# **Natural Language Generation from RDF Triples**

**Tathagata Dey** 

Department of Computer Science and Engineering, Indian Institute of Technology Bombay tathagata@cse.iitb.ac.in

#### Abstract

001 RDF to text generation is one of the cuttingedge research problems in Natural Language 003 Processing. The problem statement includes the generation of coherent, consistent and faithful human-understandable text from a given set of RDF triples. There has been considerable progress in this problem, however, none 007 of them could achieve perfection. One of the major challenges faced by the models was the graphical structure and information storing of 011 RDFs. In our work, we have tried different architectures to achieve this goal. We have produced results of prompting on GPT models and described our own concept of pipelined approach. The new contributions from our work, evaluation metrics and generated sentences show that in future this idea can be the state-of-the-art approach to this problem.

#### 1 Introduction

019

021

034

One of the prime objectives to work with Natural Language Processing is to automate tasks which consume which follow a pattern and require essential human resources and energy. Although, there are numerous difficulties to deal with in building such a system. Data-to-text generation is one of the most popular problem statements of modern NLP. The twenty-first century is a data-driven world. Starting from the research world to ending development models, everything is made of data. The research world nowadays follows the data observation approach, where any hypothesis is tested with lots of real-world data, hence computer science is leaving a long impact on many other domains. On the other hand, real-world systems are also consisting of huge amounts of data these days. Webpages, software, databases etc. are always full of information. Hence, processing this information, and representing it in a better format can increase the efficiency of systems to quite a few extent.

> A very useful and popular way to store and deal with data is RDF. RDF is Resource Description





Framework. It is a triple consisting of <i>Subject</i> , <i>Relation</i> and <i>Object</i> .	042 043
<dumdum airport,="" kolkata="" servescity,=""></dumdum>	044
In the example shown above, <i>Dumdum Airport</i> is the subject and <i>Kolkata</i> is the object which is	045 046
connected by the relation <i>servesCity</i> . The sentence generated from this RDF can be expressed as	047 048
Kolkata is served by Dumdum Airport.	049
The objective of converting RDFs to natural lan- guage can be useful in many different cases. Hence,	050 051
research has been progressing on this for a long	052
visualised as shown in figure 1.	053 054
The objective of this task is to generate a sen-	055
example, the generated sentence can look like as follows,	057 058
Dumdum Airport, which is also known as Netaji Subhas Chandra Bose Airport serves the city of Kolkata.	059 060 061
There has not been a significant amount of con- tribution by researchers in this domain. Over a	062 063
decade, different researchers with different models have tried to solve this problem. However, human evaluation show that the coherence and faithfulness	064 065 066

of the generated sentences are yet to be improved a

067

lot.

Charl	es B. Winstead
Born	May 25, 1891
Died	August 3, 1973 (aged 82) Albuquerque, New Mexico
Cause of death	pneumonia
Nationality	American
Occupation	FBI Agent
Employer	FBI
Known for	Shooting John Dillinger
Title	Special Agent

Figure 2: Wikipedia Infobox

# 2 Motivation

069

071

077

079

082

094

100

101

102

RDF triples are useful to represent many different data structures across the web and also in our local systems. RDF builds a schema to represent information in general. It can be used to define the information of a table, Wikipedia, knowledge graphs and much more.

Many of us are familiar with the box shown in figure 2 ((Liu et al., 2017)). It is the info box that appears on Wikipedia pages and is also described mostly in the first paragraph of that page. This table can be represented in the form of RDF triples. For example,

```
<Charles B. Winstead, nationality,
American>
```

is an RDF triple which represents a particular piece of information. Hence, converting this table into a paragraph can also be solved using RDFs. Similarly, numerous different objectives can be achieved by the use of RDFs.

### **3** Literature Survey

There has not been a significant amount of contribution in this domain. So, the problem is not completely solved yet. However, some notable contributions are there which should be mentioned in detail.

Graph to Text Generation. (Gao et al., 2020)
has worked on this problem using two encoders.
They used a Bidirectional Graph-based Meta-Paths
Encoder and a Bidirectional Graph Convolutional
Networks Encoder. The decoding part was performed by a usual transformer decoder. Figure 3
shows the architecture that has been used in this work.

The authors performed entity masking on the dataset. Entity masking replaces the *NOUNs* with their entity which teaches the model to not focus on names and form sentences in the same way when the same entity is given as input. The work has been performed on WebNLG 2017 challenge dataset and it achieved a BLEU score of 57.09 and a METEOR score of 43. It is regarded as one of the state-of-the-art performances. However, in many cases, the generated sentences are not necessarily coherent.

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

There have been some other works with graphbased frameworks. Such as (Zhao et al., 2020) paper shows a way of dual encoding to tackle the problem of graph-based information. The authors first developed a GCN Encoder to capture the graphical information and inter-relation between nodes. Then the data is passed through an LSTM encoder to order the sentence formation. Dual encoding helps in both understanding the information and structuring them.

The decoding is performed by an attention-based LSTM decoder. They have achieved a BLEU score of 36.73 on unseen data with a METEOR score of 0.37. The authors also performed a human evaluation to judge the coherence and consistency of the generated text. The evaluation metrics provide over 90% score in both *Coverage* and *Faith*.

(Puduppully et al., 2019) designed a model with content planning and surface realisation to achieve the goal. The model has a record encoder which takes input as an unordered table format of data. The encoder uses a multi-layer perceptron to determine the information of encoding of the records. The second part of the architecture is a content selection gate which analyses the dependencies between records. For example, in the marks of the student dataset, there can be a dependency between total marks and subjectwise marks. This dependency is planned to represent through a content selection gate. Next content planning surface realises the sentential forms.

**Pre-trained Language Models**. (Ribeiro et al., 2021) used the BART language model and T5 to achieve the goal in different objective setups. T5 and BART are both based on Transformer encoder-decoder architecture (Vaswani et al., 2017). The authors first used the T5 model with prefix tuning. The prefix added was *translate graph to text*. They also performed the same task on the BART model. The supervised task adaptation with fine-tuning helped the models to understand the pattern in the



Figure 3: Architecture of (Gao et al., 2020)

generated task. With BART and T5 they were ableto achieve 43.97, 53.67 BLEU scores and 38.61,42.26 METEOR scores respectively.

## 4 Task Definition

154

155

156

157

158

159

160

161

163

164

165

167

170

171

172

173

The problem statement of conversion of RDF to text is defined as follows. For input there is a set of n RDFs as  $s = \{r_1, r_2, ..., r_n\}$  where each RDF is a triplet of Subject, Relation and Object form, i.e.,  $r_i = \langle s_i, r_i, o_i \rangle$ :  $\forall i$  where  $s_i, r_i, o_i$  refers to Subject, Relation and Object of the ith RDF triple. The objective of the task is to generate a concise text to represent information denoted in the set s. The text t has to be a set of sentences. Every sentence  $w_i \in t$  has to be coherent and consistent towards the whole output. If t is defined as  $t = \{w_1, w_2, ..., w_k\}$  then  $k \le n$ . This means the generated text is supposed to contain sentences in less or equal numbers to the rdf triplet numbers. The objective also signifies that generating more sentences to convey information described in less number of triplets is not an efficient system at all,

In some cases, the RDFs are also seen as graphbased information. So, we must comprehend 176 the definition from this perspective too. In this 177 case, the set of n RDFs is denoted in the form of a graph, say graph G. Now, G is defined as, 179 G = (V, E) where V and E represents the set 180 of vertices and edges respectively. Here, V =181  $\{s_1, s_2, ..., s_n, o_1, o_2, ..., o_3\}$  where  $\forall i : s_i$  and  $o_i$ represents Subject and Object in the nth RDF triple. On the other hand,  $E = \{r_1, r_2, ..., r_n\}$  where 184  $\forall i : r_i$  denotes the relation in triple *i*. Now, in 185 some cases the *subject* and *object* regarding a particular relation may not be  $s_i$  and  $o_j$  respectively. 187

Subject for relation  $r_k$  may be Object for relation  $r_p$  and vice versa. Hence, sometimes, the set V is regarded as a set of entities and all the Subject and Object nodes are regarded as entity nodes.

188

189

190

191

192

193

194

195

196

197

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

# **5** Datasets

There are many datasets available to perform this task. A very popular dataset is the WebNLG challenge 2017 corpus (Gardent et al., 2017)<sup>1</sup>. This corpus is designed for the task of RDF-to-text conversion. The dataset contains multiple files of rdf from different genres, such as airports, universities, books and so on. Each file is a JSON file containing *<triple, texts>* format. There can be multiple texts for the same triple to train the model for different objectives. Each triple is a set of *<Subject, Relation, Object>* structure. The enriched version of WebNLG 2017 challenge (Castro Ferreira et al., 2018) is also published.

The table 1 shows the number of train, validation and test instances in the dataset. There is no overlap between WebNLG 2017 challenge and the supplementary data, which combined forms the Enriched dataset.

Dart (Nan et al., 2021) is another popular dataset <sup>2</sup> which contains around 82191 instances of RDF triples. However, in this task, we have used WebNLG corpus only to perform the training and measure the evaluations.

### 6 Methodology

Many different models can be used to solve this problem Sequence to sequence models are very

<sup>&</sup>lt;sup>1</sup>https://gitlab.com/shimorina

<sup>&</sup>lt;sup>2</sup>https://github.com/Yale-LILY/dart

Dataset	Train	Validation	Test
WebNLG 2017 Challenge	18102	2268	2495
Supplementary Data	13867	1762	1727
WebNLG 2017 Challenge Enriched	31969	4030	4222

Table 1: WebNLG Challenge Dataset

Aarhus_	_Air	port   cityServed	"Aarhus,	mtriple Denmark"	The Aarhus	is the	airport	of Aarhus,	lex Denmark.	comment good
		aarhus airport city Se	rved aarhu	s denmark	The Aarhus i	s the ai	rport of Aa	arhus, Denn	nark.	

Figure 4: Preprocessing of Dataset

popular in text generation tasks. These models are based on the Transformer-based encoder-decoder architecture (Vaswani et al., 2017). T5 (Raffel et al., 2020) is another popular model which is designed with some customization on the transformer decoder.

### 6.1 Preprocessing

219

220

224

227

228

232

233

238

240

241

242

243

244

245

246

247

248

249

250

251

Preprocessing is done on the WebNLG corpus. The RDF triples and the sentences are extracted from the json files. The steps of preprocessing include *Punctuation Removal, Removing Underscores* and finally *Lexicalisation*. As we can see in figure 4, the unstructured data contained punctuations, *Underscores* and even combined words in relation, which are then all dealt with the proper step-bystep modifications.

Now our dataset contains two columns of *mtriples* and *lex* with preprocessed data which is perfect to be fed into the model. The subject, relation and objects are separated with separate tags such as **<H>** tag for the subject, **<R>** tag for the relation and **<T>** tag for the object.

For example, the triple

<Aarhus\_Airport, servesCity, Aarhus\_Denmark>

is preprocessed as

<H> Aarhus Airport <R> city Serves <T> Aarhus Denmark

#### 6.2 Sequence to Sequence Models

We have used a sequence-to-sequence model to perform the task. We chose T5 model (Text-to-Text-Transfer-Transformer) (Raffel et al., 2020). T5 is trained to perform any text-to-text conversion. It is also acclaimed to achieve state-of-the-art results in these tasks. It is an **11 billion parameter** model which performs in GLUE, SuperGLUE and many of the other benchmark challenges.

T5 is a transformer-based model with modifications on the decoder blocks. The model is trained on the **Colossal Clean Crawled Corpus** (Dodge et al., 2021) which is also known as **Common Crawl** dataset. This dataset is created by crawling webpages and then removing unnecessary lines. The corpus contains refined English sentences and it is of size **700 GB**. T5 is also trained on BERT (Devlin et al., 2019) objective of the Masked Language Model. This training is useful for translation, summarization and generation tasks.

The T5 model is fine-tuned on the WebNLG corpus with prefix-tuning. The training is done with *Adafactor* optimizer at a learning rate of  $e^{-3}$ , batch size of 1 and *Cross Entropy Loss*. The losses per epoch are shown in figure 5.

However, the results for a sequence-to-sequence model are not sound. The reason is the architecture of a sequence-to-sequence model and graph information structure, We have already seen that RDF triples are a knowledge graph, i.e., they store information in the form of a graph. But the T5 model takes sequential input. Hence, to fit in the input the graph is flattened out, which misses some necessary pieces of information. This problem is shown in figure 6.

### 6.3 Prompting in OpenAI Models

We have tried prompting on OpenAI models GPT 3.5- turbo (Ye et al., 2023) and Text-da-Vinci. GPT 3.5 turbo is a model by OpenAI of the series GPT 3.5 which also powers ChatGPT. This model can be accessed through API keys and accepts textual 253

254

255

256

257



Figure 5: Loss per 1000 steps of T5 Model

input. The maximum token capability of this model is 4096 and is trained on data till September 2021. On the other hand, text-da-vinci is designed specifically for the task of following instructions. The 003 version also supports longer context than previous versions of text-da-vinci. It is popularly regarded as one of the best when used for zero-shot prompting on specific tasks.

We have performed using various prompts such as,

- 1. "Convert to English sentence:"
- 2. "Graph to English conversion:"
- 3. "Convert RDF to natural language:"

, and so on. We found the best results with "Convert RDF to natural language:" which is further used over the whole corpus.

#### 6.4 Pipeline Architecture

296

297

307

309

311

313

314

315

318

319

Our Computation For Indian Language Technology Lab of IIT Bombay is working on a new pipelined approach. In this idea, we are dividing the problem statement into two subtasks. The motivation behind this idea is the inability of sequence-to-sequence learners to interpret graphical information.

Often sequence-to-sequence models fail to find the interrelationship between different entities. As we flatten out a graph and pass it to a sequence learner, some of the dependencies get disregarded. For a sequence, the factor of correlation is a function of context length, however, a particular object can be in context with many subjects simultaneously which cannot be represented in a sequence. Hence, models like T5 fail in some cases. However, when we talk about a single triple, an object can be related to one subject only, Hence, a sequence is sufficient to convey the information. So, the sequence-to-sequence models are seen to generate sound pieces of texts on 1-triples.

322

323

324

325

326

327

328

329

330

331

332

333

We have divided the problem statement into two parts. The first is about generating text from the 1-triples. And in the next step, all the generated sentences are passed through a concise text generator to generate summed-up information.

Figure 7 shows the architecture of this pipeline. The example below shows

<dumdum airport,="" kolkata="" servescity,=""></dumdum>	334
<kolkata, bengal="" capitalof,="" west="">  </kolkata,>	335
Dumdum Airport, alsoKnownAs, Netaji	336
Subhas Chandra Bose Airport>	337

1. Dumdum Airport serves the city of Kolkata.	338
2. Kolkata is the capital of West Bengal.	339
3. Dumdum Airport is also known as Netaji	340
Subhas Chandra Bose Airport.	341

 $\downarrow$ 

 $\downarrow$ 

Dumdum Airport, which is also known as	342
Netaji Subhas Chandra Bose Airport	343
serves the city of Kolkata. It is the	344
capital of West Bengal.	345

For the first stage, the T5 model is more fine-tuned.	346
However, for the second stage a summariser is	347
used for now. A lot of progress is yet to be	348
achieved in stage 2.	349

ł



Figure 6: Problem in Sequence-to-sequence model

### 7 Results

351

352

The results also denote the progress we had over time in understanding the right path. First, we tried T5 to train on the webNLG corpus. However, the results not being satisfactory led us to try new approaches.

#### 7.1 Sequence-to-sequence Model

The sequence-to-sequence model fails in many 357 cases to generate a coherent sentence. For single triples the generated sentences are most of the time faithful and coherent however for multi-triples, the model fails. The generated sentences are not always complete for large number of triples. Also 362 in some cases, it is observed that the model only captures the first and the last triple information and misses in between ones. The WebNLG corpus carries as large as seven triples together. However, the performance of the model starts dropping from three triple onwards. The reason is suspected to be the graph structure of the information. The figures 8, 9, 10 and 11 represent the decrease in mteric scores with increasing RDF complexity. The BLEU (Papineni et al., 2002), METEOR

(Banerjee and Lavie, 2005) and ROUGE (Lin, 2004) score metrics are shown in table 2. Some of the examples of generated outputs are also shown in figure 12.

374

375

376

378

379

381

382

383

384

385

386

387

389

390

391

392

393

394

### 7.2 Prompting

Prompting in OpenAI models led to the generation of coherent and faithful text from the RDF triples. Table 3 shows that the text-davinci-003 model achieved a higher score in all of the metrics. The higher BLEU and ROUGE scores signify the overlap between the generated text and reference text. The METEOR score also represents the coherence in the generated text. Figure 13 shows different examples of a given input, expected output and the generated sentences. The input here is separated by |. We can see that both models perform soundly on the input data. However, the texts from text-davinci-003 model seem to be more coherent in terms of lexicalization.

# 7.3 Pipeline Approach

The pipeline approach has a lot to evolve. For now, in stage 2 only a BART summariser has been used.



Figure 7: Pipeline approach of CFILT Lab

Model Name	BLEU Score	METEOR Score	ROUGE-F1	ROUGE-L
T5	0.338373	0.331908	0.447568	0.359282

Table 2: Evaluation metrics of T5



Figure 8: Bleu Score Distribution over test corpus of WebNLG challenge



Figure 10: Rouge-f1 Score Distribution over test corpus of WebNLG challenge



Figure 9: Meteor Score Distribution over test corpus of WebNLG challenge



Figure 11: Rouge-L Score Distribution over test corpus of WebNLG challenge

Model Name	BLEU Score	METEOR Score	ROUGE-F1	ROUGE-L
GPT 3.5-turbo	0.605460	0.648986	0.678951	0.505117
Text-davinci-003	0.664163	0.651951	0.712198	0.540051

Table 3: Evaluation metrics on openAI

The results sometimes are correct while many 396 times not coherent. The metric scores are shown in table 4. The METEOR score suggests that the generated sentences are sound but do not have a 400 perfect overlap with reference. One of the prime reasons can be the inability of summariser models. 401 402 The summariser is not fine-tuned due to the unavailability of data. Besides, the pre-trained 403 summariser models are not trained to perform this 404 kind of task. The sentences here are full of lots of 405 information and concising them is not a 406 summarising objective. Hence, the model seems to 407 408 fail in some cases. Some of the example of generated sentences are shown in figure 14. 409

# 8 Conclusion

410

425

In this work, different approaches have been taken 411 to solve the RDF to natural language generation 412 challenge. However, not all of them are found to 413 be suitable. Sequence-to-sequence models didn't 414 perform so well due to the graphical structure of 415 416 information in RDFs. However, a newer approach has been taken to use a two-staged pipeline to 417 perform the task. As the result suggests, it is 418 performing much better than the 419 sequence-to-sequence model. Also, prompting on 420 GPT has been proven to be much useful. Critical 421 analysis of generated results also depicts the 422 coherence in the sentences. However, there is a lot 423 of progress to be made in this domain. 424

### 9 Future Work

The two-staged pipeline has not been perfectly 426 designed yet. Stage one consists of 427 sequence-to-sequence learner which will convert 428 single RDF triples into a coherent faithful 429 sentence. Our model works nearly perfectly in that 430 regard. However, stage two, where the concise text 431 generator is supposed to get all the sentences and 432 generate a concise paragraph is not at the 433 state-of-the-art yet. We have used a BART 434 summariser trained on Facebook data. However, 435 this model is not trained to tackle such 436 information-rich sentences and then compress 437 them. No summariser is designed for this task. 438

Rather a fine-tuned summariser could have 439 performed much better. 440 However, CFILT Lab is trying to use prompting 441 for stage two. Few shot learning on language 442 models can perform much better than summarisers 443 in this objective. 444 Acknowledgements 445 The study and research have been performed in 446 **Computation For Indian Language Technology** 447 Lab of IIT Bombay under the guidance of Prof. 448 Pushpak Bhattacharyya. Also thankful to sincere 449 collaboration with Mr Utkarsh Ranjan over the 450 research time regarding this problem statement. 451 References 452 Satanjeev Banerjee and Alon Lavie. 2005. METEOR: 453 An automatic metric for MT evaluation with improved 454 correlation with human judgments. In Proceedings of 455 the ACL Workshop on Intrinsic and Extrinsic 456 Evaluation Measures for Machine Translation and/or 457 Summarization, pages 65–72, Ann Arbor, Michigan. 458 Association for Computational Linguistics. 459 Thiago Castro Ferreira, Diego Moussallem, Emiel 460 Krahmer, and Sander Wubben. 2018. Enriching the 461

Krahmer, and Sander Wubben. 2018. Enriching the WebNLG corpus. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 171–176, Tilburg University, The Netherlands. Association for Computational Linguistics.

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. 2021. Documenting large webtext corpora: A case study on the colossal clean crawled corpus.

Hanning Gao, Lingfei Wu, Po Hu, and Fangli Xu. 2020. Rdf-to-text generation with graph-augmented structural neural encoders. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 3030–3036. International Joint Conferences on Artificial Intelligence Organization. Main track.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. The WebNLG

Model Name	BLEU Score	METEOR Score	ROUGE-F1	ROUGE-L
Pipeline Approach	0.438041	0.467084	0.518574	0.38243

Table 4: Evaluation metrics on Pipeline Approach

484	challenge: Generating text from RDF data. In
485	Proceedings of the 10th International Conference on
486	Natural Language Generation, pages 124–133,
487	Santiago de Compostela, Spain. Association for
488	Computational Linguistics.
489	Chin-Yew Lin. 2004. ROUGE: A package for
490	automatic evaluation of summaries. In <i>Text</i>
491	<i>Summarization Branches Out</i> , pages 74–81, Barcelona,
492	Spain. Association for Computational Linguistics.
493	Tianyu Liu, Kexiang Wang, Lei Sha, Baobao Chang,
494	and Zhifang Sui. 2017. Table-to-text generation by
495	structure-aware seq2seq learning.
496	Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau,
497	Abhinand Sivaprasad, Chiachun Hsieh, Xiangru Tang,
498	Aadit Vyas, Neha Verma, Pranav Krishna,
499	Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz
500	Rahman, Ahmad Zaidi, Mutethia Mutuma, Yasin
501	Tarabar, Ankit Gupta, Tao Yu, Yi Chern Tan,
502	Xi Victoria Lin, Caiming Xiong, Richard Socher, and
503	Nazneen Fatema Rajani. 2021. Dart: Open-domain
504	structured data record to text generation.
505	Kishore Papineni, Salim Roukos, Todd Ward, and
506	Wei-Jing Zhu. 2002. Bleu: A method for automatic
507	evaluation of machine translation. In <i>Proceedings of</i>
508	<i>the 40th Annual Meeting on Association for</i>
509	<i>Computational Linguistics</i> , ACL '02, page 311–318,
510	USA. Association for Computational Linguistics.
511	Ratish Puduppully, Li Dong, and Mirella Lapata. 2019.
512	Data-to-text generation with content selection and
513	planning.
514	Colin Raffel, Noam Shazeer, Adam Roberts, Katherine
515	Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei
516	Li, and Peter J. Liu. 2020. Exploring the limits of
517	transfer learning with a unified text-to-text transformer.
518	Leonardo F. R. Ribeiro, Martin Schmitt, Hinrich
519	Schütze, and Iryna Gurevych. 2021. Investigating
520	pretrained language models for graph-to-text
521	generation. In <i>Proceedings of the 3rd Workshop on</i>
522	<i>Natural Language Processing for Conversational AI</i> ,
523	pages 211–227, Online. Association for Computational
524	Linguistics.
525	Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob
526	Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz
527	Kaiser, and Illia Polosukhin. 2017. Attention is all you
528	need.
529	Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai
530	Shao, Shichun Liu, Yuhan Cui, Zeyang Zhou, Chao
531	Gong, Yang Shen, Jie Zhou, Siming Chen, Tao Gui,
532	Qi Zhang, and Xuanjing Huang. 2023. A
533	comprehensive capability analysis of gpt-3 and gpt-3.5
534	series models.

Chao Zhao, Marilyn Walker, and Snigdha Chaturvedi.5352020. Bridging the structural gap between encoding536and decoding for data-to-text generation. In537Proceedings of the 58th Annual Meeting of the538Association for Computational Linguistics, pages5392481–2491, Online. Association for Computational540Linguistics.541

Trial 1 Input: abilene regional airport city Served abilene texas Predicted: Abilene Regional Airport serves the city of Abilene, Texas. Expected: Abilene, Texas is served by the Abilene regional airport.

Trial 2 Input: allama iqbal international airport runway Length 3360.12 Predicted: The runway length of Allama Iqbal International Airport is 3360.12 in length Expected: The runway length at Allama Iqbal International Airport is 3,360.12.

Trial 3 Input: amarillo texas area Total 233.9 (square kilometres) Predicted: Amarillo, Texas has a total area of 233.9 SqK.D Expected: The total area of Amarillo Texas is 233.9 SqKm.

Trial 4 Input: arròs negre region valencian community Predicted: Arros negre comes from the region of the Valencian Community, Spain. Expected: Arros negre comes from the region of the Valencian Community.

Figure 12: Generated Outputs on T5

Trial 1 Input: Synthpop | stylisticOrigin | Pop\_music

Expected Output: Synthpop is a form of pop music.

Chatgpt Generated: Synthpop has its stylistic origin in pop music.

text-davinci-003 Generated: Synthpop is a style of music which originates from Pop mu sic.

Trial 4 Input: Arlington,\_Texas | isPartOf | Texas , Texas | country | United\_States , , Texas | largestCity | Houston

Expected Output: Arlington in the United States is located in the state of Texas wher e the largest city is Houston.

Chatgpt Generated: Arlington, Texas is a part of the state of Texas. Texas is a count ry within the United States. Houston is the largest city in Texas.

text-davinci-003 Generated: The city of Arlington, Texas is part of the U.S. state of Texas, which is itself part of the United States. Texas' largest city is Houston.

Trial 6

Input: Antwerp\_International\_Airport | owner | Flemish\_Region , Antwerp\_Internation
al\_Airport | operatingOrganisation | "Flemish department of Mobility and Public Works
" , , Antwerp\_International\_Airport | cityServed | Antwerp , , Antwerp\_Internatio
nal\_Airport | elevationAboveTheSeaLevel\_(in\_metres) | 12.0

Expected Output: Antwerp International Airport is owned by the Flemish Region and ope rated by the Flemish department of mobility and public works. It serves the city of A ntwerp and is 12 metres above sea level.

Chatgpt Generated: The Antwerp International Airport is owned by the Flemish Region a nd is operated by the Flemish department of Mobility and Public Works. It serves the city of Antwerp and sits at an elevation of 12.0 metres above sea level.

text-davinci-003 Generated: The Flemish Region is the owner of the Antwerp Internatio nal Airport. It is operated by the Flemish department of Mobility and Public Works an d serves the city of Antwerp. The elevation of the airport above the sea level is 12. 0 metres.

Figure 13: Prompting on OpenAI Models

Input: BBC | locationCity | Broadcasting\_House && Bananaman | starring | Bill\_Oddie &
& Bananaman | network | BBC && Bananaman | firstAired | "1983-10-03"^^xsd:date && BBC
| city | Broadcasting\_House && Bananaman | starring | Bill\_Oddie && Bananaman | broa
dcastedBy | BBC && Bananaman | firstAired | "1983-10-03"

Output: Bananaman is a TV series shown on the BBC.Bill Oddie stars in Bananaman.The l ocation of the BBC is Broadcasting House in London. Bananaman first aired on 3 Octobe r 1983.

Trial: 2

Input: German\_Empire | anthem | Die\_Wacht\_am\_Rhein && German\_Empire | anthem | Die\_Wa
cht\_am\_Rhein

Output: The Empire's anthem is Die Wacht am Rhein. The Empire's flag is made up of th ree colours: blue, white and red. The flag is flown over the Empire's capital city, B erlin.

Trial: 3

Input: University\_of\_Burgundy | numberOfUndergraduateStudents | 16800 && University\_o
f\_Burgundy | staff | 2900 && University\_of\_Burgundy | numberOfStudents | 27400 && Uni
versity\_of\_Burgundy | campus | Dijon && University\_of\_Burgundy | numberOfUndergraduat
eStudents | 16800 && University\_of\_Burgundy | staff | 2900 && University\_of\_Burgundy
| numberOfStudents | 27400 && University\_of\_Burgundy | campus | Dijon

Output: Dijon is the campus of the University of Burgundy.The University has 16800 un dergraduate students.It employs 2900 people and has 27400 students. It is located in the town of Dijon in the Ardennes.

Trial: 4

Input: Darinka\_Dentcheva | influencedBy | Andrzej\_Piotr\_Ruszczyński && Darinka\_Dentch eva | citizenship | United\_States && Darinka\_Dentcheva | birthDate | 1958-01-01 && Da rinka\_Dentcheva | knownFor | Stochastic\_programming && Darinka\_Dentcheva | influenced By | Andrzej\_Piotr\_Ruszczyński && Darinka\_Dentcheva | citizenship | United\_States && Darinka\_Dentcheva | birthDate | 1958-01-01 && Darinka\_Dentcheva | knownFor | Stochast ic\_programming

Output: The birth date of Darinka Dentcheva is 1958-01-01. She is a citizen of the Un ited States. She was influenced by Andrzej Piotr and is known for Stochastic programm ing.

Trial: 5

Input: BBC | locationCity | Broadcasting\_House && Bananaman | starring | Bill\_Oddie &
& Bananaman | creator | Steve\_Bright && Bananaman | network | BBC && Bananaman | firs
tAired | "1983-10-03"^^xsd:date && BBC | city | Broadcasting\_House && Bananaman | sta
rring | Bill\_Oddie && Bananaman | creator | Steve\_Bright && Bananaman | broadcastedBy
| BBC && Bananaman | firstAired | "1983-10-03"

Output: Bananaman is a TV series shown on the BBC.The location of the BBC is Broadcas ting House in London.Bill Oddie stars in Bananaman.The series first aired on 3 Octobe r 1983.

Figure 14: Generated sentences from Pipeline Approach