MLBMIKABR at “Discharge Me!”: Concept Based Clinical Text Description Generation

Abir Naskar
Computing and Information Systems
University of Melbourne
Parkville VIC 3052
AUSTRALIA
anaskar@student.unimelb.edu.au

Jane Hocking
Population and Global Health
University of Melbourne
Parkville VIC 3052
AUSTRALIA
j.hocking@unimelb.edu.au

Patty Chondros
General Practice and Primary Care
University of Melbourne
Parkville VIC 3052
AUSTRALIA
p.chondros@unimelb.edu.au

Douglas Boyle
General Practice and Primary Care
University of Melbourne
Parkville VIC 3052
AUSTRALIA
dboyle@unimelb.edu.au

Mike Conway
Computing and Information Systems
University of Melbourne
Parkville VIC 3052
AUSTRALIA
mike.conway@unimelb.edu.au

Abstract

This paper presents a method called Concept Based Description Generation, aimed at creating summaries (“Brief Hospital Course” and “Discharge Instructions”) using source (“Discharge” and “Radiology”) texts. We propose a rule-based approach for segmenting both the source and target texts. In the target text, we not only segment the content but also identify the concept associated with each segment based on text patterns. Our methodology involves creating a combined summarized version of each text segment, extracting important information, and then fine-tuning a Large Language Model (LLM) to generate aspects. Subsequently, we fine-tune a new LLM using a specific aspect, the combined summary, and a list of all aspects to generate detailed descriptions for each task. This approach integrates segmentation, concept identification, summarization, and language modeling to achieve accurate and informative descriptions for medical documentation tasks.

1 Introduction

The “Discharge Me!” (Xu et al., 2024) task within the BioNLP workshop at the Annual Meeting of the Association for Computational Linguistics (ACL) 2024 aims to automate the generation of “Brief Hospital Course” and “Discharge Instructions” sections in discharge notes. These notes are derived from a subset of the MIMIC-IV-Note (Johnson et al., b) and MIMIC-IV-ED (Johnson et al., a) datasets. Hosted on Codabench, the competition provides defined training, validation, and testing sets comprising 109,168 emergency department admissions.

Document statistics from the training data (Table 2 of Appx. A.2) reveal that the average size of the “text” field in Discharge data exceeds 4200 tokens, with over 65,000 documents containing more than 2000 tokens. In this task’s dataset, each discharge summary includes a “Brief Hospital Course” section, typically situated after patient history and current treatments, and a “Discharge Instructions” section, commonly found towards the note’s conclusion. Evaluation metrics such as BLEU-4 (Papineni et al., 2002), ROUGE-1, ROUGE-2, ROUGE-L (Lin, 2004), BERTScore (Zhang et al., 2020), Meteor (Banerjee and Lavie, 2005), AlignScore (Zha et al., 2023), and MEDCON (Yim et al., 2023) focus on assessing the textual similarity and factual correctness of the generated text.

Our approach focused on managing source data

Proceedings of the 23rd Workshop on Biomedical Language Processing, pages 740–747
August 16, 2024.

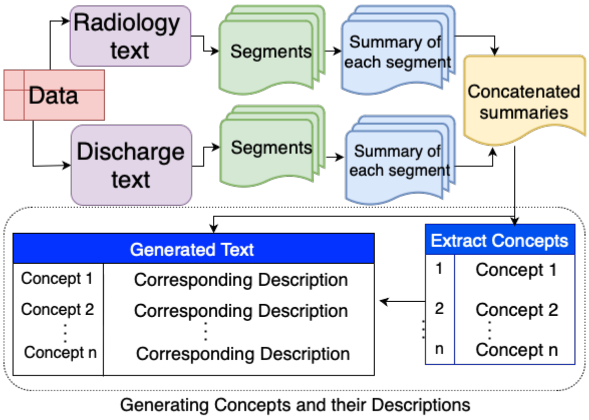

Figure 1: Overview of **Concept Based Description Generation** for generating the *"Brief Hospital Course"* and *"Discharge Instructions"* from Discharge and Radiology text.

size and generating target output systematically to capture crucial information effectively. We aimed to condense lengthy documents while retaining essential details by using a rule-based segmentation method and appropriate prompts for extracting summaries without compromising important information. This strategy allowed us to compress large documents while preserving necessary data for target text generation.

We aim to generate the target text in a structured format by creating summaries that describe specific topics related to the task. These topics are referred to as "concepts". A concept can encompass various subjects such as patient instructions, medication details, disease information, etc. An example of such a concept is illustrated in Appx. A.1.

Hence, next part of our approach unfolds in two phases. First, we predict the concepts relevant to summarized text. Then, leveraging these concepts, we generate descriptions from the same input text. This process allows us to tailor responses effectively, as the concepts are inherently task-specific, enhancing the accuracy and relevance of our generated content.

2 Related Works

Document summarization, including Query-focused Summarization (QFS), has made significant progress in recent decades. QFS targets specific query information, providing concise answers from retrieved documents. BayeSum by Daumé III and Marcu (2006) leverages multiple documents for state-of-the-art results in query-focused summarization. Vig et al. (2022) explored neural approaches, highlighting their versatility. Baumel et al. (2018) addressed challenges in extractive methods for effective QFS. These efforts showcase diverse strategies in advancing query-focused document summarization.

Varadarajan and Hristidis (2006) introduced query-specific document summarization methods. Additionally, Fu et al. (2020) explored concept extraction in clinical contexts, automatically identifying predefined clinical concepts from unstructured text.

HEPOS by Huang et al. (2021) introduces an innovative encoder-decoder attention mechanism for scalable long document summarization, processing ten times more tokens than traditional models. Moro and Ragazzi (2022) developed the semantic self-segmentation approach to overcome memory limitations in transformer architectures, particularly beneficial in law domains. Grail et al. (2021) proposed a hierarchical propagation layer to enhance reasoning in long document summarization. Pang et al. (2023) suggested a hierarchical inference framework improving summarization models' performance on lengthy texts. Koh et al. (2022) conducted a survey for evaluating research progress and future directions in long document summarization.

Efforts focus on training Large Language Models (LLMs) with medical data, like MIMIC-III (Johnson et al., 2015). Asclepius-R (Kweon et al., 2024) is a specialized clinical large language model trained on synthetic clinical notes created from publicly available case reports, designed to handle patients' clinical notes while addressing privacy and accessibility challenges. BioMistral (Labrak et al., 2024) is tailored for biomedical text, showing superior performance in question-answering and across languages, supporting research in healthcare.

3 Methods

Our pipeline, outlined in Figure 1, consists of four stages and uses three of the six provided data files. The *"text"* fields from the *"radiology"* and *"discharge"* files serve as source documents, while the *"discharge_instructions"* and *"brief_hospital_course"* fields from the *"discharge_target"* data are used for target summaries.

The first stage involves segmenting documents into distinct segments using predefined rules for both source and target data. In the second stage, we use open Large Language Models (LLMs) to generate segment-wise summaries, extracting essential information from each segment of the source text and combining these into a condensed version. For the target texts, we determine the concept for each segment by training LLMs to predict and describe concepts from the summarized text. This process is repeated for both the *"Discharge Instruction"* and *"Brief Hospital Course"* tasks. Finally, the extracted concepts and descriptions are combined and presented as the output for each task.

3.1 Source Document Segmentation

Two datasets serve as the source texts for our analysis: Discharge and Radiology data. A sample snippet from each dataset is illustrated in Figure 2 of Appx. A.1. To segment these documents effectively, we employed distinct strategies tailored to the common patterns found within each dataset. Discharge texts typically contain substantial content and exhibit various types of noise, such as multiple spaces, spaces between new lines, and multiple equal signs used as dividers. Our initial step involves noise removal from both text types. For the Discharge data, we segment the text sections using three consecutive new line characters ("\n\n\n"). Subsequently, we examine specific characters—such as colon (":"), double star ("**"), hash ("#"), and dash ("-")—present in the first line of each segment. If these characters are absent, we merge the segment with its preceding one.

Similar segmentation processes are applied to the Radiology texts, albeit with slight modifications. Here, we split the documents using double new line characters ("\n\n") and search solely for the colon (":") character in the first line of each segment. Segments lacking this character are merged with the previous segment.

3.2 Target Document Segmentation and extracting Concepts

Segmenting target document texts involves both dividing the text and identifying concepts for each segment. Figure 3 of Appx. A.1 shows segments paired with their corresponding concepts, which describe the segment's content. This pattern, though not universal, is common in many documents. Before segmentation, we reduce noise by removing multiple spaces, single spaces, or periods between new lines, and replacing multiple equal signs with a new line character.

Subsequently, we split the entire text of both types using two newline characters ("\n\n"). For the "discharge_instructions" text, we identify common keywords like "Activity", "Medications", "What was done?", "Why was I admitted to the hospital?" etc. as target concepts for generated text. Segments are retained if the first line is in all capital letters or if a colon character is present in the first line of each segment. Otherwise, segments are merged with their preceding segment. The concept key is extracted from text is the portion before the colon character in the first line of each segment, or from text in capital letters at the beginning of each segment. In cases where no concept key is found, we assign the concept as $Uncategorized_i$ where i enumerates uncategorized cases. The first segment's concept is set as $Start$ if no concept is identified using the aforementioned method.

A similar methodology is applied to the "brief_hospital_course" text, where we lack a predefined list. After denoising the text, we split the document as before and retain segments if the first line contains a colon, starts with a hash or greater than sign (">"), or is in all capital letters. The concept for each segment is determined by text preceding a colon, dash, or full stop sign in the first line, or by text in all capital letters.

3.3 Summarization

The primary objective of our summarization process is to condense document size while retaining crucial information. To achieve this goal, we avoid running the summarization model on the entire document due to the risk of potential loss of important details, especially in longer documents. Instead, we employ a specific prompt before each segment, which is given in Table 3 of Appx. A.3.

The summarization model is then applied solely to the source text and to each text segment from both the discharge and radiology reports. The summaries generated for each segment corresponding to each report are combined, resulting in a new concise report for each type of report. This helps reduce the document size with minimal information loss.

3.4 Concept Generation

Concepts play a pivotal role in our generation task, as they define the structure of the generated text for each query. To facilitate this, we train a Large Language Model (LLM) for each task—generating "brief_hospital_course" and "discharge_instructions". These models take the condensed versions of discharge and radiology texts

Metric	Performance
BLEU	0.04
ROUGE-1	0.21
ROUGE-2	0.1
ROUGE-L	0.13
BERTScore	0.18
Meteor	0.30
AlignScore	0.20
MEDCON	0.19
Overall	0.17

Table 1: Performance of Proposed model on Phase-2 test data

as input and generate a list of concepts extracted through our earlier methods.

In our approach, the prompt (Table 4 of Appx. A.4) provided to the model includes the summarized text along with a query aimed at identifying all concepts present in the text. The model's response provides all identified concepts, with each concept listed on a new line for clarity and organization. This method ensures that the generated text aligns with the extracted concepts, shaping the output according to the underlying structure of the input data.

3.5 Concept Based Description Generation

The concepts we extract from the previous section serve as directives for our model, guiding it to generate concise and relevant descriptions. These extracted concepts provide a roadmap for the generator, indicating the specific topic it should focus on. Our approach involves training a model that takes summarized text as input, along with a comprehensive list of concepts extracted using the method described earlier. We then train a Large Language Model (LLM) for each task, with each task designed to answer a question based on the input text.

We provide response as in Table 4 of Appx. A.4. The model's response is expected to be a description of that particular concept only. Once we have obtained all concepts and their corresponding descriptions for each task, we combine them to generate the final output text. In the combination process, we can exclude concepts such as "uncategorized" or "start" to enhance the naturalness of the output. This method allows us to generate well-structured and explanatory output for each task, resulting in a coherent and informative final text.

4 Results and Discussion

The "Discharge Me!" dataset comprises training (68,785 samples), validation (14,719 samples), phase I testing (14,702 samples), and phase II testing (10,962 samples) datasets, all sourced from MIMIC-IV submodules. It is worth noting that the phase II testing dataset, set for release on April 12th, 2024, will serve as the final evaluation test set. Due to time constraints, we were unable to execute our model on the entire dataset, ultimately utilizing a subset of 10,000 samples for training purposes. During the generation of summaries using the BioMistral-7B model (Labrak et al., 2024), we set the temperature to 0.0 to ensure determinism in our results. This controlled setting aimed to maintain consistency in the generated outputs during our experiments.

We proceeded to fine-tune the same model separately for four distinct tasks: concept generation and subsequent description generation corresponding to each concept, aimed at generating the "Brief Hospital Course" and "Discharge Instructions" sections. This fine-tuning process involved adopting the instruction tuning method, implemented using the Supervised Fine-tuning Trainer [1]. We configured the parameters as follows: $max\_seq\_length = 4096$, $learning\_rate = 2e-4$. To reduce memory size, we utilized 4-bit quantization, and for Low-Rank Adaptation (Mangrulkar et al., 2022), we set $rank = 64$, $alpha = 16$, and $dropout = 0.1$. The model underwent training for 5 epochs, employing a training batch size of 4 and an evaluation batch size of 20. These parameter settings were chosen to strike a balance between training efficiency and model performance across the different tasks.

In Table 1, we present the metrics proposed by the task organizers, which were calculated based on the outputs generated by our model. These metrics encompass a comprehensive evaluation of the model's performance across various dimensions specified by the Discharge Me task. The calculations were performed using Codabench. Additionally, for transparency and reproducibility, the task organizers have provided a Python script[2] for scoring, ensuring consistency and facilitating further analysis of our model's results.

[1] https://huggingface.co/docs/trl/en/sft_trainer

[2] https://github.com/Stanford-AIMI/discharge-me/tree/main/scoring

5 Conclusion

In conclusion, this paper describes our participation in the DischargeMe shared task, which involved generating summaries of hospital course and discharge reports using MIMIC-IV data. Our approach included data segmentation, concept identification and description, prompt-based summarization, and training models for concept extraction and description generation. We used pre-trained and fine-tuned Large Language Models (LLMs) to produce structured, informative summaries. Future work will compare different models and prompts and explore advanced data segmentation techniques to improve accuracy and efficiency.

6 Limitations

Our work faces challenges such as the necessity to summarize each text segment, limitations of rule-based methods, handling long segments with threshold limits, and dependence on the model and prompt used. Time constraints have hindered comprehensive comparisons of different models and prompts. We plan to develop a new model for better segmenting lengthy documents. The success of our generation task depends on accurate concept generation, as poor summarization impacts overall quality. These challenges highlight the complexity of the task and need for ongoing research and improvement.

References

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Tal Baumel, Matan Eyal, and Michael Elhadad. 2018. Query Focused Abstractive Summarization: Incorporating Query Relevance, Multi-Document Coverage, and Summary Length Constraints into seq2seq Models. *arXiv preprint*. ArXiv:1801.07704 [cs].

Hal Daumé III and Daniel Marcu. 2006. Bayesian Query-Focused Summarization. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 305–312, Sydney, Australia. Association for Computational Linguistics.

Sunyang Fu, David Chen, Huan He, Sijia Liu, Sungrim Moon, Kevin J. Peterson, Feichen Shen, Liwei Wang, Yanshan Wang, Andrew Wen, Yiqing Zhao, Sunghwan Sohn, and Hongfang Liu. 2020. Clinical Concept Extraction: A Methodology Review. *Journal of Biomedical Informatics*, 109:103526.

Quentin Grail, Julien Perez, and Eric Gaussier. 2021. Globalizing BERT-based Transformer Architectures for Long Document Summarization. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1792–1810, Online. Association for Computational Linguistics.

Luyang Huang, Shuyang Cao, Nikolaus Parulian, Heng Ji, and Lu Wang. 2021. Efficient Attentions for Long Document Summarization. *arXiv preprint*. ArXiv:2104.02112 [cs].

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Leo Anthony Celi, Roger Mark, and Steven Horng. a. MIMIC-IV-ED.

Alistair Johnson, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. b. MIMIC-IV-Note: Deidentified Free-text Clinical Notes.

Alistair Johnson, Tom Pollard, and Roger Mark. 2015. MIMIC-III Clinical Database.

Huan Yee Koh, Jiaxin Ju, Ming Liu, and Shirui Pan. 2022. An Empirical Survey on Long Document Summarization: Datasets, Models, and Metrics. *ACM Computing Surveys*, 55(8):154:1–154:35.

Sunjun Kweon, Junu Kim, Jiyoun Kim, Sujeong Im, Eunbyeol Cho, Seongsu Bae, Jungwoo Oh, Gyubok Lee, Jong Hak Moon, Seng Chan You, Seungjin Baek, Chang Hoon Han, Yoon Bin Jung, Yohan Jo, and Edward Choi. 2024. Asclepius-R : Clinical Large Language Model Built On MIMIC-III Discharge Summaries.

Yanis Labrak, Adrien Bazoge, Emmanuel Morin, Pierre-Antoine Gourraud, Mickael Rouvier, and Richard Dufour. 2024. Biomistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains. *Preprint*, arXiv:2402.10373.

Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Sourab Mangrulkar, Sylvain Gugger, Lysandre Debut, Younes Belkada, Sayak Paul, and Benjamin Bossan. 2022. Peft: State-of-the-art Parameter-Efficient Fine-Tuning Methods. `https://github.com/huggingface/peft`.

Gianluca Moro and Luca Ragazzi. 2022. Semantic Self-Segmentation for Abstractive Summarization of Long Documents in Low-resource Regimes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11085–11093.

Bo Pang, Erik Nijkamp, Wojciech Kryscinski, Silvio Savarese, Yingbo Zhou, and Caiming Xiong. 2023. Long Document Summarization with Top-down and Bottom-up Inference. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1267–1284, Dubrovnik, Croatia. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Ramakrishna Varadarajan and Vagelis Hristidis. 2006. A System for Query-specific Document Summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, CIKM '06, pages 622–631, New York, NY, USA. Association for Computing Machinery.

Jesse Vig, Alexander Fabbri, Wojciech Kryscinski, Chien-Sheng Wu, and Wenhao Liu. 2022. Exploring Neural Models for Query-Focused Summarization. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1455–1468, Seattle, United States. Association for Computational Linguistics.

Justin Xu, Zhihong Chen, Andrew Johnston, Louis Blankemeier, Maya Varma, Jason Hom, William J. Collins, Ankit Modi, Robert Lloyd, Benjamin Hopkins, Curtis Langlotz, and Jean-Benoit Delbrouck. 2024. Overview of the First Shared Task on Clinical Text Generation: RRG24 and "Discharge Me!" In *The 23rd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, Bangkok, Thailand. Association for Computational Linguistics.

Wen-wai Yim, Yujuan Fu, Asma Ben Abacha, Neal Snider, Thomas Lin, and Meliha Yetisgen. 2023. Aci-bench: a Novel Ambient Clinical Intelligence Dataset for Benchmarking Automatic Visit Note Generation. *Scientific Data*, 10(1):586. Publisher: Nature Publishing Group.

Yuheng Zha, Yichi Yang, Ruichen Li, and Zhiting Hu. 2023. AlignScore: Evaluating Factual Consistency with A Unified Alignment Function. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11328–11348, Toronto, Canada. Association for Computational Linguistics.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. *arXiv preprint*. ArXiv:1904.09675 [cs].

A Appendix

A.1 Source and Target Text Segments

As described in Sections 3.1 and 3.2, Figures 2 and 3 illustrate the segmentation of source and target texts, respectively. In Figure 2, different segments are denoted by various colors. In Figure 3, all segments are separated, with the blue parts representing the concepts and the corresponding text in yellow serving as their descriptions.

A.2 Training Data Statistics

The training data statistics are presented in Table 2. It shows the minimum, maximum, and average document token lengths for both source and target texts arranged against "hadm_id". The table also categorizes the documents based on token length into three groups: less than 500, between 500 and 2000, and more than 2000 tokens. From the table, we can observe that the source discharge texts are typically very long, whereas the target discharge instruction texts are usually short.

A.3 Prompt for Summarization

In Table 3, we provide the prompt used for summarizing all text segments to ensure minimal loss of important information. We use this prompt with BioMistral-7B-DARE to generate summaries for each segment.

A.4 Prompt for Fine-tuning Model

In Table 4, we provide the prompt used to train the LLM to generate the "Concept" and the corresponding description. This table depicts two prompt templates. The first template is used for extracting concepts from the summarized source text. The second template generates a description of a particular concept given the same summarized source text and the list of all extracted concepts.

	Min token length	Max token length	Avg token length	<500	>500 <2000	>2000
Radiology	34	48980	1491.7	19762	34184	14839
Discharge	510	21087	4263	0	3186	65599
Discharge instruction	12	8935	332.6	57328	11378	79
Brief hospital course	13	6959	635.6	32068	35654	1063

Table 2: Data statistics for the given training corpus, calculated using Mistral Tokenizer (Labrak et al., 2024)

(1)

....
Major Surgical or Invasive Procedure:
None

History of Present Illness:
___ yo M with CLL on ibrutinib, HTN, T2DM, SSS s/p PPM (A-paced, LRL appears ~50), hx of AF on Xarelto, CAD s/p ___ LHC with DES (? x2) placed in mid-LAD, now presenting with constant ___ chest pain on left radiating down arm. Chest pain is described as a central substernal chest "pressure", like a "fist into the chest".

Pertinent Results:
___ 11:05PM BLOOD WBC-11.9* RBC-3.25* Hgb-8.9* Hct-27.1* MCV-83 MCH-27.4 MCHC-32.8 RDW-14.2 RDWSD-42.9 Plt ___ ___ 11:05PM BLOOD Neuts-45.0 ___ Monos-7.0 Eos-1.4 Baso-0.5 Im ___ AbsNeut-5.36 AbsLymp-5.41* AbsMono-0.83* AbsEos-0.17 AbsBaso-0.06 ___ 11:05PM BLOOD Hypochr-NORMAL Anisocy-1+ Poiklo-1+Macrocy-NORMAL Microcy-1+ Polychr-NORMAL Ovalocy-1+Tear

Discharge Medications:
1. albuterol sulfate 2.5 mg /3 mL (0.083 %) inhalation QID:PRN
....

(2)

```
INDICATION:  ___ woman with severe C. diff colitis and acute mental status
changes with hypercarbia,  the left arm pain, assess for acute intracranial
process.

TECHNIQUE:  Contiguous axial images were obtained through the brain without
intravenous contrast.  Coronal and sagittal reformations were prepared.

FINDINGS:  There is no acute intracranial hemorrhage, edema, mass effect,
or major vascular territorial infarction.  There is no shift of normally
midline structures. ….
```

Figure 2: The rule based text segmentation for 1. *Discharge text* 2. *Radiology text*. Two consecutive segments are marked by different colors.

Prompt for Summarization
You are an intelligent clinical language model. Below is a snippet of patient's discharge summary and a following instruction from healthcare professional. Write a response that appropriately completes the instruction. The response should provide the accurate answer to the instruction, while being concise. [Discharge Report Begin] {text_segment} [Discharge Report End] [Instruction Begin] Summarize the text in very concise form, only keep the important information. [Instruction End]

Table 3: Prompt for summarization task of each text segment

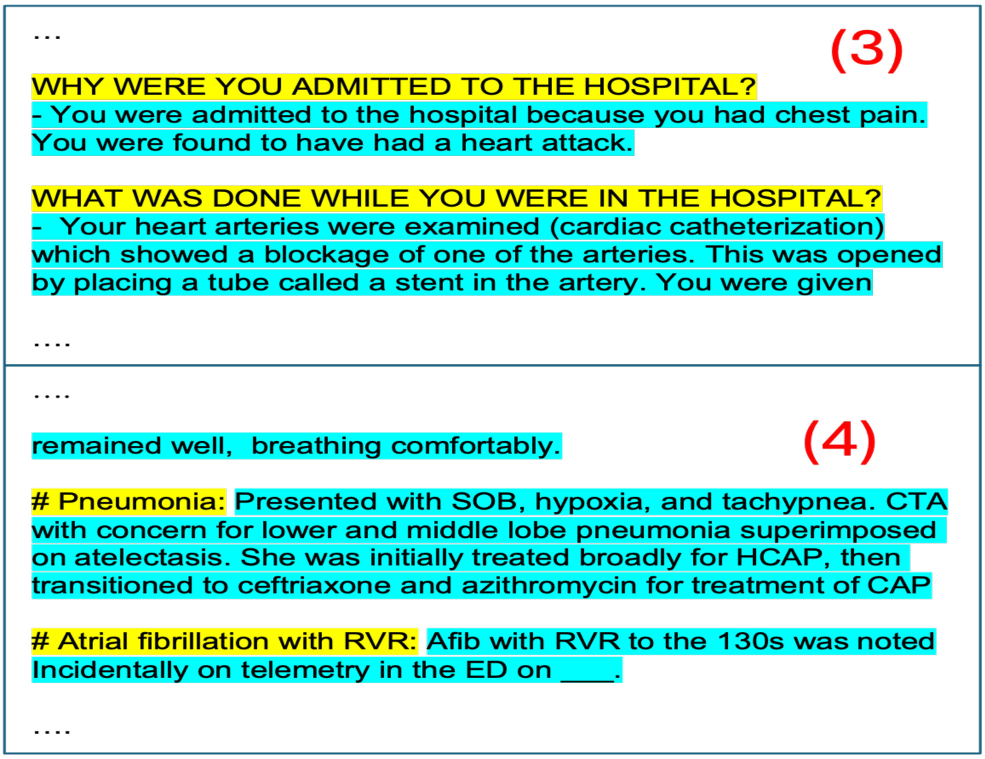

Figure 3: The rule based text segmentation for 3. *Discharge Instruction* 4. *Brief Hospital Course*. The Concepts are marked in yellow followed by corresponding description in blue.

Prompt for Extraction of Concept
Instruction: Below is a input context which contains the summaries of discharge and radiology reports followed by a question. Generate the response for the question using the context. ### Input: {Summarised Source Text} ### Question: What are the possible aspects for {discharge instruction / brief hospital course} in the above document?
Prompt for Generation of Description Corresponding to Concept
Instruction: Below is a input context which contains the summaries of discharge and radiology reports followed by a question. Generate the response for the question using the context. ### Input:{Summarised Source Text} ### Concepts: {List of Concepts} ### Question: Describe the concept C_i based on the above text.

Table 4: Prompt Template for Fine-tune a LLM to generate Concept and Corresponding Description