

# JMultiWOZ: 日本語タスク指向型対話データセットの構築

大橋厚元<sup>1\*</sup> 平井龍<sup>1\*</sup> 飯塚慎也<sup>2</sup> 東中竜一郎<sup>1</sup><sup>1</sup> 名古屋大学大学院情報学研究科 <sup>2</sup> 名古屋大学情報学部{ohashi.atsumoto.c0, hirai.ryu.k6, iizuka.shinya.a8}@s.mail.nagoya-u.ac.jp  
higashinaka@i.nagoya-u.ac.jp

## 概要

深層学習を活用したタスク指向型対話システムの研究において、対話データセットは重要である。近年英語圏では、多数のマルチドメインタスク指向型対話データセットが提案されており、タスク指向型対話システムの発展が著しい。一方で、日本語によるマルチドメインタスク指向型対話データセットはいまだ存在せず、英語圏と比較してタスク指向型対話の研究は限られている。本研究では、日本語によるタスク指向型対話システムの研究開発の促進を目指し、日本語初のマルチドメインタスク指向型対話データセット JMultiWOZ を構築する。

## 1 はじめに

近年、タスク指向型対話システムの研究において、ニューラルモデルをベースとした手法が活発に導入されており、タスク指向型対話システムの発展が著しい [1, 2, 3, 4]。これらニューラルモデルの実装には、タスク指向型対話データセットが必須であり、英語における多数のシングルドメインの対話データセット [5, 6, 7, 8, 9] が利用されてきた。最近では、より複雑なマルチドメインの対話に対応するため、MultiWOZ [10] が提案された。MultiWOZ は、7つのドメイン（観光名所、ホテル、レストラン、タクシー、電車、警察、病院）にまたがる対話が収録されている大規模データセットである。さらに MultiWOZ を先駆けとして、SGD [11], STAR [12], ABCD [13] といった大規模対話データセットが提案されており、これらをベンチマークとしたタスク指向型対話モデルの研究が活発化している。また中国語においても、CrossWOZ [14], RiSAWOZ [15], BiTOD [16] など、大規模マルチドメイン対話データセットが多く提案されており、中国語を含む多言語のタスク指向型対話システムの研究が促進されてい

\* Equal contribution

表1 JMultiWOZ の統計

尺度	合計	Train	Dev	Test
対話数	4,254	3,654	300	300
ターン数	61,325	52,637	4,425	4,263
トークン数	1,103,974	945,492	80,670	77,812
語彙数	11,065	10,238	2,911	2,847

る。一方、日本語においては、マルチドメインタスク指向型対話データセットは未だ存在しない。

本研究では、日本語初のマルチドメインタスク指向型対話データセット JMultiWOZ (Japanese Multi-Domain Wizard-of-Oz) を提案する。JMultiWOZ は旅行タスクに関する6つのドメイン（観光名所、宿泊施設、飲食店、買い物施設、タクシー、天気）にまたがる対話を合計4,254対話収録している。表1に、JMultiWOZ における train/dev/test セットの内訳を示す。

本稿では、JMultiWOZ の構築手順とデータの統計量を概説する。また、データベースの検索条件推定タスクを設定し、これを用いた評価によって、JMultiWOZ のベンチマークとしての将来性を示す。今後、対話行為 [17] 及び対話状態 [18] のアノテーションを付与し、タスク指向型対話における多数のタスク（意図推定、対話状態追跡、行動決定、そして対話生成）のためのベンチマークを提供する予定である。

## 2 データ構築

JMultiWOZ は、日本国内の9都市（札幌、仙台、東京、横浜、名古屋、京都、大阪、福岡、那覇）のいずれかへの旅行者が、観光情報を収集しながら旅行を計画する対話を収録したコーパスである。各対話には、6つのドメイン（観光名所、宿泊施設、飲食店、買い物施設、タクシー、天気）に関するタスクが1つ以上含まれている。Wizard-of-OZ 法 [19] を用い、各対話は、人間の対話者2人が旅行者（ユーザ）役と

情報提供者（オペレータ）役に分かれて実施した。本コーパスは、(1) オントロジーの定義、(2) オペレータ役が旅行情報を取得するために使用するバックエンドデータベースの構築、(3) 対話ゴールの設計と作成、(4) 対話収集の4ステップで構築した。

## 2.1 オントロジーの定義

タスク指向型対話における**オントロジー**とは、バックエンドデータベースの構造を表現するものである。具体的には、バックエンドデータベースに含まれるエンティティ<sup>1)</sup>が持つスロット（エンティティの名称や住所などの属性情報）を定義する。本研究では、既存研究のオントロジー [10, 14] を参考にして、ドメインごとにオントロジーを定義した。付録の表 5 に全ドメインのオントロジーを示す。

## 2.2 バックエンドデータベースの構築

オントロジーに基づいて、対話中にオペレータ役が旅行情報を取得するために使用するバックエンドデータベースをドメインごとに構築した。実在するエンティティを用いて対話のリアリティを高める目的から、観光名所、宿泊施設、飲食店、買い物施設、タクシーのデータベース構築では、政府や各都市の市町村が公開している観光名所や宿泊施設のリスト（付録の表 6 を参照）を使用した。このリストの内、web サイトを公開している施設のみをデータベースに含めるべきエンティティとして抽出し、エンティティに含めるべきスロットの情報を、各 web サイトから目視で取得した。なお、タクシードメインでは、エンティティの単位をタクシー会社とした。最終的に構築されたデータベースにおける各ドメインのエンティティ数としては、観光名所が 447 件、宿泊施設が 884 件、飲食店が 952 件、買い物施設が 445 件、タクシーが 167 件となった。天気ドメインのデータベース構築においては、エンティティの単位を日付とし、各都市につき 365 日分の天気情報を疑似的に作成した。

## 2.3 対話ゴールの設計と生成

対話ゴールは、オペレータ役との対話を通して達成されるべきユーザ役の目標であり、各対話につき 1 つ設定される。各ゴールは、1 つ以上の *informable* スロット（希望の予算や目的地など、ユーザが探すべきエンティティの検索条件）と、1 つ以上の

**表 2** 対話ゴールの例。具体的な条件（“名古屋”や“高め”など）が設定されているスロットと、“?” が設定されているスロットは、それぞれ *informable* スロットと *requestable* スロット（2.3 節を参照）である。

ドメイン	スロット
観光名所	都市 = “名古屋”
	駐車場の有無 = “有り (無料)”
	Wi-Fi = “有り (無料)”
	地区 = “?”
	最寄り駅 = “?”
	定休日 = “?”
宿泊施設	値段帯 = “高め”
	Wi-Fi の有無 = “有り (無料)”
	近隣の駅からの所要時間 = “?”
	タイプ = “?”

*requestable* スロット（電話番号や住所など、ユーザが取得すべきエンティティの属性情報）で構成される。各対話ゴールは、2.1 節で定義されたオントロジーから、ドメインとスロットをランダムにサンプルして作成した。対話の長さや複雑さにバリエーションを持たせるため、1 つのゴールには 1~3 つのドメインが含まれ、各ドメインにつき合計 2~7 つのスロットが含まれるように設計した。なお都市の *informable* スロットについては、現実性を考慮し、対話ゴール内のドメイン間で共有される。また、現実で一般的に予約を行うようなドメイン（例えば、所望の飲食店や宿泊施設を見つけた場合、日付や人数の条件を用いて予約することが多い）においては、予約のための条件を *informable* スロットとして、ランダムに追加した。後段の対話収集に先立ち、合計 5,000 種類の対話ゴールを作成した。作成された対話ゴールの具体例を表 2 に示す。

## 2.4 対話収集

バックエンドデータベースとランダムに生成された対話ゴールを用いて対話を収集した。Zhu ら [14] の対話収集用プラットフォーム<sup>2)</sup>をベースとしてユーザ役とオペレータ役の対話作業用 web UI を実装し、対話収集に使用した。クラウドソーシングサービスとしてランサーズ<sup>3)</sup>を使用し、ユーザ役とオペレータ役の対話作業者をそれぞれ募集した。作業員自身のプロフィールに関するアンケート（年代や性別等）と対話作業によって得られるデータの公開に関する同意書に回答した作業員のみが対話作業

1) 具体的な観光名所や宿泊施設

2) <https://github.com/thu-coai/CrossWOZ>

3) <https://www.lancers.jp/>

- あなたは**名古屋**への旅行を計画しています。
- 当日訪れる**観光名所**を探してください。駐車場を**無料**で利用できる場所にしてください。Wi-Fiが**無料**で利用できる場所にしてください。
- 条件に合う観光地が見つかったら、**エリア**、**最寄り駅**、**定休日**を聞いてください。
- 当日泊まる**宿泊施設**を探してください。予算は**高め**が希望です。Wi-Fiが**無料**で利用できる場所にしてください。
- 条件に合う宿泊施設が見つかったら、**近辺の駅からの所要時間**、**タイプ**を聞いてください。

図 1 表 2 の対話ゴールを表す説明文。青文字は informable スロットを、緑文字は requestable スロットをそれぞれ示している。

に参加した<sup>4)</sup>。作業者は、対話作業のマニュアルを読み、対話作業のデモンストレーション動画（web UI の操作方法や対話の進め方などを動画で解説したもの）を視聴し、作業の流れを学習してから対話作業を実施した。多様なユーザ発話を確保するため、ユーザ役が参加できる対話作業回数を最大 100 対話とし、同じオペレータ役との対話回数を最大 20 対話とした。一方オペレータ役は一貫した振る舞いをするのが好ましいため、対話作業数に制限を設けなかった。最終的に、ユーザ役 46 名とオペレータ役 17 名が対話作業に参加した。以降で、ユーザ役とオペレータ役それぞれのタスクを説明する。

**ユーザ役のタスク** ユーザ役のタスクは、対話ゴールに設定されている informable スロットを適切にオペレータ役に伝え、requestable スロットの情報をオペレータ役から取得することである。ユーザ役は、表 2 のような形式の対話ゴールの代わりに、各スロットを説明したテンプレート文を読むことで、対話ゴールを把握する。表 2 の対話ゴールを説明するテンプレート文を図 1 に示す。

**オペレータ役のタスク** オペレータ役のタスクは、対話相手であるユーザ役から伝えられた情報に合ったエンティティを旅行情報データベースから検索し、見つかったエンティティに関する情報をユーザ役に伝えることである。オペレータ役が使用する対話作業用 web UI を付録の図 3 に示す。この UI を用いてバックエンドデータベースの検索や検索結果からの情報の読み取り等の操作を行う。オペレータ役が各ターンに入力したエンティティの検索条件はアノテーションとして記録される。既存コーパスでは、アノテーションにおける表記ゆれ等の問題 [20, 21] が報告されているため、本研究では、ア

4) 事前に、倫理的観点における承認を所属組織から得た。

表 3 学習セットにおける、マルチドメインタスク指向型対話データセットの比較。下 2 行は 1 対話あたりの平均ドメイン数と平均ターン数をそれぞれ示している。

尺度	MultiWOZ	CrossWOZ	JMultiWOZ
言語	英語	中国語	日本語
ドメイン数	7	5	6
スロット数	57	72	79
合計対話数	8,438	5,012	3,654
合計ターン数	115,424	84,692	52,637
平均ドメイン数	1.8	3.24	2.02
平均ターン数	13.7	16.9	14.4

ノテーションを正規化することを目的とし、プルダウンメニューから入力候補を選択するようにし、作業者が手動で検索条件を入力することを避けた。さらに、オペレータ役の作業の質を高めるため、事前に 3~5 対話の練習作業を行い、問題のある作業に対しては、著者らからのフィードバックを受けた。フィードバックを繰り返し、問題が無くなった作業者のみが本番の対話作業に参加した。

上記手順により、4,508 対話が収集された。ここで、データセットを構築する上でノイズとなる対話を、以下 2 つの手順によって修正または除外した：

- 各対話終了時、対話中に発生した問題を報告するためのアンケートを作業者に回答させた。アンケートで問題が報告されていた対話は著者らが目視で確認し、誤字など軽微なミスについては手作業で修正した。作業者が対話ゴールを取り違えて対話を進めている場合など、重大な問題がある対話はデータセットから除外した。
- 各対話の対話ゴールと対話最終時点でのデータベースの検索条件を比較し、いずれかのスロットの値が一致していない対話は除外した。

この修正によって、最終的に合計 4,254 対話からなるデータセットが構築された。ここで、dev/test セットとして 300 対話ずつ、4,254 対話からランダムに抽出し、残る 3,654 対話を train セットとした（各セットの統計は表 1 を参照）。

## 2.5 統計量と既存データセットとの比較

表 3 は、JMultiWOZ と、英語と中国語における主要なマルチドメインタスク指向型対話データセット MultiWOZ と CrossWOZ の統計量を示している。ドメイン数、スロット数、平均ドメイン数、平均ターン数が同程度であることから、JMultiWOZ の対話の

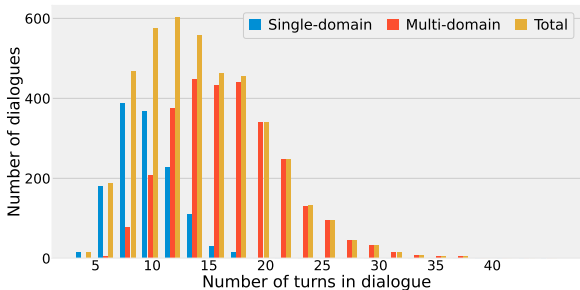


図 2 対話長の分布. 1 ドメインのみ含まれる対話 (single-domain) と 2 ドメイン以上含まれる対話 (multi-domain) に分けて集計されている.

複雑さは、既存のデータセットと比較して遜色ないと言える. 図 2 は、全 4,254 対話を 1 ドメインのみ含まれる対話 (single-domain) と複数ドメインが含まれる対話 (multi-domain) に分けた際の、対話ターン数の分布を示している. いずれの対話でも長さや複雑さが多様であることが分かる.

### 3 ベンチマークとしての JMultiWOZ

JMultiWOZ が現時点でタスク指向型対話ベンチマークとしてどの程度有用であるかを調査するため、JMultiWOZ を用いてタスク指向型対話モデルを学習し評価する. 本研究では、対話タスクとして、データベースの検索条件推定タスクを設定し評価する. 検索条件推定は、各ターンにオペレータ役が入力したデータベース検索条件 (2.4 節を参照) を対話履歴から推定するタスクである. 一般的なベンチマークである対話状態追跡 (Dialogue State Tracking; DST) [22, 23] と類似したタスクであるが、ユーザが陽に示した要求のみを扱う点において異なる.

検索条件推定の評価尺度には、DST の主要な評価尺度である Joint Goal Accuracy (JGA) を使用する. JGA では、各ターンのデータベース検索条件に含まれるスロットの値を全て正確に推定できたかが評価される. 実験では、検索条件推定のためのモデルとして T5DST [24] を採用し (詳細は付録の A.1 節を参照)、全ての学習データを用いた評価 (full-shot) と、初見ドメインへの適応能力の評価 (zero-shot cross domain) を行った.

#### 3.1 Full-shot

Full-shot 評価実験では、全ての train セットで学習された検索条件推定モデルを、全ての test セットを用いて評価した. 評価結果としては、JGA が 90.3% と高かった. 参考値として、DST タスクにお

表 4 JMultiWOZ における zero-shot cross domain 評価の結果. スコアは全て Joint Goal Accuracy (JGA) を示す.

観光名所	宿泊施設	飲食店	買い物	タクシー	天気
60.6	34.5	51.7	75.1	45.4	62.6

ける、英語と中国語の主要なマルチドメインタスク指向型対話データセットである MultiWOZ<sup>5)</sup> [10] と CrossWOZ [14] の JGA も評価したところ、それぞれ 56.9% と 53.6% であった. DST タスクでは、索条件推定タスクとは異なり、暗示的な対話状態の更新 (MultiWOZ ではユーザがオペレータの推薦条件に肯定した場合、その条件を用いて対話状態が更新される) 等の難しさがあり、それがこの精度の差となって表れたと考えられる. 今後、JMultiWOZ についても対話状態のアノテーションを実施し、DST タスクの精度を確認したい.

#### 3.2 Zero-shot Cross Domain

ゼロショットドメイン間 (zero-shot cross domain) 転移学習 [25, 26] における検索条件推定タスクの精度を評価した. この実験では、train セットの内、目標となる 1 ドメインを除いた 5 ドメインのみが含まれるデータでモデルを学習し、評価時は test セットに含まれる目標ドメインのみを使用した.

評価結果を表 4 に示す. 宿泊施設、飲食店、タクシーでの JGA は、full-shot 評価と比較し大幅に低下しており、また、観光名所、買い物など、他のドメインと共通のスロットを多く持つドメインでも JGA の低下がみられた. そのため、検索条件推定タスクは、full-shot においては比較的容易なタスクである一方、zero-shot cross domain においては十分難しく、取り組むべき課題であると考えられる.

### 4 おわりに

本研究では、日本語におけるマルチドメインタスク指向型対話データセット JMultiWOZ を構築し、その構築手順と統計量を報告した. また、データベースの検索条件推定タスクによる評価を行い、タスク指向型対話ベンチマークとしての将来性を示した. 本データセットは、各ターンに対する対話行為と対話状態のアノテーションなど、データを整備したうえで、バックエンドデータベースと共に公開する予定である.

5) 本研究では、アノテーションエラーが修正された MultiWOZ2.2 [21] を用いた.

## 謝辞

本研究は、JST ムーンショット型研究開発事業、JPMJMS2011 の支援を受けたものです。また、対話収集用プラットフォームを公開していただいた Zhu ら [14] に感謝申し上げます。

## 参考文献

- [1] Yichi Zhang, Zhijian Ou, and Zhou Yu. Task-Oriented Dialog Systems That Consider Multiple Appropriate Responses under the Same Context. In **Proc. AAAI**, pp. 9604–9611, 2020.
- [2] Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. A simple language model for task-oriented dialogue. In **Proc. NeurIPS**, pp. 20179–20191, 2020.
- [3] Wanwei He, Yinpei Dai, Yinhe Zheng, Yuchuan Wu, Zheng Cao, Dermot Liu, Peng Jiang, Min Yang, Fei Huang, Luo Si, Jian Sun, and Yongbin Li. Galaxy: A generative pre-trained model for task-oriented dialog with semi-supervised learning and explicit policy injection. In **Proc. AAAI**, No. 10, pp. 10749–10757, 2022.
- [4] Yixuan Su, Lei Shu, Elman Mansimov, Arshit Gupta, Deng Cai, Yi-An Lai, and Yi Zhang. Multi-Task Pre-Training for Plug-and-Play Task-Oriented Dialogue System. In **Proc. ACL**, pp. 4661–4676, 2022.
- [5] Matthew Henderson, Blaise Thomson, and Jason D. Williams. The Second Dialog State Tracking Challenge. In **Proc. SIGDIAL**, pp. 263–272, 2014.
- [6] Layla El Asri, Hannes Schulz, Shikhar Sharma, Jeremie Zumer, Justin Harris, Emery Fine, Rahul Mehrotra, and Kaheer Suleman. Frames: a corpus for adding memory to goal-oriented dialogue systems. In **Proc. SIGDIAL**, pp. 207–219, 2017.
- [7] Tsung-Hsien Wen, David Vandyke, Nikola Mrkšić, Milica Gašić, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. A Network-based End-to-End Trainable Task-oriented Dialogue System. In **Proc. EACL**, pp. 438–449, 2017.
- [8] Mihail Eric, Lakshmi Krishnan, Francois Charette, and Christopher D. Manning. Key-value retrieval networks for task-oriented dialogue. In **Proc. SIGDIAL**, pp. 37–49, 2017.
- [9] Pararth Shah, Dilek Hakkani-Tür, Gokhan Tür, Abhinav Rastogi, Ankur Bapna, Neha Nayak, and Larry Heck. Building a Conversational Agent Overnight with Dialogue Self-Play. **arXiv preprint arXiv:1801.04871**, 2018.
- [10] Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling. In **Proc. EMNLP**, pp. 5016–5026, 2018.
- [11] Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. Towards Scalable Multi-Domain Conversational Agents: The Schema-Guided Dialogue Dataset. In **Proc. AAAI**, Vol. 34, pp. 8689–8696, 2020.
- [12] Johannes EM Mosig, Shikib Mehri, and Thomas Kober. Star: A schema-guided dialog dataset for transfer learning. **arXiv preprint arXiv:2010.11853**, 2020.
- [13] Derek Chen, Howard Chen, Yi Yang, Alexander Lin, and Zhou Yu. Action-based conversations dataset: A corpus for building more in-depth task-oriented dialogue systems. In **Proc. NAACL**, pp. 3002–3017, 2021.
- [14] Qi Zhu, Kaili Huang, Zheng Zhang, Xiaoyan Zhu, and Minlie Huang. CrossWOZ: A large-scale Chinese cross-domain task-oriented dialogue dataset. **TACL**, pp. 281–295, 2020.
- [15] Jun Quan, Shian Zhang, Qian Cao, Zizhong Li, and Deyi Xiong. RiSAWOZ: A large-scale multi-domain Wizard-of-Oz dataset with rich semantic annotations for task-oriented dialogue modeling. In **Proc. EMNLP**, pp. 930–940, 2020.
- [16] Zhaojiang Lin, Andrea Madotto, Genta Indra Winata, Peng Xu, Feijun Jiang, Yuxiang Hu, Chen Shi, and Pascale Fung. Bitod: A bilingual multi-domain dataset for task-oriented dialogue modeling. **arXiv preprint arXiv:2106.02787**, 2021.
- [17] David R. Traum. Speech Acts for Dialogue Agents. **Foundations of Rational Agency**, pp. 169–201, 1999.
- [18] Jason Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. The dialog state tracking challenge. In **Proc. SIGDIAL**, pp. 404–413, 2013.
- [19] J. F. Kelley. An iterative design methodology for user-friendly natural language office information applications. **ACM Trans. Inf. Syst.**, pp. 26–41, 1984.
- [20] Mihail Eric, Rahul Goel, Shachi Paul, Abhishek Sethi, Sanchit Agarwal, Shuyang Gao, Adarsh Kumar, Anuj Goyal, Peter Ku, and Dilek Hakkani-Tur. MultiWOZ 2.1: A consolidated multi-domain dialogue dataset with state corrections and state tracking baselines. In **Proc. LREC**, pp. 422–428, 2020.
- [21] Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Jianguo Zhang, and Jindong Chen. MultiWOZ 2.2: A dialogue dataset with additional annotation corrections and state tracking baselines. In **Proc. NLP4ConvAI**, pp. 109–117, 2020.
- [22] Jason D Williams, Antoine Raux, and Matthew Henderson. The dialog state tracking challenge series: A review. **Dialogue & Discourse**, pp. 4–33, 2016.
- [23] Léo Jacqmin, Lina M. Rojas Barahona, and Benoit Favre. “do you follow me?”: A survey of recent approaches in dialogue state tracking. In **Proc. SIGDIAL**, pp. 336–350, 2022.
- [24] Zhaojiang Lin, Bing Liu, Seungwhan Moon, Paul Crook, Zhenpeng Zhou, Zhiguang Wang, Zhou Yu, Andrea Madotto, Eunjoon Cho, and Rajen Subba. Leveraging Slot Descriptions for Zero-Shot Cross-Domain Dialogue StateTracking. In **Proc. NAACL**, pp. 5640–5648, 2021.
- [25] Chien-Sheng Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher, and Pascale Fung. Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems. In **Proc. ACL**, pp. 808–819, 2019.
- [26] Zhaojiang Lin, Bing Liu, Andrea Madotto, Seungwhan Moon, Zhenpeng Zhou, Paul Crook, Zhiguang Wang, Zhou Yu, Eunjoon Cho, Rajen Subba, and Pascale Fung. Zero-shot dialogue state tracking via cross-task transfer. In **Proc. EMNLP**, pp. 7890–7900, 2021.
- [27] Zhaojiang Lin, Andrea Madotto, Genta Indra Winata, and Pascale Fung. MinTL: Minimalist transfer learning for task-oriented dialogue systems. In **Proc. EMNLP**, pp. 3391–3405, 2020.
- [28] Chia-Hsuan Lee, Hao Cheng, and Mari Ostendorf. Dialogue State Tracking with a Language Model using Schema-Driven Prompting. In **Proc. EMNLP**, pp. 4937–4949, 2021.
- [29] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. **JMLR**, pp. 1–67, 2020.
- [30] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In **Proc. NAACL**, pp. 483–498, 2021.

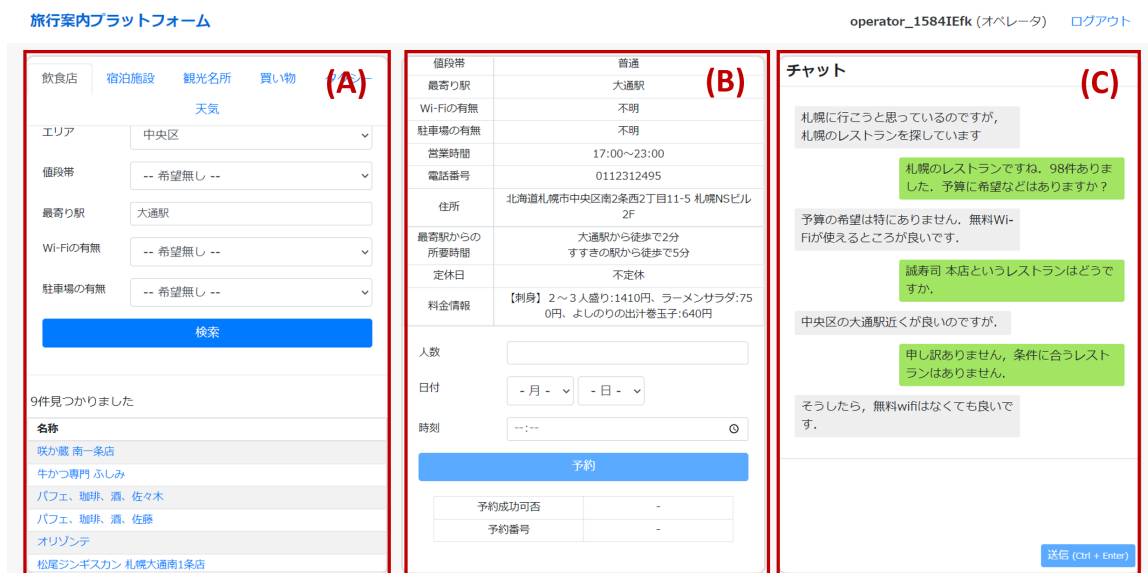
## A 付録

**表 5** 各ドメインのオントロジーとして定義されているスロットの一覧. 上付き数字はそれぞれのスロットが属するドメインを示している. \*: 全ドメイン, 1: 観光名所, 2: 宿泊施設, 3: 飲食店, 4: 買い物施設, 5: タクシー, 6: 天気.

都市*	名称 <sup>1</sup> , ジャンル <sup>1,3,4</sup> , エリア <sup>1,2,3,4,6</sup> , 最寄り駅 <sup>1,2,3,4</sup> , wifiの有無 <sup>1,2,3</sup> , 駐車場の有無 <sup>1,2,3,4</sup> , 営業時間 <sup>1,3,4</sup> , 電話番号 <sup>1,2,3,4,5</sup> , 住所 <sup>1,2,3,4</sup> , 近隣の駅からの所要時間 <sup>1,2,3,4</sup> , 定休日 <sup>1,3,4</sup> , 大人入場料 <sup>1</sup> , 子供入場料 <sup>1</sup> , 料金情報 <sup>1,2,3</sup> , 施設名 <sup>2</sup> , タイプ <sup>2</sup> , 値段帯 <sup>2,3</sup> , レストランの有無 <sup>2</sup> , チェックイン開始時刻 <sup>2</sup> , チェックアウト終了時刻 <sup>2</sup> , 予約番号 <sup>2,3,5</sup> , 人数 <sup>2,3</sup> , 日付 <sup>2,3,5,6</sup> , 宿泊日数 <sup>2</sup> , 店名 <sup>3</sup> , 時刻 <sup>3</sup> , 店舗名 <sup>4</sup> , タクシー会社名 <sup>5</sup> , キャッシュレス対応 <sup>5</sup> , ジャンボタクシー対応 <sup>5</sup> , 出発時刻 <sup>5</sup> , 出発地 <sup>5</sup> , 目的地 <sup>5</sup> , 天気 <sup>6</sup> , 最低気温 <sup>6</sup> , 最高気温 <sup>6</sup>
-----	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**表 6** バックエンドデータベース構築時に参照した, 各ドメインにおけるエンティティリストのソースサイト

ドメイン	サイト名
観光名所	札幌観光情報サイト ようこそさっぽろ, 仙台観光情報サイト - せんだい旅日和, 東京の観光公式サイト, 横浜市観光情報サイト, 名古屋市観光情報「名古屋コンシェルジュ」, 京都観光 Navi, 大阪公式観光情報 OSAKA-INFO, 福岡市公式シティガイド よかなび, 沖縄那覇の観光情報サイト — NAHANAVI
宿泊施設	旅行者向け Go To トラベル事業公式サイト
飲食店	Go To Eat 北海道キャンペーン, みやぎ 美味しいとこ Go To Eat キャンペーン, GO TO Eat キャンペーン Tokyo 食事券対象店検索, Go To Eat かながわ 食事クーポン, Go To Eat キャンペーンあいち, 京都 Go To Eat キャンペーン お食事券, Go To Eat 大阪キャンペーン, Go To Eat キャンペーン福岡, Go To Eat キャンペーンおきなわ
買い物施設	旅行者向け Go To トラベル事業公式サイト
タクシー	一般社団法人北海道ハイヤー協会, 宮城県タクシー協会仙台地区総支部, 東京のタクシー [一般社団法人 東京ハイヤー・タクシー協会], 神奈川県タクシー協会, 名古屋タクシー協会, 一般社団法人 京都府タクシー協会, 一般社団法人 大阪タクシー協会, 福岡市タクシー協会公式サイト 福岡タクシーナビ, 一般社団法人 沖縄県ハイヤー・タクシー協会



**図 3** オペレータ役の対話作業用 web UI. (A): バックエンドデータベースの検索欄. (B): エンティティの属性情報が表示される欄. (C): チャット欄.

### A.1 モデル実装の詳細

検索条件推定のモデルとしては, 近年多くの研究がなされている事前学習済み言語モデルを用いた DST モデル [27, 28] の内, T5DST [24] を採用した. T5DST は, 言語モデル T5 [29] を用いて, 対話履歴と, あるスロットの説明文から, そのスロットの値を推定するモデルである. Full-shot 評価実験では, 多言語版 T5 の mT5 [30] (300M バージョン<sup>6)</sup>) を使用した. Zero-shot cross domain 評価実験では, 日本語版 T5 (t5-base-japanese<sup>7)</sup>) を使用した. JGA の計測には, Lin ら [24] の評価スクリプトを用いた. 学習時には, バッチサイズを 32 とし, 最大 3 エポック学習した. 学習率は, 初期値  $5e-5$  から学習ステップごとに線形減衰させた. テスト時には, dev セットでの損失が最も低いチェックポイントを使用し, 生成時のビームサイズは 1 とした.

6) <https://huggingface.co/google/mt5-small>

7) <https://huggingface.co/sonois/t5-base-japanese>