Towards Seamless Human-Robot Dialogue through a Robot Action Ontology

Diego Reforgiato Recupero[1,*], Lorenzo Boi[1]

[1]Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy

Abstract

This research paper introduces a novel methodology enabling the Zora humanoid robot to effectively engage in dynamic interactions by responding to user queries and complementing its responses with appropriate gestures. Notably, these inquiries may extend beyond mere questions to encompass action commands articulated by the user, which the robot proficiently recognizes and executes. The integration of a Large Language Model enhances the system's capabilities, particularly in the domain of question-answering. To bolster the recognition and execution of action commands, we have employed a robot action ontology established in previous research endeavors. This ontology defines relevant classes and individuals, forming the basis for a nuanced understanding of user-inputted action commands. Further refinement involves the generation of succinct three-word strings for each action, ensuring semantic alignment with the user's verbal instructions. Importantly, our system operates in two distinctive modes: STATELESS and STATEFUL. In STATEFUL mode, the robot possesses awareness of its present posture, allowing it to execute action commands only when they align with its current state. This adaptive feature enhances the overall effectiveness of the system, catering to the dynamic nature of human-robot interactions and promoting a seamless and contextually aware dialogue between the NAO humanoid robot and its users.

Keywords

Action Robot Ontology, Human-Robot Interaction, Natural Language Processing, Large Language Models

1. Introduction

Within the expansive domain of robotics and Artificial Intelligence (AI), recent progress has sparked the emergence of a myriad of applications centered around robotic systems. A prevailing sentiment is taking root, positing a 50% likelihood of AI outpacing human capabilities across all tasks within the next 45 years. This projection, highlighted by authors in [1], envisions a future where automation gradually extends its reach, ultimately encompassing the entirety of human occupations within a span of 120 years.

Notably, social robots have swiftly ascended to prominence and are now being deployed in diverse countries, each deployment tailored to serve a wide array of purposes. These sophisticated machines are designed with a singular overarching objective: to elevate human interaction by fostering more effective and efficient engagements. This shift towards the integration of social robots into various societal frameworks underscores a transformative wave

Workshop on Actionable Knowledge Representation and Reasoning for Robots, May 27, 2024, Crete, Greece

*Corresponding author.

diego.reforgiato@unica.it (D. R. Recupero); l.boi23@studenti.unica.it (L. Boi)

0000-0001-8646-6183 (D. R. Recupero)

in how technology interfaces with human interaction, with the ultimate goal of creating more seamless and enriched connections between individuals and artificial entities.

The landscape of robotics applications has undergone a profound transformation with the emergence of Large Language Models (LLMs), presenting a paradigm shift in the augmentation of robotic systems' capabilities [2]. Notably, the infusion of LLMs, exemplified by cutting-edge models like OpenAI's GPT-4, into the realm of robotics research and development has ushered in a new era of possibilities. This integration has not only propelled advancements in human-robot interaction but has also significantly influenced cognitive processing and autonomous decision-making within robotic frameworks.

An exemplary application of LLMs in the field of robotics revolves around natural language understanding, a critical component that empowers robots to interpret and respond to human commands with unparalleled accuracy [3, 4]. This heightened linguistic proficiency introduces a transformative element, enabling robots to engage in more intuitive and user-friendly interactions. Users can now communicate with robots using everyday language, fostering a communication interface that is not only simplified but also significantly broadening the accessibility of robotic technologies. This inclusivity extends the reach of robotics to individuals with varying levels of technical expertise, democratizing the utilization of these advanced technologies and making them more approachable to a diverse user base.

This paper introduces an innovative methodology designed to harness the question-answering capabilities inherent in LLMs to foster dynamic and engaging conversations between users and the Zora humanoid robot. Zora is essentially a NAO robot of Softbank Robotics[1], expanded by Zora Robotics through the incorporation of a middleware software layer [5, 6, 7]. This layer empowers individuals without programming skills to customize various behaviors of the robot. Additionally, it provides the capability to comprehensively manage the robot, including controlling its movements, utilizing all its sensors, adjusting network settings, selecting from eight language options, and configuring various settings. Zora has been successfully used as case study in several research tasks [8, 9, 10, 11].

Our approach involves the intelligent parsing of the user's responses into sub-sentences, which are then articulated by the robot. A key feature of our approach lies in its ability to discern if a given sub-sentence implies an action aligning with the robot's ontology. In such instances, the robot seamlessly executes the action; otherwise, it gracefully defaults to the standard *Animated Say* animation.

Furthermore, our system extends the capability for users to issue action commands directly. In these scenarios, the system leverages the semantic similarity between sentence embeddings and three-word strings derived from the robot action ontology introduced in [12]. This nuanced process empowers the robot to decipher the user's intent and execute the appropriate action, fostering a fluid and natural interaction between the user and the robot. This dual functionality not only enhances the responsiveness of the robot but also amplifies the overall user experience, creating an intuitive and seamless dialogue between humans and the Zora humanoid robot.

The remainder of this paper is structured as follows. In Section 2, we delve into an exploration of pertinent works in the field, providing a comprehensive overview of the existing literature and contextualizing our research within the broader landscape. Section 3 offers an

[1]https://www.softbankrobotics.com/

in-depth discussion on the chosen ontology, elucidating the rationale behind its selection and its integral role in shaping the foundation of our proposed approach. Moving forward, Section 4 meticulously details the architectural framework underpinning our innovative methodology, providing a comprehensive elucidation of its components and functionalities. This section not only outlines the architectural intricacies but also elucidates how our proposed approach operates cohesively to achieve its objectives. Finally, in Section 5, we present a summary of our findings. This section also serves as a platform for delineating future directions, offering insights into the potential trajectories and advancements that our research may inspire.

2. Related Work

The development and implementation of robot ontologies have garnered substantial attention in recent years, with researchers exploring various aspects to enhance the capabilities and intelligence of robotic systems. In this section, we provide a comprehensive review of related works, categorizing them based on key themes such as ontology design, applications of robot ontologies, and their integration with advanced technologies. Researchers have made significant strides in the design and construction of ontologies tailored specifically for robotic systems. Olivares-Alarcos et al. [13] introduced an ontology aimed at collaborative robotics and adaptation, centered on two key concepts: collaboration and plan adaptation. This ontology is designed to establish a dependable framework for human-robot collaboration by enabling robots to formalize and reason about plan adaptations and collaborations within unstructured collaborative robotic scenarios. Additionally, the presented ontology contributes to the increased reusability of domain terminology, empowering robots to articulate their knowledge across diverse collaborative and adaptive situations. Mitrevsk et al. [14] presented and examined a strategy to generalize parameterized execution models for manipulation actions across various objects, relying on an object ontology. Specifically, the approach involves the robot transferring a known execution model to objects belonging to related classes as defined by the ontology, contingent upon the absence of evidence suggesting the model's unsuitability. This method harnesses ontological knowledge as an initial set of information, subsequently refined based on the robot's own experiences. The effectiveness of this algorithm was validated through experimentation involving two actions: grasping and stowing everyday objects.

The application domains of robot ontologies are vast and varied. Bernardo et al. [15] introduced a framework that seamlessly combines a domain ontology, specifically a home environment ontology, with a task planner. This integration facilitates the translation of objectives originating from a given agent, whether it be a robot or a human, into actionable tasks for a robotic agent. Building on this, the same authors proposed an innovative framework focused on enhancing the motion planning of a manipulator robot through semantic knowledge-based reasoning, as detailed in their subsequent work [16]. In this later framework, the Semantic Web Rule Language was employed for inferring new knowledge based on the existing understanding of the environment and the robotic system. Ontological knowledge, including semantic maps, was generated using a deep neural network trained to identify and categorize objects within the robotic agent's operational environment. This ontological knowledge then facilitated the deduction of manipulation constraints, and an environment corresponding to the agent's manipulation workspace was formulated. This environment description enabled the task planner to interpret and generate collision-free paths for the robot's effective motion planning.

Advancements in technologies such as natural language processing (NLP) and machine learning have been seamlessly integrated with robot ontologies. Zhang and collaborators [17] created a tutorial system designed for the resolution of kinematics problems, utilizing a combination of neural networks and ontology. Initially, they introduced an ontology tailored to articulate the knowledge associated with kinematics, thereby aiding the robot in comprehending kinematic problems. Subsequently, to establish alignment between natural language text and the ontology, the authors presented an innovative tagging scheme. This scheme, rooted in the kinematic problem understanding model within named entity recognition, enhances the system's ability to correlate and interpret natural language expressions with kinematics-related knowledge.

Other researchers [12] have introduced a robot action ontology tailored for an NAO humanoid robot. This ontology is integrated into a Natural Language Processing engine, which engages in machine reading of user-input text in natural language. The primary objective is to discern actionable commands for the robot's execution. For every conceivable action within the robot's repertoire, the authors meticulously modeled a corresponding element in the ontology. This modeling extends to encompass a detailed list of both compatible and non-compatible actions associated with each modeled action. Furthermore, the system adeptly manages compound expressions, such as "move your arms up" as well as multiple expressions embedded within a single sentence. The robot comprehends and executes these nuanced commands effectively.

Our paper capitalizes on the established ontology to discern the full spectrum of potential actions that the robot can undertake. The subsequent section will delve into a more comprehensive discussion of the intricacies and specifics of this ontology.

3. The Robot Action Ontology

The robot ontology[2], sourced from [12] and applicable to the Zora robot, is divided into two primary sections. The initial segment encompasses entities related to the robot and its actions, while the latter pertains to reasoning. Given Zora's humanoid design, the `Robot` entity serves as the starting point, with the `RobotBody` entity defined as its subclass, representing the robot as a whole. Subordinate entities, such as `RobotHead`, `RobotEye`, `RobotHand`, `RobotArm`, and `RobotLeg`, delineate specific parts of the robot. For each entity, corresponding individuals are defined; for example, `RobotHand` has resulting individuals like `ZoraLeftHand` and `ZoraRightHand`. The `RobotAction` entity encompasses classes with actions related to the robot, falling under the `Robot` entity. These actions include those using a single body part (e.g., opening the right hand), both sides of the body (e.g., moving both arms up), and actions involving the entire body (e.g., walking, sitting). The first category is represented by the `BaseAction` entity, which includes subclasses like `HeadAction`, `HandAction`, `ArmAction`, and `LegAction`. The `CompoundAction` entity includes movements involving both sides of the body, categorized as `CompoundArmsAction`, `CoumpoundEyesAction`, and `CompoundHandsAction`. The `SimpleAction` entity encompasses movements involving the robot as a whole: `Posture`, `Rotation`, and `Walk`. The `Posture` entity includes individuals related to positions the robot can assume, such as `Stand`, `Sit`, `SitRelax`,

[2]Publicly accessible at https://www.w3id.org/zoraActions

Crouch, LyingBack, and LyingBelly. The Rotation entity includes rotation types, such as a complete turn (pirouette), half turn, and a quarter turn to the right and left. The Walk entity defines walk styles: forward, backward, side to the left, and right.

The second part of the ontology is tailored for machine reading, reasoning, and identifying action commands in natural language expressions provided by humans to enable the robot to execute actions. This part contains keywords and synonyms for recognizing robot body parts and actions outlined in the ontology. The primary entity is Word, with subclasses BodyPartWord and ActionWord. In BodyPartWord, words for robot body parts are defined as individuals, each with a representative keyword and a list of synonyms. For instance, the individual arm has a representative keyword arm and synonyms like biceps, forearm, forelimb, and appendage. Similarly, ActionWord entities define words for specific actions, each with a representative keyword and a list of synonyms capturing various expressions for the given action. For example, the individual stand has a representative keyword stand and synonyms such as get up, stand up, rise up, etc. The individuals of BodyPartWord are linked to individuals of the class RobotBody and its subclasses through the uses and related inverse isUsedBy property. Similarly, ActionWord entities are linked to the class RobotAction and its subclasses using the same properties. The ontology comprises 33 individuals of ActionWord and BodyPartWord, formally encompassing all possible robot actions and their associated objects and adjectives, if needed.

4. Architecture

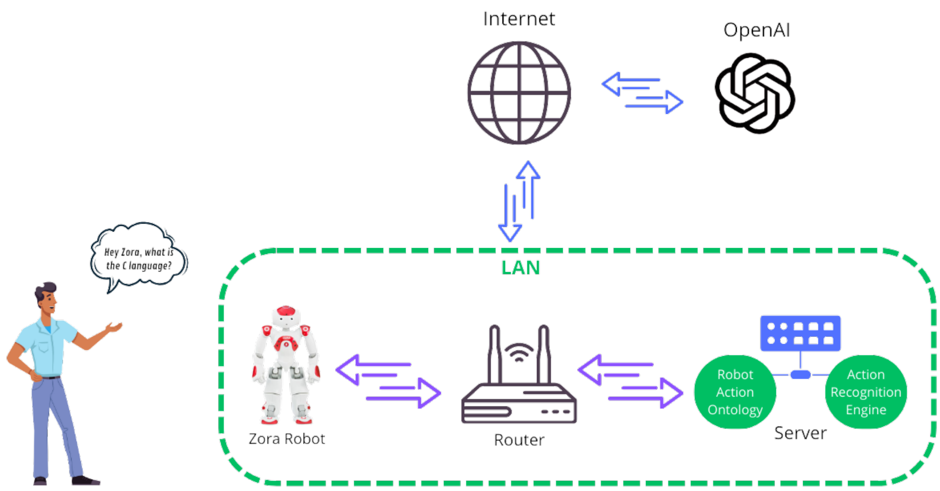

Figure 1: Architecture of the proposed approach.

The schematic representation of the methodology proposed in this paper is depicted in Figure 1. The NAO robot is positioned within the same local area network as a server hosting an Action Recognition Engine, which has been developed to interact with elements from the robot action ontology. Efficient communication between the Zora and the server is established through a router.

The execution flow involves the Zora robot running a pre-designed Choregraphe[3] program, orchestrating a series of actions. Initially, the robot initiates interaction by prompting the user to speak, awaiting the user's verbal response. Upon receiving the user's speech, the robot records the uttered content, transmitting the recorded audio to the OpenAI Speech-to-text capability[4] for efficient conversion to text. The resultant text is promptly returned to the robot by OpenAI Speech-to-text.

Following this, the robot forwards the obtained text to the Action Recognition Engine hosted on the server. The robot then awaits the engine's output, ready to interpret the results. If the output identifies a command action, the robot proceeds to execute it. However, if the text is not recognized as an action command, the robot forwards it to OpenAI ChatGPT to generate an appropriate response using its question-answering capabilities.

The robot retrieves the response from OpenAI ChatGPT and sends it back to the Action Recognition Engine for further processing. Subsequently, the robot receives a dictionary from the Action Recognition engine, each entry containing a pair: the sub-sentence extracted from the response generated by ChatGPT and the corresponding action to be performed. This communication and processing pipeline ensures effective interaction and response generation, integrating both speech recognition and natural language understanding capabilities.

The Action Recognition Engine, operational on the server, carries out a series of tasks to interpret and respond to the input received from the robot. Initially, it awaits the arrival of text from the robot. Once received, the engine engages in computing the semantic similarity between the received text and all potential actions outlined in the action command ontology we have used[5].

To achieve this, the **text representing e⊠ch ⊠ction** is transformed into embeddings through the utilization of the *bert-base-nli-mean-tokens* Sentence Transformer[6]. The subsequent semantic similarities between the text and all defined actions are then sorted in decreasing order. If the first element in the sorted list is higher than a predetermined empirical threshold of 0.8, then the engine effectively communicates the identified action, corresponding to the first element in the list, back to the robot for immediate execution.

However, in cases where no match is found with any action in the ontology, the engine conveys a specific code, signaling the absence of a match, and awaits additional text input from the robot. Following this, the Action Recognition Engine retrieves the text from the robot, which, in turn, is expected to return a response generated by ChatGPT based on the user's input. The engine then divides this response into sub-sentences, constructing a dictionary where each entry corresponds to a sub-sentence and is linked with the closest action based on semantic similarity.

For entries where the semantic similarity falls below the fixed threshold, the action associated with the underlying sub-sentence defaults to the *Animated Say* action of the NAO robot. This intricate process underscores the sophisticated semantic analysis, precise action determination, and adept response generation capabilities embedded within the proposed framework.

[3] https://www.robotlab.com/choregraphe-download-resources
[4] https://platform.openai.com/docs/guides/speech-to-text
[5] https://github.com/Fspiga13/Humanoid-Robot-Obeys-Human-Action-Commands-through-a-Robot-Action-Ontology
[6] https://huggingface.co/sentence-transformers/bert-base-nli-mean-tokens

To establish the **text representing each action**, as mentioned earlier, an analysis of the robot action ontology was conducted. The process began with the identification of `BodyPartWord` individuals and their corresponding `keywords` and `synonyms` from the ontology. For example, the `arm` entity falls under the category of `BodyPartWord` with associated `keywords` and `synonyms` such as *arm, appendage, bicep, forearm, forelimb.*

Subsequently, an exploration of classes that are subclasses of both `BaseAction` and `SimpleAction` was undertaken, encompassing classes like `ArmAction`, `HandAction`, `HeadAction`, `LegAction`, `Walk`, and `Posture`. The individuals within these classes, such as `LegDown`, `LegUp`, `HeadDown`, `HeadForward`, `HeadLeft`, `HeadRight`, `HeadUp`, `ArmDown`, `ArmForward`, `ArmSide`, `ArmUp`, `Crouch`, `LyingBack`, `LyingBelly`, `WalkBackward`, were extracted and collectively denoted as set S.

To formulate a comprehensive string representation for each action, the objects associated with the action were gathered. For each $s \in S$, elements $w \in W$ and $b \in B$ were extracted satisfying relationships such as {s, `involves`, t}, {t, `uses`, w}, {t, `bodySide`, b}, and {w, `is_a`, `BodyPartWord`}. Simultaneously, for a holistic understanding of the action, all elements $a \in A$ were collected for each $s \in S$ in relationships like {s, `uses`, a} and {a, `is_a`, `ActionWord`}.

The final step involved extracting the `keyword` and `synonym` values for each element from sets W, B, and A. This process enabled the formulation of all conceivable combinations for each potential action recognized by the ontology, incorporating the body parts from W, the body side from B, and the remaining action words from A. The result is a detailed and comprehensive set of representations for actions within the defined ontology.

Taking the `ArmDown` instance as an illustrative example, we can extrapolate the sets W, B, and S to further elucidate the multifaceted nature of this action. For the set W, encompassing the objects involved in the action, we have {arm, bicep, forearm, hand, claw, paw, etc.}. The set B, representing the potential sides of the body, is defined as $B = \{left, right\}$, signifying the left and right sides. Additionally, the set S, capturing the verbs or actions associated with the movement, is exemplified as $S = \{down, drop, lower, etc.\}$.

This analysis extends to the relationships within the ontology, involving triples such as `ArmDown involves ZoraLeftArm`, `ArmDown involves ZoraRightArm`, `ZoraLeftArm uses arm`, `ZoraLeftArm uses hand`, `ZoraRightArm uses arm`, `ZoraRightArm uses hand`, and `ArmDown uses down`.

This intricate network of relationships showcases the interplay between body parts, sides, and associated actions within the defined ontology. The objects involved in the action, the potential sides of the body engaged, and the specific movements or actions are systematically analyzed and cataloged, providing a detailed understanding of the semantic nuances associated with the `ArmDown` instance. This methodology ensures a comprehensive representation of the actions within the broader context of the action command ontology.

By combining values extracted from the sets W, B, and S, a diverse array of three-word strings is generated, each encapsulating a unique aspect of the `ArmDown` instance. These combinations serve as representations of the potential actions within the ontology, where W contributes the involved body parts, B offers the potential sides of the body, and S brings forth the specific actions or movements associated with the instance.

The crucial next step involves assessing the similarity of these generated strings with the user's input text. Through a comparison, the combination exhibiting the highest similarity to the user's input is singled out. This identified combination effectively serves as the robot's interpreted action, aligning with the user's input.

This process underscores the nuanced understanding of the user's command, with the robot discerning the most relevant action based on the amalgamation of body parts, body sides, and associated actions within the robot action ontology. The approach ensures a dynamic and context-aware response, enhancing the robot's ability to accurately execute commands in alignment with user expectations.

In the realm of action commands, the system incorporates a dual-mode functionality known as STATELESS and STATEFUL. The operational distinction between these modes significantly influences how the robot responds to and executes human expressions interpreted as action commands.

In the STATELESS mode, the robot promptly executes each human expression correctly identified as an action command and subsequently reverts to its default posture. This mode prioritizes the independent execution of individual commands without retaining awareness of the robot's ongoing posture.

Conversely, the STATEFUL mode introduces a heightened level of awareness regarding the robot's current posture. In this mode, the robot carefully evaluates the compatibility of each command with its existing state before execution. Unlike the STATELESS mode, the robot, in STATEFUL mode, refrains from reverting to its default posture after executing a command. This nuanced approach allows for a more context-aware and seamless interaction with the robot.

This capability enables users to provide a sequence of action commands to the robot. For instance, in STATEFUL mode, a user might instruct the robot to stand on its left leg, follow up with a question answered using ChatGPT, and then issue another command, such as walking. However, the execution of the last action is withheld due to its incompatibility with the robot's current state, notably the raised left leg. The determination of such incompatibilities is derived from the predefined robot action ontology, ensuring that the robot's actions align with its physical capabilities and constraints. This dual-mode operation enhances the adaptability and responsiveness of the robot to user instructions in various interaction scenarios.

5. Conclusions and Future Work

The methodology outlined in this paper exemplifies a harmonious integration between a robot and a robot action ontology, employing an LLM, with ChatGPT being the chosen model in our implementation, to enhance user interactions. Our approach involves extracting classes and individuals from the ontology, subsequently constructing three-word strings. These strings serve as a basis for comparison with the user's input, facilitating the identification of actions. The outcome is a fluid and natural interaction, empowering users to engage with the robot in a manner that closely mirrors human conversation.

During these interactions, the robot dynamically gestures in accordance with the identified action - whether derived from the user's input or generated text through question-answering by the LLM. This integration not only enhances the robot's responsiveness but also contributes to a more lifelike and engaging interaction experience for users.

For those interested in replicating or exploring our methodology further, the scripts developed for both the Action Recognition Engine and the Choregraphe script are openly available in a public repository[7]. This transparency encourages collaboration, experimentation, and improvement within the robotics community.

Furthermore, to provide a tangible example of the interaction experience, we have shared a public video[8] showcasing the seamless engagement between the robot and users. This demonstration not only illustrates the practical application of our approach but also serves as a visual representation of the potential and versatility inherent in integrating a robot with a language model for enriched user experiences.

In future work, our goal is to conduct a thorough and systematic evaluation of every subset of the system. Leveraging the ontology's list of actions, our task will involve assessing the accurate comprehension and mapping of natural language text, whether expressed by the user or within ChatGPT's response, by the action recognition engine to the respective robot action.

References

[1] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, O. Evans, Viewpoint: When will ai exceed human performance? evidence from ai experts, Journal of Artificial Intelligence Research 62 (2018) 729–754.

[2] C. Zhang, J. Chen, J. Li, Y. Peng, Z. Mao, Large language models for human–robot interaction: A review, Biomimetic Intelligence and Robotics 3 (2023) 100131. URL: https://www.sciencedirect.com/science/article/pii/S2667379723000451. doi:https://doi.org/10.1016/j.birob.2023.100131.

[3] N. Yoshikawa, M. Skreta, K. Darvish, S. Arellano-Rubach, Z. Ji, L. Bjørn Kristensen, A. Z. Li, Y. Zhao, H. Xu, A. Kuramshin, A. Aspuru-Guzik, F. Shkurti, A. Garg, Large language models for chemistry robotics, Autonomous Robots 47 (2023) 1057–1086. URL: https://doi.org/10.1007/s10514-023-10136-2. doi:10.1007/s10514-023-10136-2.

[4] A. Zeng, B. Ichter, F. Xia, T. Xiao, V. Sindhwani, Demonstrating large language models on robots, in: K. E. Bekris, K. Hauser, S. L. Herbert, J. Yu (Eds.), Robotics: Science and Systems XIX, Daegu, Republic of Korea, July 10-14, 2023, 2023. URL: https://doi.org/10.15607/RSS.2023.XIX.024. doi:10.15607/RSS.2023.XIX.024.

[5] D. R. Recupero, Technology enhanced learning using humanoid robots, Future Internet 13 (2021) 32. URL: https://doi.org/10.3390/fi13020032. doi:10.3390/FI13020032.

[6] R. Alonso, E. Concas, D. R. Recupero, A flexible and scalable social robot architecture employing voice assistant technologies, in: B. N. D. Carolis, C. Gena, A. Lieto, S. Rossi, A. Sciutti (Eds.), Proceedings of the Workshop on Adapted intEraction with SociAl Robots, cAESAR 2020, Cagliari, Italy, March 17, 2020, volume 2724 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 36–40. URL: https://ceur-ws.org/Vol-2724/paper10.pdf.

[7] D. R. Recupero, D. Dessì, E. Concas, A flexible and scalable architecture for human-robot interaction, in: I. Chatzigiannakis, B. E. R. de Ruyter, I. Mavrommati (Eds.), Ambient Intelligence - 15th European Conference, AmI 2019, Rome, Italy, November 13-15, 2019, Proceedings, volume 11912 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 311–317.

[7] https://github.com/loriboi/zoraProject

[8] https://www.youtube.com/watch?v=hEC9EHhjVe4&feature=youtu.be

URL: https://doi.org/10.1007/978-3-030-34255-5_21. doi:10.1007/978-3-030-34255-5_21.

[8] R. Alonso, A. Bonini, D. R. Recupero, L. D. Spano, Exploiting virtual reality and the robot operating system to remote-control a humanoid robot, Multim. Tools Appl. 81 (2022) 15565–15592. URL: https://doi.org/10.1007/s11042-022-12021-z. doi:10.1007/S11042-022-12021-Z.

[9] M. Atzeni, D. R. Recupero, Multi-domain sentiment analysis with mimicked and polarized word embeddings for human-robot interaction, Future Gener. Comput. Syst. 110 (2020) 984–999. URL: https://doi.org/10.1016/j.future.2019.10.012. doi:10.1016/J.FUTURE.2019.10.012.

[10] N. Cauli, D. R. Recupero, Video action recognition and prediction architecture for a robotic coach (short paper), in: S. Consoli, D. R. Recupero, D. Riboni (Eds.), Proceedings of the First Workshop on Smart Personal Health Interfaces co-located with 25th International Conference on Intelligent User Interfaces, SmartPhil@IUI 2020, Cagliari, Italy, March 17, 2020, volume 2596 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 69–77. URL: https://ceur-ws.org/Vol-2596/paper6.pdf.

[11] F. Gerina, S. M. Massa, F. Moi, D. R. Recupero, D. Riboni, Recognition of cooking activities through air quality sensor data for supporting food journaling, Hum. centric Comput. Inf. Sci. 10 (2020) 27. URL: https://doi.org/10.1186/s13673-020-00235-9. doi:10.1186/S13673-020-00235-9.

[12] D. Reforgiato Recupero, F. Spiga, Knowledge acquisition from parsing natural language expressions for humanoid robot action commands, Information Processing & Management 57 (2020) 102094. URL: https://www.sciencedirect.com/science/article/pii/S0306457319303322. doi:https://doi.org/10.1016/j.ipm.2019.102094.

[13] A. Olivares-Alarcos, S. Foix, S. Borgo, Guillem Alenyà„ Ocra – an ontology for collaborative robotics and adaptation, Computers in Industry 138 (2022) 103627. URL: https://www.sciencedirect.com/science/article/pii/S0166361522000227. doi:https://doi.org/10.1016/j.compind.2022.103627.

[14] A. Mitrevsk, P. G. Plöger, G. Lakemeyer, Ontology-assisted generalisation of robot action execution knowledge, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE Press, 2021, p. 6763–6770. URL: https://doi.org/10.1109/IROS51168.2021.9636791. doi:10.1109/IROS51168.2021.9636791.

[15] R. Bernardo, J. Sousa, P. Gonçalves, Planning robotic agent actions using semantic knowledge for a home environment, Intelligence & Robotics (2021). doi:10.20517/ir.2021.10.

[16] R. Bernardo, J. M. Sousa, P. J. Gonçalves, A novel framework to improve motion planning of robotic systems through semantic knowledge-based reasoning, Computers & Industrial Engineering 182 (2023) 109345. URL: https://www.sciencedirect.com/science/article/pii/S0360835223003698. doi:https://doi.org/10.1016/j.cie.2023.109345.

[17] J. Zhang, J. Yuan, J. Xu, S. Ban, X. Zan, J. Zhang, Kinrob: An ontology based robot for solving kinematic problems, Int. J. Know.-Based Intell. Eng. Syst. 26 (2022) 299–312. URL: https://doi.org/10.3233/KES-218162. doi:10.3233/KES-218162.