Human Robot Interaction through an ontology-based dialogue engine

Human Robot Interaction through an ontology-based dialogue engine

Conference Paper · January 2024

DOI: 10.1145/3610978.3640642

Alberto Lillo
Università degli Studi di Torino

Marco Stranisci
Università degli Studi di Torino

Cristina Gena
Università degli Studi di Torino

Human Robot Interaction through an ontology-based dialogue engine

Alessandro Saracco
alessandro.saracc@edu.unito.it
University of Turin, Italy

Alberto Lillo
alberto.lillo@unito.it
University of Turin, Italy

Marco Stranisci
marcoantonio.stranisci@unito.it
University of Turin, Italy

Cristina Gena
cristina.gena@unito.it
University of Turin, Italy

ABSTRACT

This paper outlines the evolution of the Sugar, Salt, Pepper project for high level functioning children affected by autism, focusing on the development of a dialogue system that relies on an ontology-based knowledge base. The ontology offers a formal representation of knowledge and interrelationships within the movie domain. The dialogue system addresses issues related to predefined answers, emphasizing adaptability for multi-platform use, particularly in the context of the social robot Pepper. The research covers detailed phases of construction and development, highlighting implementation choices and challenges faced. This work tries to make an advancement in the development of sophisticated and intuitive human-robot interaction systems, capable of adapting to user needs and delivering increasingly accurate and consistent responses.

CCS CONCEPTS

• **Information systems → Ontologies**; • **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Ontology, Human-Robot Interaction, Dialog, Machine Learning

ACM Reference Format:
Alessandro Saracco, Alberto Lillo, Marco Stranisci, and Cristina Gena. 2024. Human Robot Interaction through an ontology-based dialogue engine. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640642

1 INTRODUCTION

Ontologies play a crucial role in human-robot interaction by providing a clear framework for defining context [15, 18], user preferences, robot capabilities, and communication modes. They may address challenges related to language and context diversity, fostering a shared framework for information exchange. The integration of ontologies in human-robot interaction has led to significant advancements in healthcare, domestic robotics, and manufacturing, enhancing the efficiency and intuitiveness of interactions [13].
Sugar, Salt & Pepper [7, 8] is a collaborative project focusing on a living laboratory integrating skills related to educational and social robotics [9] to address the specific needs of highly functioning (Asperger) children with autism. The laboratory, designed and led by a multidisciplinary team, involved children between the ages of 11 and 13, who participated in the lab sessions once a week for four months in which exchanges and their interactions with the robot Pepper were tested, with a focus on language, communication, emotions, and the enhancement of social skills. An additional goal was to provide young participants with a space to increase their communication and socialization skills with each other and strengthen the acquisition of strategies and autonomy related to daily activities, such as preparing a snack, with the help of Pepper, configured as a highly motivating and engaging tool. After conducting a set of lab sessions, children were asked to complete a questionnaire (Attribution of Mental State [14]) to assess the mental states attributed to the robot. The results indicated a generally low score, emphasizing the necessity for improved conversational strategies to navigate the complexities of social interaction.
The identified issues with Pepper's dialogue system included: Predefined Dialogue Strategies: The existing system predetermined possible questions and answers, limiting flexibility during interactions; Robustness and Accuracy: Pepper struggled to consistently handle and respond to questions outside the usual topics covered.

The study concluded that refining dialogue strategies and enriching Pepper's knowledge base was crucial. Specifically, integrating a knowledge base in the form of an ontology was proposed. This ontology, due to its structured information management properties, could enhance the robot's ability to navigate and reason about various topics, improving conversation management and understanding of user input.
This paper introduced the development of a knowledge base (KB) in the form of an ontology. This integration aimed to elevate the robot's intelligence and reasoning abilities, enabling adaptation to user needs and personalized interactions. Unlike dialogue systems that are purely based on a stochastic approach (e.g., ChatGPT), systems based on ontologies [16] are better suited for educational and therapeutic applications, since they provide controlled and transparent outputs. The project's goal was to define an ontology in the movie domain, which was one of the conversation topics proposed by the robot during the social moments of the laboratory, emphasizing the importance of avoiding solely predefined responses for more effective and engaging conversations.

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640642

2 APPROACH

The goal of the project concerned the definition of a knowledge domain, structured in the form of an ontology, to enrich the knowledge base of the Pepper robot, in order to make it able to establish a conversation with the human interlocutor that would avoid the robot responding solely by means of predefined and static sentences, thus making the conversation more effective and engaging. To accomplish this goal, we identified the ontological domain related to the movie field.

The project consists of four main phases: i) Ontology modelling, ii) Knowledge retrieval, iii) Dialog engine, and iv) Rest API services. In the following we present in details the description of these steps.

2.1 Ontology modelling

This project planned to start from a pre-existing formal ontology, so we started from the ontology mentioned in Gena et al. [6] and reworked it according to what was needed for our project.

In the ontology definition phase, the primary goal was to create a structured representation of the movie domain so that the robot could understand and interact with relevant concepts and information. The ontology was conceived as a set of classes, properties, and relationships that reflect the conceptual structure of the domain.

Initially, an in-depth analysis of the domain was conducted to identify the main relevant categories and concepts in order to establish a hierarchical approach for structuring the classes. The main goal was to capture the complexity of the movie domain through a hierarchy of classes that reflects the nature of the relationships between key elements. Below is an example (Table 1)

Classes	≡ *equivalentClass*	⊑ *subClassOf*
Person	`dbo:Person`	`Thing`
Artist	`dbo:Artist`	`Person`
Director	`dbo:Actor`	`Artist`

Table 1: Ontology's classes hierarchy example

Next, relationships were established between the different classes to model the conceptual connections existing in the domain. Table 2 shows an example.

Obj Property	≡ *equivalentProperty*	**Domain**	**Range**
hasDirector	`dbo:director`	Movie	Director

Table 2: Ontology's object property example

To ensure a complete and accurate representation, attributes were also included in the classes to capture specific details. Table 3 shows an example.

Data Property	≡ *equivalentProperty*	**Domain**	**Range**
movie_poster	`wdt:P154`	Movie	xsd:string

Table 3: Ontology's data property example

2.1.1 *Ontology alignement.*

This process was designed to significantly enhance the completeness and coherence of the ontology, allowing for a more accurate and detailed representation of the movie domain. The ontology alignment operation proved to be of significant importance in the development of the ontology, as it allowed the data in the main structure to be enriched with information from established sources such as Wikidata [3] and DBpedia [1] resources widely used to enrich ontologies and knowledge bases with data from the semantic web.

The alignment operation was carried out by consistently associating the classes and properties in the ontology with the respective entities and properties in the external sources, using the `IRI` (Internationalised Resource Identifier) as linking keys. Through these operations we were able to work on a structured and solid ontology through which we could implement the code for populating it and its subsequent use.

2.2 Knowledge Retrieval

This phase describes the tasks conducted from data extraction to the subsequent population of the ontology, in order to provide the robot with a knowledge base. The entire component was programmed in Python language, as it is widely used and supported for tasks of this application nature.

A list of the relevant steps in the development process is described in the following:

Definition of SPARQL queries for the extraction of data from Wikidata and DBpedia all the SPARQL queries needed to extract the data of interest were defined.

Modelling pipeline development concerning

(1) **Execution of SPARQL queries** on the enpoint of Wikidata and DBpedia;
(2) **Data extraction and storage**;
(3) **Modelling**, which concerns the cleaning of the data (duplicated values, nulls, dirty characters etc.). The partial result generated an intermediate data structure, which consists of a matrix of this form (Table reftab:partial-result-matrix-example);

Movie	**Actor**	**Gender**	**Date of Birth**
Q104123	Q2680	M	1955-03-19
Q104123	Q80938	M	1954-02-18
...	...	...	...

Table 4: Partial extracted data matrix example

(4) **Label assignment**: starting with an element identifier, labels and/or sitelinks were retrieved using the Wikidata API. The new final matrix containing the extracted and manipulated data with the respective retrieved labels was then defined(Table 5).

Movie	Actor	Label	Gender	Date of Birth
Q104123	Q2680	Bruce Willis	M	1955-03-19
Q104123	Q80938	John Travolta	M	1954-02-18
...	...	...	...	...

Table 5: Final extracted data matrix example

Creation of relational database. We designed a relational database using MySQL, with tables dedicated to specific entities. Tables were designed with primary keys and relationships, providing a solid structure for the integration with the ontology.
Data entry. We automated the process of entering data into the database using predefined SQL queries, in Python. The code was structured to handle large amounts of data efficiently, inserting over 2,700,000 records, of which about 300,000 were related to movies.
Mapping. We used the `Ontop` plug-in for Protègè [4] to map ontology classes, properties and relationships to relational database tables. The mapping was essential to link the extracted data with the semantic structure of the ontology.
Integration on GraphDB. Leveraging on the above mapping, we integrated the ontology with the relational database on GraphDB, enabling fast and accurate access to information.

These steps made possible to transform data from external sources, often complex and heterogeneous, into a structured and coherent KB, ensuring consistency and integrity of the extracted data.

2.2.1 *Issues.*

During development, for the Knowledge Retrieval module, we faced several difficulties, including managing the variable quality of data from Wikidata and DBpedia, and dealing with challenges related to variable information density. These challenges were addressed by implementing scalable strategies, splitting complex operations into batches, and addressing specific problems such as extracting labels from Wikidata and handling limitations in DBpedia results. In addition, we optimised code performance through caching and intermediate data storage strategies.
On the other hand, in the development of the dialogue engine, several challenges emerged. In particular, Named Entity Recognition (NER), for identifying key elements in users' sentences, showed limitations with the Italian language having been trained with less data than in other languages (such as English). In addition, the complexity of entities (e.g., compound names of movies) made the task more complex.
Consequently, the adoption of strategies such as recognizing grammatical dependencies and sentence cleaning improved the process, enabling more accurate SPARQL queries. Another issue involved the development of a custom sentiment analysis model, which proved challenging. An approach using a `Multinomial Naive Bayes` classifier and `TF-IDF` was inconsistent, and the scarcity of labeled data in Italian posed challenges in creating a precise model. Better results were achieved with the use of the BERT-based model [10], despite the complexity of linguistic nuances and the lack of specific data for that language.

2.3 Dialog Engine

This development phase forms the heart of the conversation between the user and the robot. The main goal was to avoid the classical question-answer predefined scheme, trying instead to make the conversation more dynamic and robot-centered, to prevent the user from asking things the robot could not answer.
The entire module is written in Python and AIML (Artificial Intelligence Markup Language)[1]. After an initial testing phase, it emerged that the most promising approach was to split the conversation into different components, characterised by different specific structures and machine learning models.

2.3.1 *Dialogue phases.*

The conversation was then divided into the following components, following this order:

(1) The **AIML module** for the initial stage, wherein the user interacts with the robot through the AIML language, asking questions that cover the initial stages of a conversation, facilitating the initial interaction and helping to establish a stronger connection with the user. Furthermore, the robot captures relevant information (as for instance demographic data) during this interaction, which will be used in the later stages of the dialogue;
(2) The **Profiling Module** phase is dedicated to the user profiling on the movie domain. At first, a genre prediction model is used to predict the user's **most likely preferred genre** (based on the information gathered from the previous module), then information about both user and robot preferences (such as favorite movie, character, etc.) is exchanged, which is used to adapt the system's responses to the user's preferences, creating a more personalised dialogue experience;
(3) In the **Question Answering Module**, the robot asks the user questions, based on previously shared preferences (as for instance based on the user's favorite movie, the robot may ask what the director of that film is). The user and the robot take turns asking and answering questions. This approach adds entertainment to the dialogue and stimulates the active interest of the user;
(4) **Survey Form** concludes the dialogue phase. The robot asks questions about the interaction that took place (for instance, about the degree of satisfaction with how the robot understood the answers). The user responds verbally, by providing a numerical rating (on a Likert scale of 1 to 5) of the overall experience with the robot.

2.3.2 *AI models integration.*

The AI models were defined with the aim of creating a natural, engaging and personalised conversational environment, making the interaction more *intelligent* in understanding and responding to the user's needs. The models implemented are:
Genre prediction model, used in the profiling module, designed to predict the most probable user's preferred movie genre, based on gender and user's age (asked in the AIML module at 2.3). It uses a decision tree-based algorithm trained on the MovieLens 20M [2] dataset, underwent several preprocessing stages initially to filter the necessary data and transform it into a format suitable for the

[1] http://www.aiml.foundation/doc.html

project's needs. Model training resulted in an accuracy of 87%, with an overall Gini index of decision tree nodes equal to 0.19;
Quiz model, which uses a neural network with, (1) an input layer, a *bag of words* vector related to the user's query obtained through a natural language processing approach (NER Tagging), (2) two hidden layers with ReLU activation

$$\mathsf{ReLU}(x) = \max(0, x)$$

and (3) an output layer for the prediction of the relevant SPARQL query. The model is trained on a dataset containing intents, which constitute the different topics or themes related to the domain queries, and provides answers by extracting the data via SPARQL queries to the ontology;
Sentiment Analysis Model: In order to guarantee a more accurate result, we used a BERT-based model specific to the Italian language. Sentiment analysis gained crucial importance, as it became essential to understand whether each user sentence expressed a positive or negative intent.

Thanks to this implementation and the models used, it was possible to create a dynamic conversation while maintaining the robot-based orientation. In other words, the robot guides the conversation towards the domain of its knowledge, but without resorting to predefined and static responses, making the interaction more fluid and adaptable to the user's needs.

2.4 Rest API services

In order to make the dialogue engine scalable and shareable across different platforms, we defined Rest API services to handle user and robot input and output directly and synchronously. In detail, the user's input is processed by Pepper, then sent via a POST request to the dialogue engine which will send the output to the dedicated endpoint so that Pepper will read it and return it to the user. Through this implementation, we were able to structure a communication management system that is efficient and synchronised, but above all usable on any device via API requests.

2.5 Test and Results

The main idea underlying this preliminary testing was geared towards evaluating the effectiveness of the used strategy. In addition, the experimentation focused on different demographic parameters of the users and the differences in the evaluations between the interaction modes. The design of the data collection involved the integrated survey module at the end of each interaction (see Survey form in 2.3), with methodically and hierarchically structured questions. The participants, 52 neuro-typical users (75% male and 25% female), provided numerical ratings on a Likert scale of 1 to 5, expressing detailed opinions on various stages of the interaction, which, due to space constraints, are not reported in details.
The results of the evaluations were analysed in detail, revealing interesting trends. The questions about the robot's understanding of the users' preferences received positive average ratings (AV: 3.05, SD: 1.10), with moderate variability. Similarly, the questions related to the interactions showed positive average ratings (AV: 3.13, SD: 1.14), indicating satisfactory user experiences.
However, the question on the understanding of user questions showed a lower distribution, indicating difficulties in the process of translating the questions into SPARQL language (AV: 2.33, SD: 1.10). This suggests a specific area that may require further refinement in the interaction process (as also described in [5]).
The overall evaluation of the entire dialogue received positive average ratings (AV: 3.25, SD: 0.93), with a more concentrated distribution between moderate and high ratings. The lack of extremely low ratings suggests that, despite some critical issues, most users had an overall satisfactory experience.
Further analysis was conducted on subgroups of users according to age, gender, educational qualification, and type of interaction. Although some differences emerged, these were not found to be statistically significant, indicating an overall similar evaluation between the different groups.

3 CONCLUSION AND FUTURE WORK

Despite some identified challenges, the project demonstrated positive potential in HRI, and the collected preliminary evaluations provided a solid basis for future optimisations. We have identified a number of specific challenges that we will address in future work. Analysis of the questionnaire data revealed a wide range of user opinions at different stages of interaction with the robot. The evaluations varied significantly, with some categories eliciting strongly contrasting reactions. This diversity of opinions provides ground for future investigations to understand the reasons behind these evaluations and to improve users' experiences with the robot.
We believe that this work provides an initial solution for moving beyond dialog generation systems, marking a significant stride toward developing increasingly sophisticated and adaptable HRI systems. Indeed, structuring a dialogue based on ontology (where a set of concepts and their relationships are delineated), allows for a more predictable and controlled output. This approach ensures greater consistency in context, being particularly useful in contexts characterized by the presence of specific rules, and it should be preferred when machine responses need to be controlled and predictable. The ontology can be adapted for specific needs of a domain, allowing for greater customization and flexibility than a large generative models.
Looking to the future, there are multiple directions in which this project could evolve further. First, it is essential to continue working on adaptive mechanisms to enrich the robot with more knowledge and reasoning capabilities, enabling it to adapt autonomously to the needs of each user and his/her preferred choices [12]. Specifically, the following directions can be considered: *i)* Improving the knowledge base on which the robot relies. This includes a thorough mapping of our ontology with existing authoritative resources [17] and the integration of knowledge from addisional sources [11] ; *ii)* Improving AI models for better adapting the system to the user features; *iii)* Improving communication with the robot and exploring advanced ways of interacting with the robot by improving speech recognition steps (Speech-to-Text) or leveraging sensor for purposes such as facial expression analysis; *iv)* Further optimizations continue to optimize the system to reduce response times and increase consistency in responses.

REFERENCES

[1] [n. d.]. DBpedia. https://www.dbpedia.org/.
[2] [n. d.]. MovieLens 20M Dataset. https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset.

[3] [n. d.]. Wikidata. https://www.wikidata.org/.

[4] 2015. Ontop 1.15. https://protegewiki.stanford.edu/wiki/Ontop_1.15 Accessed on November,30th, 2023.

[5] Debayan Banerjee, Mohamad Yaser Jaradeh, Sushil Awale, Nandana Mihindukulasooriya, Sören Auer, and Ricardo Usbeck. 2023. The Scholarly QALD Challenge. In *International Semantic Web Conference*.

[6] Cristina Gena, Rossana Damiano, Claudio Mattutino, Alessandro Mazzei, Stefania Brighenti, Matteo Nazzario, Andrea Meirone, Camilla Quarato, Elisabetta Miraglio, Giulia Ricciardiello, Francesco Petriglia, Federica Liscio, Giuseppe Piccinni, Loredana Mazzotta, Cesare Pecone, and Valeria Ricci. 2022. Ontologies and Open Data for Enriching Personalized Social Moments in Human Robot Interaction. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 151–154. https://doi.org/10.1145/3511047.3537690

[7] C Gena, C Mattutino, S Brighenti, A Meirone, F Petriglia, L Mazzotta, F Liscio, M Nazzario, V Ricci, C Quarato, et al. 2021. Sugar, Salt & Pepper-Humanoid robotics for autism. In *CEUR WORKSHOP PROCEEDINGS*, Vol. 2903. CEUR-WS, 1–4.

[8] Cristina Gena, Claudio Mattutino, Andrea Maieli, Elisabetta Miraglio, Giulia Ricciardiello, Rossana Damiano, and Alessandro Mazzei. 2021. Autistic Children's Mental Model of an Humanoid Robot. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 128–129.

[9] Cristina Gena, Claudio Mattutino, Gianluca Perosino, Massimo Trainito, Chiara Vaudano, and Davide Cellie. 2020. Design and Development of a Social, Educational and Affective Robot. In *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems, EAIS 2020, Bari, Italy, May 27-29, 2020*. IEEE, 1–8. https://doi.org/10.1109/EAIS48028.2020.9122778

[10] Xiaodong Gu, Kang Min Yoo, and Jung-Woo Ha. 2021. Dialogbert: Discourse-aware response generation via learning to recover and rank utterances. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 12911–12919.

[11] Nicolas Heist and Heiko Paulheim. 2019. Uncovering the semantics of Wikipedia categories. In *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18*. Springer, 219–236.

[12] Anthony Jameson, Silvia Gabrielli, Per Ola Kristensson, Katharina Reinecke, Federica Cena, Cristina Gena, and Fabiana Vernero. 2011. How can we support users' preferential choice?. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Extended Abstracts Volume, Vancouver, BC, Canada, May 7-12, 2011*, Desney S. Tan, Saleema Amershi, Bo Begole, Wendy A. Kellogg, and Manas Tungare (Eds.). ACM, 409–418. https://doi.org/10.1145/1979742.1979620

[13] Sumaira Manzoor, Yuri Goncalves Rocha, Sung-Hyeon Joo, Sang-Hyeon Bae, Eun-Jin Kim, Kyeong-Jin Joo, and Tae-Yong Kuc. 2021. Ontology-based knowledge representation in robotic systems: A survey oriented toward applications. *Appl. Sci. (Basel)* 11, 10 (May 2021), 4324.

[14] Laura Miraglia, Giulia Peretti, Federico Manzi, Cinzia Di Dio, Davide Massaro, and Antonella Marchetti. 2023. Development and validation of the Attribution of Mental States Questionnaire (AMS-Q): A reference tool for assessing anthropomorphism. *Frontiers in Psychology* 14 (2023). https://doi.org/10.3389/fpsyg.2023.999921

[15] Edson Prestes, Joel Luis Carbonera, Sandro Rama Fiorini, Vitor A. M. Jorge, Mara Abel, Raj Madhavan, Angela Locoro, Paulo Goncalves, Marcos E. Barreto, Maki Habib, Abdelghani Chibani, Sébastien Gérard, Yacine Amirat, and Craig Schlenoff. 2013. Towards a core ontology for robotics and automation. *Robotics and Autonomous Systems* 61, 11 (2013), 1193–1204. https://doi.org/10.1016/j.robot.2013.04.005 Ubiquitous Robotics.

[16] Aarne Ranta. 2023. Multilingual Text Generation for Abstract Wikipedia in Grammatical Framework: Prospects and Challenges. *Logic and Algorithms in Computational Linguistics 2021 (LACompLing2021)* (2023), 125–149.

[17] Barbara Tillett. 2005. What is FRBR? A conceptual model for the bibliographic universe. *The Australian Library Journal* 54, 1 (2005), 24–30.

[18] Alessandro Umbrico, Andrea Orlandini, and Amedeo Cesta. 2020. An Ontology for Human-Robot Collaboration. *Procedia CIRP* 93 (2020), 1097–1102. https://doi.org/10.1016/j.procir.2020.04.045 53rd CIRP Conference on Manufacturing Systems 2020.