

LD-SDS: Towards an Expressive Spoken Dialogue System based on Linked-Data

Alexandros Papangelis
Speech Technology Group,
Toshiba Research Europe
alex.papangelis@crl.toshiba.co.uk

Yannis Stylianou
Speech Technology Group,
Toshiba Research Europe,
and Computer Science
Department, University of
Crete, GREECE
yannis.stylianou@crl.toshiba.co.uk

Panagiotis Papadakos
Institute of Computer Science,
FORTH-ICS, GREECE
papadako@ics.forth.gr

Yannis Tzitzikas
Institute of Computer Science,
FORTH-ICS, GREECE, and
Computer Science
Department, University of
Crete, GREECE
tzitzik@ics.forth.gr

Margarita Kotti
Speech Technology Group,
Toshiba Research Europe
margarita.kotti@crl.toshiba.co.uk

Dimitris Plexousakis
Institute of Computer Science,
FORTH-ICS, GREECE, and
Computer Science
Department, University of
Crete, GREECE
dp@ics.forth.gr

ABSTRACT

In this work we discuss the related challenges and describe an approach towards the fusion of state-of-the-art technologies from the Spoken Dialogue Systems (SDS) and the Semantic Web and Information Retrieval domains. We envision a dialogue system named LD-SDS that will support advanced, expressive, and engaging user requests, over multiple, complex, rich, and open-domain data sources that will leverage the wealth of the available Linked Data. Specifically, we focus on: a) improving the identification, disambiguation and linking of entities occurring in data sources and user input; b) offering advanced query services for exploiting the semantics of the data, with reasoning and exploratory capabilities; and c) expanding the typical information seeking dialogue model (slot filling) to better reflect real-world conversational search scenarios.

Keywords

Spoken Dialogue Systems, Semantic Web, Linked Data, Advanced Query Services, Exploratory Search

1. INTRODUCTION

Conversational systems is a thriving research area with many commercial applications, such as intelligent personal assistants, e.g. Microsoft's Cortana, Apple's Siri, and Amazon's Echo among others. In addition, Amazon Lex and Watson Conversation have made the functionality of building, testing, and deploying chatbots publicly available. However, most deployed systems still utilise the rather cumbersome and time-consuming process of hand-crafted rules and finite state machines for dialogue management, paired with

domain specific and sometimes static databases.

Moving away from commercial applications, dialogue managers are typically statistical and require a lot of data – either crowd-sourced or simulated – to train their internal models, which is a challenging task. Various toolkits are available for developing such systems, e.g.: OpenDial [25] is a web-based tool that allows a user to create a slot filling dialogue system automatically, using about fifteen probabilistic rules; Olympus is a complete framework for implementing spoken dialogue systems (SDS) [4]; PyDial [30] is a toolkit for developing and training statistical multi-domain SDS based on the slot filling paradigm. Lately, [34] have proposed a way to reduce the amount of training data by encoding domain-specific knowledge and using developer-provided action templates.

Information-seeking (usually modeled as slot filling, see Section 2.3) is a very common use case for SDS, where a database is inquired for a specific item given a set of hard restrictions [6, 32]. However, databases in such systems are usually small and domain-specific. Efforts to advance such task-based systems include hybrid approaches, for example, in [19] slot filling is combined with a chatbot. The aim is to have more natural and engaging systems that switch between task-based and chatbot style dialogue, leading to more satisfied users. Another way to improve the slot filling dialogue experience is via adding memory; for example, in [35] limited contextual dialogue memory is used to jointly optimise the signals of user slot filling, intent prediction and system action prediction. Memory is essential, especially for commercial applications, as underlined by [24]. In their system, the user can refer back to an earlier state in the dialogue, which is essential e.g. when comparing alternatives or researching a complex subject. Recently, efforts are being made towards multi-domain SDS, as well as SDS that work with large knowledge graphs that combine independent domain-specific databases with other knowledge bases (e.g. Freebase) [1, 11, 20, 21]. A general trend, therefore, is to move to conversational open-domain question answering (QA). Large-scale QA systems like IBM's DeepQA rely on multiple sources to form their response: Wikipedia, other knowledge bases, dictionaries, news articles, books, etc. However, the dialogue is usually limited to one turn.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SCAI '17 Amsterdam, Netherlands

© 2017 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06...\$15.00

DOI: 10.1145/1235

As detailed above, current systems are usually: i) statistical on small domains; or ii) hand-crafted on large/open domains, mostly offering QA, rather than a natural conversation. In this work, we propose an architecture that combines the benefits of SDS and QA, and allows users to retrieve meaningful information by conversing with the system. We achieve this by taking advantage of Linked-Data, Semantic Web, and Information Retrieval technologies. Specifically, our Linked-Data SDS (LD-SDS) can fetch items from available open Linked-Data sources and support advanced exploratory and query services that will help users better understand their own goals and criteria, which are not always explicit and predefined. We here present a prototype of this system, which in the future will have memory, via exploiting a session-based model for exploratory search. Our vision is to enable users to make general queries that lead to more natural conversations.

The rest of the paper is organized as follows: Section 2 provides the related background regarding the exploratory nature of most information needs and current exploratory approaches, Linked-Data and integration of data from different resources, and the information seeking dialogue models. Section 3 discusses the proposed architecture and the prototype we have developed, while Section 4 presents the related challenges. Finally, Section 5 concludes the paper.

2. BACKGROUND

2.1 Information Needs & Exploration

In general, we can classify information needs into two very broad categories: a) *precision-oriented*; and b) *recall-oriented*. Precision-oriented information needs typically are not time consuming and the goal is to locate one resource and/or its attributes like a telephone number, an address, etc. On the other hand, recall-oriented needs frequently are time expensive, and the goal is to locate and analyze/compare a set of resources and/or their attributes, or aggregated attributes or interrelationships. Recall-oriented tasks aim at *decision making* over one or more criteria and have an exploratory nature, like search tasks in the medical, legal, consulting, patent, or academic fields, and consumer-related tasks like car buying, travel planning or even species identification. The following key attributes of exploratory tasks have been identified in [33]: a) they are associated with the goals of learning and/or investigation, b) they are general rather than specific, c) they are open-ended, d) they target multiple items, e) they involve uncertainty, f) they elicit through ill-structured information problems, g) they are dynamic, h) they are lengthy, i) they are multi-faceted, j) they are complex, and finally k) they are accompanied by other information and cognitive behaviors, like sense-making.

According to Marchionini [15] the majority of information needs are *recall-oriented*. In the same direction, Broder [5] categorizes queries as navigational (e.g. “Porsche site”), informational (e.g. “what is the best mobile phone”) and transactional for performing a task (e.g. “book a hotel”). According to the same author, the queries that are related to recall-oriented needs (i.e. the informational and transactional queries) correspond to 80% of queries (50% for the first and 30% for the latter query category). Conversational needs have an exploratory nature and are recall-oriented.

Despite this, current information systems like general purpose web search engines mainly focus on single query precision-

oriented needs. Only a small number of prototype information systems provide means for supporting recall-oriented ones (e.g. [17, 18, 23]). Most of these systems offer their exploratory features (e.g. overviews of available objects, their active attributes/values, counts, etc.) on top of faceted search, an interaction framework based on a multi-dimensional classification of data objects that allows a guided, yet unconstrained way of browsing and exploring the information space through a simple user interface. Faceted search is currently the de facto standard in e-commerce (e.g. eBay, booking.com), and its popularity and adoption is increasing in several other domains [28]. Features of this framework include: (a) display of current results in multiple categorization schemes either statically or dynamically mined through structured or unstructured data sources [17] (called facets, or dimensions, or attributes, or slots) through the selection of hard constraints, (b) display of facets and values leading to non-empty results only, (c) display of the count information for each value (i.e. the number of results the user will get by selecting that value), and (d) the ability to refine the focus gradually, since it is a session-based interaction paradigm.

Preference-enriched Faceted Search (PFS) [29] is an extension of faceted search that supports the expression of soft constraints (i.e. preferences) that can impose a ranking over the facets, values and objects of the provided information space. PFS is aligned with the principles of faceted search, allowing users to define explicitly the desired preference structure in a gradual and flexible manner, supporting also set-valued attributes and hierarchically organized values. In brief, PFS adds actions that allow the expression of various and even conflicting preferences for ranking facets, values, and objects. Such preference actions include best/worst values, relative preferences (e.g. I prefer A to B), around/not around and between actions. PFS can exploit a number of policies for composing preference actions over different facets (e.g. priority, skyline¹, etc.). Hippalus is a system that implements PFS, as described in Section 3.

2.2 Linked-Data and Data Integration

A big number of datasets (or sources) has been published according to the principles of Linked Data and this number keeps increasing. The ultimate objective of LOD (Linked Open Data) is linking and integration, for enabling discovery and integrated query answering and analysis. Linked Data refers to a method of publishing structured data, so that they can be interlinked and become more useful through semantic queries, founded on HTTP, RDF² and URIs. As proposed in [3] the major principles of Linked Data are the use of HTTP URIs as names for things, the use of standards for providing useful information and the inclusion of links to other URIs to help the exploration of even more resources.

Therefore, one of the major principles of Linked Data, i.e. interlinking of URIs, favor integration by proposing to the publishers to establish relationships with URIs from other datasets. The linking of datasets is essentially realized through the existence of common URIs, referring to

¹Database operator that filters out results, keeping only those objects that are not worse than any other object on all criteria

²<https://www.w3.org/RDF/>

schema (defined through RDF Schema³ and OWL⁴) or data elements. SPARQL is a standard query language for retrieving and manipulating RDF data across diverse data sources. The need for better linking and contextualization is also evident from proposals that rate open data, such as the 5-star Open Dat⁵, as well as ratings for vocabulary use [13]. According to [7], the LOD cloud already contains over 9,000 datasets and billions of RDF triples.

Note that before taking advantage of the available LOD sources there is a preparation phase that generates the semantic layer. During this phase and depending on the exploitation scenario and the properties of those data sources, we have to decide whether we should build a semantic warehouse that will host the gathered triples (e.g. as in [27, 16]) or a mediated access layer (e.g. as in [10]). A dedicated warehouse has the additional benefit of gaining reliability and efficiency – for example the SPARQL endpoint provided by DBpedia is not very stable or fast.

This semantic data layer can be used for both analyzing what the user says as well as for feeding the responses of the SDS. Specifically, it can be exploited for identifying named entities, for getting their descriptions or related entities through the semantic data layer either from the user’s response or other unstructured data sources. This is sometimes referred as semantic enrichment [8]. In addition, the semantic layer can be used for producing summaries by estimating the more important elements or associations [9].

2.3 Information-Seeking Dialogue Models

Dialogue models for information-seeking applications have been traditionally modeled with the slot filling paradigm [26]. According to that, the goal of the dialogue is to extract enough information from the user in order to form a database query that yields acceptable results, e.g. according to the number of items returned or some other metric. The database’s attributes, therefore, are called slots, and the process of forming the query through dialogue is called slot filling. According to [26], each slot s can take one value v_s from a given set of values: $v_s \in V_s, s \in S$, where V_s may represent a set of strings, events, numerical values, general entities, etc., and S is the set of slots. An *ontology* describes each slot and its values, as well as other attributes of the database’s schema.

Without loss of generality, we can assume that $V_s, \forall s \in S$ is a set of symbols, somehow extracted by Spoken Language Understanding (SLU), representing something meaningful for each slot. Usually, it is assumed that the goal of the user can be expressed as a tuple of slot-value pairs: $\langle s_0 = v_0, \dots, s_k = v_k \rangle$, a representation that may be somewhat restrictive if we hope to train SDS for realistic applications. Indeed, in recent work, especially for dialogue management, a simple version of slot filling is typically adopted to model human - machine dialogues, e.g. [11, 21, 32, 34].

Such approaches to model information-seeking dialogues have been largely driven by the trend to view dialogue as an optimisation problem and address it with (Partially Observable) Markov Decision Processes (POMDP). Specifically, a POMDP Dialogue Manager (DM) typically receives an n-best list of language understanding hypotheses, which are

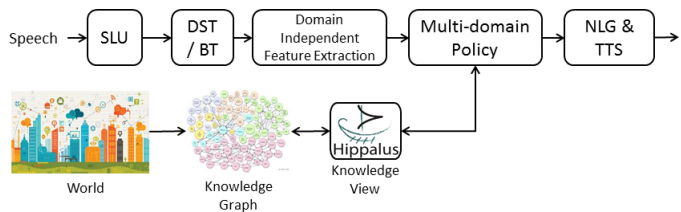


Figure 1: The proposed LD-SDS architecture. TTS refers to Text To Speech synthesis.

used to update the belief state (reflecting an estimate of the user’s goals). Using Reinforcement Learning (RL), the system selects a response that maximises the long-term return of the system. This response is typically selected from an abstract action space and has to be converted to text through language generation.

We here extend the information-seeking model used in most (PO)MDP and Deep Learning approaches to dialogue and connect to live semantic knowledge bases, aiming to create a more realistic paradigm that will drive our research forward. In the following section we describe LD-SDS and then proceed to outline some of the challenges we are facing.

3. LD-SDS ARCHITECTURE & PROTOTYPE

Moving towards a more realistic information-seeking dialogue paradigm, it is important to allow users the freedom they need to express their complex intents. These intents do not always express hard restrictions (constraints) but often express preferences that users may or may not be willing to relax as the dialogue progresses. Such preferences may refer to the importance of attributes over other attributes (e.g. *location* is much more important than *has-free-wifi* when searching for accommodation), or may refer to preferred values of a given attribute (e.g. prefer *central* over *northern* locations but *northern* may still be okay under certain circumstances), etc.

We therefore implement a number of operators $Op = \{<, \leq, >, \geq, =, \neq, \textit{around}, \textit{not around}, \textit{between}, \textit{not between}, \textit{prefer}\}$, for two types of constraints: hard and soft. Hard constraints restrict the information space, while soft constraints express preferences and impose a ranking on the available options. Of course, not all constraints are applicable to all types of slots (e.g. if a slot’s values are not ordinal). Instead of forcing each slot to have one value, therefore, we allow constraints of the form $s \textit{ op } v_s$, where op is an operator in Op except *between* and *prefer*, for which we have: $s \textit{ op } v_s^1 (v_s^1, v_s^2)$. The *prefer* operator is defined for slots in two ways: *prefer X over Y*, which means that slot X is more important than slot Y (this affects the ranking of the results, among other things); and *prefer X and Y*, which means that both slots X and Y are preferred over the rest of the slots available. In a similar manner *prefer* can be defined for a specific slot’s values. Besides *not* which is explicitly defined as a different operator to make belief tracking easier, other logical operators such as *and*, *or* are handled by the SLU and mapped into a set of constraints using the operators described above.

Another aspect of (the typical version of) the slot filling paradigm that we extend is the nature of each slot’s values. Specifically, we allow values to be defined in a hierarchical manner (e.g. *location* with regions, sub-regions, neighbour-

³<http://www.w3.org/TR/rdf-schema/>

⁴<http://www.w3.org/TR/owl2-overview/>

⁵<http://5stardata.info/en/>

hoods, etc), and we allow slots to take multiple values, from a given set (e.g. *hotel amenities*). Slots with hierarchical or multiple values directly impact SLU and dialogue state / belief tracking (DST / BT). Formally, a slot with hierarchical values can be defined as $s^h = v^h, v^h \in V^H$, where V^H are the nodes of a hierarchy (e.g. a tree). We treat such slots as regular slots (i.e. that take a single value) but handle belief tracking in a way that accounts for the value hierarchy. Multi-valued slots can be defined as $s^m = v^m, v^m \subseteq V^m$, where V^m is the set of acceptable values for slot s^m .

To process the complex intent of the user, we connect our SDS with Hippalus, an exploratory search system that materializes the Preference-enriched Faceted Search [29] over semantic views gathered from different data sources through SPARQL queries. Hippalus is a publicly accessible web system⁶ that supports the previously defined hard and soft restriction actions that allow the user to order facets, values, and objects. All the above functionality is offered in an efficient way, by using the algorithms described in [29]. The information base that feeds Hippalus is represented in RDF/S (using a schema adequate for representing objects described according to dimensions with hierarchically organized and set-valued attributes). For loading and querying such information, Hippalus supports a number of triple stores. The performed actions are internally translated to statements in the preference language described in [29], and are then sent to the server through HTTP requests. The server analyzes them, using the language’s parser, and checks their validity. If valid, they are passed to the appropriate preference algorithm. Finally, the respective preference bucket (a preference bucket holds incomparable objects regarding the given soft-constraints) order is computed and the ranked list of objects according to preference is sent to the user’s browser. The answer is also enriched with a number of metrics computed over the response (preference score, pair-wise wins of buckets’ objects regarding preference, etc.) that can be exploited by the Belief Tracking and Natural Language Generation components (e.g. for identifying the most important slots and breaking ties through user input). Figure 1 shows the overall architecture of the system.

Example Interaction

Figure 2 shows a screen shot of the working LD-SDS prototype, on a knowledge base about Japanese Hotels (382 hotels across Japan, 69 system actions, and 28 slots some of which are hierarchical and some are multi-valued). To showcase our conversational information navigation paradigm, we walk through a real interaction with our prototype LD-SDS in Table 3. Belief tracking is performed as described in the following section; parallel to this we also process the recognised dialogue acts to identify constraints or preferences. In the example we also note very briefly the gist of the response from Hippalus.

4. CHALLENGES

Conversational search is much more complex than traditional (e.g. command line) search, as the algorithms need to take into account what has happened so far in the interaction, need to include common knowledge, and knowledge about the specific user. A SDS should be able to alleviate some of this burden, and this is one of the aims of

⁶<http://www.ics.forth.gr/isl/Hippalus/>

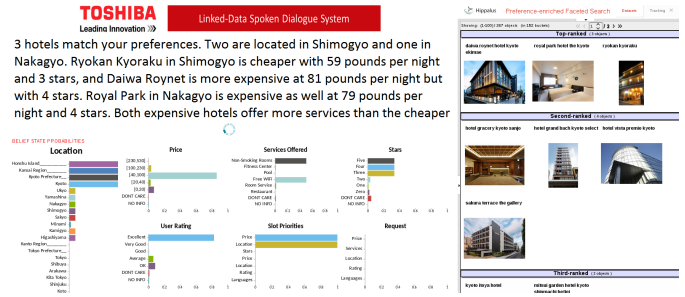


Figure 2: Screenshot of LD-SDS. Parts of the belief space on the left, Hippalus output on the right.

System: Hello, welcome to Toshiba’s Japanese Hotels spoken dialogue system. How may I help you?
User: I’m looking for a hotel in Kyoto but not in Minami where they offer free Wi-Fi and have non smoking rooms. (ASR confidence: 0.75154209) <i>Minami is a region in Kyoto.</i>
Constraints: [type=hotel, location=kyoto, location≠minami, amenities={free-wifi, non-smoking-rooms}]
Preferences: []
Hippalus: [267 items in 2 buckets; {60, 207} items in buckets]
S: What price range are you looking for?
U: I want something around 70 pounds and with more than two stars. (0.7065863)
C: [pricerange ~ 70, stars > 2]
P: []
H: [188 items in 45 buckets; {25, 103, 13, ...} items in buckets]
S: What kind of user rating would you prefer the hotel to have?
U: I’d like excellent ratings. (0.92533112)
C: []
P: [ratings=excellent : best]
H: [188 items in 27 buckets; {9, 11, 8, ...} items in buckets; identified slots that can affect ranking and reduce items in first bucket: {stars, location, price}]
S: Which of the following criteria are important for you? stars, price, or location?
U: I prefer location and price. (0.95948964)
C: []
P: [location >> all, price >> all]
H: [188 items in 152 buckets; 3 items in first bucket]
S: 3 hotels match your preferences. Two are located in Shimogyo and one in Nakagyo. Ryokan Kyoraku in Shimogyo is cheaper with 59 pounds per night and 3 stars, and Daiwa Roynet is more expensive at 81 pounds per night but with 4 stars. Royal Park in Nakagyo is expensive as well at 79 pounds per night and 4 stars. Both expensive hotels offer more services than the cheaper one.
U: Thank you, goodbye. (0.97125274)

Table 1: Example interaction between a human user and our LD-SDS prototype. In the interest of space, the notes under each dialogue turn briefly show items that correspond to new information. The belief state is updated accordingly (see Section 4). ASR: Automatic Speech Recognition.

our work. Moreover, the increased expressiveness necessary for information navigation is more natural in spoken rather

than written language, and this may be one way to get SDS more widely adopted and perhaps on par with command line search or apps. In this context, there are a number of challenges that need to be taken into account, all of which fall in our current and future work plans:

Language Understanding. In order to process input that reflects complex user intents (and may have different meanings for different users or context, etc.), besides the usual challenges, SLU needs to have access to multiple kinds of knowledge, across many domains. As the system is connected to live knowledge bases, it is important for SLU to be able to adapt over time, as well as handle out of domain input gracefully.

Belief Tracking. Translating the identified user intentions from SLU into a belief state is not trivial, even for slot filling models with one or two operators (e.g. $=, \neq$). As an initial approach to belief tracking under this expanded paradigm, we follow the simple principles outlined on Table 2 in conjunction with an existing belief tracker. While this is straightforward for regular slots, when we have slots with hierarchical values or multi-valued slots, we need a different kind of belief update. Specifically, for hierarchical slots we need to recursively perform the belief update, while still following the basic principles of Table 2. As the constraints become more complex, traversing the hierarchy of values becomes non-trivial. In our prototype, we traverse the hierarchy once for each constraint (relevant to a specific hierarchical slot) and then combine the updates into a single belief update by taking the average for each value. When updating multi-valued slots, we simply divide the probability mass across each value that was mentioned (and not negated), although this may not be optimal.

$< x, \leq x$	bias update towards values greater / equal to or greater than x
$> x, \geq x$	bias update towards values less / equal to or less than x
$= x, \neq x$	bias update towards values equal or not equal to x
around x , not around x	bias update towards values around or not around x - e.g. within one std. dev.
between $[x, y]$, not between $[x, y]$	bias update towards values between or outside of $[x, y]$

Table 2: Basic principles for our initial belief tracker that is based on the tracker provided by [30].

Dialogue Policy. Robust, scalable and multi-domain policy models are necessary, that take into account what has happened in the dialogue (e.g. belief state), what has happened in similar dialogues with other users (global dialogue history), and also what is currently happening in the real world (e.g. response from knowledge bases). As an initial approach, we plan to apply our Deep Q-Network (DQN) trained multi-domain policy network [21] on this extended paradigm. The core idea is that the policy model is trained on domain-independent features extracted for each slot at run-time and can thus operate even if the slots (or their values) change dynamically. Such features for example include slot entropy (w.r.t. its values), number of database matches if a slot has a specific value, distribution of values over the database, etc. A policy operating in such a feature space,

therefore, is independent of domain-specific slots and values.

Language Generation. As also noted by other researchers (e.g. [2] when analyzing human to human data), a very important capability of the LD-SDS is to be able to summarize the current state of the results and compare various items on various aspects. Other challenges include the ability to describe items information for which may reside in multiple live semantic knowledge bases, whose schemas may change over time. Natural Language Generation (NLG) in LD-SDS is currently done via templates that allow us to compare up to 3 items, on aspects mentioned by the user in the constraints and preferences. Our next step is to collect data and train models for NLG, following works such as [12, 14, 22, 31], and enrich Hippalus’ answer with relevant metrics.

Semantic Layer. The integration of available information is a hot and important topic. Although a big number of datasets has been published according to the principles of Linked Data, there is no evidence regarding the connectivity of the current LOD cloud or its quality. Aspects to consider include the complementarity, the discovery and selection of datasets, the integration, novelty and provenance of information, the evolution of datasets, the scalability and the efficiency of the approaches. Creating appropriate semantic views based on the user input, that offer advanced query and exploratory services by exploiting entity mining, disambiguation methods, available unstructured data sources and the history and profile of the user, is a challenging task.

Evaluation. As there is no clear way to define task success in this setting, traditional metrics like precision and recall need to be modified and appropriate baselines have to be carefully selected. General dialogue quality estimators may still be used to capture the conversational aspect of the interaction, but we need metrics to also measure the quality of information retrieved by the system (such as metrics for interactive or session-based information retrieval [28]).

5. CONCLUSIONS

In this paper we have motivated the need for more expressive SDS that can leverage the wealth of Linked Data (semantic knowledge bases in general) and state-of-the-art exploratory search services. Towards this direction we have presented our prototype LD-SDS which operates on an expanded dialogue paradigm (compared to paradigms in the literature) and connects to live semantic knowledge bases. Specifically, it adopts and exploits a more expressive data model (multi-valued attributes, hierarchically organized values), supports more complex questions/commands (less than, around), supports a session-based interaction model appropriate for (gradual) decision making that supports preferences (i.e. soft constraints for ranking the available options) that exploits the expressiveness of the data model (preference inheritance is supported, scope-based resolution of conflicts), takes initiatives (based on the focus and the distribution of values) for deciding what “clarification questions” it should ask the user. Last, we have demonstrated the feasibility of this approach for hotel selection.

There are several directions that are worth further research. One is to investigate how to exploit global scale semantic indexes, like those proposed in [16]. Another is training statistical models (deep learning and others) for the various parts of the SDS outlined in Section 4; we are currently working on this, starting with data collection and designing and training an appropriate simulated user.

6. REFERENCES

- [1] A. Aghaebrahimian and F. Jurčiček. Constraint-based open-domain question answering using knowledge graph search. In *Proceedings of the 26th International Conference on Computational Linguistics*, 2016.
- [2] L. E. Asri, H. Schulz, S. Sharma, J. Zumer, J. Harris, E. Fine, R. Mehrotra, and K. Suleman. Frames: A corpus for adding memory to goal-oriented dialogue systems. *18th SIGDial*, 2017.
- [3] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pages 205–227, 2009.
- [4] D. Bohus, A. Raux, T. K. Harris, M. Eskenazi, and A. I. Rudnicky. Olympus: an open-source framework for conversational spoken language interface research. In *Proceedings of the workshop on bridging the gap: Academic and industrial research in dialog technologies*, pages 32–39. ACL, 2007.
- [5] A. Broder. A Taxonomy of Web Search. *SIGIR Forum*, 36(2):3–10, 2002.
- [6] H. Cuayáhuitl, S. Yu, A. Williamson, and J. Carse. Scaling up deep reinforcement learning for multi-domain dialogue systems. In *2017 International Joint Conference on Neural Networks*, pages 3339–3346, May 2017.
- [7] I. Ermilov, J. Lehmann, M. Martin, and S. Auer. Lodstats: The data web census dataset. In *International Semantic Web Conference*, pages 38–46. Springer, 2016.
- [8] P. Fafalios, M. Baritakis, and Y. Tzitzikas. Exploiting linked data for open and configurable named entity extraction. *International Journal on Artificial Intelligence Tools*, 24(2):1540012, 2015.
- [9] P. Fafalios, P. Papadakos, and Y. Tzitzikas. Enriching textual search results at query time using entity mining, linked data and link analysis. *International Journal of Semantic Computing*, 8(04):515–544, 2014.
- [10] P. Fafalios and Y. Tzitzikas. X-ENS: Semantic Enrichment of Web Search Results at Real-Time. In *SIGIR’13*, pages 1089–1090, Dublin, Ireland, 2013.
- [11] M. Gašić, N. Mrkšić, L. M. Rojas-Barahona, P.-H. Su, S. Ultes, D. Vandyke, T.-H. Wen, and S. Young. Dialogue manager domain adaptation using gaussian process reinforcement learning. *Computer Speech & Language*, 45:552 – 569, 2017.
- [12] D. Gkatzia, O. Lemon, and V. Rieser. Natural language generation enhances human decision-making with uncertain information. In *The 54th Annual Meeting of the Association for Computational Linguistics*, page 264, 2016.
- [13] K. Janowicz, P. Hitzler, B. Adams, D. Kolas, I. Vardeman, et al. Five stars of linked data vocabulary use. *Semantic Web*, 5(3):173–176, 2014.
- [14] R. Lebrecht, D. Grangier, and M. Auli. Neural text generation from structured data with application to the biography domain. *arXiv:1603.07771*, 2016.
- [15] G. Marchionini. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, 2006.
- [16] M. Mountantonakis, N. Minadakis, Y. Marketakis, P. Fafalios, and Y. Tzitzikas. Quantifying the connectivity of a semantic warehouse and understanding its evolution over time (accepted for publication). *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2016.
- [17] P. Papadakos, N. Armenatzoglou, S. Kopidaki, and Y. Tzitzikas. On exploiting static and dynamically mined metadata for exploratory web searching. *Knowledge and information systems*, 30(3):493–525, 2012.
- [18] P. Papadakos and Y. Tzitzikas. Hippalus: Preference-enriched faceted exploration. In *EDBT/ICDT Workshops*, volume 172, 2014.
- [19] I. Papaioannou and O. Lemon. Combining chat and task-based multimodal dialogue for more engaging HRI: A scalable method using reinforcement learning. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 365–366, March 2017.
- [20] A. Papangelis and Y. Stylianou. Multi-domain spoken dialogue systems using domain-independent parameterisation. In *Domain Adaptation for Dialogue Agents*, 2016.
- [21] A. Papangelis and Y. Stylianou. Single-model multi-domain dialogue management with deep learning. In *International Workshop for Spoken Dialogue Systems*, 2017.
- [22] O. Press, A. Bar, B. Bogin, J. Berant, and L. Wolf. Language generation with recurrent generative adversarial networks without pre-training. *arXiv preprint arXiv:1706.01399*, 2017.
- [23] B. Qarabaqi and M. Riedewald. User-driven refinement of imprecise queries. In *IEEE 30th International Conference on Data Engineering, Chicago, ICDE 2014, IL, USA, March 31 - April 4, 2014*, pages 916–927, 2014.
- [24] H. Schulz, J. Zumer, L. E. Asri, and S. Sharma. A frame tracking model for memory-enhanced dialogue systems. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 219 – 227, August 2017.
- [25] S. Stoyanchev, P. Lison, and S. Bangalore. Rapid prototyping of form-driven dialogue systems using an open-source framework. In *Proceedings of the 17th SIGDial*, pages 216–219, September 2016.
- [26] M. Surdeanu and H. Ji. Overview of the english slot filling track at the tac2014 knowledge base population evaluation. In *Proc. Text Analysis Conference*, 2014.
- [27] Y. Tzitzikas, C. Allocca, C. Bekiari, Y. Marketakis, P. Fafalios, M. Doerr, N. Minadakis, T. Patkos, and L. Candela. Unifying heterogeneous and distributed information about marine species through the top level ontology marinetto. *Program*, 50(1):16–40, 2016.
- [28] Y. Tzitzikas, N. Manolis, and P. Papadakos. Faceted exploration of rdf/s datasets: A survey. *J. Intell. Inf. Syst.*, 48(2):329–364, Apr. 2017.
- [29] Y. Tzitzikas and P. Papadakos. Interactive exploration of multi-dimensional and hierarchical information spaces with real-time preference elicitation. *Fundamenta Informaticae*, 122(4):357–399, 2013.
- [30] S. Ultes, L. M. Rojas Barahona, P.-H. Su, D. Vandyke, D. Kim, I. n. Casanueva, P. Budzianowski, N. Mrkšić, T.-H. Wen, M. Gasic, and S. Young. PyDial: A Multi-domain Statistical Dialogue System Toolkit. In *Proceedings of ACL 2017, System Demonstrations*, pages 73–78, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [31] T.-H. Wen, M. Gašić, N. Mrkšić, L. M. Rojas-Barahona, P.-H. Su, D. Vandyke, and S. Young. Multi-domain neural network language generation for spoken dialogue systems. In *Proceedings of NAACL-HLT*, pages 120–129, 2016.
- [32] T.-H. Wen, D. Vandyke, N. Mrkšić, M. Gasic, L. M. Rojas-Barahona, P.-H. Su, S. Ultes, and S. Young. A network-based end-to-end trainable task-oriented dialogue system. In *Proceedings of the European Chapter of the Association for Computational Linguistics*, pages 438–449, April 2007.
- [33] B. M. Wildemuth and L. Freund. Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of Human-Computer Interaction and Information Retrieval*, page 4. ACM, 2012.
- [34] D. J. Williams, K. Asadi, and G. Zweig. Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 665–677, August 2017.
- [35] X. Yang, Y. N. Chen, D. Hakkani-Tür, P. Crook, X. Li, J. Gao, and L. Deng. End-to-end joint learning of natural language understanding and dialogue manager. In *2017 IEEE ICASSP*, pages 5690–5694, March 2017.