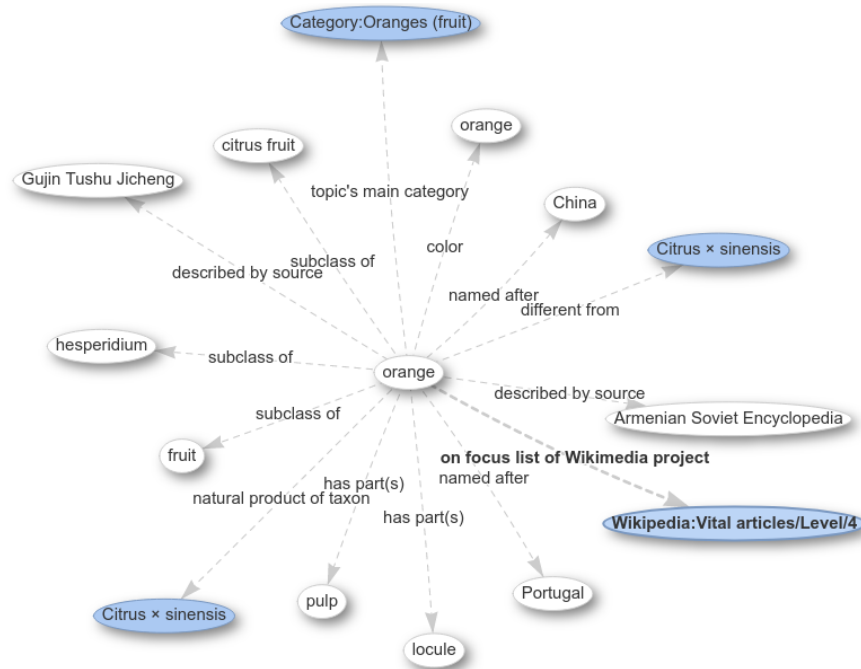[7]https://www.wikidata.org/wiki/Q13191

Figure 4: A view of the orange class in WikiData, visualized using the Explorer function of the built-in Query Builder tool.

describing some of recent knowledge-driven robotic systems we described a simple case-study, which served as a motivating example, and as an investigation of the problem domain. The results of this investigation indicated that while some of the robotic systems could partially perform the PEL task, with the RoboSherlock's ontology being the most representative, all of the systems examined could be extended with a more complete and precise object representation method. Furthermore we discovered that all of the large, "common-sense" knowledge graphs investigated lack some common-sense qualities of an orange, such as shape, weight, and size.

As an initial effort, this paper does not consider a number of issues, which are deferred and used as further motivation for the work that follows this paper. Some of the open questions are:

- What type of entity should PEL link to? In which cases should the system link to instances and in which cases should the linking be performed to classes?
- How can sensory experiences from different modalities be combined to give rise to emergent features?
- How do the challenges of the named entity linking task of NLP (e.g. scalability, name variations, multiple languages) translate to the task of PEL?

Future work will focus on addressing each of the above mentioned issues, as well as on providing a formal definition of the problem, followed by a structured analysis of the relevant knowledge available on the Semantic Web resources, a thorough and systematic examination of how knowledge-driven robotic systems solve the different levels of PEL and on analyzing the different approaches that could be taken to solve the problem of Perceived-Entity Linking.

References

[1] N. Guarino, D. Oberle, S. Staab, What Is an Ontology?, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, International Handbooks on Information Systems, Springer, Berlin, Heidelberg, 2009, pp. 1–17. URL: https://doi.org/10.1007/978-3-540-92673-3_0. doi:10.1007/978-3-540-92673-3_0.

[2] A. Olivares-Alarcos, D. Beßler, A. Khamis, P. Goncalves, M. K. Habib, J. Bermejo-Alonso, M. Barreto, M. Diab, J. Rosell, J. Quintas, J. Olszewska, H. Nakawala, E. Pignaton, A. Gyrard, S. Borgo, G. Alenyà, M. Beetz, H. Li, A review and comparison of ontology-based approaches to robot autonomy, The Knowledge Engineering Review 34 (2019). URL: https://www.cambridge.org/core/journals/knowledge-engineering-review/article/review-and-comparison-of-ontologybased-approaches-to-robot-autonomy/C0EF5777394264ABA95B28DBEA215FFF. doi:10.1017/S0269888919000237, publisher: Cambridge University Press.

[3] E. Aguado, R. Sanz, Using Ontologies in Autonomous Robots Engineering, 2021, pp. https://www.intechopen.com/online–first/using. doi:10.5772/intechopen.97357.

[4] S. Harnad, The symbol grounding problem, Physica D: Nonlinear Phenomena 42 (1990) 335–346. URL: https://www.sciencedirect.com/science/article/pii/0167278990900876. doi:10.1016/0167-2789(90)90087-6.

[5] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American 284 (2001) 34–43. URL: https://www.jstor.org/stable/26059207, publisher: Scientific American, a division of Nature America, Inc.

[6] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, DBpedia: A Nucleus for a Web of Open Data, in: K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, P. Cudré-Mauroux (Eds.), The Semantic Web, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2007, pp. 722–735. doi:10.1007/978-3-540-76298-0_52.

[7] R. Speer, J. Chin, C. Havasi, ConceptNet 5.5: An Open Multilingual Graph of General Knowledge, Proceedings of the AAAI Conference on Artificial Intelligence 31 (2017). URL: https://ojs.aaai.org/index.php/AAAI/article/view/11164. doi:10.1609/aaai.v31i1.11164, number: 1.

[8] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (2014) 78–85. URL: https://dl.acm.org/doi/10.1145/2629489. doi:10.1145/2629489.

[9] K. Taylor, A. Haller, M. Lefrancois, S. Cox, R. Garcıa-Castro, D. Le-Phuoc, J. Lieberman, C. Stadler, The Semantic Sensor Network Ontology, Revamped (????).

[10] L. Fischer, S. Hasler, J. Deigmoller, T. Schnurer, M. Redert, U. Pluntke, K. Nagel, C. Senzel, J. Ploennigs, A. Richter, J. Eggert, Which tool to use? Grounded reasoning in everyday environments with assistant robots (????).

[11] J. Young, V. Basile, L. Kunze, E. Cabrio, N. Hawes, Towards lifelong object learning by integrating situated robot perception and Semantic Web mining, Frontiers in Artificial Intelligence and Applications 285 (2016). URL: https://ora.ox.ac.uk/objects/uuid:5f8d458f-efad-4677-a292-083c319cb835, iSBN: 9781614996712 Publisher: IOS Press.

[12] M. Waibel, M. Beetz, J. Civera, R. D'Andrea, J. Elfring, D. Gálvez-López, K. Häussermann,

R. Janssen, J. Montiel, A. Perzylo, B. Schießle, M. Tenorth, O. Zweigle, R. V. De Molengraft, RoboEarth, IEEE Robotics & Automation Magazine 18 (2011) 69–82. doi:10.1109/MRA.2011.941632, conference Name: IEEE Robotics & Automation Magazine.

[13] M. Tenorth, M. Beetz, KnowRob: A knowledge processing infrastructure for cognition-enabled robots, The International Journal of Robotics Research 32 (2013) 566–590. URL: https://doi.org/10.1177/0278364913481635. doi:10.1177/0278364913481635, publisher: SAGE Publications Ltd STM.

[14] M. Beetz, F. Balint-Benczedi, N. Blodow, D. Nyga, T. Wiedemeyer, Z.-C. Marton, RoboSherlock: Unstructured information processing for robot perception, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Seattle, WA, USA, 2015, pp. 1549–1556. URL: http://ieeexplore.ieee.org/document/7139395/. doi:10.1109/ICRA.2015.7139395.

[15] Mohammed Diab, Aliakbar Akbari, Muhayy Ud Din, Jan Rosell, PMK—A Knowledge Processing Framework for Autonomous Robotics Perception and Manipulation, 2019. URL: https://www.mdpi.com/1424-8220/19/5/1166, titleTranslation:.

[16] D. Vrandečić, Wikidata: a new platform for collaborative data collection, in: Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion, Association for Computing Machinery, New York, NY, USA, 2012, pp. 1063–1064. URL: https://dl.acm.org/doi/10.1145/2187980.2188242. doi:10.1145/2187980.2188242.

[17] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, 2020. URL: http://arxiv.org/abs/2004.10934. doi:10.48550/arXiv.2004.10934, arXiv:2004.10934 [cs, eess].

[18] F. Scharffe, D. Fensel, Correspondence Patterns for Ontology Alignment, in: A. Gangemi, J. Euzenat (Eds.), Knowledge Engineering: Practice and Patterns, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2008, pp. 83–92. doi:10.1007/978-3-540-87696-0_10.

[19] J. P. Gunderson, L. F. Gunderson, Robots, Reasoning, and Reification, Springer US, Boston, MA, 2009. URL: https://link.springer.com/10.1007/978-0-387-87488-3. doi:10.1007/978-0-387-87488-3.

[20] B. Collins, Introduction to Data Association (????). URL: https://www.cse.psu.edu/~rtc12/CSE598C/datassocPart1.pdf.

[21] A. Chella, S. Coradeschi, M. Frixione, A. Saf, Perceptual Anchoring via Conceptual Spaces (2004) 6.

[22] W. Liu, A. Daruna, M. Patel, K. Ramachandruni, S. Chernova, A survey of semantic reasoning frameworks for robotic systems, Robotics and Autonomous Systems (2022) 104294. URL: https://www.sciencedirect.com/science/article/pii/S092188902200183X. doi:10.1016/j.robot.2022.104294.

[23] G. H. Lim, I. H. Suh, H. Suh, Ontology-Based Unified Robot Knowledge for Service Robots in Indoor Environments, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 41 (2011) 492–509. doi:10.1109/TSMCA.2010.2076404.

[24] D. Beßler, R. Porzel, M. Pomarlan, A. Vyas, S. Höffner, M. Beetz, R. Malaka, J. Bateman, Foundations of the Socio-physical Model of Activities (SOMA) for Autonomous Robotic Agents, 2020. URL: http://arxiv.org/abs/2011.11972. doi:10.48550/arXiv.2011.11972, arXiv:2011.11972 [cs].

[25] L. Kunze, T. Roehm, M. Beetz, Towards semantic robot description languages, in: 2011 IEEE International Conference on Robotics and Automation, IEEE, Shanghai, China, 2011, pp. 5589–5595. URL: http://ieeexplore.ieee.org/document/5980170/. doi:10.1109/ICRA.2011.5980170.

[26] F. B'alint-Bencz'edi, J.-H. Worch, D. Nyga, N. Blodow, P. Mania, Z. M'arton, M. Beetz, RoboSherlock: Cognition-enabled Robot Perception for Everyday Manipulation Tasks, ArXiv (2019). URL: https://www.semanticscholar.org/paper/RoboSherlock%3A-Cognition-enabled-Robot-Perception-B%27alint-Bencz%27edi-Worch/b02ee5c14d4c51faa3f2165294c975dc53cd6864.

[27] D. Nyga, F. Balint-Benczedi, M. Beetz, PR2 looking at things — Ensemble learning for unstructured information processing with Markov logic networks, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Hong Kong, China, 2014, pp. 3916–3923. URL: http://ieeexplore.ieee.org/document/6907427/. doi:10.1109/ICRA.2014.6907427.

[28] T. Pellissier Tanon, G. Weikum, F. Suchanek, YAGO 4: A Reason-able Knowledge Base, in: A. Harth, S. Kirrane, A.-C. Ngonga Ngomo, H. Paulheim, A. Rula, A. L. Gentile, P. Haase, M. Cochez (Eds.), The Semantic Web, volume 12123, Springer International Publishing, Cham, 2020, pp. 583–596. URL: http://link.springer.com/10.1007/978-3-030-49461-2_34. doi:10.1007/978-3-030-49461-2_34, series Title: Lecture Notes in Computer Science.