
srWOZ: Towards Task-Oriented Dialogue Modelling in Serbian

Andelka Zečević¹[0000-0002-5656-007X],
Jovana Rađenović^{2,5}[0000-0002-2707-3870], Milica Gašić^{1,3}[0000-0003-0318-9147],
Songbo Hu⁴[0000-0003-1695-459X], Ranka Stanković²[0000-0001-5123-6273],
Mihailo Škorić²[0000-0003-4811-8692]

¹ Serbian Academy of Sciences and Arts, Mathematical Institute, Serbia,
andjelkaz@mi.sanu.ac.rs

² University of Belgrade, Faculty of Mining and Geology, Serbia,
{jovana.radjenovic|ranka.stankovic|mihailo.skoric}@rgf.bg.ac.rs

³ Heinrich Heine University Düsseldorf, Germany, gasic@hhu.de

⁴ University of Cambridge, Language Technology Lab, United Kingdom,
sh2091@cam.ac.uk

⁵ University of Belgrade, Faculty of Mathematics, Serbia

Scientific paper

DOI: 10.18485/judig.2026.2.ch3

Abstract

The rapid adoption of large language models and agentic frameworks has increased the demand for high-quality conversational datasets for the development and evaluation of dialogue systems. Among these, task-oriented dialogue systems have emerged as a particularly important class of conversational systems, assisting users in accomplishing specific goals through natural language interaction. Despite their growing popularity and practical importance, research and development of task-oriented dialogue systems for Serbian remain limited due to the scarcity of adequate language resources and annotated dialogue datasets.

In this paper, we present *srWOZ*, a task-oriented multi-turn conversational dataset for Serbian. Its creation is inspired by the existing *Multi³WOZ* dataset, a multilingual multi-domain multi-parallel dataset of culturally adapted task-oriented dialogues. In order to create *srWOZ*, we followed the original *Multi³WOZ* language adaptation methodology. First, we adapted the underlying ontology to the Serbian cultural and institutional context, covering seven domains: attractions, hospitals, police services, hotels, restaurants, taxis, and public transportation, while also performing slot-value normalization. Second, we adapted outline-based

dialogue generation component by adapting dialogue schemas and dialogue writing modules. Finally, we collected dialogues from native Serbian speakers with a bachelor degree in Serbian language using a bottom-up outline-based method. This process resulted in an initial dataset comprising 812 dialogues.

By addressing the lack of conversational resources for Serbian, *sr-WOZ* contributes to the broader effort of extending dialogue technologies to low-resource linguistic settings by highlighting the process and the difficulty and best practices of such an endeavour.

Keywords: dialogue system, task-oriented conversations, cultural adaptation

1. Introduction

Conversational assistants have become a common feature of many software applications and are increasingly used in everyday life. Voice assistants, customer-support bots as well as assistants powered by large language models (LLMs) are used for a wide range of tasks, such as booking appointments, checking schedules, and managing orders. For applications that require more than a fluent response, natural language generation alone is not sufficient. Task-oriented dialogue (ToD) systems must maintain information across multiple turns, identify and keep track of the user’s goals, access relevant information from external sources, and guide the interaction towards a specific outcome (Jurafsky and Martin, 2026).

LLMs have demonstrated strong capabilities in natural language understanding and generation, but their responses are not always factually reliable. They may produce fluent statements that are unsupported or incorrect, commonly referred to as hallucinations (Sahoo et al., 2024). Task-oriented dialogue systems provide a setting in which this issue can be addressed, at least in part, by maintaining an explicit dialogue state through the conversations. The concepts tracked by dialogue states are organised in ontologies and databases (Heck et al., 2020), allowing dialogue systems to access this structured domain knowledge during dialogue processing and ensuring that the information used by the system is explicitly represented and controlled (Young et al., 2013; Vukovic et al., 2026).

Another important limitation concerns the resources used to train LLMs. These resources are often neither publicly available nor

fully documented, making it difficult to determine which resources and information have contributed to the training of a given model (Bommasani et al., 2025). This lack of transparency is particularly relevant for languages with fewer digital resources, where it may be difficult to establish whether the training data adequately represents natural language use or has been created or reviewed by native speakers (Yao et al., 2024). Consequently, the linguistic quality and cultural appropriateness of the data underlying a model cannot always be reliably assessed. Independently developed and well-documented resources are therefore important for building and evaluating language technologies.

The availability of ToD modelling resources remains highly uneven across languages. Research on task-oriented dialogue has been strongly shaped by English-language datasets, most notably *MultiWOZ* (Budzianowski et al., 2018) and its subsequent extensions (Zang et al., 2020). For other languages, multilingual resources have often been created by translating existing English datasets. Although this approach provides a relatively efficient way of extending dataset coverage, translated dialogues may retain linguistic patterns and conversational conventions of the source language. They may also rely on entities, places, and services that are unfamiliar or inappropriate in the target cultural context, potentially limiting their usefulness for representing language and culture specific dialogue behaviour (Razumovskaia et al., 2022).

Serbian illustrates this resource gap. In recent years, efforts have been made to develop resources for Serbian natural language processing, including annotated and parallel corpora, morphosyntactic tools, and pretrained language models (Škorić and Janković, 2024; Marovac et al., 2023; Škorić and Petalinkar, 2025). Nevertheless, to the best of our knowledge, there is no publicly available dataset specifically designed for task-oriented dialogue in Serbian. Consequently, Serbian ToD systems currently require either adapting resources developed for other languages or creating new data. However, simply translating dialogue data does not guarantee that the resulting resource adequately represents the target language and context. Native speaker involvement is therefore important for controlling the quality of the data, ensuring that dialogues are natural and linguistically appropriate for the target language. For task-oriented dialogue in particular, native speakers can also ensure that entities, places, ser-

vices, and conversational situations are appropriately adapted to the target cultural context.

In this work, we present the initial development of *srWOZ*, a multi-turn, task-oriented dialogue dataset for Serbian. Instead of translating an existing dialogue corpus, *srWOZ* follows the culturally adapted, outline-based methodology and annotation framework introduced in *Multi³WOZ* (Hu et al., 2023). The dataset was developed by localising the ontology and associated database to Belgrade while covering seven domains: attractions, hospitals, police services, hotels, restaurants, taxis, and public transport. Native Serbian speakers then produced utterances based on dialogue outlines that represent user and system turns. Since these outlines are shared with the four existing *Multi³WOZ* languages, *srWOZ* is structurally parallel to these resources while retaining Serbian-specific linguistic and cultural characteristics. To the best of our knowledge, it is the first publicly available Serbian task-oriented dialogue resource developed directly in the target language with cultural adaptation. The main contributions of this work are:

- (i) a Belgrade-localised ontology and database covering seven task-oriented dialogue domains;
- (ii) the initial development of a Serbian multi-turn dialogue dataset, with dialogues written and annotated by native speakers with bachelor or master degree of Faculty of Philology, University of Belgrade.

The remainder of the paper is organised as follows. Section 2 reviews related work on task-oriented and multilingual dialogue datasets. Section 3 describes the construction of *srWOZ*: the cultural adaptation of the ontology and database, and outline-based dialogue generation together with the annotation process, data statistics and a linguistic analysis of the collected data. Section 4 concludes and outlines directions for future work.

2. Related Work

MultiWOZ (Budzianowski et al., 2018) has become a central benchmark for multi-domain task-oriented dialogue, providing around 10k Wizard-of-Oz dialogues across seven domains. Several subsequent versions addressed inconsistencies in dialogue-state and

utterance annotations, including *MultiWOZ 2.1* (Eric et al., 2020), *MultiWOZ 2.2* (Zang et al., 2020), *MultiWOZ 2.3* (Han et al., 2021) and *MultiWOZ 2.4* (Ye et al., 2022). The original MultiWOZ dataset was built using paid crowd workers. This enabled large-scale data collection, but it did have its toll on quality, which is why this highly influential dataset has had so many incarnations. This is the reason why, in this work, in the context of a low-resource language setup, we focus on native speakers with a degree in Serbian language.

Other influential resources extended the scope of task-oriented dialogue in different directions: Schema-Guided Dialogue introduced a broader set of services and schema-based representations (Rastogi et al., 2020), while *CrossWOZ* provided a large-scale Chinese multi-domain dataset (Zhu et al., 2020). These resources support research on core ToD components such as dialogue state tracking and response generation, as well as end-to-end dialogue modelling and evaluation. These resources have also supported the development of dialogue-system toolkits such as *ConvLab-3* (Zhu et al., 2023), as well as research on LLMs as task-oriented dialogue agents (Hudeček and Dusek, 2023).

Developing task-oriented dialogue systems for languages other than English remains challenging given the limited availability of high-quality training and evaluation data. Some multilingual resources address individual Natural Language Understanding (NLU) components, such as intent detection and slot labelling, without providing complete dialogues; examples include *MASSIVE* (FitzGerald et al., 2023) and *Multi3NLU++* (Moghe et al., 2023). Other datasets extend full task-oriented dialogue resources to multiple languages, often through translation or post-editing of English data. *Multi²WOZ* (Hung et al., 2022) and *GlobalWOZ* (Ding et al., 2022) follow this general direction. These approaches differ in how translation, localization, and human post-editing are incorporated into the data creation process. *BiToD* (Lin et al., 2021) takes a different approach by developing a bilingual English–Chinese dataset with a bilingual knowledge base rather than deriving target-language dialogues through translation. While these resources substantially broaden multilingual coverage, translation-based approaches can introduce translation artefacts and provide limited cultural and linguistic adaptation.

An alternative is to separate the underlying dialogue structure from its language-specific realization. Majewska et al. (2023)

propose an outline-based approach in which language-agnostic dialogue schemata guide native speakers in producing natural target-language dialogues. *Multi³WOZ* (Hu et al., 2023) extends this approach at scale, providing multi-domain, multi-parallel dialogues in Arabic, English, French, and Turkish. Its data collection process combines shared dialogue structures with language-specific realization and localization to entities and conventions of the target cultures. *srWOZ* follows this outline-based methodology and applies it to Serbian.

3. srWOZ

The Serbian version of the *MultiWOZ* dataset was developed following a culturally adapted and language-specific data creation framework inspired by the methodology introduced in *Multi³WOZ*. Since direct translation of the original English dialogues can lead to grammatical inaccuracies, unnatural phrasing, lexical inconsistencies etc., it was not employed in the dataset creation process. Instead, the dataset was manually created by the authors of this paper and student volunteers, all native Serbian speakers, enabling the creation of more natural, contextually appropriate, and culturally relevant task-oriented conversations.

The development pipeline for *Multi³WOZ* consisted of three main steps: (1) normalization of annotations from *MultiWOZ v2.3* with canonical values, (2) cultural adaptation and (3) collecting linguistically adapted dialogues. Unlike the original *Multi³WOZ* construction process, no slot-value normalization was required for *srWOZ*, as the dialogue outlines and ontology obtained from *Multi³WOZ* already employ normalized canonical slot values.

The remainder of this section is divided into four subsections. The first discusses cultural adaptation, while the second focuses on adapting the dialogue dataset creation process. The third presents the dataset statistics, and the fourth provides comments on the qualitative evaluation.

3.1. Domain Data Collection and Adaptation

As with other localized versions of *MultiWOZ*, each language-specific version is associated with a specific city, such as Cambridge in the original *MultiWOZ*, Dubai for Arabic, Paris for French, and

Ankara for Turkish in *Multi³WOZ*. In our case, Belgrade was selected as the reference city for *srWOZ*.

The database was modified as part of the localization process. The resulting database preserves the same set of domains as in all MultiWOZ-based versions, including attractions, hospitals, police services, hotels, restaurants, taxis, and public transportation. In *Multi³WOZ*, slot-value localization and redistribution are explicitly allowed, which in some language versions leads to structural modifications of the knowledge base (e.g., removal of certain attributes such as postcodes in the Arabic version). In our case, we observed that the configuration adopted in the French version is also well-suited to the Serbian setting, and we therefore retained a similar design without altering the original slot structure. However, we did change certain slot values, e.g., Indian to European restaurants, as swapping values is also allowed as part of the localization process. The database was populated predominantly through manual processes.

When constructing databases for different languages in *Multi³WOZ*, controlled entity replacement using one-to-one entity mapping was proposed. Entities across databases for different languages can be mapped to each other only if all categorical slot values attributed to each entity are identical. Steps for constructing the Serbian database are shown in Figure 1.

Upon analyzing the required data, the first step to acquiring useful information was performed programmatically using Python and publicly available OpenStreetMap¹ (OSM) web services. This step involved obtaining the geographic extent of Belgrade, Serbia, through the Nominatim geocoding service. The returned bounding box coordinates were used to define the spatial area of interest for subsequent queries. After defining the study area, the Overpass API was used to retrieve OSM objects located within the boundaries of Belgrade, using an appropriate key-value pair (e.g. for hotels ‘tourism=hotel’, for restaurants ‘amenity=restaurant’). For each retrieved object, available attribute information was extracted, including the name, address, postal code, district, contact details, website, cuisine type, price range, and geographic coordinates (latitude and longitude). As some of the collected information was deemed insufficient, additional manual collection was performed.

¹ An open-source geographic database that provides detailed information about points of interest and urban infrastructure.

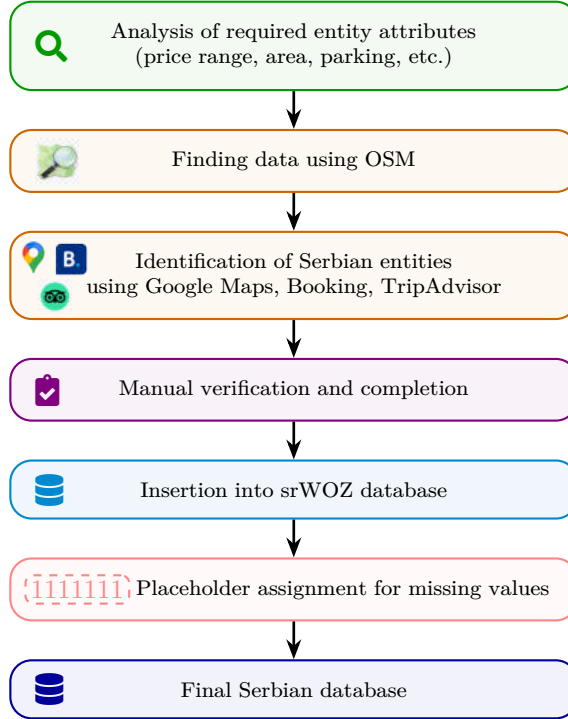


Figure 1: Workflow for constructing the Serbian database.

Additional information was collected and cross-checked using publicly available sources such as Google Maps,² Booking.com,³ TripAdvisor,⁴ and other relevant online resources. When specific information required by the database schema could not be reliably obtained (e.g., telephone numbers), placeholder values were introduced in order to preserve the completeness and structural consistency of the database.

Although the constructed database is grounded in real-world entities, it should not be interpreted as a fully accurate representation of the referenced establishments and services. The resource was designed primarily to support dialogue system research and benchmark development. Consequently, some entries may contain approximations, simplified representations, placeholder values, or information that may have changed over time.

² <https://maps.google.com>

³ <https://www.booking.com>

⁴ <https://www.tripadvisor.com>

After constructing the database, the ontology was also modified to reflect the changes. Overall, the adaptation process ensures that the dialogues remain semantically equivalent to the original interactions while appearing authentic within the target cultural context.

3.2. Outline-Based Dialogue Generation

The creation of the srWoZ dataset followed the outline-based annotation pipeline for multilingual ToD datasets introduced in [Majewska et al. \(2023\)](#). The central component of this pipeline is **dialogue schemata** organized around domains that provides intention descriptions of each dialogue turn and the slots of information it should contain. Following these instructions, annotators can express themselves in their native language while adopting the slot values to the target culture and geographical context. According to the analysis of participants surveys reported in the original study, task-oriented dialogues created using this approach, achieved higher naturalness and better reflect cultural specifics than translation-based datasets, without compromising annotation speed or data quality.

As stated, the *Multi³WOZ* dataset covers seven different domains: hotels, restaurants, attractions, hospitals, police services, taxi, and public transportation. These domains are described in terms of above mentioned ontologies and spread across *single-domain* and *multi-domain* schemata. For example, a dialogue focused on obtaining the address of a particular attraction is considered as a single-domain dialogue, whereas a dialogue that first addresses the location of an attraction and subsequently involves arranging transportation to that location represents a multi-domain dialogue. For the initial version of the srWoZ dataset we covered three domains - hotels, restaurants, and attractions - and restricted the annotation to single-domain dialogue schemata only. This design choice allowed closely monitoring of the annotation process and the familiarization with the annotation platform. The precise number of dialogue schemata per domain is provided in the table below (Table 1).

The annotation process was conducted using a platform originally developed by the University of Cambridge ([Hu et al., 2023](#)). For the purpose of the srWoZ creation, we adopted the platform’s terms of use regarding annotator compensation, as all annotators participated on a voluntary basis. We retained the licensing terms of the

domain	#dialogues
hotels	595
restaurants	1,127
attractions	130

Table 1: Number of dialogues per domain

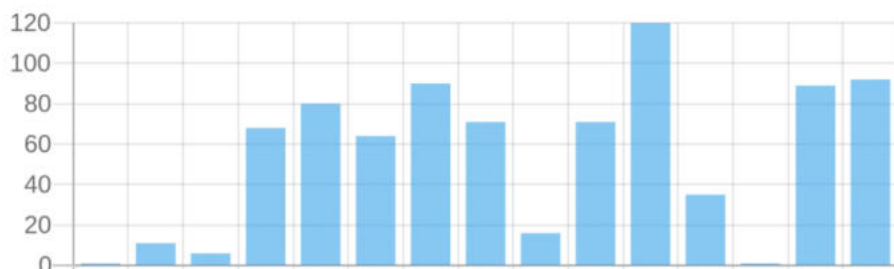


Figure 2: Number of dialogue annotations per user

original dataset and ensured compliance with the local General Data Protection Regulation (GDPR) requirements. Additionally, we establish a web-based instance to enable remote access by the annotators⁵.

A total of 15 annotators participated in the dataset creation. All annotators were bachelor or master students at the Faculty of Philology, University of Belgrade. They were native speakers of Serbian and fluent in English, which was necessary for understanding the dialogue schemata and their instructions. The statistics on the number of dialogues created by each annotator is shown below (Figure 2) as well as the example of a populated dialogue schema (Figure 3).

The annotation period lasted approximately two months, starting in May 2026. During this period, weekly online sessions were organized to provide an opportunity to discuss annotation-related and linguistic challenges encountered during the process. The observations and discussion outcomes were noted and will be consolidated into annotation guidelines to support the future extension of the srWoZ dataset and to ensure greater consistency in subsequent annotation phases.

⁵ The platform is available at <http://147.91.183.138/>

User Turn 1

Express your intent to search for a restaurant with the following property(ies):

price budget =

area or place =



Dobar dan! U potrazi sam za ekskluzivnijim restoranom koji se nalazi u zapadnom delu grada.

Assistant Turn 1

Ask for the following information about the restaurant:



Dobar dan! Da li Vas zanima neki poseban tip kuhinje?

Figure 3: Dialogue annotation following outline-based methodology

3.3. Dataset Statistics

Following the data collection process described in the previous section, a total of 812 dialogues were collected. Figure 4 presents the distribution of dialogue lengths measured in terms of the number of turns. The average dialogue length is 9.07 turns, while the longest dialogue contains 24 turns. Moreover, more than 70% of the dialogues contain at least 10 turns, indicating a relatively high level of conversational complexity within the dataset.

Overall, the dataset comprises 7,368 turns. The average turn length, measured in tokens, is 8.14 for user turns and 9.27 for assistant (wizard) turns. Figure 5 illustrates the corresponding token-length distributions. As expected, assistant turns tend to be more verbose than user turns.

As stated above, the domains covered in the initial version of the srWOZ dataset are restaurants, hotels, and attractions, represented exclusively through single-domain dialogues. Each domain is characterized by an ontology that defines the corresponding dialogue act types and possible dialogue slots. The analysis of the srWoZ

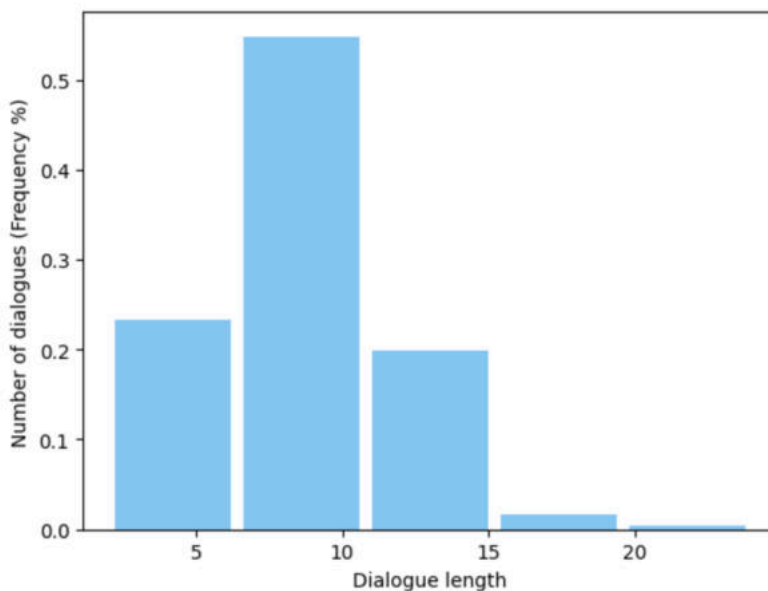


Figure 4: Dialogue length distribution in terms of number of turns

dataset follows the systematization introduced by (Budzianowski et al., 2018), which is summarized in Figure 6. For example, the dialogue acts *inform* and *request* appear across all three domains. In the hotel domain, the dialogue act *inform* represents a turn expressing the intent to search for a hotel, optionally specifying one or more desired properties through the corresponding slots, while *request* represents an assistant turn requesting specific hotel properties. Table 2 presents the distribution of dialogue act types across the three domains included in the current version of the srWoZ dataset.

3.4. Dataset Qualitative Evaluation

To date, we have not conducted a comprehensive qualitative evaluation of the dataset. However, during the regular weekly on-line sessions, various linguistic and annotation-related questions that could affect the quality of the dataset were discussed. These primarily concerned the annotation process itself, translation choices (e.g., *guesthouse* being translated as *pansion*), and dialogue style (e.g., formal, less formal, or casual language). As all dialogues were annotated using the Latin script, we also identified some inconsistencies in the use of diacritics (e.g., *veceras* instead of *večeras*) as well as occasional

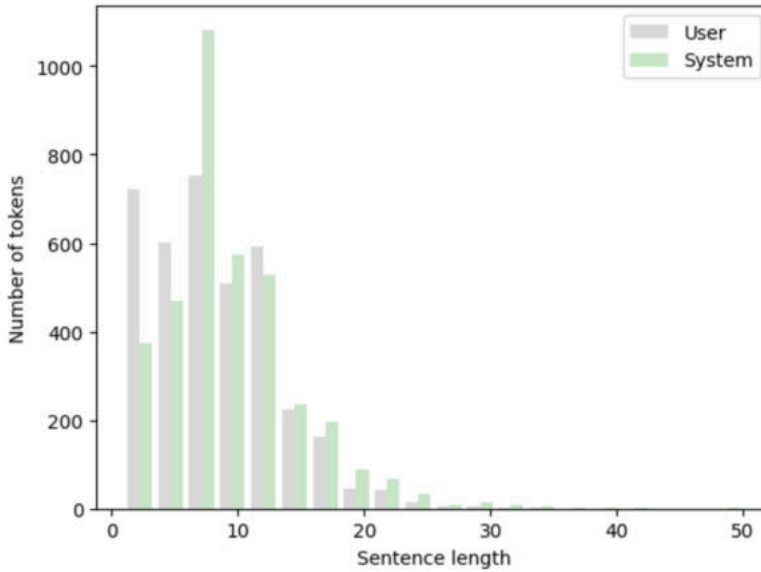


Figure 5: Turn length distribution in terms of tokens

act type	inform* / request* / select ¹²³ / recommend ¹²³ / not found ¹²³ request booking info ¹²³ / offer booking ¹²³⁵ / inform booked ¹²³⁵ / decline booking ¹²³⁵ welcome* / greet* / bye* / reqmore*
slots	address* / postcode* / phone* / name ¹²³⁴ / no of choices ¹²³⁵ / area ¹²³ / pricerange ¹²³ / type ¹²³ / internet ² / parking ² / stars ² / open hours ³ / departure ⁴⁵ destination ⁴⁵ / leave after ⁴⁵ / arrive by ⁴⁵ / no of people ¹²³⁵ / reference no. ¹²³⁵ / trainID ⁵ / ticket price ⁵ / travel time ⁵ / department ⁷ / day ¹²³⁵ / no of days ¹²³

Figure 6: Dialogue acts and types per domain covered by *Multi³WOZ* ontologies. The upper script indicates which domains it belongs to. *: universal, 1: restaurant, 2: hotel, 3: attraction, 4: taxi, 5: train, 6: hospital, 7: police

typographical errors. In addition, during the initial annotation iterations, we observed some partial replacements of addresses and phone numbers resulting from the automatic ontology mapping. This issue was addressed and the mapping procedure was improved. All identified issues and observations have been documented and will be included into the annotation guidelines for the next annotation phase. Below (Figure 7), you can find an example of a dialogue (the dialogue scheme *SNG1116.json*) from the *attraction* domain.

A comprehensive quality-control procedure will be conducted upon completion of the annotation phase of the entire dataset to en-

Dialogue Act	Restaurant	Hotel	Attraction
Inform	5,289	3,662	462
Request	1,573	837	214
Not Found	407	162	15
Recommend	254	180	53
Select	252	201	18
Booking-Inform	368	488	5
Booking-Request	184	260	0
Booking-Not Found	215	245	0
Booking-Book	426	426	0
general-reqmore	902	594	131
general-bye	1,279	621	142
general-thank	1,020	559	138
general-welcome	389	169	41
general-greet	117	109	31

Table 2: Distribution of dialogue act types across the restaurant, hotel, and attraction domains.

User: Imam nešto slobodnog vremena i htela bih da osmotrim gradsku arhitekturu. Šta mi predlažeš?
Assistant: Na koji deo grada želite da se fokusirate?
User: Na centar.
Assistant: Odličan izbor! U tom slučaju predlažem zgradu hotela "Moskva". Ulaz je potpuno besplatan.
User: Na kojoj se adresi nalazi ovaj hotel?
Assistant: U pitanju je adresa Balkanska 1.
User: Hvala puno!
Assistant: Nema na čemu!

Figure 7: Dialogue *SNG1116.json*

sure a systematic and consistent evaluation. This evaluation will be aligned with protocols suggested for other languages and will include a more thorough assessment of linguistic consistency, annotation quality, and adherence to the defined ontology and annotation guidelines.

4. Conclusion and Future Work

In this paper, we introduced **srWoZ**, the first dataset of task-oriented dialogues in Serbian. Datasets such as srWoZ remain essential for the systematic evaluation and comparison of dialogue

models, as well as for the training and fine-tuning of culturally and linguistically adapted conversational systems. As part of future work, we plan to extend the dataset by covering the remaining domains, including hospitals, police stations, and transportation services, and by annotating longer and more complex multi-domain dialogues. We also intend to conduct a more comprehensive linguistic analysis of the dataset to further assess its quality, diversity, and representativeness. Finally, we aim to publicly release both the dataset and the underlying ontologies to support the research community and foster further advances in Serbian dialogue systems and conversational AI.

Acknowledgements

This research was supported by Science Fund of the Republic of Serbia, GRANT #7276, Text Embeddings- Serbian Language Applications (TESLA).

Language Resources

Zang, Xiaoxue and Rastogi, Abhinav and Sunkara, Srinivas and Gupta, Raghav and Zhang, Jianguo and Chen, Jindong. 2020. *MultiWOZ 2.2: A Multi-Domain Dialogue Dataset*. Version 2.2; Apache License 2.0. Accessed: 2026-07-30.

References

- Rishi Bommasani, Kevin Klyman, Shayne Longpre, Betty Xiong, Sayash Kapoor, Nestor Maslej, Arvind Narayanan, and Percy Liang. 2025. [Foundation Model Transparency Reports](#). In *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '24, page 181–195. AAAI Press.
- Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. [MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Process-*

- ing*, pages 5016–5026, Brussels, Belgium. Association for Computational Linguistics.
- Bosheng Ding, Junjie Hu, Lidong Bing, Mahani Aljunied, Shafiq Joty, Luo Si, and Chunyan Miao. 2022. [GlobalWoZ: Globalizing MultiWoZ to Develop Multilingual Task-Oriented Dialogue Systems](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1639–1657, Dublin, Ireland. Association for Computational Linguistics.
- Mihail Eric, Rahul Goel, Shachi Paul, Abhishek Sethi, Sanchit Agarwal, Shuyang Gao, Adarsh Kumar, Anuj Goyal, Peter Ku, and Dilek Hakkani-Tur. 2020. [MultiWOZ 2.1: A Consolidated Multi-Domain Dialogue Dataset with State Corrections and State Tracking Baselines](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 422–428, Marseille, France. European Language Resources Association.
- Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, Swetha Ranganath, Laurie Crist, Misha Britan, Wouter Leeuwis, Gokhan Tur, and Prem Natarajan. 2023. [MASSIVE: A 1M-Example Multilingual Natural Language Understanding Dataset with 51 Typologically-Diverse Languages](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4277–4302, Toronto, Canada. Association for Computational Linguistics.
- Ting Han, Ximing Liu, Ryuichi Takanabu, Yixin Lian, Chongxuan Huang, Dazhen Wan, Wei Peng, and Minlie Huang. 2021. [MultiWOZ 2.3: A Multi-domain Task-Oriented Dialogue Dataset Enhanced with Annotation Corrections and Co-Reference Annotation](#). In *Natural Language Processing and Chinese Computing: 10th CCF International Conference, NLPCC 2021, Qingdao, China, October 13–17, 2021, Proceedings, Part II*, page 206–218, Berlin, Heidelberg. Springer-Verlag.
- Michael Heck, Carel van Niekerk, Nurul Lubis, Christian Geishauser, Hsien-Chin Lin, Marco Moresi, and Milica Gašić. 2020. [TripPy: A Triple Copy Strategy for Value Independent Neural Dialog State Tracking](#). In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 35–44, 1st virtual meeting. Association for Computational Linguistics.

- Songbo Hu, Han Zhou, Mete Hergul, Milan Gritta, Guchun Zhang, Ignacio Iacobacci, Ivan Vulić, and Anna Korhonen. 2023. [Multi 3 WOZ: A Multilingual, Multi-Domain, Multi-Parallel Dataset for Training and Evaluating Culturally Adapted Task-Oriented Dialog Systems](#). *Transactions of the Association for Computational Linguistics*, 11:1396–1415.
- Vojtěch Hudeček and Ondrej Dusek. 2023. [Are Large Language Models All You Need for Task-Oriented Dialogue?](#) In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 216–228, Prague, Czechia. Association for Computational Linguistics.
- Chia-Chien Hung, Anne Lauscher, Ivan Vulić, Simone Ponzetto, and Goran Glavaš. 2022. [Multi2WOZ: A Robust Multilingual Dataset and Conversational Pretraining for Task-Oriented Dialog](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3687–3703, Seattle, United States. Association for Computational Linguistics.
- Daniel Jurafsky and James H. Martin. 2026. [Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models](#), 3rd edition. Online manuscript released August 19, 2026.
- Zhaojiang Lin, Andrea Madotto, Genta Indra Winata, Peng Xu, Fei-jun Jiang, Yuxiang Hu, Chen Shi, and Pascale Fung. 2021. [BiToD: A Bilingual Multi-Domain Dataset For Task-Oriented Dialogue Modeling](#). In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*.
- Olga Majewska, Evgeniia Razumovskaia, Edoardo Ponti, Ivan Vulić, and Anna Korhonen. 2023. [Cross-Lingual Dialogue Dataset Creation via Outline-Based Generation](#). *Transactions of the Association for Computational Linguistics*, 11:139–156.
- Ulfeta Marovac, Aldina Avdić, and Nikola Milošević. 2023. [A Survey of Resources and Methods for Natural Language Processing of Serbian Language](#). *arXiv preprint arXiv:2304.05468*.
- Nikita Moghe, Evgeniia Razumovskaia, Liane Guillou, Ivan Vulić, Anna Korhonen, and Alexandra Birch. 2023. [Multi3NLU++: A Multilingual, Multi-Intent, Multi-Domain Dataset for Natural Lan-](#)

- guage Understanding in Task-Oriented Dialogue. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 3732–3755, Toronto, Canada. Association for Computational Linguistics.
- Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. 2020. [Towards Scalable Multi-Domain Conversational Agents: The Schema-Guided Dialogue Dataset](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:8689–8696.
- Evgeniia Razumovskaia, Goran Glavaš, Olga Majewska, Edoardo Ponti, Anna Korhonen, and Ivan Vulic. 2022. [Crossing the Conversational Chasm: A Primer on Natural Language Processing for Multilingual Task-Oriented Dialogue Systems](#). *Journal of Artificial Intelligence Research*, 74:1351–1402.
- Pranab Sahoo, Prabhash Meharia, Akash Ghosh, Sriparna Saha, Vinija Jain, and Aman Chadha. 2024. [A Comprehensive Survey of Hallucination in Large Language, Image, Video and Audio Foundation Models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 11709–11724, Miami, Florida, USA. Association for Computational Linguistics.
- Mihailo Škorić and Nikola Janković. 2024. [New Textual Corpora for Serbian Language Modeling](#). *Infotheca: Journal for Digital Humanities*, 24(1):71–96.
- Mihailo Škorić and Saša Petalinkar. 2025. [Quality Textual Corpora and New South Slavic Language Models](#). In *JUDIG – Proceedings*, pages 337–348, Belgrade.
- Renato Vukovic, Carel van Niekerk, Michael Heck, Benjamin Ruppik, Hsien-chin Lin, Shutong Feng, Nurul Lubis, and Milica Gašić. 2026. [Text-to-SQL Task-oriented Dialogue Ontology Construction](#). *Transactions of the Association for Computational Linguistics*, 14:893–917.
- Feng Yao, Yufan Zhuang, Zihao Sun, Sunan Xu, Animesh Kumar, and Jingbo Shang. 2024. [Data contamination can cross language barriers](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17864–17875, Miami, Florida, USA. Association for Computational Linguistics.
- Fanghua Ye, Jarana Manotumruksa, and Emine Yilmaz. 2022. [Multi-WOZ 2.4: A Multi-Domain Task-Oriented Dialogue Dataset with](#)

- Essential Annotation Corrections to Improve State Tracking Evaluation. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 351–360, Edinburgh, UK. Association for Computational Linguistics.
- Steve Young, Milica Gašić, Blaise Thomson, and Jason D. Williams. 2013. [POMDP-Based Statistical Spoken Dialog Systems: A Review](#). *Proceedings of the IEEE*, 101(5):1160–1179.
- Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Jianguo Zhang, and Jindong Chen. 2020. [MultiWOZ 2.2 : A Dialogue Dataset with Additional Annotation Corrections and State Tracking Baselines](#). In *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, pages 109–117, Online. Association for Computational Linguistics.
- Qi Zhu, Christian Geishauser, Hsien-chin Lin, Carel van Niekerk, Baolin Peng, Zheng Zhang, Shutong Feng, Michael Heck, Nurul Lubis, Dazhen Wan, Xiaochen Zhu, Jianfeng Gao, Milica Gašić, and Minlie Huang. 2023. [ConvLab-3: A Flexible Dialogue System Toolkit Based on a Unified Data Format](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 106–123, Singapore. Association for Computational Linguistics.
- Qi Zhu, Kaili Huang, Zheng Zhang, Xiaoyan Zhu, and Minlie Huang. 2020. [CrossWOZ: A Large-Scale Chinese Cross-Domain Task-Oriented Dialogue Dataset](#). *Transactions of the Association for Computational Linguistics*, 8:281–295.

srWOZ: ka modelovanju dijaloga usmerenih na zadatak u srpskom jeziku

Sažetak

Nadasve brzo odomaćivanje velikih jezičkih modela i agentskih okvira otvorilo je potrebu za kreiranjem visokokvalitetnih skupova podataka koji se mogu koristiti za razvoj i evaluaciju sistema za dijalog. Među njima se posebno izdvajaju *sistemi za dijalog usmereni na zadatak* (engl. *task-oriented dialogue systems*) kao važna klasa konverzacijskih sistema koja pomaže korisnicima da ostvare konkretne zadatke i ciljeve kroz interakcije na prirodnom jeziku. Uprkos njihovoj rastućoj popularnosti i praktičnom značaju, istraživanje i razvoj ovakvih sistema za srpski jezik i dalje su ograničeni zbog nedostatka odgovarajućih jezičkih resursa i anotiranih skupova dijaloških podataka.

U ovom radu predstavljamo *srWOZ*, skup podataka za srpski jezik koji sadrži *višekružne* dijaloge usmerene na zadatak. Njegovo kreiranje inspirisano je postojećim skupom podataka *Multi³WOZ*, višejezičnim, višedomenskim i višeparalelnim skupom kulturno prilagođenih dijaloga orijentisanih na zadatak. Za potrebe razvoja skupa podataka *srWOZ*, pratili smo originalnu metodologiju jezičke adaptacije primenjenu u okviru projekta *Multi³WOZ*. Najpre smo prilagodili osnovnu ontologiju srpskom kulturnom i institucionalnom kontekstu, obuhvativši sedam domena: turističke atrakcije, bolnice, policijske službe, hotele, restorane, taksi prevoz i javni prevoz, uz normalizaciju vrednosti slotova. Zatim smo prilagodili komponentu za generisanje dijaloga zasnovanu na shemama (engl. *outline-based*), adaptacijom odgovarajućih dijaloških shema i modula za zapisivanje dijaloga. Konačno, prikupili smo dijaloge od strane izvornih govornika srpskog jezika primenom *odozdo-nagore* pristupa zasnovanog na shemama (engl. *bottom-up outline-based method*). Ovim procesom formiran je inicijalni skup podataka sa 817 dijaloga.

Nadomešćujući problem nedostatka resursa za modelovanje dijaloga na srpskom jeziku, *srWOZ* doprinosi širim naporima jezičke zajednice usmerenim ka proširenju dijaloških tehnologija na jezička okruženja sa ograničenim resursima poput srpskog jezičkog okruženja.

Ključne reči: sistemi za dijalog, dijalog orijentisan na zadatak, kulturna adaptacija

Works are licensed under the terms and conditions of the Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>

