UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

itee PhD
information technology
electrical engineering

DIETI UNINA

Marco Grazioso

Building task-oriented dialogue systems for industry: a graph-based framework

Tutor: Prof. Francesco Cutugno

co-Tutor: Ph.D. Valentina Russo

Cycle: XXXVI

Year: Third

itee PhD
information technology
electrical engineering

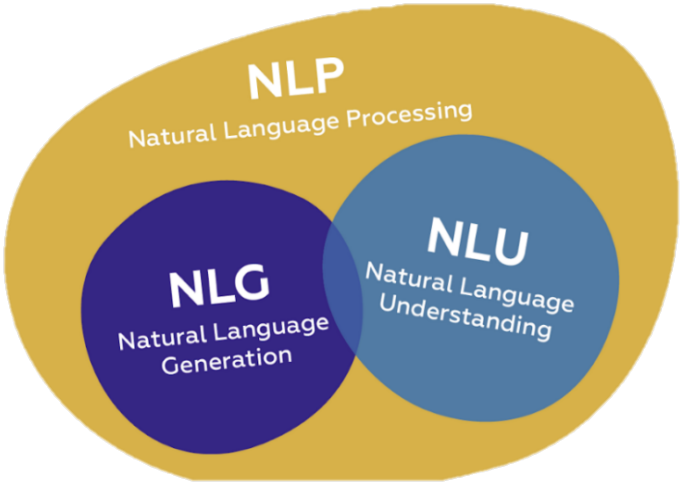
Background information

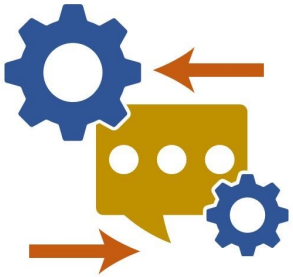
- **Master Degree**: Computer Science
- **Research group/laboratory**: Urban/Eco Research center
- **PhD start- end dates** : 01/11/2020 – 31/01/2024
- **Scholarship type**: company funded scholarship
- **Partner company**: Logogramma s.r.l
- **Period abroad**: Heriot-Watt University, Edinburgh,UK
  from 04/02/2023 to 31/07/2023

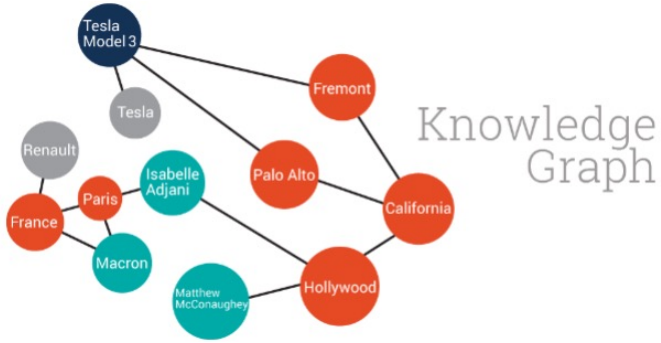
Summary of study activities

- **7 courses** (or PhD school).
  - Human Language Technologies
  - Game Engine Architectures and Interactive Experiences
  - Neural networks and deep learning
  - Lectures on Computational Linguistics 2022
  - Lectures on Computational Linguistics 2023
- **63 seminars.** most of them on AI, DNN, NLP, and Dialogue systems
- **Last conferences**
  - Italian Conference on Computational Linguistics (CLIC-it), Venice, November 30 - December 2, 2023 (poster presentation)

Dialogue Systems

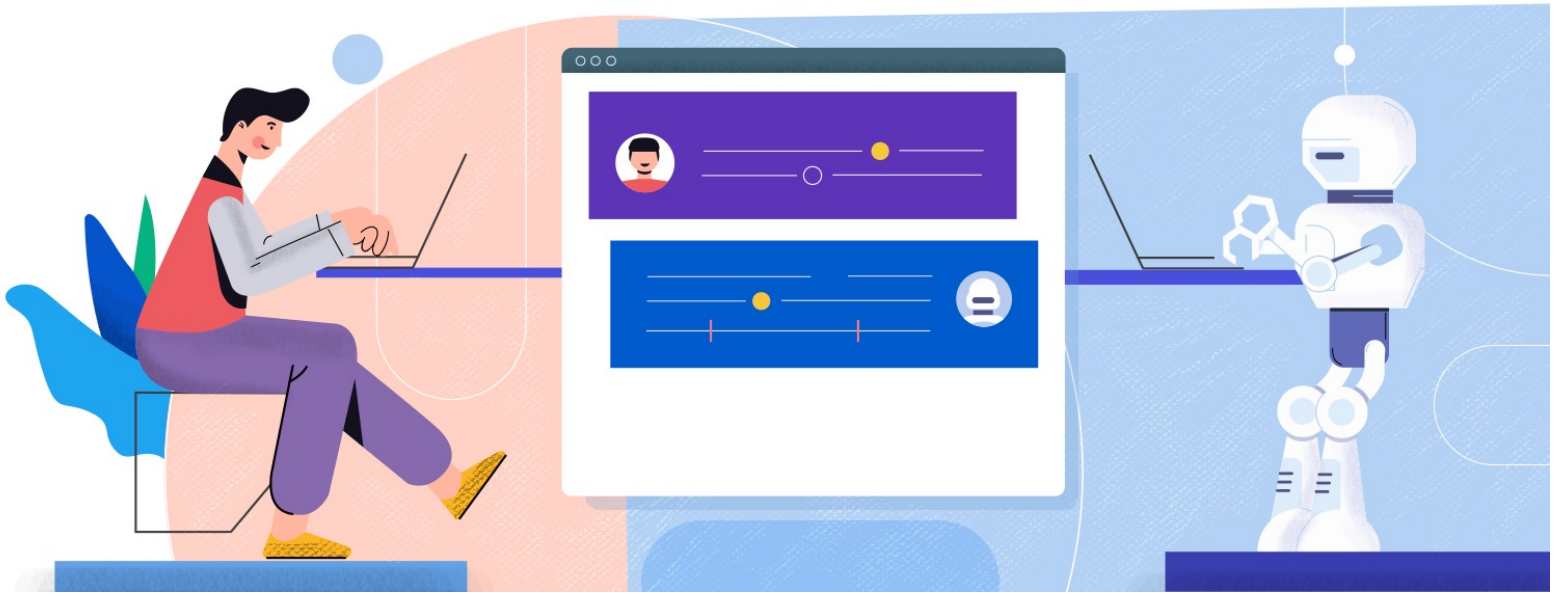






Dialogue manager

- **Dialogue state tracking**
- **Dialogue policy**



Research results

- Implementation of a domain-independent dialogue management graph structure handling complex dialogue flows
- *Implementation of a graph structure supporting the training of stat-of-the-art intent and entity recognition models*
- *Implementation of a graph structure for the dialogue state tracking task*
- *Company distributed the solution obtaining positive results*

Research products

[P1]	M. Grazioso, A. Suglia, An Analysis of Visually Grounded Instructions in Embodied AI Tasks, **9th Italian Conference on Computational Linguistics**, Venice, Italy, Nov.-Dec. 2023, CEUR-WS
[P2]	M. Grazioso, A. S. Podda, S. Barra, F. Cutugno Natural interaction with traffic control cameras through multimodal interfaces, **International Conference on Human-Computer Interaction** Washington DC, USA, Jul. 2021, pp. 501 - 515, Springer, DOI: 10.1007/978-3-030-77772-2_33.
[P3]	V. Russo, A. Mancini, M. Grazioso, M. Di Bratto, Graph-based representations of clarification strategies supporting automatic dialogue management, **Italian Journal of Computational Linguistics**, vol. 8 (1), pp. 1013-1062, 2022, DOI: 10.4000/ijcol.984
[P4]	A. Origlia, M. Grazioso, M. L. Chiacchio, F. Cutugno 3d avatars and semantic models annotations for introductory cultural heritage presentations, **2022 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage**, Rome, Italy, Jun. 2022, Vol-3243, CEUR-WS
[P5]	A. Origlia, M. L. Chiacchio, M. Grazioso, F. Cutugno, Increasing visitors attention with introductory portal technology to complex cultural sites, **International Journal of Human-Computer Studies**, vol. 180, 2023, DOI: 10.1016/j.ijhcs.2023.103135.

Research products

[P6]	F. A. D'Asaro, L. Raggioli, S. Malek, M. Grazioso, S. Rossi, An application of a runtime epistemic probabilistic event calculus to decision-making in e-health systems, **Theory and Practice of Logic Programming**, vol. 23(5), pp. 1013-1062, 2023, DOI:10.1017/S1471068422000382.
[P7]	M.Campi, V.Cera, F.Cutugno, A. Di Luggo, P. Giulierini, M. Grazioso, A. Origlia, D. Palomba, Virtual Canova: a Digital Exhibition Across MANN and Hermitage Museums, **Representation Challenges - New Frontieres of AR and AI Research for Cultural Heritage and Innovative Design**, vol. 1, pp. 253-260, 2022, DOI: 10.3280/oa-845-c219.

The company adopted the framework as the base for their solutions.

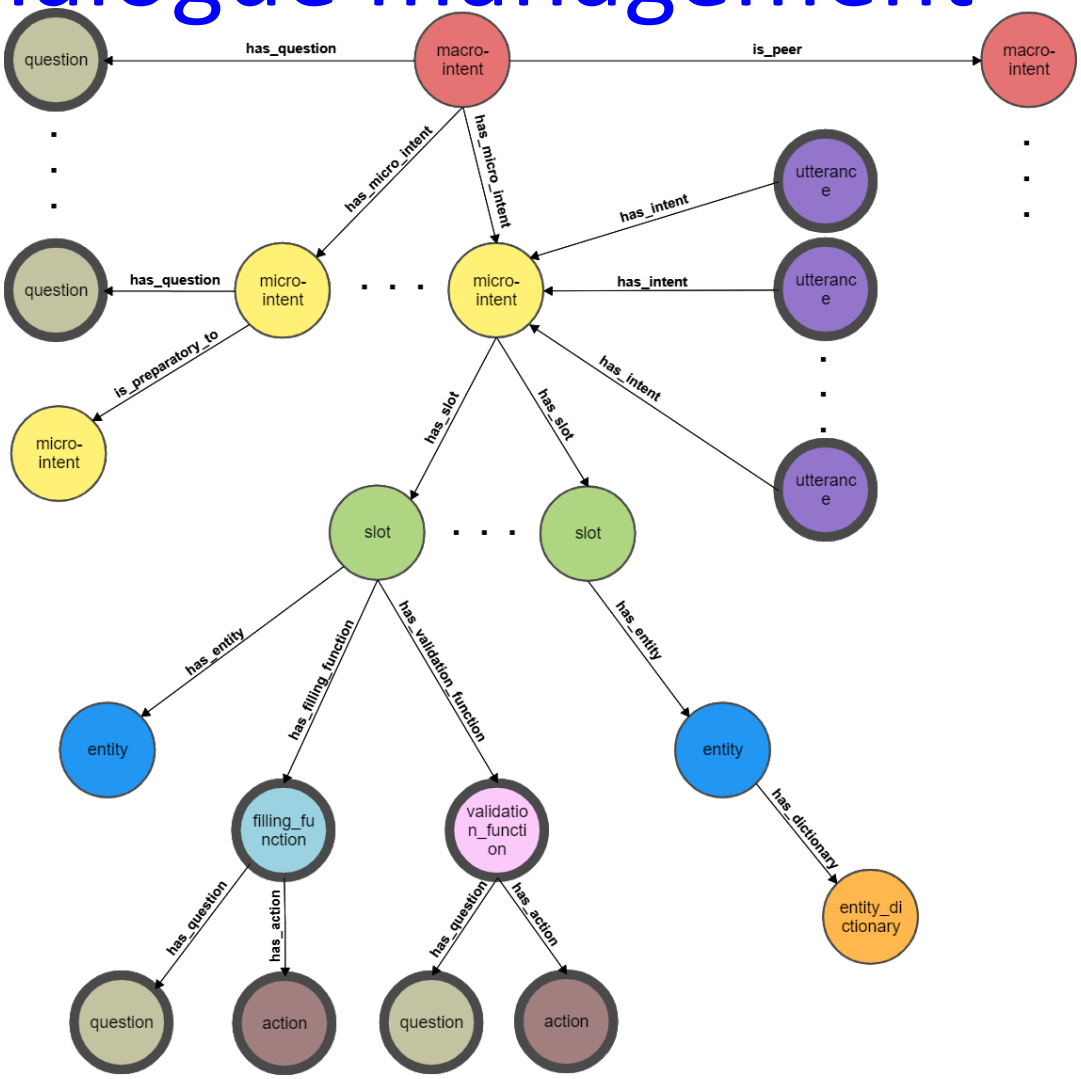
PhD thesis overview

Development of a joint structure supporting dialogue management, model training, and complex dialogue scenarios

- Objective
  - *domain-independent dialogue management*
  - *domain-specific language understanding*
  - *management of different pragmatic scenarios*
- Methodology
  - *graph-based representation*
  - *Pre-trained language models for domain adaptation*

Graph-based dialogue management

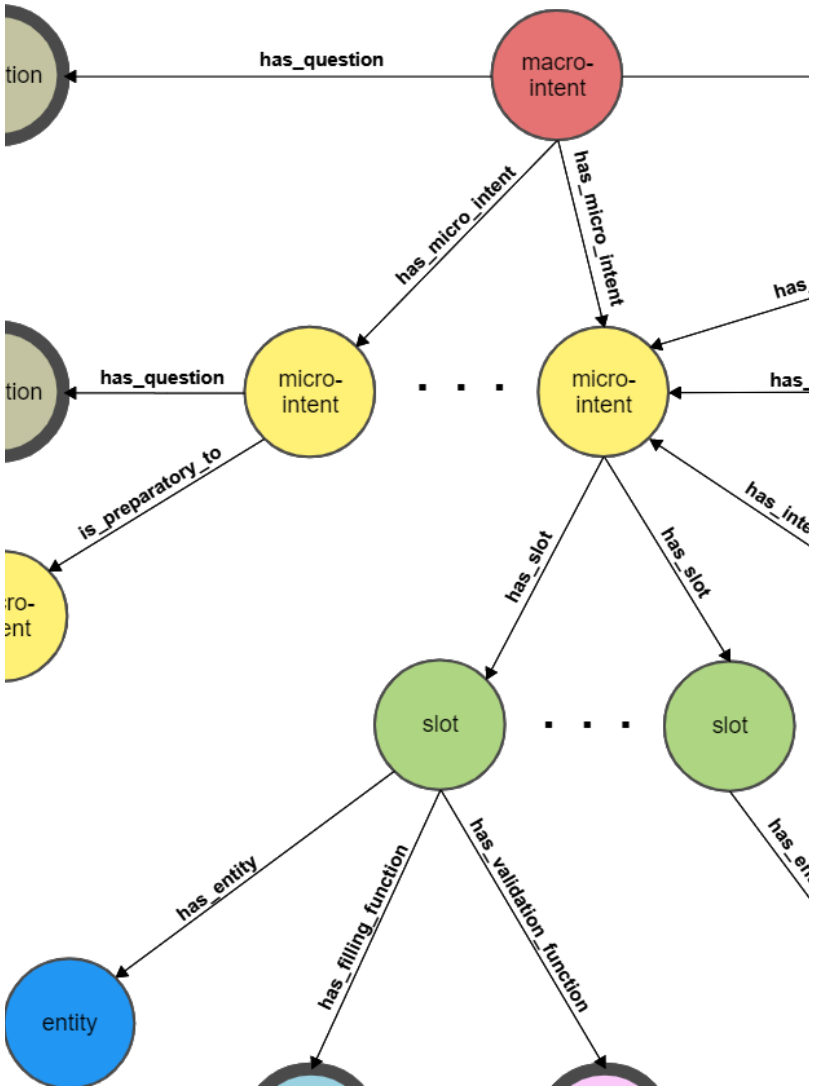
- *Overall graph representation[1]*



[1] Russo, V., Mancini, A., Grazioso, M., & Di Bratto, M. (2022). Graph-based representations of clarification strategies supporting automatic dialogue management. *IJCoL. Italian Journal of Computational Linguistics*, *8*(8-1).

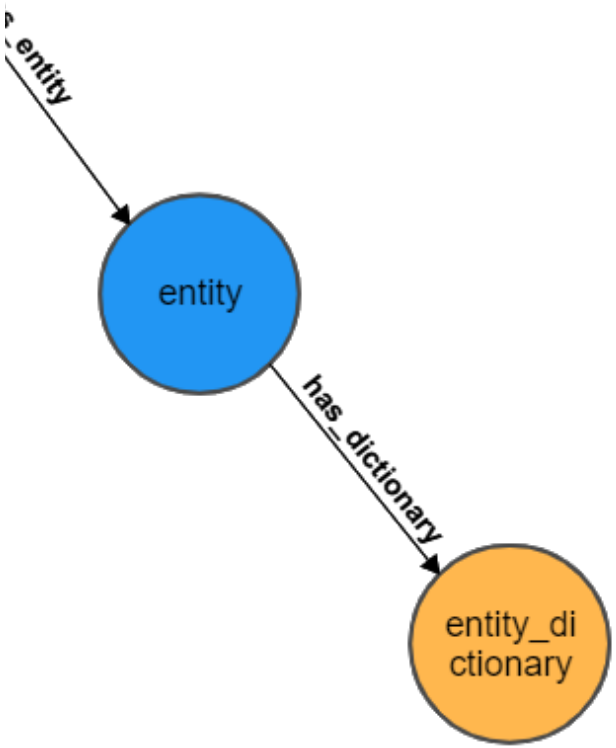
Domain modelling

- *Intents represent user intention*
- *They could be specialised into micro-intents*
- *Each intent has his own slots*
- *Slots are related to an entity type*



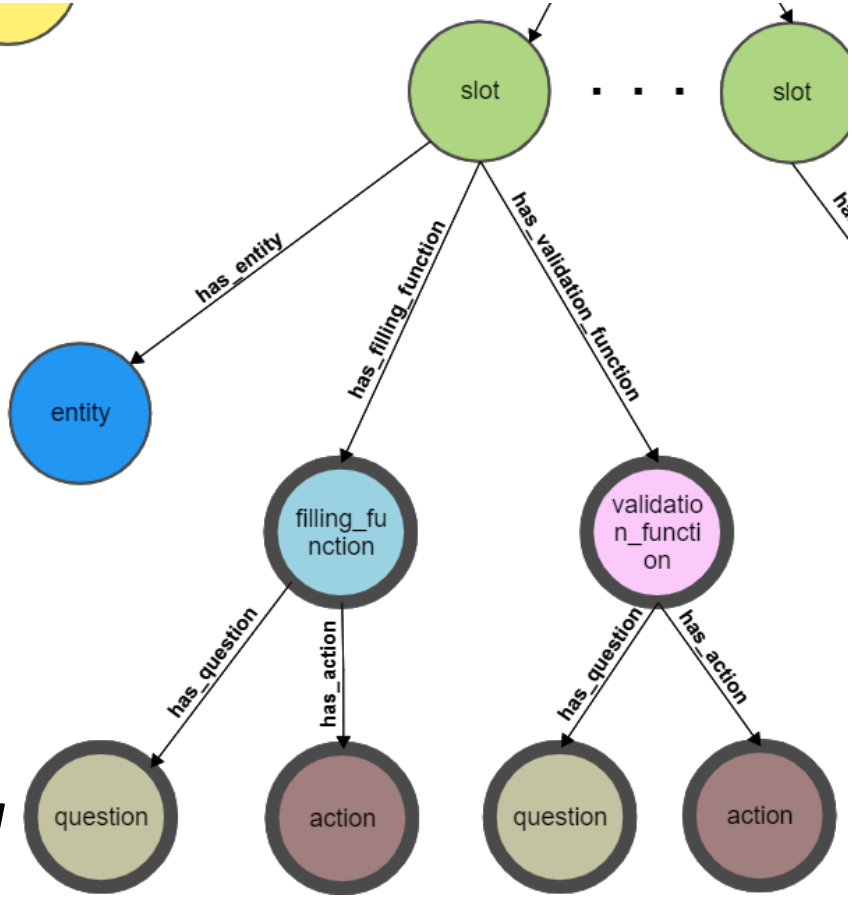
Domain modelling

- *Entities can be related to dictionaries*
- *Dictionaries can be populated from different sources*

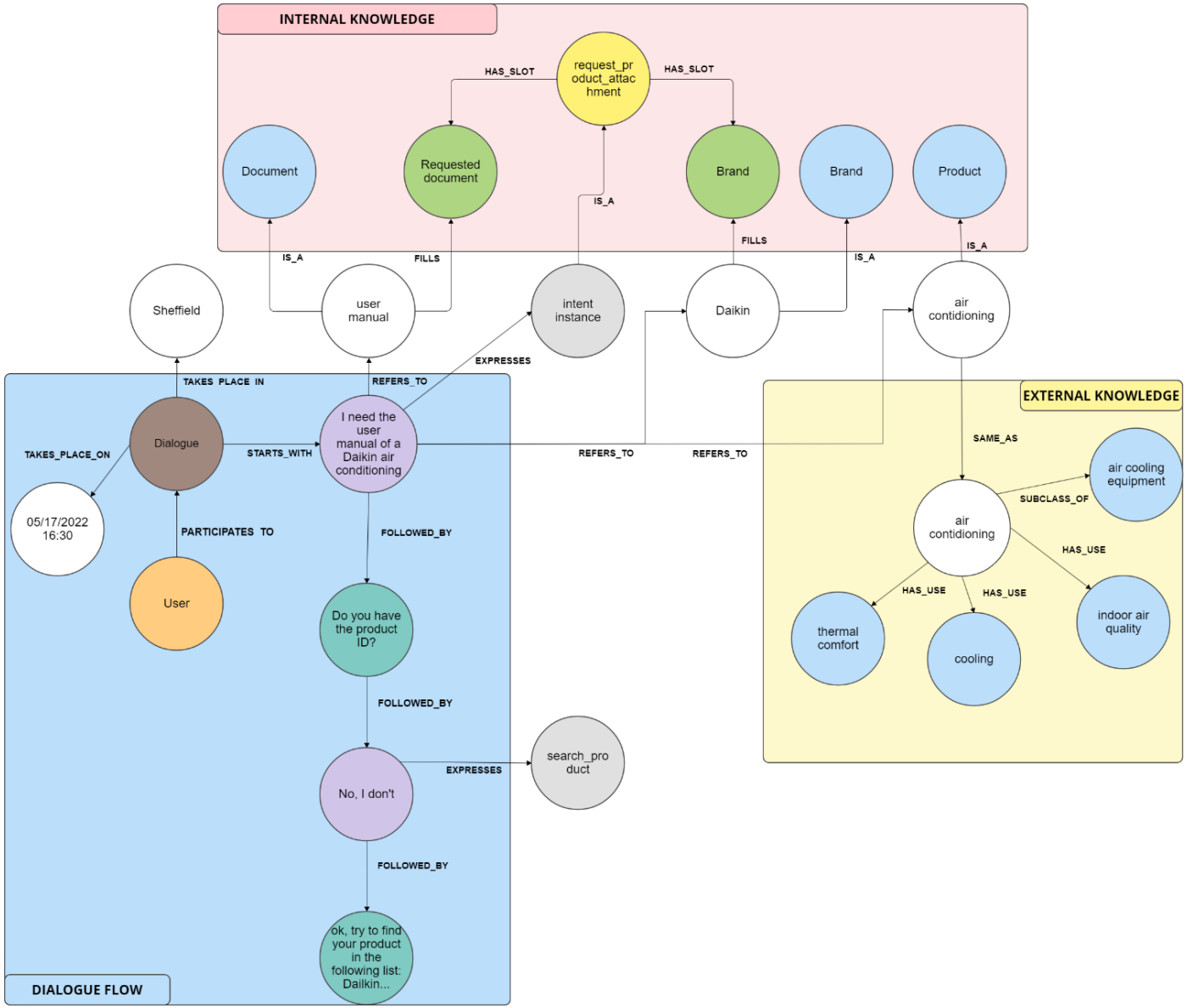


Dialogue modelling

- *Filling functions define how the slot must be filled*
- *Validation functions define how the value must be validated*
- *Actions define the operations to be performed*

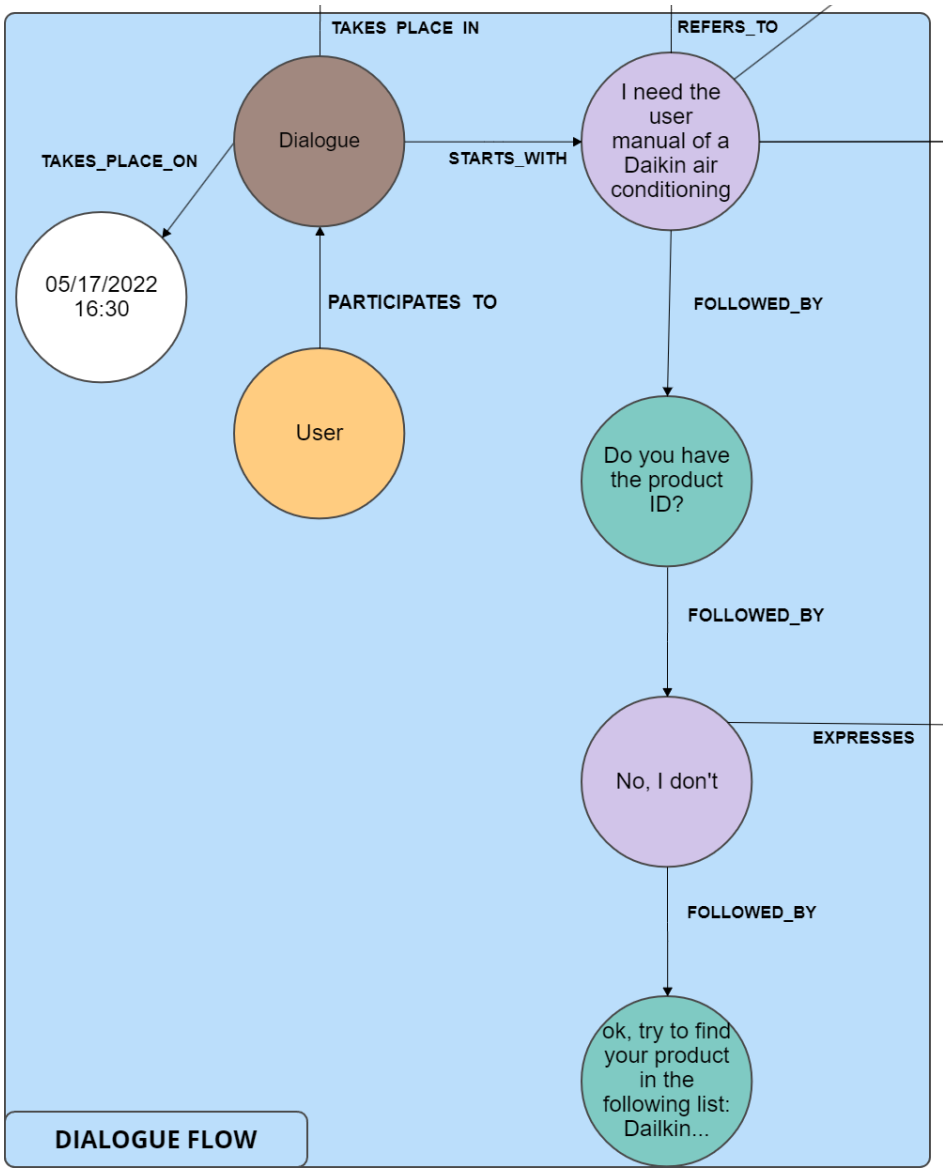


Dialogue state modelling



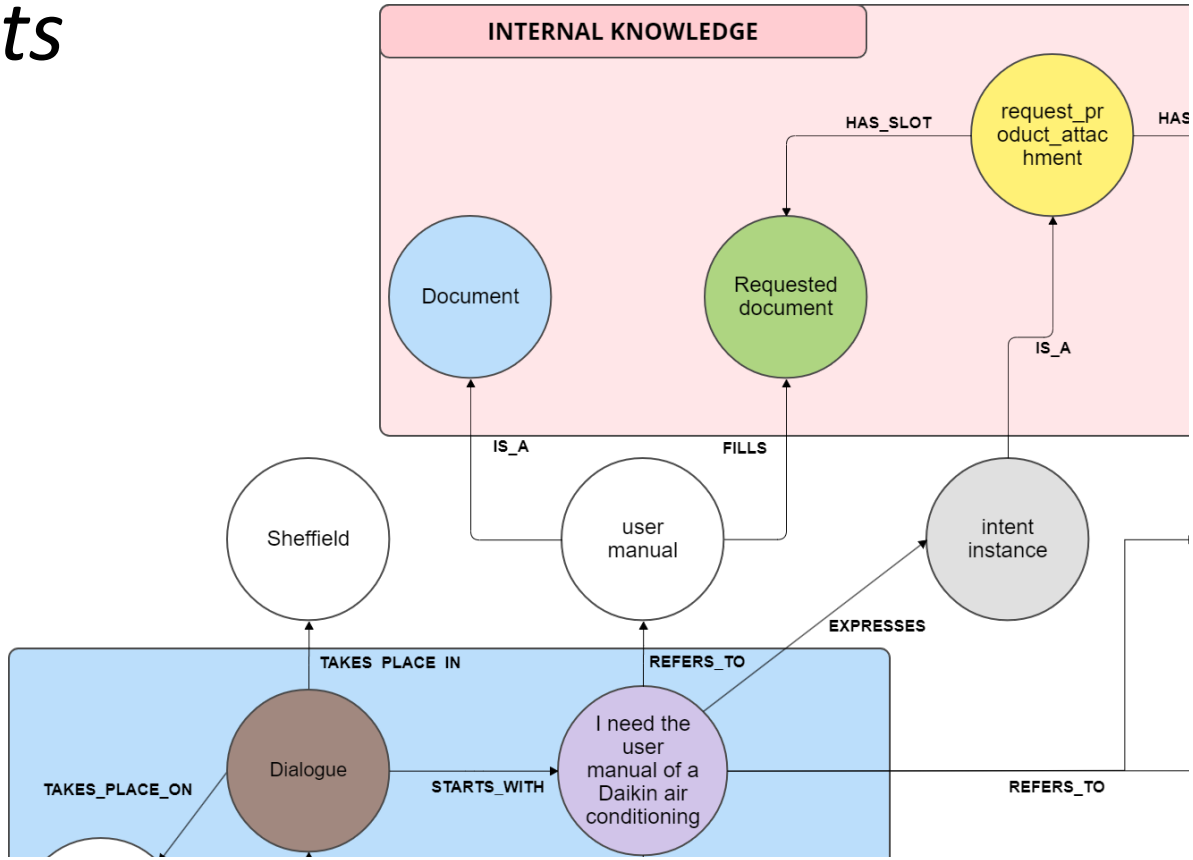
Dialogue flow

- *The dialogue is represented through a chain of nodes*



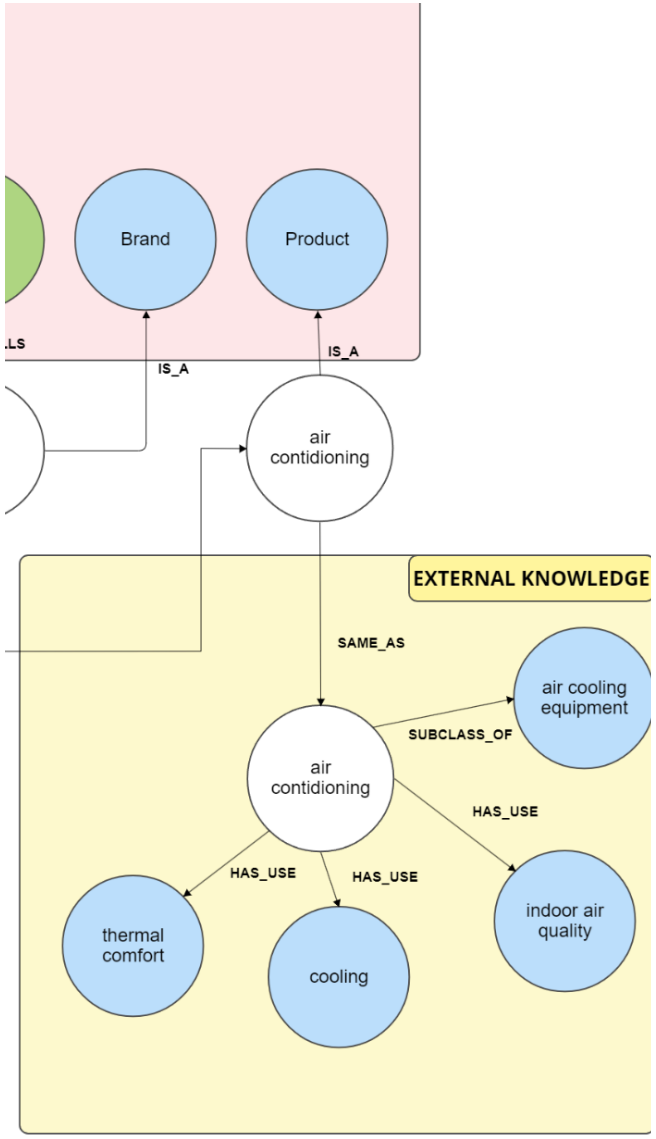
Domain connection

- *Recognised intents and entities are connected to the domain nodes*



External knowledge

- *Recognised entities can be in a relationship with data coming from other sources (WikiData and similar) (Entity Linking)*



Dialogue state enabled task

- *Previously mentioned items through the FOLLOWED_BY relationship*
- *Knowledge about the user through User node*
- *Knowledge about the world through the external knowledge connection*
- *Conflict search by navigating the utterances sequence*
- *Pattern search. Recurrent conversation patterns can be found and then used to instantiate specific dialogue policies.*

On field results

- *Human-human conversation in retail field*

Total conversations t ime (minutes)	Total number of conversations	Average turns per conversation
204:34:00	72	29.40

- *Human-machine conversation in retail field*

Total number of conversations	Average turns per conversation
656	6.12

Next step: situational dialogue

Preliminary results

- *Starting from a state-of-the-art model*
- *We trained a NLU model*
- *We conducted an error analysis to investigate situational issues*

Model	Language processing error rate
FILM	0.196
Ours	0.117

Error rate on all language processing tasks on unseen validation set.

Preliminary results

Challenges

- **Commonsense knowledge**
- **Visual context**
- **Multi-modality**
- **Interactive skills**

Error type	Subtype	Rate
Referential ambiguity	Mismatching	**40/821**
	Underspecification	**24/821**
	Others	**32/821**
Target object search	Object not visible	**171/821**
	Spatial understanding	**106/821**
Others interaction errors		**218/821**


Thanks
For
Your
Attention