VOnDA: A Framework for Ontology-Based Dialogue Management

Bernd Kiefer and Anna Welker and Christophe Biwer

Abstract We present VOnDA, a framework to implement the dialogue management functionality in dialogue systems. Although domain-independent, VOnDA is tailored towards dialogue systems with a focus on social communication, which implies the need of a long-term memory and high user adaptivity. For these systems, which are used in health environments or elderly care, margin of error is very low and control over the dialogue process is of topmost importance. The same holds for commercial applications, where customer trust is at risk. VOnDA’s specification and memory layer relies upon (extended) RDF/OWL\(^1\), which provides a universal and uniform representation, and facilitates interoperability with external data sources, e.g., from physical sensors.

1 Introduction

Natural language dialogue systems are becoming more and more popular, be it as virtual assistants such as Siri or Cortana, as Chatbots on websites providing customer support, or as interface in human-robot interactions in areas ranging from human-robot teams in industrial environments [17] over social human-robot-interaction [1] to disaster response [12].

A central component of most systems is the dialogue manager, which controls the (possibly multimodal) reactions based on external triggers and the current internal state. When building dialogue components for robotic applications or in-car assistants, the system needs to take into account inputs in various forms, first and foremost the user utterances, but also other sensor input that may influence the dialogue, such as information from computer vision, gaze detection, or even body and environment sensors for cognitive load estimation.

In the following, we will describe VOnDA, an open-source framework initially developed to implement dialogue strategies for conversational robotic and virtually embodied agents. The implementation mainly took place in the context of the ALIZ-E and PAL projects, where a social robotic assistant supports diabetic children managing their disease. This application domain dictates some requirements that led to the decision to go for a rule-based system with statistical selection and RDF/OWL underpinning.

Firstly, it requires a lot of control over the decision process, since mistakes by the system are only tolerable in very specific situations, or not at all. Secondly, it is vital to be able to maintain a relationship with the user over a longer time period. This requires a long-term memory which can be efficiently accessed by the dialogue system to exhibit familiarity with the user in various forms, e.g., respecting personal preferences, but also making use of knowledge about conversations or events that were part of interactions in past sessions. For the same reason, the system needs high adaptability to the current user, which means adding a significant number of variables to the state space. This often poses a scalability problem for POMDP-based approaches, both in terms of run-time performance, and of probability estimation, where marginal cases can be dominated by the prominent

\(^1\) Resource Description Framework https://www.w3.org/RDF/
Web Ontology Language https://www.w3.org/OWL/
situation. A third requirement for robotic systems is the ability to process streaming sensor data, or at least use aggregated high-level information from this data in the conversational system.

Furthermore, data collection for user groups in the health care domain is for ethical reasons even more challenging than usual, and OWL reasoning offers a very flexible way to access control.

VOnDA therefore specifically targets the following design goals to support the system requirements described before:

- Flexible and uniform specification of dialogue semantics, knowledge and data structures
- Scalable, efficient, and easily accessible storage of interaction history and other data, like real-time sensor data, resulting in a large information state
- Readable and compact rule specifications, facilitating access to the underlying RDF database, with the full power of a programming language
- Transparent access to standard programming language constructs (Java classes) for simple integration with the host system

VOnDA is not so much a complete dialogue management system as rather a fundamental implementation layer for creating complex reactive systems, being able to emulate almost all traditional rule- or automata-based frameworks. It provides a strong and tight connection to a reasoning engine and storage, which makes it possible to explore various research directions in the future.

In the next section, we review related work that was done on dialogue frameworks. In section 3, we will give a high-level overview of the VOnDA framework, followed by a specification language synopsis. Section 5 covers some aspects of the system implementation. Section 6 describes the application of the framework in the PAL project’s integrated system. The paper concludes with a discussion of the work done, and further directions for research and development.

2 Related Work

The existing frameworks to implement dialogue management components roughly fall into two large groups, those that use symbolic information or automata to specify the dialogue flow (IrisTK [18], RavenClaw [3], Visual SceneMaker [7]), and those that mostly use statistical methods (PyDial [20], Alex [8]). Somewhat in between these is OpenDial [13], which builds on probabilistic rules and a Bayesian Network.

For reasons described in the introduction, VOnDA currently makes only limited use of statistical information. A meaningful comparison to purely learned systems like PyDial or Alex therefore becomes more complex, and would have to be done on an extrinsic basis, which we cannot yet provide. We studied comparable systems focusing mainly on two aspects: the specification of behaviours, and the implementation of the dialogue memory / information state.

The dialogue behaviours in IrisTK and SceneMaker are specified using state charts (hierarchical automata). Additional mechanisms (parallel execution, history keeping, exception mechanisms like interruptive edges) make them more flexible and powerful than basic state charts, but their flexibility and generalisation capabilities are limited.

RavenClaw [3] uses so-called task trees, a variant of flow charts that can be dynamically changed during run-time to implement dialogue agents for different situations in the dialogue, and an agenda, which selects the appropriate agent for the current dialogue state. The resemblance to agent-based architectures using preconstructed plans is striking, but the improved flexibility also comes at the cost of increased complexity during implementation and debugging.

OpenDial [13] tries to combine the advantages of hand-crafted systems with statistical selection, using probabilistic rules which can be viewed as templates for probabilistic graphical models. The parameters for the models can be estimated using previously collected data (supervised learning), or during the interactions with reinforcement learning techniques. Being able to specify structural
knowledge for the statistical selection reduces the estimation problem if only a small amount of data is available, and allows to explicitly put restrictions on the selection process.

3 High-Level System Description

VOnDA follows the Information State / Update paradigm [19]. The information state represents everything the dialogue agent knows about the current situation, possibly containing information about dialogue history, the belief states of the participants, situation data, etc., depending on the concrete system. Any change in the information state will trigger a reasoning mechanism of some sort, which may result in more changes in the information state, or outputs to the user or other system components.

VOnDA implements this paradigm by combining a rule-based approach with statistical selection, although in a different way than OpenDial. The rule specifications are close to if-then statements in programming languages, and the information state is realised by an RDF store and reasoner with special capabilities (HFC [10]), namely the possibility to directly use $n$-tuples instead of triples. This allows to attach temporal information to every data chunk [9, 11]. In this way, the RDF store can represent dynamic objects, using either transaction time or valid time attachments, and as a side effect obtain a complete history of all changes. HFC is very efficient in terms of processing speed and memory footprint, and has recently been extended with stream reasoning facilities. VOnDA can use HFC either directly as a library, or as a remote server, also allowing for more than one database instance, if needed. The initial motivation for using an RDF reasoner was our research interest in multi-session, long-term interactions. In addition, this also allows processing incoming facts in different layers. Firstly, there is the layer of custom reasoning rules, which also comprises streaming reasoning, e.g., for real-time sensor data, and secondly the reactive rule specifications, used mainly for agent-like functionality that handles the behavioural part. This opens new research directions, e.g., underpinning the rule conditions with a probabilistic reasoner.

The RDF store contains the terminological and the dynamic knowledge: specifications for the data types and their properties, as well as a hierarchy of dialogue acts, semantic frames and their arguments, and the data objects, which are instantiations of the data types. The data type specifications are also used by the compiler to infer the types for property values (see section 4), and form a declarative API to connect new components, e.g., for sensor or application data.

We are currently using the DIT++ dialogue act hierarchy [4] and shallow frame semantics along the lines of FrameNet [16] to interface with the natural language understanding and generation units.
Our dialogue act object currently consist of a dialogue act token, a frame and a list of key-value pairs as arguments to the frame (Offer(Transporting, what=tool, to=workbench)). While this form of shallow semantics is enough for most applications, we already experience its shortcomings when trying to handle, for example, social talk. Since the underlying run-time core is already working with full-fledged feature matrices, only a small syntax extension will be needed to allow for nested structures.

A set of reactive condition-action rules (see figure 4) is executed whenever there is a change in the information state. These changes are caused by incoming sensor or application data, intents from the speech recognition, or expired timers. Rules are labelled if-then-else statements, with complex conditions and shortcut logic, as in Java or C. The compiler analyses the base terms and stores their values during processing for dynamic logging. A rule can have direct effects, like changing the information state or executing system calls. Furthermore, it can generate so-called proposals, which are (labelled) blocks of code in a frozen state that will not be immediately executed, similar to closures.

All rules are repeatedly applied until a fixed point is reached where no new proposals are generated and there is no information state change in the last iteration. Subsequently, the set of proposals is evaluated by a statistical component, which will select the best alternative. This component can be exchanged to make it as simple or elaborate as necessary, taking into account arbitrary features from the data storage.

A VOnDA project consists of an ontology, a custom extension of the abstract Agent class (the so-called wrapper class), a client interface to connect the communication channels of the application to the agent, and a set of rule files that are arranged in a tree, using import statements. The blue core in Figure 2 is the run-time system which is part of the VOnDA framework, while all elements above are application specific parts of the agent. A Yaml project file contains all necessary information for compilation: the ontology, the wrapper class, the top-level rule file and other parameters, such as custom compile commands.

The ontology contains the definitions of dialogue acts, semantic frames, class and property specifications for the data objects of the application, and other assertional knowledge, such as specifications for “forgetting”, which could be modeled in an orthogonal class hierarchy and supported by custom deletion rules in the reasoner.

Every rule file can define variables and functions in VOnDA syntax which are then available to all imported files. The methods from the wrapper class are available to all rule files.

The current structure assumes that most of the Java functionality that is used inside the rule files will be provided by the Agent superclass. There are, however, alternative ways to use other Java classes directly, with support for the same type inference as for RDF classes.

Fig. 2 A schematic VOnDA agent
4 Dialogue Specification Language

VOnDA’s rule language at first sight looks very similar to Java/C++. However, there are a number of specific features which make it convenient for the implementation of dialogue strategies. Maybe the most important one is the handling of RDF objects and classes, which can be treated similarly to those of object oriented programming languages, including the (multiple) inheritance and type inference that are provided by the RDF class hierarchies.

```java
1  user = new Animate;
2  user.name = "Joe";
3  set_age:
4      if (user.age <= 0) {
5          user.age = 15;
6      }
```

Figure 3 contains an example of VOnDA code, and how it relates to RDF type and property specifications, schematically drawn on the right. The domain and range definitions of properties are picked up by the compiler and used in various places, e.g., to infer types, do automatic code or data conversions, or create ‘intelligent’ boolean tests, such as the one in line 4, which will expand into two tests, one testing for the existence of the property for the object, and in case that succeeds, a test if the value is greater than zero. If there is a chain of more than one field resp. property access, every part is tested for existence in the target code, keeping the source code as concise as possible. Also, for reasons of brevity, the type of a new variable needs not be given if it can be inferred from the value assigned to it.

New RDF objects can be created with `new`, similar to Java objects; they are immediately reflected in the database, as are all changes to already existing objects.

Many operators are overloaded, especially boolean operators such as `<=`, which compares numeric values, but can also be used to test if an object is of a specific class, for subclass tests between two classes, and for subsumption of dialogue acts.

```java
1  if (!saidInSession(#Greeting(Meeting)) {
2      timeout("wait_for_greeting", 7000){ //Wait 7 secs before taking initiative
3          if (! receivedInSession(#Greeting(Meeting))
4            propose("greet") {
5                da = #InitialGreeting(Meeting);
6                if (user.name) da.name = user.name;
7                emitDA(da);
8            }
9        }
10    }
11
12    if (receivedInSession(#Greeting(Meeting))
13      propose("greet_back") { // We assume we know the name by now
14        emitDA(#ReturnGreeting(Meeting, name={user.name}));
15    }
16 }
```

Fig. 4 VOnDA code example

There are two statements with a special syntax and semantics: `propose` and `timeout`. `propose` is VOnDA’s current way of implementing probabilistic selection. All (unique) propose blocks that are in active rule actions are collected, frozen in the execution state in which they were encountered, such
as closures known from functional programming languages. When rule processing stops, a statistical component picks the “best” proposal and its closure is executed.

Timeouts also generate closures, but with a different purpose. They can be used to trigger proactive behaviour, or to check the state of the system after some time period, or in regular intervals. A timeout will only be created if there is no active timeout with that name.

Figure 4 also contains an example of the short-hand notation for shallow semantic structures (starting with #). Since they predominantly contain constant (string) literals, this is the default when specifying such structures. The special syntax in user={user.name} allows to insert the value of expressions into the literal, similar to an eval.

This section only described the most important features of VOnDA’s syntax. For a detailed description, the reader is referred to the user documentation.

5 Compiler / Run-Time Library

The compiler turns the VOnDA source code into Java source code using the information in the ontology. Every source file becomes a Java class. Although the generated code is not primarily for the human reader, a lot of care has been taken in making it still understandable and debuggable. The compile process is separated into three stages: parsing and abstract syntax tree building, type checking and inference, and code generation.

The VOnDA compiler’s internal knowledge about the program structure and the RDF hierarchy takes care of transforming the RDF field accesses into reads from and writes to the database. Beyond that, the type system, resolving the exact Java, RDF or RDF collection type of (arbitrary long) field accesses, automatically performs the necessary casts for the ontology accesses.

The run-time library contains the basic functionality for handling the rule processing, including the proposals and timeouts, and for the on-line inspection of the rule evaluation. There is, however, no blueprint for the main event loop, since that depends heavily on the host application. It also contains methods for the creation and modification of shallow semantic structures, and especially for searching the interaction history for specific utterances. Most of this functionality is available through the abstract Agent class, which has to be extended to a concrete class for each application.

There is functionality to directly communicate with the HFC database using queries, in case the object view is not sufficient or too awkward. The natural language understanding and generation components can be exchanged by implementing existing interfaces, and the statistical component is connected by a message exchange protocol. A basic natural language generation engine based on a graph rewriting module is already integrated, and is used in our current system as a template based generator. The example application also contains a VoiceXML based interpretation module.

Debugger / GUI

VOnDA comes with a GUI [2] that helps navigating, compiling and editing the source files belonging to a project. It uses the project file to collect all the necessary information.

Upon opening a project, the GUI displays the project directory (in a file view). The user can edit rule files from within the GUI or with an external editor like Emacs, Vim, etc. and can start the compilation process. After successful compilation, the project view shows what files are currently used, and marks the top-level and the wrapper class files. A second tree view (rule view) shows the rule structure in addition to the module structure. Modules in which errors or warnings were reported during compilation are highlighted, and the user can quickly navigate to them using context menus.

Additionally, the GUI can be used to track what is happening in a running system. The connection is established using a socket to allow remote debugging. In the rule view, multi-state check boxes are

---

2 https://github.com/bkiefer/vonda/blob/master/doc/master.pdf
used to define which rules should be observed under which conditions. A rule can be set to be logged under any circumstances, not at all or if its condition evaluated to true or to false. Since the rules are represented in a tree-like structure, the logging condition can also be set for an entire subgroup of rules, or for a whole module. The current rule logging configuration can be saved for later use.

The logging view displays incoming logging information as a sortable table. A table entry contains a time stamp, the rule’s label and its condition. The rule’s label is coloured according to the final result of the whole boolean expression. Each base term of the condition is coloured accordingly, or greyed out if short-cut logic led to premature failure or success of the expression. Inspecting the live system helps pin-point problems when the behaviour is not as expected. The log shows how the currently active part of the information state is processed, and the window offers easy navigation using the mouse from the rule condition to the corresponding source code.

6 Applications

VOnDA is used in the integrated system of the EU project PAL [15], which uses human-robot interaction to support children with diabetes type 1 in coping with their disease. Children interact with a real NAO robot, or with an Android app that connects to the core system and exhibits a virtual character that is as similar to the robot as possible, also in its behaviour.

The dialogue component, which is largely responsible for the agent’s behaviour, is implemented using the VOnDA framework. In addition, HFC, the RDF store that VOnDA builds upon, is the main database of the system, storing all relevant information and being the central data exchange hub. The system runs as a cloud-based robotic solution, spawning a new system instance for every user. It has been successfully tested with more than 40 users at a time on a medium sized virtual machine with only moderate load factors, giving a positive indication of the scalability of HFC and the VOnDA approach.

There are two helper modules integrated into the dialogue component which quite extensively exploit the connection between the database and the rule part, namely the Episodic Memory and the Targeted Feedback. While the targeted feedback reacts to current events in the running session, like entering a bad or good glucose value, or the current achievement of a task, the episodic memory aggregates data from the past and eventually converts them into so-called episodes that are used for interactions in subsequent sessions. Both are only triggered if relevant changes in the database occur,

---

3 Softbank Robotics [https://www.ald.softbankrobotics.com](https://www.ald.softbankrobotics.com)
4 4 core Xeon E5-2683@2.00GHz, 16 GB RAM
VOnDA has also been used in a recent project aiming to implement a generalised, ontology-based approach to open-domain talk [21]. The Smoto system uses an additional HFC server running WordNet [14, 6] as semantic database, thereby gaining knowledge about semantic concepts that can be used in the dialogue and to find appropriate reactions on arbitrary user input.

7 Discussion and Further Work

We believe that there are still many interesting application areas for hybrid statistical and handcrafted systems, e.g., if they are relatively small, or there is little domain-specific data available. Many currently deployed systems that build on much simpler technology like VoiceXML can certainly profit from hybrid approaches such as OpenDial or VOnDA.

VOnDA is under active development. We designed it such that it can be integrated in most applications and opens many ways for improvements and additions. As a rule-based framework that is close to being a programming language, VOnDA is able to completely emulate the automata-based frameworks. In fact, we are currently working on a graphical editor à la SceneMaker and the precompilation of hierarchical state charts into VOnDA code. We hope this will facilitate the implementation of new applications for inexperienced users and help with rapid prototyping, while retaining the greater flexibility and modularization capabilities. In this way, we combine the intuitive way of specifying simple strategies with the full flexibility of the framework.

VOnDA could also be used to implement modules that simulate the agents of RavenClaw. To get a functionality similar to RavenClaw’s agenda, its action selection module would have to be implemented as a dialogue state tracker, activating the most probable agent at each dialogue step.

Using the well-established RDF/OWL standard as specification layer makes it very easy to add or change application specific data structures, especially because of the existing tool support. We already use the reasoning facilities for type and partially for temporal inference, but given the possibility of attaching also confidence or credibility information to the RDF data, a more integrated probabilistic approach with soft preconditions could be implemented, e.g., on the basis of Dempster-Shafer theory [5]. Moreover, additional meta knowledge, such as trustworthiness or validity periods could be declared using multiple inheritance, which opens many interesting research directions.

Other next steps will be the addition of default adaptors for obviously needed external modules like automatic speech recognition, more flexible language understanding, and the like. We will also work on the improvement of the GUI, including features such as a watch window and/or a timeline to track changes of specific values in the database, and a tool that analyses the dependencies between rules on the basis of the conditions’ base terms.

From the research perspective, there are two very interesting lanes: integrating probabilistic reasoning as a first-class option, which is directly integrated with the rule conditions, and adding an additional layer to facilitate the implementation of BDI-like agents, to study the connections and dependencies between conversational and non-conversational behaviours.

Source Code and Documentation

The VOnDA core system can be downloaded at git@github.com:bkiefer/vonda.git. The main page has detailed instructions for the installation of external dependencies. The debugger currently lives in a separate project: git@github.com:yoshegg/rudibugger.git. Both projects are licensed under the Creative Commons Attribution-NonCommercial 4.0 International License, and are free for all non-
commercial use. A screen cast showing the GUI functionality and the running PAL system is available at https://youtu.be/nSotEVZUEyw.

Acknowledgements

The research described in this paper has been funded by the Horizon 2020 Framework Programme of the European Union within the project PAL (Personal Assistant for healthy Lifestyle) under Grant agreement no. 643783.

This paper is dedicated to our colleague and friend Hans-Ulrich Krieger, the creator of HFC. Hope you found peace, wherever you are.

References

1. ALIZ-E project. http://www.aliz-e.org/ (2010). URL http://www.aliz-e.org/
2. Biwer, C.: rudibugger - Graphisches Debugging der Dialogmanagementtechnologie VOnDA. Bachelor’s Thesis, Saarland University (2017). DOI 10.13140/RG.2.2.36556.31368
3. Bohus, D., Rudnicky, A.I.: The RavenClaw dialog management framework: Architecture and systems. Computer Speech & Language 23(3), 332–361 (2009)
4. Bunt, H., Alexandersson, J., Choe, J.W., Fang, A.C., Hasida, K., Petukhova, V., Popescu-Belis, A., Traum, D.R.: ISO 24617-2: A semantically-based standard for dialogue annotation. In: LREC, pp. 430–437. Citeseer (2012)
5. Dempster, A.P.: The Dempster-Shafer calculus for statisticians. International Journal of approximate reasoning 48(2), 365–377 (2008)
6. Fellbaum, C.: WordNet. Wiley Online Library (1998)
7. Gebhard, P., Mehlmann, G., Kipp, M.: Visual SceneMaker—a tool for authoring interactive virtual characters. Journal on Multimodal User Interfaces 6(1-2), 3–11 (2012)
8. Jurčiček, F., Dušek, O., Platek, O., Žilka, L.: Alex: A statistical dialogue systems framework. In: International Conference on Text, Speech, and Dialogue, pp. 587–594. Springer (2014)
9. Krieger, H.U.: A temporal extension of the Hayes/ter Horst entailment rules and an alternative to W3C’s n-ary relations. In: Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS), pp. 323–336 (2012)
10. Krieger, H.U.: An efficient implementation of equivalence relations in OWL via rule and query rewriting. In: Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on, pp. 260–263. IEEE (2013)
11. Krieger, H.U.: A detailed comparison of seven approaches for the annotation of time-dependent factual knowledge in rdf and owl. In: Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation (2014)
12. Kruiff-Korbayová, I., Colas, F., Gianni, M., Pirri, F., de Greeff, J., Hindriks, K., Neerincx, M., Ögren, P., Svoboda, T., Worst, R.: TRADR project: Long-term human-robot teaming for robot assisted disaster response. KI-Künstliche Intelligenz 29(2), 193–201 (2015)
13. Lison, P., Kennington, C.: Developing spoken dialogue systems with the OpenDial toolkit. SEMDIAL 2015 goDIAL p. 194 (2015)
14. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM 38(11), 39–41 (1995)
15. PAL: The PAL Project Website. http://www.pal4u.eu/ (2018). Last access: 11.12.2018
16. Ruppenhofer, J., Ellsworth, M., Petruck, M.R., Johnson, C.R., Scheffczyk, J.: FrameNet II: Extended theory and practice. Institut für Deutsche Sprache, Bibliothek (2016)
17. Schwartz, T., Zinnikus, I., Krieger, H.U., Bürckert, C., Folz, J., Kiefer, B., Hevesi, P., Lüth, C., Pirkl, G., Spieldenner, T., et al.: Hybrid teams: flexible collaboration between humans, robots and virtual agents. In: German Conference on Multiagent System Technologies, pp. 131–146. Springer (2016)
18. Skantze, G., Al Moubayed, S.: IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In: Proceedings of the 14th ACM international conference on Multimodal interaction, pp. 69–76. ACM (2012)
19. Traum, D.R., Larsson, S.: The information state approach to dialogue management. In: Current and new directions in discourse and dialogue, pp. 325–353. Springer (2003)
20. Ultes, S., Barahona, L.M.R., Su, P.H., Vandyke, D., Kim, D., Casanueva, I., Budzianowski, P., Mrkšić, N., Wen, T.H., Gašić, M., et al.: Pydial: A multi-domain statistical dialogue system toolkit. Proceedings of ACL 2017, System Demonstrations pp. 73–78 (2017)
21. Welker, A.: Wissensbasiertes Dialogmanagement für Social Talk. Bachelor’s Thesis, Saarland University (2017). DOI 10.13140/RG.2.2.16423.65441