

Hotel Scribe: Generating High Variation Hotel Descriptions

Saad Mahamood

trivago N.V.
Düsseldorf, Germany
saad.mahamood@trivago.com

Maciej Zembruski

trivago N.V.
Düsseldorf, Germany
maciej.zembruski@trivago.com

Abstract

This paper describes the implementation of the Hotel Scribe system. A commercial Natural Language Generation (NLG) system which generates descriptions of hotels from accommodation metadata with a high level of content and linguistic variation in English. It has been deployed live by *trivago* for the purpose of improving coverage of accommodation descriptions and for Search Engine Optimisation (SEO). In this paper, we describe the motivation for building this system, the challenges faced when dealing with limited metadata, and the implementation used to generate the highly variate accommodation descriptions. Additionally, we evaluate the uniqueness of the texts generated by our system against comparable human written accommodation description texts.

1 Introduction

The hotel search business is a highly competitive market in which websites attempt to align the accommodation needs of a given user with the available marketplace of prices/deals offered by hoteliers and other accommodation providers. It is imperative that users are able to find the type of accommodation they are seeking and find relevant information associated to a given accommodation in the form of images, text, maps, and infographics.

One key piece of textual information is an *accommodation description*, which provides a given user with detailed information about the accommodation and the facilities that it offers for their guests. Within *trivago* these descriptions have been typically written manually by humans either with the use of freelancers or the hotelier providing a description themselves. As the global market size of hotels has continuously grown over the past few years this has meant an ever increasing

inventory of accommodations requiring a description. Other problems such as the cost of employing freelancers across multiple languages, the lack of consistency, and the time-lag in providing updated descriptions for accommodations meant that an automated solution for generating accommodation descriptions was needed. Finally, with the importance of SEO for consumers searching for a specific accommodation from a search engine, there was also a need to generate accommodation descriptions with a high level of content and linguistic variation to make generated description as distinct as possible.

In this paper, we will describe relevant past work that has been done using NLG for commercial applications and the use of variation in NLG. In later sections we will describe our approach for generating accommodation descriptions with variation, our methodology for evaluating the effectiveness of the implemented system, and discuss the results that we have obtained. In the final section we conclude the paper based upon the findings we have obtained.

2 Background

2.1 Commercial NLG

Over the past 5-10 years there has been a substantial increase in the number of commercial based NLG solutions (Dale, 2019). Commercial NLG applications have appeared in differing domains such as Weather (Sripada et al., 2014), Automated Journalism (Caswell and Dörr, 2018), Oil & Gas Industry (Reiter, 2017), Healthcare (Harris, 2008), and Financial reporting (Danlos et al., 2011). This increase has been in part due to the rise of companies such as *Narrative Science*, *Automated Insights*, *Arria NLG*, *Yesop*, and *Ax Semantics*. Whilst there has been some interest among these companies in exploring aspects of

multimodality (Mahamood et al., 2014), referring expression generation (REG) (Reiter, 2017), and morphology inflection (Madsack et al., 2018), the majority of these commercial NLG offerings leverage little-to-no rich linguistic concepts such as aggregation and REG that have been developed by the research community (Dale, 2019). Preferring instead an approach of using ‘smart templates’ to generate routine texts from data (Dale, 2019), which in some domains the actual implementation can be “embarrassingly simple” (Graefe, 2016).

Building commercial NLG systems can introduce a set of challenges that may not be found in typical academic implementations. These challenges can range from issues with ensuring reliability/accuracy of the output text generated (to minimise legal liability) (Harris, 2008), reusability & configurability (Sripada et al., 2014; Reiter, 2017), absence of appropriate data (Caswell and Dörr, 2018), and finally the need for scalability in generating large volumes of output (Harris, 2008). Additionally challenges can be found during the corpus analysis phase with target corpora not being available entirely (Sripada et al., 2014), access issues due to data privacy (Harris, 2008), or a lack of consistency in style due the corpora being authored by multiple authors (Sripada et al., 2004). The totality of these numerous challenges means that for some commercial applications there is a need to focus on simplicity rather than implement cutting-edge techniques (Harris, 2008).

2.2 NLG & Textual Variation

Variation in NLG systems is the process in which there is a variance in content and/or lexical expressions chosen by a given system when generating textual output. Traditionally, output texts created through the use of ‘canned text’ rules tended to be fairly simplistic and show almost no variation (Theune et al., 2001). Systems, however, must be built with variation as part of their core-design specification. This can be achieved with the use of probabilistic variation and/or through the use of more parametrised variation, which adapts the type of variation based on a given context (van Deemter et al., 2005).

The desirability of having variation in the generated output text very much depends on the nature of the application. Systems where users are only likely to read at most a single generated text have less need for variation than compared to applications where users are expected to read many

generated text over a period time (Reiter and Sripada, 2002). However, in such systems where there is lexical variation there should also be consistency in terminology to avoid confusing readers with differing definitions (Reiter and Sripada, 2002). Work by Foster et al. (2007) has shown that when texts containing variation are evaluated they are strongly preferred and appreciated by human evaluators.

2.3 NLG for Hotel & Restaurant Descriptions

There has been very examples of NLG systems for generating descriptions about hotels. The only recent example is the SuRe system (Tien et al., 2015). The SuRe system generated an abstractive summary that summarised the positive and negative factors for a given accommodation solely from user reviews. Factors such as whether users thought in aggregate that the accommodation’s location, food, etc. was a positive aspect or not. Like many applied NLG systems, it used a standard data-to-text pipeline architecture, but with the key difference being the need to first identify and aggregate opinions from user reviews about features within the given accommodation before performing pipeline based text generation with a surface realiser. However, a majority of the human evaluators found that the system have low levels of grammar quality and the output to be repetitious.

In contrast generating descriptions of restaurants has garnered far more interest within the NLG research community. There has been a considerable amount of past work in generating descriptions or recommendations of restaurants with effort also applied to generating recommendations with stylistic variations (Oraby et al., 2017). The most recent work in this domain has been due to the E2E shared task utilising a crowdsourced dataset of 50k meaning representation instances in the restaurant domain (Novikova et al., 2017). This shared task had brought about the training and construction of a number of systems using either machine learning (ML), rule-based, or template-based approaches to generating short one to three sentence descriptions of a given restaurant. Whilst ML approaches outperformed rule or template based approaches several issues were identified such as hallucinations in the output text, short output length, and low levels of output diversity and syntactic complexity (Dušek et al., 2020).

3 Corpus Analysis

Before deciding on the implementation approach for our system, we first conducted a corpus analysis by examining five existing human written accommodation descriptions (example shown in Table 1) chosen at random. The purpose of this exercise was to gain a better understanding of possible gaps or shortcomings with the available accommodation metadata. Each fact within the human authored corpora is checked to see if there was underlying data present to generate the same or a similar statement as well using an automatic approach.

Overall there were two main issues that were identified after analysing the human authored corpora: The *first* was the lack of specificity in accommodation metadata and the *second* was incomplete data coverage.

The lack of specificity was a significant challenge. Whilst the human authored corpora might mention the name of an in-house hotel restaurant, the actual metadata for a given accommodation would only have binary data points, which indicate the presence of a hotel restaurant and/or café. The encoding of hotel amenities as binary information meant that the metadata would lack details that make it challenging to describe accommodation facilities with significant detail. The second main issue was the incompleteness in data coverage. The analysis of the corpora showed that the corpora contained details that were not represented within the accommodation metadata. This lack of data coverage also affected transportation and places of interest (POI) data as well. In particular, there were issues in regards to the lack of data coverage for nearby transportation stations and a lack of data on nearby POIs for a given accommodation.

As we were building an initial first version of the Hotel Scribe system these data quality issues were not addressed prior to implementation.

4 System Implementation

The Hotel Scribe system was implemented using a standard Data-to-Text NLG pipeline architecture (Reiter, 2007), using SimpleNLG (Gatt and Reiter, 2009) for English surface realisation. Data is read by the system from a database and is mapped onto ontological concepts which represents a taxonomical structure of the hotel accommodation domain. The taxonomical structure of the ontology was devised manually with the assistance of a domain ex-

pert and specifies with increasing levels of granularity all the possible types of entities that can exist in the accommodation domain. For example a Casa Rural *is a* form of self catering which *is a* form of accommodation type.

The key difference from a standard NLG Data-to-Text application was how the Document Planner and Microplanner components of the pipeline functioned. Both of these modules incorporated several variation strategies to help increase the amount of variation in the generated output texts. These strategies included the following types:

- Semantic variation – Varying what content to talk about.
- Content ordering variation – Varying the order of how content is placed.
- Aggregation variation – Varying how and when concepts should be aggregated in a single sentence or not.
- Linguistic variation – Variation in how the concepts are expressed in language.

All four of these variation strategies were executed in both the Document Planner and Microplanner probabilistically. Linguistic variation included variation at a phrase/word level and also REG variation such as varying how an accommodation is referred to in the generated text (e.g. “Hilton Hotel London”, “Hilton”, “accommodation”, “hotel”), which was specified in the 31 syntactic rule specification files to lexicalise the ontological concepts. The combination of these strategies enabled a high level of output variation. For example, the sentence generated by the system to describe 24 hour front desk check-in/check-out services has over 6,000 unique variants.

In addition to implementing the variation strategies the secondary challenge for this system was to make it scalable to generate a large number of texts in a short period of time. This was addressed by making the system multi-threaded therefore capable of generating a number of texts in parallel from a single request. The system was also made scalable and could be deployed to multiple machines at the same time, through the deployment onto scalable cloud computing resources. This in turn allowed many multiple text generation requests to be processed simultaneously. These two optimisations allowed the system to generated up to

| Hotel Scribe Text | Freelancer Text |
|---|---|
| <p>Located in London, the four-star Park Plaza Victoria London is near to Victoria Station.</p> <p>Parents of children should note that there are a number of child friendly amenities including childcare facilities and baby cribs. Meal options are accessible in this hotel through an on-site café and a restaurant. Express check-in/check-out can be done at the 24 hour service counter from 14:00 for check-ins and as late as 12:00 for check-out. In terms of water based facilities this residence includes a pool for guests. Parking amenities comprises of a close by car park, with valet parking. There is complimentary Wifi connectivity within the hotel in both public and in-room hotel areas. For guests travelling on business, this residence features a business centre and conference/meeting rooms.</p> <p>Rooms within this accommodation feature facilities such as a hairdryer, desk, minibar and a telephone. In-room entertainment is available for guests, which is provided by cable TV. Additionally, some rooms come with views of the city.</p> | <p>Situated in the heart of London, just a three-minute walk from Apollo Victoria Theatre, Park Plaza Victoria London boasts a modern, fully equipped fitness centre.</p> <p>A 49-inch flat-screen TV with cable channels, tea and coffee-making facilities, and a desk are provided in each room at Park Plaza Victoria London. The standard apartments boast a kitchenette with a microwave and fridge, while the larger apartments include a balcony.</p> <p>Park Plaza Victoria London offers guests a range of business services. The executive lounge boasts free Wi-Fi and daily newspapers, while the 24-hour front desk offers concierge services and luggage storage.</p> <p>Venetian-Italian dishes are prepared in the on-site TOZI Restaurant, while cocktails can be sipped at Lounge Bar. This hotel also serves breakfast daily, and allows guests to order room service 24 hours a day.</p> <p>Guests staying at here are just a 14-minute walk away from Buckingham Palace, and five minutes' walk from bustling Victoria Station. This hotel is just over a kilometre from Tate Britain.</p> |

Table 1: Text examples for the same hotel between the Hotel Scribe and a freelancer written description.

600,000+ accommodation description texts within a matter of a few days.

5 Evaluation

To better understand the effectiveness and value of the texts generated by the Hotel Scribe we undertook an evaluation of the system. We compared the texts generated by the Hotel Scribe system against accommodation descriptions written from three different sources. The first was accommodation descriptions written by freelancers, the second was from those written directly by the hoteliers, and the third was descriptions written by a direct commercial competitor¹. Corpora for the commercial competitor were collected by automatically scraping accommodation description texts from their website across a random set of cities.

To make the comparison between these four different sources of corpora we used two different evaluation metrics. The first metric was the use of a commercial anti-plagiarism software² to measure the amount of repetition and thus the amount of variation between the four types of corpora. For each source we selected at random around a 1,000 different texts into a private index. Next we selected at random another 200 texts which were used to compute the average “percentage matched” metric for each text against the given private index.

¹Anonymous due to commercial sensitivity.

²Copyscape – <https://www.copyscape.com/>

Similarly, we calculated the Levenshtein edit distance between each of the 200 randomly selected texts and the documents in the private index and we select the lowest edit distance for each of the 200 texts. The calculated average of the minimal edit distances was also a proxy metric for estimating the level of text variation. This evaluation approach is similar to the one undertaken by Foster et al. (2007) to measure the level of textual variation.

Given the level of variation implemented in the Hotel Scribe system, the system was also evaluated for the level of factual correctness as part of a general quality assurance check. This was done with a team of seven human evaluators with each evaluating 13 different accommodation descriptions that were chosen at random to make an evaluated total of 91 descriptions. For each description, the evaluator would count the total number of facts present in the given description and then check to see if the same corresponding fact was also present in the corresponding accommodation webpage. From this a count of the number of *incorrect* and *correct* facts could be derived for each annotator for each of the 91 accommodation descriptions. These counts of *incorrect* and *correct* facts enabled the calculation of an accuracy score at a per accommodation description level and also an average accuracy score for over all descriptions.

| Text Source Type | Avg. Copyscale “percent matched” (lower is better) | Avg. Levenshtein edit distance (higher is better) |
|---------------------------|---|--|
| Freelancer Descriptions | 4.09 (σ 10.66) | 647.88 (σ 69.79) |
| Hotelier Descriptions | 0.842 (σ 6.26) | 602.48 (σ 341.64) |
| Competitor Descriptions | 32.43 (σ 25.81) | 345.35 (σ 150.68) |
| Hotel Scribe Descriptions | 44.02 (σ 11.39) | 296.25 (σ 104) |

Table 2: Results for both Copyscale and Levenshtein edit distance metrics. Standard deviation is shown in brackets.

6 Results

The results are shown above in table 2. From the results obtained it’s sufficiently clear that the accommodation descriptions written by the Freelancers and Hoteliers contain considerably more variation than texts generated by the Hotel Scribe system and those written by the commercial competitor. This is not unexpected as both Freelancers and Hoteliers were unconstrained from limiting themselves from writing their descriptions from only the database accommodation metadata and were free to use external information resources. Therefore resulting in descriptions that are much richer in detail and more unique in comparison to the automatically generated texts.

What was interesting is the finding showing the near-comparable performance between the Hotel Scribe system and the direct commercial competitor. Whilst texts from the commercial competitor out-performed our system for both the Copyscape and Levenshtein edit distance metrics the difference is small. This is a considerable result given the limitations in data as described in section 3 and the short development time of only a few months. However, the large gap in performance between the commercial competitor and the texts written by humans (Freelancer and Hoteliers) may possibly indicate that they are also using an automated approach to generate their texts or a hybrid approach with humans post-editing the texts. However, this cannot be known for certain.

The Hotel Scribe system was also evaluated for factual correctness and the average score between the seven judges was 84%. Some of the discrepancies were due to errors in the input data and others were due to software bugs in system, which were subsequently fixed in later revisions of the system.

7 Conclusion

In this paper we described an approach for generating accommodation descriptions with a large num-

ber of textual variations and evaluated this against other types of corpora. Whilst, the system does not have the level of performance as human written corpora in terms of uniqueness and variation the fact that it has a near-comparable performance to a direct competitor is highly encouraging.

The discrepancy in performance between our system and human written corpora indicates a greater need to have more detailed accommodation metadata with greater *breadth* and *depth*, which will enable our system to generate more unique descriptions about the amenities/facilities found in a given accommodation. We have put our system into production to cover accommodations that have no existing human written description as shown in figure 1. Going forward, we will continue to refine its capabilities and performance in the future.

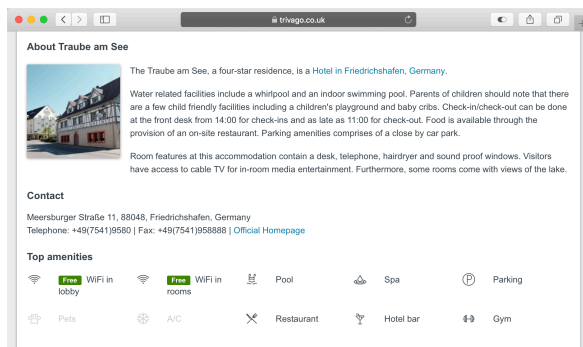


Figure 1: A live example of an accommodation description displayed on the trivago website.

Acknowledgements

Many thanks for the following people who provided considerable help and support during the development of this project: *Annachiara Foschi, Anja Klassen, Doreh Taghavidinani, Janine Cushnie, Allyson Clark, Agata Abram, Roman Orishko, and Nick Dale.*

References

- David Caswell and Konstantin Dörr. 2018. [Automated Journalism 2.0: Event-driven narratives](#). *Journalism Practice*, 12(4):477–496.
- Robert Dale. 2019. [NLP commercialisation in the last 25 years](#). *Natural Language Engineering*, 25(3):419–426.
- Laurence Danlos, Frédéric Meunier, and Vanessa Combet. 2011. EasyText: an Operational NLG System. In *ENLG 2011 - 13th European Workshop on Natural Language Generation*.
- Kees van Deemter, Emiel Krahmer, and Mariët Theune. 2005. Real versus Template-Based Natural Language Generation: A False Opposition? *Computational Linguistics*, 31(1):15–23.
- Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. 2020. Evaluating the state-of-the-art of end-to-end natural language generation: The E2E NLG Challenge. *Computer Speech & Language*, 59:123–156.
- Mary Ellen Foster and Michael White. 2007. Avoiding repetition in generated text. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 33–40. Association for Computational Linguistics.
- Albert Gatt and Ehud Reiter. 2009. SimpleNLG: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, pages 90–93.
- Andreas Graefe. 2016. Guide to Automated Journalism. *Tow Center for Digital Journalism*.
- Mary Dee Harris. 2008. Building a large-scale commercial NLG system for an EMR. In *Proceedings of the Fifth International Natural Language Generation Conference*, pages 157–160.
- Andreas Madsack, Alessia Cavallo, Johanna Heining, and Robert Weißgraeber. 2018. AX Semantics’ Submission to the CoNLL-SIGMORPHON 2018 Shared Task. In *Proceedings of the CoNLL SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, pages 43–47.
- Saad Mahamood, William Bradshaw, and Ehud Reiter. 2014. Generating Annotated Graphs using the NLG Pipeline Architecture. In *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, pages 123–127.
- Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017. The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206, Saarbrücken, Germany. Association for Computational Linguistics.
- Shereen Oraby, Sheideh Homayon, and Marilyn Walker. 2017. Harvesting creative templates for generating stylistically varied restaurant reviews. In *Workshop on Stylistic Variation at EMNLP 2017*.
- Ehud Reiter. 2007. An architecture for data-to-text systems. In *Proceedings of the 11th European Workshop on Natural Language Generation (ENLG’07)*.
- Ehud Reiter. 2017. A Commercial Perspective on Reference. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 134–138.
- Ehud Reiter and Somayajulu Sripada. 2002. Human Variation and Lexical Choice. *Computational Linguistics*, 28:545–553.
- Somayajulu Sripada, Neil Burnett, Ross Turner, John Mastin, and Dave Evans. 2014. A case study: NLG meeting weather industry demand for quality and quantity of textual weather forecasts. In *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, pages 1–5.
- Somayajulu G Sripada, Ehud Reiter, Ian Davy, and Kristian Nilssen. 2004. Lessons from deploying NLG technology for marine weather forecast text generation. In *Proceedings of the 16th European Conference on Artificial Intelligence*, pages 760–764.
- M. Theune, E. Klabbers, J. Odijk, J.R. de Pijper, and E. Krahmer. 2001. From Data to Speech: A General Approach. *Natural Language Engineering*, 7(1):47–86.
- Minh Tien, François Portet, and Cyril Labbé. 2015. [Hypertext Summarization for Hotel Review](#).