

# Integration of a Semantic Storytelling Recommender System in Speech Assistants

María González-García<sup>1,\*†</sup>, Julián Moreno-Schneider<sup>1†</sup>, Malte Ostendorff<sup>1†</sup> and Georg Rehm<sup>1†</sup>

<sup>1</sup>Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH), Alt-Moabit 91c, 10559 Berlin, Germany

## Abstract

Nowadays, practically there is no customer service that does not use speech assistants or smart voice assistants in almost every area of the society. Depending on the area, these assistants are used to develop different tasks. Considering that, in this work, we present a semantic storytelling approach executed through the combination of a question answering system and a recommender system applied in the context of speech assistants. Our contribution to this context is to provide additional information semantically related to the user's request and QA system answer. Apart from the underlying technology, a prototypical graphical user interface (as a chatbot) has been developed to demonstrate the functionality and some tests and evaluations have been accomplished to check the performance of the approach.

## Keywords

Semantic Storytelling, Speech Assistant, Recommender System

## 1. Introduction

Speech assistants or smart voice systems are becoming a very popular technology in recent times and can be found in more diverse and specific areas. The customer service that does not use this type of system is practically non-existent, either partially (being helped in some part of the process by human intervention) or totally, ending the process completely automatically (independently). In recent years some approaches are introducing the usage of artificial intelligence techniques to automate parts or the entire process to avoid costly resource consuming techniques. There are plenty of examples of this kind of artificial intelligence voice assistant in different areas, for example within the garment industry [1], H&M or Shepora offers chatbots for recommendation, or news portals and television networks such as the CNN chatbot or airlines companies [2] such as KLM Royal Dutch Airlines, Lufthansa or Ryanair.

All these voice assistants have the peculiarity that, although they use AI, their functionality is limited to a specific domain or to a small amount of information (specific databases of the companies themselves), achieving very good results in these limited areas. This trend has continued in recent years, generalizing in domains or situations, that is, voice assistants such as

---


In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): *Proceedings of the Text2Story'23 Workshop, Dublin (Republic of Ireland), 2-April-2023*

\*Corresponding author.

✉ maria.gonzalez\_garcia@dfki.de (M. González-García); julian.moreno\_schneider@dfki.de (J. Moreno-Schneider); malte.ostendorff@dfki.de (M. Ostendorff); georg.rehm@dfki.de (G. Rehm)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Google, Amazon Alexa or Apple Siri have become systems capable of integrating perfectly into people's daily lives thanks to their large ability to answer specific questions. However, they are still not capable of adequately reasoning out a complete story based on previous searches carried out by users. At this point, ChatGPT<sup>1</sup> comes into play. This AI chatbot, launched by OpenAI, was built on top of OpenAI's GPT-3 family of large language models and has been fine-tuned using supervised and reinforcement learning techniques. ChatGPT can accomplish different tasks such as text generation, text completion, QA, summarization, etc. These models are improving over previous approaches, but coherent story generation is still not fully achieved.

This is where our system appears, whose main objective is to offer the user semantically related extra information to a specific answer previously obtain (also automatically) from a question answering (QA) system. In summary, the main contributions are: i) Develop a specific approach for semantic storytelling composed of two modules, a QA system, and a recommender system; ii) Provide additional information considering the semantic relations that exist between two documents; iii) Release the whole code of our approach, which is available at GitLab<sup>2</sup>.

## 2. Related Work

The way a story is told directly influences the final result of the story. As introduced in [3], the term semantic storytelling refers to the automatic (or semi-automatic) generation of stories, where a story is considered a natural language text containing a complete, correct and unambiguous story.

As aforementioned, our semantic storytelling approach is composed of a QA system and a recommender system, therefore, some works related to these topics have been analyzed. For example, the recommender system used in this work is based on the work of [4] that builds an automatic classifier for semantic relations between text segments, designs custom annotation guidelines, annotates a dataset and performs evaluations with Transformer language models. Our work complements this classifier building a complete system together with the QA system.

Other examples of the development of these systems are explained below. [5] presents an interactive recommendation approach that recommends movies and visualizes them in an animated comic strip fashion. In [6], a storytelling approach is followed to build a music recommender system using the structured and linked data offered by the Semantic Web. Other example is the work of [7] that describes a new approach that recommends cultural-touristics paths taking advantages of the user and item profile knowledge. Or the work of [8], that provides a tailor-made story based on the context in which the user is located. These works develop recommender systems based on the user preferences, in contrast, our recommender system is based on the semantic relations between documents to recommend an answer.

## 3. Semantic Storytelling Recommender System

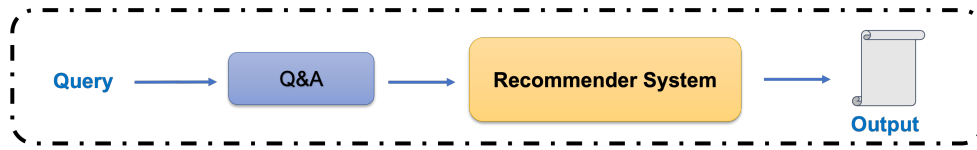
The main goal of our system is to assist users in obtaining additional information to concrete answers through speech assistants. This section describes the technical details of our system,

---

<sup>1</sup><https://openai.com/blog/chatgpt>

<sup>2</sup><https://gitlab.com/speaker-projekt/chatbotdemo>

which is mainly composed of two modules: Question Answering (QA) and Recommender System, as it is shown in Figure 1.



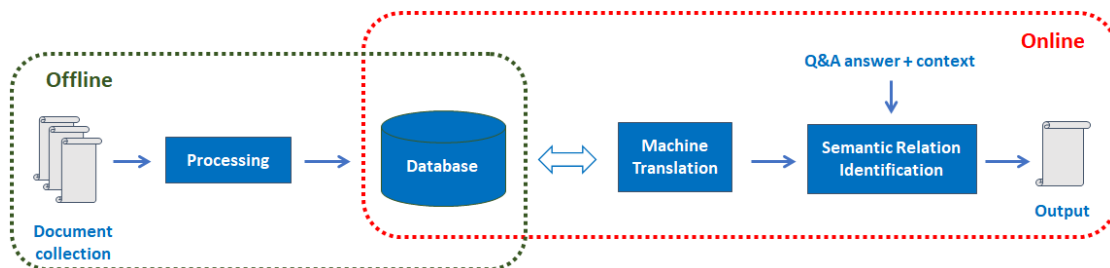
**Figure 1:** Semantic Storytelling Recommender System Architecture.

### 3.1. Question Answering

The QA module is an adapted version of an existing system that worked in English to be able to work with documents in German. As a basis for this development, we have used the Haystack<sup>3</sup> system, a fairly established QA framework that has modules that include the latest technologies in the field of Information Retrieval, such as Sparse (BM25, TF-IDF) or Dense (transformer-based) retrievers, and in the field of Information Extraction, such as FARMReader (based on roberta-base<sup>4</sup>). The Haystack system consists of a large number of predefined and trained QA modules that can be used directly to implement a QA pipeline. The predefined pipelines work exclusively in English, so we had to adapt it for German language using components for German in the indices where the information is stored and the German ELECTRA base model<sup>5</sup>.

### 3.2. Recommender System

Let us assume the following situation. A user has provided a concrete question  $Q$ , and the QA module has already provided a suitable answer  $A$  for it. The goal of this module is to identify and to suggest new content for the user that could be semantically related to the question and the answer. To accomplish this goal, the module is composed of an offline processing and an online processing, as they are going to be described below (Figure 2).



**Figure 2:** Architecture of the Recommender System Module.

<sup>3</sup><https://haystack.deepset.ai/>

<sup>4</sup><https://huggingface.co/deepset/roberta-base-squad2>

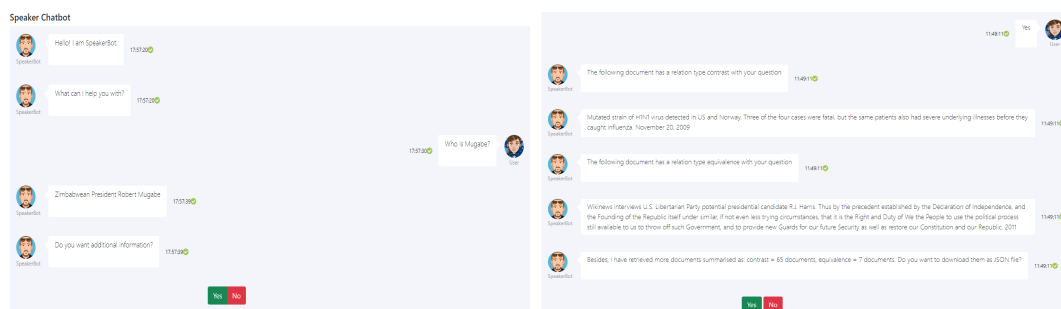
<sup>5</sup><https://huggingface.co/deepset/gelectra-base>

**Offline Processing: Generation of the Semantic Relations Database** In this part of the process, firstly, we get all the origin documents from the dataset (every document is identified by its position in the database). Secondly, each document is selected and combined one by one with the rest of the documents to predict the semantic relations between each one of them. According to [4], 12 semantic relations (none, identity, equivalence, causal, contrast, temporal, conditional, description, attribution, fulfillment, summary and purpose) are going to be considered. And finally, each document and the semantic relations between all documents are stored.

**Online Processing: Finding the Related Segments** The information stored in the offline processing together with the QA answer and the context allow us to look for the semantic relations between the response of the QA service and the documents of our dataset. Moreover, if the language of the request is German, before looking for the semantic relations, the documents of the dataset will be translated through a machine translation module to extend the recommender system module to this language at least until an annotated German dataset will be built.

## 4. Graphical User Interface (GUI): Chatbot

The graphical user interface (see Figure 3) has been designed as a simple chatbot interface with voice input capabilities, apart from the usual text input, in which the extra information provided by the semantic storytelling service has been included as an answer visually different.



**Figure 3:** Chatbot (Front-End) (Higher resolution images available at <https://gitlab.com/speaker-projekt/chatbotdemo/-/tree/main/media>).

The system allows interacting through two different modes: text and voice. Textual input is the most common and intuitive mode and allows using directly the input for all subsequent processes. As for the voice input, the user can record a message that is recognized by an automatic speech recognition systems (ASR), specifically, the Hensoldt Analytics Speech-to-text for English<sup>6</sup> (available at the European Language Grid platform<sup>7</sup>). After that, the chatbot shows the text to find out if it has being recognized properly. In this case, the chatbot behaves as if a textual input was written. On the contrary, it provides the possibility of rewriting the question.

<sup>6</sup><https://live.european-language-grid.eu/catalogue/tool-service/20891/overview/>

<sup>7</sup><https://www.european-language-grid.eu>

Once the text is available, it is processed by an intent recognition module. This module uses simple rules to classify the query into five classes: GREETINGS, GOODBYE, ASSERTION, THANKS and QUESTION. If the intent is classified into one of these classes: GREETINGS, GOODBYE, ASSERTION and THANKS, the response of the chatbot is one of those shown in Table 1. Otherwise if the intent is classified into the QUESTION class, the chatbot calls the QA module and, if it is required, the recommender system module. In this case, the chatbot shows two examples as additional information that contain the type of semantic relation between the QA response and a document of the dataset and the document itself. Moreover, if there are more semantic relations, the chatbot also shows the number and the type of these relations as well as offering the possibility of downloading all this information in a JSON format file.

**Table 1**

Requests and Responses associated with four of the classes considered in our system.

Intent	Input	Response
GREETINGS	Hello; Hi; Hey	Hello;Hi; Welcome
GOODBYE	Bye; Good bye; See you later	Have a nice time;Bye; See you soon
ASSERTION	Yes; Yes, please; Of course	What can I help you with?; What is your question?
THANKS	Thanks; Thank you; That’s helpful; Awesome; Great; No, thanks; No thanks	Happy to help!; Any time; You’re welcome; My pleasure

## 5. Experiments

The evaluation of the semantic storytelling approach is accomplished by several experiments, which aim to explore the suitability of the approach and help us to gain an understanding of what we can achieve in the long run. Two different datasets have been used during our experiments, the dataset explained in [4] for training the recommender system and the German News Dataset<sup>8</sup> for evaluating the performance of the recommender system in German.

### 5.1. Experiment for the recommendation model

Two experts have evaluated the performance of the recommendation model. Due to the fact that the German News dataset is too big, we used a sample set of it that allows us to analyze 1026 semantic relations between each two documents (or parts of a document). In this sense, it is important to remark that we have only used the most relevant semantic relation between documents. Besides, considering that the recommender system uses an English model and we want to test its performance in German language, we also used a machine translation module, concretely, the Argos Translate library<sup>9</sup>. Once the documents are translated, we predict the semantic relations between them and after that we calculate their accuracy (see Table 2) .

<sup>8</sup><https://www.kaggle.com/datasets/pqbsbk/german-news-dataset>

<sup>9</sup><https://github.com/argosopentech/argos-translate>

**Table 2**

Predicted semantic relations and their accuracy.

Semantic relation	None	Causal	Temporal	Equivalence	Contrast	Avg
Number	812	75	69	57	13	1026
True	632	45	41	31	2	751
False	180	30	28	26	11	275
Accuracy	0.78	0.6	0.59	0.54	0.15	0.73

After performing a qualitative analysis of the predicted semantic relations, there are several remarks that are worth mentioning. As expected, the sample set is unbalanced and the number of *none* semantic relations is much bigger than the rest of the represented semantic relations ( $812 > 214$ ). Only 5 out of 12 (the semantic relations recognized in [4]) have been identified. But, in general, and considering the errors that can be propagated from the translations and that we are using a sample set, the outcomes seems promising.

## 5.2. Other experiments

Apart from this experiment, the QA module (working in German) and the semantic storytelling approach through the GUI have also been evaluated. Regarding the QA evaluations, we have used the GermanQuAD dataset [9] and we have evaluated the Retriever, obtaining *Recall* : 0.972, *MeanAvgPrecision* : 0.877, and the Reader (using gelectra model), obtaining *Top – N – Accuracy* : 97.42, *ExactMatch* : 72.42 and *F1 – Score* : 90.35. Although we are still in the process of intensively test and evaluate the GUI and the way the semantic storytelling integrates into it, we have been able to test it to prove the functionality of both. Future plans include to accomplish further user satisfaction evaluations.

## 6. Conclusions and Future Work

We have developed a storytelling approach starting from a QA system and a recommender system. Our first experiment of the recommender model showed that thanks to the automatic translation module it is possible to use the model without having to train it with new data in different languages. Due to time constrains, we could not include in this work the evaluations regarding the complete semantic storytelling approach through the GUI. These new evaluations are our short-time future work. On the long-term we are planning to compare our results with a recommender system using a model trained on German data (that we are planning to annotate). Besides, the end-to-end integration of all components currently under development is foreseen, as well as the adaptation of the approach to new domains.

## Acknowledgments

This work has received funding from the German Federal Ministry for Economic Affairs and Climate Action (BMWK) through the project SPEAKER (no. 01MK19011).

## References

- [1] L. Syarova, Chatbot usage in e-retailing and the effect on customer satisfaction, 2022. URL: <http://essay.utwente.nl/92080/>.
- [2] S. D. Sarol, M. F. Mohammad, N. A. A. Rahman, Mobile Technology Application in Aviation: Chatbot for Airline Customer Experience, Springer Nature Singapore, Singapore, 2023, pp. 59–72. URL: [https://doi.org/10.1007/978-981-19-6619-4\\_5](https://doi.org/10.1007/978-981-19-6619-4_5). doi:10.1007/978-981-19-6619-4\_5.
- [3] G. Rehm, K. Zaczynska, J. Moreno, Semantic storytelling: Towards identifying storylines in large amounts of text content., in: Text2Story@ ECIR, 2019, pp. 63–70.
- [4] M. Raring, M. Ostendorff, G. Rehm, Semantic relations between text segments for semantic storytelling: Annotation tool - dataset - evaluation, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 4923–4932. URL: <https://aclanthology.org/2022.lrec-1.526>.
- [5] K. Wegba, A. Lu, Y. Li, W. Wang, Interactive storytelling for movie recommendation through latent semantic analysis, in: 23rd International conference on intelligent user interfaces, 2018, pp. 521–533.
- [6] S. Baumann, R. Schirru, B. Streit, Towards a storytelling approach for novel artist recommendations, in: International Workshop on Adaptive Multimedia Retrieval, Springer, 2011, pp. 1–15.
- [7] M. Casillo, M. D. Santo, M. Lombardi, R. Mosca, D. Santaniello, C. Valentino, Recommender systems and digital storytelling to enhance tourism experience in cultural heritage sites, in: 2021 IEEE International Conference on Smart Computing (SMARTCOMP), 2021, pp. 323–328. doi:10.1109/SMARTCOMP52413.2021.00067.
- [8] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, A context aware recommender system for digital storytelling, in: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), 2018, pp. 542–549. doi:10.1109/AINA.2018.00085.
- [9] T. Möller, J. Risch, M. Pietsch, Germanquad and germandpr: Improving non-english question answering and passage retrieval, CoRR abs/2104.12741 (2021). URL: <https://arxiv.org/abs/2104.12741>. arXiv:2104.12741.