Towards the Automatic Generation of Pedagogical Conversational Agents from Lecture Slides

Matthias Wölfel(✉)

Faculty of Computer Science and Business Information Systems, Karlsruhe University of Applied Sciences, Karlsruhe, Germany
matthias.woelfel@hs-karlsruhe.de

Abstract. Although corresponding technological and didactical models have been known for decades, the digitization of teaching has hardly advanced beyond simple non-interactive formats (e.g. downloadable slides are provided within a learning management system). The COVID-19 crisis is changing this situation dramatically, creating a high demand for highly interactive formats and fostering exchange between conversation partners about the course content. Systems are required that are able to communicate with students verbally, to answer their questions, and to check the students' knowledge. While technological advances have made such systems possible in principle, the game stopper is the large amount of manual work and knowledge that must be put into designing such a system and feeding it the right content.

In this publication, we present a first system to overcome the aforementioned drawback by automatically generating a corresponding dialog system from slide-based presentations, such as PowerPoint, OpenOffice, or Keynote, which can be dynamically adapted to the respective students and their needs. Our first experiments confirm the proof of concept and reveal that such a system can be very handy for both respective groups, learners and lecturers, alike. The limitations of the developed system, however, also reminds us that many challenges need to be addressed to improve the feasibility and quality of such systems, in particular in the understanding of semantic knowledge.

Keywords: Intelligent dialog systems · Voice assistants · Ontology-based information extraction · Natural language processing · Learning analytics

1 Introduction

New challenges due to the COVID-19 crisis as well as the increasing popularity of remote access to education by distance learning courses result in the abandonment of classroom teaching. Online learning is convenient, flexible, (possibly) cost-effective, tailorable to specific needs, immediate, and unrestricted. However, the shift into virtual space leads to an imbalance in communication and exchange between learners and lectures as well as between learners. Moving teaching from—real word—lecture halls to

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 216–229, 2021.
https://doi.org/10.1007/978-3-030-82565-2_18

- **video lectures** offer *no real-time return channel of any kind and is, in general, a solo act (one screen per user)*
- **webinars** cause *hardly any interaction* between learners and the lecturer—and if there is any it is quite artificial due to time delays or modality change (lecturer speaks, learners write)
- **learning platforms** are impersonal, lack human-like communication, and offer mostly *close-ended questions* (e.g. dichotomous, multiple-choice, rank order)

Because, in many cases and for various reasons, allies for discussing relevant teaching content are not available alternative solutions are required to jump in. Handcrafting such systems, however, even though in those cases where templates are already available, requires significant time and effort. Thereby the deployment of pedagogical conversational agents in educational context and training settings remains limited even though the technology is readily available and proven to be beneficial. For instance Rus et al. found that intelligent tutoring systems are more effective at improving learning outcomes if they use natural language interaction in comparison to those systems that do not incorporate natural language [23].

Many approaches in technologically supported teaching fail because it requires additional efforts in *time* (content needs to be prepared or revised), *skills* (new tools need to be learned), and *didactics* (other media requires different approaches) for the already busy lecturers. To overcome this drawback Winkler and Söllner proposed a method that helps educators to create smart personal assistants for their learning environments [28]. Their goal is to empower educators to develop their own agents without deep technological knowledge. While the former arguments about the high entry barrier hold for all kinds of technological teaching support, this barrier is particularly high for dialog-based approaches as the required skills to design and develop such systems are quite demanding: to hold a conversation with learners of any kind in natural language, to understand and answer questions, to simulate the dialogue moves of human tutors, and to follow ideal pedagogical strategies. The few systems using dialog-based approaches to support teaching are created manually rely on techniques within the field of *artificial intelligence* (AI) and *natural language processing* (NLP). Those "agents can guide the learner on what to do next, deliver didactic instruction, hold collaborative conversations, and model ideal behavior, strategies, reflections, and social interactions" [12]. Those systems have been proven useful to augment common educational practice, which is often limited to classical lecture formats and the provision of documents, by an interactive component in natural language. In this way, learners can enter into a dialogue with the language assistant at any time, who answers questions in real-time, provides references to further sources, and rate given answers. To make such systems accessible for a broader audience we propose to *replace the time-consuming process of creating the dialog content and structure manually with an automated process relying on semantically prepared lecture slides* as a content provider. The effort for the lecturer, therefore, can be reduced to a minimum.

2 Literatur Review

AI-based systems have a rather long history of providing additional services to students. While some services augment the lecture by providing transcriptions [10, 29] or translations [18] of what the lecturer has said, other services analyze learning behavior, learning activities, and attitudes during the learning process. Ideally, this can lead to a guided learning process that can adapt, in real-time, to the particular skills and requirements of each individual learner [11]. Apart from the limited amount of empirical evidence regarding the effectiveness of these AI-based learning-teaching systems, limited autonomy and little interactivity of learners with the available systems can be observed [25]. To improve natural interactions within those tools, conversational assistants and task-oriented dialog systems have been introduced [34]. Today, conversational assistants, in the field of education, enable learners to access data and services and to exchange information by simulating human-like conversations in the form of a natural language dialogue about a specific topic [9].

The use of natural language user interfaces for learning systems seems to be particularly promising as the field of NLP (which includes the methods of speech recognition, speech synthesis, and natural language understanding) has developed rapidly in recent years and is becoming an established interface for non-pedagogical use cases within the target group. According to a recent survey by Forsa Politikund Sozialforschung GmbH, this is over 95% in the target group in Germany [24], 22% of all German citizens used voice assistants in 2019. In the US more than 33% of the total population use voice assistants and 20% of the population use smart speakers.

Examples in higher education where voice assistants are used include Amazon Alexa with frequently asked questions about a course, room booking systems, and campus signposts [1]. Also in practical tests at our university, simple chat offers, with which e.g. free workrooms, timetable contents, or the daily canteen offer can be determined, have proven to be an excellent information channel with high demand by students. While the given examples so far demonstrate the use of voice assistants to manage to get around on campus and to handle administrative tasks we now turn to conversational assistants that support coursework. A literature review on pedagogical conversational agents is given in [14]: In their descriptive analysis, the distribution of publications per year shows an increasing interest in recent years, and ‘messenger-like’ conversational agents are published more than twice as often as ‘embodied’ conversational agents. Further, Hobert and von Wolff [14] conclude that more often agents target non-formal over formal learning situations. As explained by Hobert and Berens [13], there is a trend towards the use of mobile voice assistants in education.

However, besides all progress, voice assistants lack an adequate generalization of existing research results and comprehensive evaluation studies and process models. These are important, however, because the success of new technologies depends on how easy technical innovations are to use, how great their personal added value is, and how well they can be integrated into familiar routines. Graesser [12] describes a system that simulates human tutoring conversation patterns as well as alternative patterns that follow ideal pedagogies. They argue that alternative strategies to the simulation of human patterns are superior as human tutoring is not always ideal. This opens a whole new

discussion about the use of pedagogical conversational agents that need to be addressed in the future.

While hand-drafted conversational agents are in use, there is a lack of procedures that can generate conversational agents for different knowledge areas (domains) on the basis of un/semi-structured and slim information—such as lecture slides. One exception is the work by Kelsey et al. [16] who presented a framework to develop dialog-based tutors that interact naturally with learners. Their approach combines automatically generated domain-specific content with a generalized domain-independent framework. While domain-independent semantic language comprehension is still a long way off, there exist very good solutions for certain narrow areas (domains). Despite the disruptive advances in AI systems through deep learning, the dialogs and dialog structures of the voice assistants are still mostly created manually. This is due to the fact that corresponding sample dialogs with intents and entities are usually not available or only to a limited extent. Current research efforts are often focused on the aspect of *intent recognition* in dialogs rather than on the overall system [6]. Advances in dialog control based on semantic data have been achieved throughout the years.

3 Requirement Analysis

Our proposed system offers interactive, natural-language possibilities to convey learning content in a way that was previously only possible in the context of human-to-human training or by hand-crafted conversational dialog systems. Little is known, therefore, about the general requirements of learners on such a system, a first orientation is given by user stories for students, and educators as presented in [28]. The success of the voice assistant does not only depend on its technical implementation but also on its expectation, acceptance, and final use of the learners. Therefore, we conducted a brief survey among our students to get some insights into what is required and expected by students. Most respondents use chatbots several times a year, all have tried chatbots before. Interestingly, none of the asked students loves to chat with bots, but they are accepted as a means to fulfill particular goals just like normal messengers. The majority of respondents (83%) use messengers daily, all use it occasionally. Some (8%) exchange remote information through this channel exclusively. The use of messengers is based on different motivations, for instance, 41% prefer written over spoken language.

Students seem to *prefer writing over speaking as text input*. In line with the previous finding output in *written language is preferred over spoken output*. However, voice output is rated more popular than voice input. As it makes no difference in learning outcomes when students express their contributions in spoken or written form [7] both channels can be offered as an input device to fulfill personal preferences. In contrast findings by Lachert et al. suggest that explaining orally or written has an effect [15]. However, if this difference is caused by social presence is an open question and thus it stays unclear if this difference exists also for conversational tutors (in particular for those without embodiment). The popularity of written text in our questionnaire might lie in the fact that the given information from the *assistant should be accompanied by images and graphics*. Videos are less popular and equations should be presented only if required by the content of the lecture.

Asked about the 'character' students stated that the assistant should behave like a tutor (50%), or like a friend (25%), however not like a lecturer/teacher. The assistant should motivate (66%) but never praise (0%), be friendly in any case (83%) but rarely be funny (25%). The assistant should answer briefly to questions and thereafter ask whether further information is desired (66%), however, asking directly whether a topic was understood is not welcome by most users (33%).

The basic functionalities of the assistant should include answering questions about the content of the lecture (75%), replaying definitions (66%), and referring to relevant slides (66%). Advanced functionalities should include suggesting possible topics of discussion about teaching content (66%), examining content with evaluation and feedback (58%), and pointing out additional slide content (50%). Functionalities that are considered less relevant include linking to further literature (basic 25%, advanced 35%), to check contents without evaluation but with subsequent display of the correct answer (basic 25%, advanced 8%), small talk (basic 33%, advanced 8%), and to publish information about the lecturer/professor (25% each).

4 System Components

In this section, we present the system architecture and its components. Due to space constraints, we do not describe all elements of the system in detail but concentrate on the components which vary from or extend common conversational system architectures [27]. Figure 1 presents our system overview including *data sources* (on the left), *components* (in the center), and the *human-machine interfaces* (on the right). The grey boxes are those components described in more detail.

4.1 Knowledge Extractor

Building a dialog requires knowledge on how to design conversational user interfaces [21] and in-domain knowledge. While the former has to be woven into our system by applying generic patterns, the in-domain knowledge needs to be provided from external data sources. The extraction of relevant information and its underlying semantic meaning from a given source is by far the largest challenge in the semi-automatic generation of a dialog structure. The use of AI-based NLU components is absolutely necessary for this conversion process because semantic patterns have to be used together with pre-stored knowledge components.

In particular *contextualized word representations* [22] such as *keywords extraction* [33] and *named entity recognition* [19] needs to be detected automatically. Besides the detection of keywords and named entity the relationship between them is also important; e.g. 'conversational user interfaces' are a subtopic of 'natural user interfaces'.

As it is an unsolved problem by today's technology to automatically convert an unstructured text into a meaningful dialog structure we decided to use both: *structured data*, using slide templates designed particularly to provide a parseable semantic structure and large amounts of *unstructured data*. This hybrid approach allows to combines to good effect the *pedagogical knowledge*, available in a formalized ontology, and *large general text data* ('big data') when generating the dialog. Additional domain

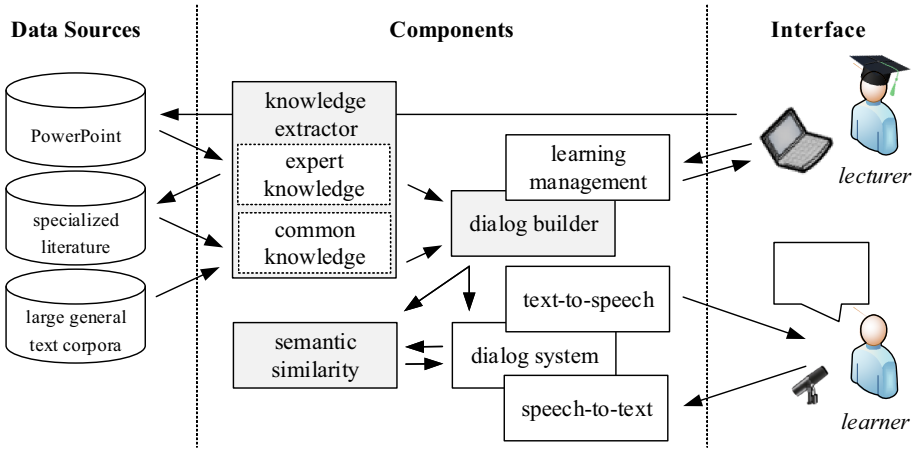

Fig. 1. System overview with all components.

knowledge (specific to the respective subject area) can be added from non-semantically annotated lecture materials (scripts, recommended textbooks if available in digital form) to extend and finetune the general knowledge base. Each layer in the ontology represents a particular view of the learning process: The *pedagogical* view is given by the structure of the slides, the *domain-specific* view defines the vocabulary and terminology, the *course-specific* view reflects the concrete realization of the learning material. The *learner-specific* aspect including age, gender, previous knowledge, and learning success. Therefore, those aspects are not represented within the provided learning material but need to be considered in the design of possible learning paths within the generated dialog structure.

Annotation of Slides

Since many lecturers prefer to use slides instead of blackboards when giving their presentations, extracting the content of the slides can be a very valuable source of information to be processed automatically. In contrast to second-hand sources, such as textbooks written by others, the lecture slides reflect exactly the outline and structure of the course, its content, and its lingua. Another advantage of lecture slides over textbooks is their stringent organization in a particular format which is quite similar to one another. These structures can be used in order to provide additional services such as the semantic ranking of lecture slides [26] or to generate concept maps [2]. The structure and information given in regular lecture slides, however, are not sufficient to find good mappings between specific text passages and the given ontology to generate useful dialogues. To support the parsing process we prepared PowerPoint slides to represent the defined ontology. Annotations include ‘Chapter’, ‘Topic’, ‘Subtopic’, ‘Keywords’, ‘Abbreviation’, ‘Definition’, ‘Example’, ‘Idea’, ‘Goal’, ‘Quote’, ‘Attention’, ‘Equation’, ‘Image’, ‘Index’, and ‘Slide Number’. Each item by itself is representing a list and thus can include more elements. If required, the given ontology can be expanded, however, the inclusion needs to be specially treated in the dialog builder. An example of an annotated PowerPoint slide is given in Fig. 2.

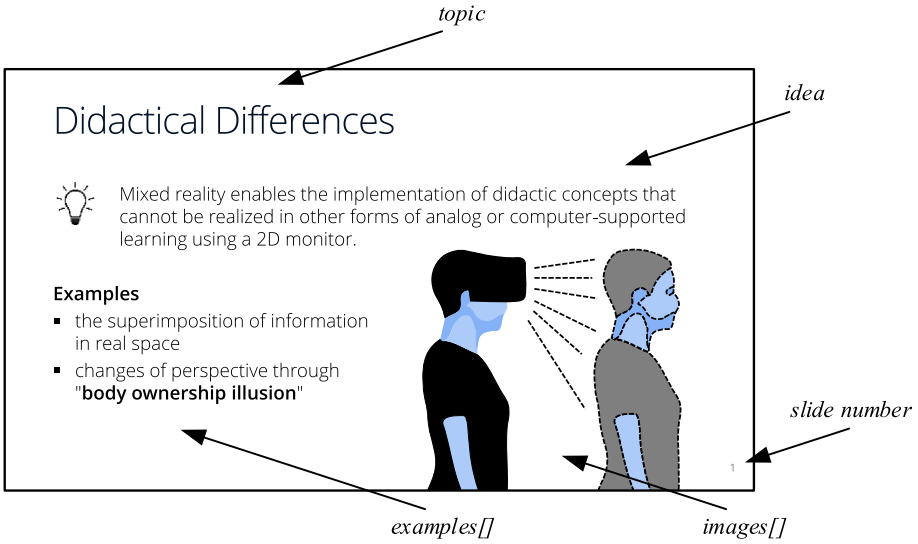

Fig. 2. Example of an annotated PowerPoint slide.

In order to make the application easy to use for lecturers, we developed semantically annotated presentation slide templates, which pre-define didactically adequate learning paths (as macro learning paths) for the lecturers through their set of slides, which gets reflected in the dialogue structure. Note that if the slide templates are created and used consistently no additional labeling is required in the production step of the slides. Therefore, no extra step is required here and no plugin to PowerPoint is required. To prepare our slides accordingly we had to edit the shape names within PowerPoint follow these simple steps:

1. click on the shape (textbox)
2. select *Format* tab
3. click on *Selection Pane*
4. rename the shape to defined the label (ID) for information extraction

The information given in those specifically prepared slides can now easily extracted and stored as a JSON-object and written into a database. To extract the information from PPTX files we used the python-pptx module which lets you access the different shape names and content. Because some relevant pieces of information such as font type or font style cannot be extracted using this tool we had to develop our own extension. The change of font type or style can be a good indicator to identify keywords as well as named entities.

Particular challenges in the parsing process appear if topics are spread over more than one slide or more than one topic (e.g. two definitions) is covered in a single slide. Thus, *intra-slide* and *inter-slides* relationships have to be considered:

- *index pages*—showing an overview of following slides, can be useful to structure subtopics but contains no content itself
- *combining continuous slides*—important topic may cover several continuous slides who might share the same title with or without potential numbering; e.g. 'Interface Design', 'Interface Design II', and 'Interface Design III'. Content coming from these slides should be considered as the same item in the content structure.
- *redundancy* —various slides might contain the same topic including the same, or even worse, contradicting information.
- *subtopic borders within slides*—some slides provide more than one topic at the same time e.g. in a comparison between two approaches. Content coming from those slides have to be considered as separate items in the content structure.

Up to this point, no user intervention is necessary. However, because not all information is already given within the slide context itself—in particular for slides containing images—we decided to allow to override or extend the given information by looking for JSON-expressions within the comments section of each slide. For instance, not any given acronym or domain-specific terminology may be reflected in the slides. The user has the possibility to modify, revise, and add content and questions not presented within the visual elements of the slides. For instance, to add keywords you can write in the comment section:

```
{
  "keywords": [
    "natural language understanding",
    "natural language processing"
  ]
}
```

Note that only JSON-formated text is parsed in the notes section while the continuous text is ignored in the parsing process. Therefore, the notes section can be used as usual.

4.2 Dialog Builder

To generate a useful dialog structure from the extracted data it is required to define transformation rules for each specific data type. Dependent on the data type the generated dialog structure has to offer prompts, hints, or forced-choice dialogue patterns and should be integrated into a more conversational dialog structure. Regardless of the defined data type, it is required to have various deviations in the dialogue to make the interactions seem more natural. In particular, the dialog system must be able to respond to several forms of spoken or written inquiries ranging from asking for the definition of a 'term' or for 'examples', and should also be able to answer questions that require higher levels of knowledge and understanding. Another challenge—specific for the task of a conversational tutor—in designing the dialog structure is the integrating of diagrams, images, and other visual media in addition to pure text.

Certain aspects of dialogue formulations within a teaching context are largely domain-independent such as chit-chat or based on phrases e.g. to ask for a definition.

We developed these elements of the dialogues as generic templates which were then filled according to our database. This library of templates can be easily extended to fit the particular needs of a data type and grow over time to get more variation for the data types already defined.

As a starting point to build our dialog system we used the open-source software RASA [3]. According to our defined rules, the dialog builder is generating a dialog syntax that follows the structure as given in the listing from which the dialog can be trained (note that our system works for German, but for better comprehensibility for the non-German speaking audience we show the dialog in English here):

```
nlu:
- intent:query_topics
examples: |
- [define]{"entity": "topic_attribute", "value":"
  definitions"}
- i look for [definition]{"entity": "topic_attribute","
  value": "definitions"}
- do you know [speech processing](topic_name) ?
[...]
- synonym:define
examples: |
- define
- specify
- determine
[...]
- lookup:topic_name
examples: |
- intuitive interface
- sensor knowledge
- conversational user interface
[...]
```

4.3 Semantic Similarity

While most of the required functionalities are readily contained in any dialog system to build a pedagogical conversational agent, the evaluation of given 'free text' answers from learners needs to be treated otherwise. To grade the quality of 'free text' answers semantic textual similarity needs to be calculated. Literature has commonly reported that pre-trained text embedding models can be used as effective feature extractors and achieve good performance on calculating semantic textual similarities. Therefore, for the time being, we have not used any kind of transfer learning to adapt to the content of the slides or to tune the hyperparameters. To compare the semantic similarity between the given answer and the reference, we directly assess the similarity of the sentence embeddings produced by the transformer model as described in [31]. The similarity between the two sentence embeddings was calculated using the cosine similarity and converted into an angular distance by arccos. For more detail see [4].

5 Features of the Pedagogical Conversational Tutor

In this section, we describe some of the features currently implemented in the assistant. While some of the features such as small talk and user feedback are content-independent other features rely on the provided lecture slides such as listing, searching, and suggesting topics, asking for definitions, or being tested about the lecture context. Additional content might be favorable but is neither content independent nor given on the slides, this might include additional information of the lecturer such as curriculum vitae, opening hours, telephone numbers, etc. Due to space constraints, we only describe topics suggestions and self-test here and leave out descriptions for small talk, user feedback, and additional information.

5.1 List, Search, Suggest Topics, and Similar Content

While listening and searching for topics is a straight forward task after the DB has been generated from the slide content, suggesting topics based on a given topic needs more attention. We handle this by relying on different measures including the ontology defined by topic and sub-topic as well as the distance between slides (assuming that similar topics appear close in the slide order).

5.2 Self Test of Lecture Content

We have investigated two forms of self-tests: multiple choice/fill-in-the-blank text (close-ended questions) and free-text answers (open-ended questions). Note that both types of questions investigated here are limited to represent knowledge embedded in a very small-sized text fitting on a single slide and often represent only a single sentence.

Multiple-Choice & Fill-in-the-Blank Text

For multiple-choice, the task is to collect answers from the choices offered in a list. It is frequently used in educational testing and easy to grade automatically. CH and Saha present a systematic review of systems to automatically generate multiple-choice questions from text [5]. In the case to generate multiple-choice questions from annotated slides we have to perform four essential steps (excluding preand post-processing): *sentence selection*, *key selection*, *distractor generation*, and *question formation*. The *sentence selection* step finds relevant sentences from which questions are formed. For each selected sentence the *key* (or *target word*), which is the correct answer to the question, has to be determined in the *key selection* step. The set of *distractors* contains possible wrong answers along with the correct answer in order to befuddle the examinee. The *question formation* is the task of transforming the declarative sentence into its interrogative form.

While sentence selection and question formation will be covered later, as it is also relevant in free-text answers, we will briefly discuss key selection and distractor generation here. Obviously, for multiple-choice as well as fill-in-the-blank text tasks the selected word or word sequence that will be blanked out is of utmost importance. In the literature methods based on frequency count, part-of-speech, semantic information, and pattern matching is addressed. We regard a word or word sequence as key if a high

semantic similarity was given in the 'Topic' as well as in the selected sentence. To find distractors is not as straight forward. For numerical questions, similar numbers can be generated, but for a word or word sequence, a more semantically driven approach is required. Wherever possible, in those cases where the found key matching a 'Subtopic', it is possible to rely on domain ontology, by drawing distractors from 'Subtopic' within the same 'Topic'. In those cases where this is not possible only open-ended questions have been formulated by the system.

Free-Text Answers

For free-text answers, the task is to collect answers which are freely formulated. It shares some steps in common with the automatic generation of close-ended questions, namely *sentence selection* and *question formation*, but has no need for *key selection* and *distractor generation.* While multiple-choice answers and grading can readily be integrated into a dialog structure, free-text answers need an extra processing step as described in Sect. 4.3. The similarity score can then be handed back to the dialog system and respective answers based on the score can be rendered; e.g. for a similarity score of 95% for a given definition you get a "you met the definition very well" while for a score of 30% you get a "maybe you should review the definition".

Sentence Selection

Not each sentence in a given text is worth to be rendered as a question. Only those sentences containing questionable facts should be selected. While the selection process for general question selection is non-trivial and different methods such as sentence length, parts-of-speech, semantic information, or the number of occurrence of a particular word, have been proposed, we regard a sentence worth to be rendered as a question based on semantic similarity and occurrence between the words used in 'Topic' and 'Definition' or 'Idea'. The sentence or sentences in the same tag showing a high correlation are selected.

Question Formulation

The selected sentence has to be rendered as a question to be useful for multiple-choice or free-text answers. Of course, the formulated question depends on the type of answer and has to be treated differently for multiple-choice or free-text. However, both types rely most of the time on rule-or pattern-based methods including appropriate 'wh-word' selection (who, what, why, when, where; including also how and how much), subject-verb-object and their relationship, dependency-based patterns, and syntactic transformation. More information on how to generate questions can be found in [20].

6 Observations and Limitations

We introduced a first working prototype of a pedagogical conversational agent to be generated from annotated lecture slides. The development has revealed that there are a couple of difficult problems that need to be addressed in order to provide a tool that can be easily applied as outlined in the introduction. Nevertheless, with the working prototype,

it was already possible to demonstrate the proof of concept. A dialog structure given 577 semantically annotated slides for the university course 'Intuitive and Perceptual Interfaces' for computer science in media bachelor students in their third year of study at the Karlsruhe University of Applied Sciences have been semi-automatically parsed and converted into the dialog structure. No post-processing of any kind has been used to improve the system.

The generated pedagogical conversational agent has been used for the evaluation. In total 19 students (12 male, 7 female, age between 18 and 27) familiar with the content of the course tried the prototype. It could be observed that the participants were generous to small mistakes and serious deficiencies in the offered possibilities in the dialog or the given answers. At first, though this finding might contradict Graesser who observed that "many students expect the computers to be accurate rather than polite" in a conversation [12]. However, by analyzing the emerged problems within the conversation we found that the accuracy of finding the intend was not very good but if the intend was resolved correctly the accuracy of the resulting answer was quite high. In those cases where the intent could not be resolved, its incorrectness was sometimes even completely ignored and something else was tried or it lead to amusement rather than frustration. It is important to note that our findings might be biased as the involved students had a strong background in computer science and some had already prior knowledge in the theory of conversational user interfaces. While all students seem to be in favor of such a system, some had concerns about data security and were reluctant if they would actually use the system in production; e.g. one student stated "It has to show a real benefit and has to offer more than simply the slide content in a new format."

It has been shown that having the instructor visible on the screen has a very limited effect on the learning outcome (although eye contact seems to be helpful) [8]. Therefore, it has been decided that the agent uses no visual representation of a scholar such as a portrait, talking head, or avatar that generate speech, actions, facial expressions, and gestures to give the system some kind of body [17]. Even though forms of visual embodiment are missing and just a dialog are offered, the system got recognized as a personality, rather than a technology. In those cases where the dialog system was recognized as a personality the conversation tended to be more emotional; e.g. "Hey, could you give me some examples. Please!" instead of "show example". Users placed more emphasis on the linguistic interpersonal content and less on the technical function. In those cases where the dialog system was perceived more as a technology the dialog was more rationally and followed the structure of syntax and keywords; e.g. just "example" even without the verb show, and lacked any form of politeness. With an increased time of using the system, the behavior of the user became significantly freer, more relaxed, and more playful. At first, the users frequently explore which answers are given for different question formulations. Often they interact with the system at first only by keywords, later on, they switch to complete questions up to a complete, free dialog in natural language.

It is worth noting that the quality of the dialog depends heavily on the provided quality and quantity of the slides. Carelessness in the conception and creation of the lecture slides is directly reflected in the generated dialogue. Inconsistency, contradictions, or wrong labels in the slides result in inconsistent dialogs which might even have a stronger

negative impact as in its original form of presentation. No measures have been applied so far to check the quality and consistency of the provided slides.

Results by Zellou and Cohn indicate that social and functional factors influence dialog structure within humans and between humans and a dialog system [32]. Therefore, conversational content should have a particular structure to flow naturally and the dialogue should be guided by the given answers of the dialog partner. It would be favorable if the dialogue would reflect the style of a tutor and personality. So far, these capabilities have not been integrated. Besides optimizing the dialog in the future it could be investigated how to create a more emotional dialog in written form; e.g. by reflecting voice characteristics in the textual representation [30].

The system presented here has to be validated and optimized in further research. To support this process the developed software solution, Powerpoint templates, and pre-trained models will be published under a Creative Commons license on GitHub.

References

1. Amazon: Alexa in higher education (2020). https://aws.amazon.com/de/education/alexa-edu/higher-education/
2. Atapattu, T., Falkner, K., Falkner, N.: A comprehensive text analysis of lecture slides to generate concept maps. Comput. Educ. **115**, 96–113 (2017)
3. Bocklisch, T., Faulkner, J., Pawlowski, N., Nichol, A.: Rasa: open source language understanding and dialogue management. arXiv preprint arXiv:1712.05181 (2017)
4. Cer, D.M., et al.: Universal sentence encoder. arXiv abs/1803.11175 (2018)
5. Ch, D.R., Saha, S.K.: Automatic multiple choice question generation from text: a survey. IEEE Trans. Learn. Technol. **13**(1), 14–25 (2020). https://doi.org/10.1109/TLT.2018.2889100
6. Damonte, M., Goel, R., Chung, T.: Practical semantic parsing for spoken language understanding. arXiv preprint arXiv:1903.04521 (2019)
7. D'Mello, S.K., Dowell, N., Graesser, A.: Does it really matter whether students' contributions are spoken versus typed in an intelligent tutoring system with natural language? J. Exp. Psychol. Appl. **17**(1), 1 (2011)
8. Fiorella, L., Stull, A.T., Kuhlmann, S., Mayer, R.E.: Instructor presence in video lectures: the role of dynamic drawings, eye contact, and instructor visibility. J. Educ. Psychol. **111**(7), 1162 (2019)
9. Følstad, A., Skjuve, M., Brandtzaeg, P.B.: Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. In: Bodrunova, S., et al. (eds) INSCI 2018. LNCS, vol. 11551, pp. 145–156. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17705-8_13
10. Fügen, C., et al.: Advances in lecture recognition: the ISL RT-06s evaluation system. In: Ninth International Conference on Spoken Language Processing (2006)
11. Gašević, D., Dawson, S., Rogers, T., Gasevic, D.: Learning analytics should not promote one size fits all: the effects of instructional conditions in predicting academic success. Internet High. Educ. **28**, 68–84 (2016)
12. Graesser, A.C.: Conversations with autotutor help students learn. Int. J. Artif. Intell. Educ. **26**(1), 124–132 (2016)
13. Hobert, S., Berens, F.: Small talk conversations and the long-term use of chatbots in educational settings – experiences from a field study. In: Følstad, A., et al. (eds.) CONVERSATIONS 2019. LNCS, vol. 11970, pp. 260–272. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39540-7_18

14. Hobert, S., Meyer von Wolff, R.: Say hello to your new automated tutor–a structured literature review on pedagogical conversational agents (2019)
15. Jacob, L., Lachner, A., Scheiter, K.: Learning by explaining orally or in written form? Text complexity matters. Learn. Instr. **68**, 101344 (2020)
16. Kelsey, E., Ray, F., Brown, D., Robson, R.: Design of a domain-independent, interactive, dialogue-based tutor for use within the GIFT framework. In: Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (Giftsym3), pp. 161–168 (2015)
17. Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., Welch, G.F.: Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In: 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 105–114. IEEE (2018)
18. Kolss, M., Wolfel, M., Kraft, F., Niehues, J., Paulik, M., Waibel, A.: Simultaneous german-English lecture translation. In: International Workshop on Spoken Language Translation (IWSLT) 2008 (2008)
19. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360 (2016)
20. Monz, C.: Machine learning for query formulation in question answering. Nat. Lang. Eng. **17**(4), 425–454 (2011)
21. Pearl, C.: Designing Voice user Interfaces: Principles of Conversational Experiences. O'Reilly Media, Inc., Sebastopol (2016)
22. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365 (2018)
23. Rus, V., D'Mello, S., Hu, X., Graesser, A.: Recent advances in intelligent systems with conversational dialogue. AI Mag. **34**, 42–54 (2013)
24. Schlobinski, P., Siever, T.: Sprachliche kommunikation in der digitalen welt. Eine repräsentative Umfrage, durchgeführt von forsa (1619–1021) (2018)
25. Seufert, S., Meier, C., Soellner, M., Rietsche, R.: A pedagogical perspective on big data and learning analytics: a conceptual model for digital learning support. Technol. Knowl. Learn. **24**(4), 599–619 (2019)
26. Wang, Y., Sumiya, K.: Semantic ranking of lecture slides based on conceptual relationship and presentational structure. Procedia Comput. Sci. **1**(2), 2801–2810 (2010)
27. Wei, C., Yu, Z., Fong, S.: How to build a chatbot: chatbot framework and its capabilities. In: Proceedings of the 2018 10th International Conference on Machine Learning and Computing, pp. 369–373 (2018)
28. Winkler, R., Söllner, M.: Towards empowering educators to create their own smart personal assistants. In: Proceedings of the 53rd Hawaii International Conference on System Sciences (2020)
29. Wölfel, M.: Robust automatic transcription of lectures. KIT Scientific Publishing (2009)
30. Wölfel, M., Schlippe, T., Stitz, A.: Voice driven type design. In: 2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 1–9. IEEE (2015)
31. Yang, Y., et al.: Multilingual universal sentence encoder for semantic retrieval (2019)
32. Zellou, G., Cohn, M.: Social and functional pressures in vocal alignment: differences for human and voice-AI interlocutors (2020)
33. Zhang, Y., Tuo, M., Yin, Q., Qi, L., Wang, X., Liu, T.: Keywords extraction with deep neural network model. Neurocomputing **383**, 113–121 (2020)
34. Zhang, Z., Takanobu, R., Zhu, Q., Huang, M., Zhu, X.: Recent advances and challenges in task-oriented dialog systems. Sci. China Technol. Sci., 1–17 (2020)