

DUNGEONS AND DRAGONS FANTASY ADVENTURE GENERATION

Halpewattage Akila Nuwantha Peiris

219381B

Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

March 2024

DUNGEONS AND DRAGONS FANTASY ADVENTURE GENERATION

Halpewattage Akila Nuwantha Peiris

219381B

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

March 2024

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science in Computer Science Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Nisansa de Silva

Signature of the Supervisor:

Date:

DEDICATION

To Aarya Nyx Peiris

ACKNOWLEDGEMENT

I would like to acknowledge and give my sincere gratitude to my supervisor, Nisansa de Silva without whom this project would not have been a success. His extensive knowledge of both Natural Language Processing and the game Dungeons and Dragons has been an immense help throughout this research.

My gratitude goes out to all the volunteers who helped make the Forgotten Realms Fandom Wiki the giant monolith of Dungeons and Dragons knowledge that it is today. This research would not have been possible without their contributions to the Forgotten Realms Fandom Wiki. I extend my gratitude to all the players and Dungeon Masters over the decades, without whom, this game would not have had the longevity nor the success it currently has.

I would also like to thank Manish Mathai, my fellow D&D enthusiast, for his contributions to this research with his D&D domain knowledge expertise.

To my Parents Ranjith and Kanthi, and my brother Chamod for their encouragement and support and for all they have done for me over the years. To my friends who have helped me in various ways during this period.

My unending gratitude to my wife Ayesha and my daughter Aarya who have been my constant pillars of support. My life, my universe, my everything.

ABSTRACT

In this thesis, we evaluate the possibility of creating an automated fantasy adventure generation system for the popular tabletop game Dungeons and Dragons. We have evaluated what the existing practices are when story generation is considered and their limitations, especially given the domain specificity. We have detailed our steps in creating the *Forgotten Realms Wiki (FRW) data set*, a Dungeons and Dragons domain-specific data set containing several sub-datasets with various degrees of pre-processing useful in different Natural Language Processing tasks. Going a step further, we have evaluated the goodness of these data via semantic similarity comparisons, establishing the uniqueness and validity of the existence of these different data sets. We have created multiple text generation models capable of free text generation as well as in-filling focused text generation. In addition, the study describes the potential pitfalls when using uni-directional models such as attention-based language models for infilling tasks. We have finally created an instruction fine-tuned Large Language Model, *Mistral of Realms Instruct 7b* using *FRW* data set. Then we have created a system comprised of Artificial Intelligence agents backed by this instruct-fine tuned model to create a system capable of generating the necessary components of a dungeons and dragons adventure module including the main narrative, character descriptions, and gameplay mechanics. We have qualitatively evaluated the goodness of the generated content with a combination of similar techniques.

Keywords: Natural Language Processing, Natural Language Generation, Story Generation, Dungeons and Dragons, Large Language Models

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Dedication	iii
Acknowledgement	v
Abstract	vii
Table of Contents	ix
List of Figures	xv
List of Tables	xvii
List of Abbreviations	xvii
List of Appendices	xxi
1 Introduction	1
1.1 Dungeons and Dragons	1
1.2 Adventure/ Module/ Campaign	1
1.2.1 Structure of a Quest	2
1.2.2 Narrative	2
1.3 Dungeon Master	3
2 Research Problem	5
3 Research Objectives	7
4 Literature Survey	9
4.1 Natural Language Generation	10
4.2 Story Generators	11
4.2.1 Automated Story Generation	11
4.2.2 Simulation Based Story Generators	12
4.2.3 Story Generation via Character Relation Modeling	14
4.2.4 Pitfalls of Automated Story Generation	15
4.3 Domain-Specific Text Generation	15
4.4 Procedural Content Generation	17
4.4.1 General Disadvantages of PCG	21

4.5	Generating Quests Based on Player Type	22
4.6	Mixed-Initiative and Semi-Automated Systems	24
4.7	Quest Graphs	27
4.8	Dungeon and Quest Generation Philosophies	29
4.8.1	Choices and Constraints vs Gameplay Experience	29
4.8.2	Meaningful Quest Engagement	30
4.8.3	Finding Balance	33
4.8.4	Non-Linear Stories	34
4.9	Sentence Infilling	35
4.9.1	Rule-Based Methods	36
4.9.2	Statistical Methods	36
4.9.3	Neural Network-Based Methods	36
4.10	Large Language Models	36
4.10.1	Large Language Models for Infilling	38
4.10.2	Autogen	38
4.10.3	Retrieval Augmented Generation	39
4.10.4	Training LLMs on a Budget	39
4.11	Wikipedia Data in NLP	40
4.11.1	Wikipedia and Other Wikis as Data Sources	40
4.11.2	Live Play Data Sets	42
5	Forgotten Realms Wiki Data Set	43
5.1	Introduction	43
5.2	Data Set Overview	43
5.3	Data Extraction and Pre-Processing	45
5.4	Data Set Breakdown	46
5.4.1	Plain Text (<i>FRW-P</i>)	46
5.4.2	Plain Text JSON (<i>FRW-J</i>)	46
5.4.3	First Paragraph JSON (<i>FRW-FJ</i>)	47
5.4.4	Link Graphs (<i>FRW-L</i> , <i>FRW-FL</i> , <i>FRW-CL</i>)	48
5.4.5	Infoboxes (<i>FRW-I</i>)	49
5.4.6	Embedding Data Sets	52

5.5	Use Case Analysis 1: Semantic Similarity Comparison	53
5.5.1	Issues with Category Links for Semantic Similarity Evaluation	53
5.5.2	Generating Text Pairs for Evaluation	53
5.5.3	Hierarchical Similarity Measures	54
5.5.4	Embedding Based Similarity Measures	55
5.6	Use Case Analysis 1 Results	56
5.6.1	First Link Traversal in Forgotten Realms Fandom Wiki	56
5.6.2	Semantic Similarities	56
5.7	Use Case Analysis 2: <i>Forgotten Realms Free Text Generator</i>	59
5.7.1	Generated Text	59
5.7.2	Named Entity Classifier	61
5.8	Discussion	62
6	Manual Hierarchical Data Annotation	63
6.1	Introduction	63
6.2	Pitfalls of Manual Annotation	64
6.3	Annotation Consistency	64
6.4	SHADE Overview	64
6.5	SHADE Workflow	65
6.6	Implementation Details	66
6.6.1	Escaping the Infobox	66
6.6.2	List of Internal Links	67
6.6.3	List of Noun Phrases	68
6.6.4	Manual Input Labels	69
6.6.5	Assigning and Keeping Track of Annotations	70
6.7	Annotation with SHADE	71
7	Infilling to Create a Narrative	73
7.1	Introduction	73
7.2	Masked Text Corpus	73
7.3	Infilling Model Training	75
7.4	Results and Observations	76
7.5	Infilling with Quest Graphs	78

8	Adventure Module Generation with LLM Agents	81
8.1	Introduction	81
8.2	LLM Agents	82
8.3	Embedding Domain Knowledge into Language Models	87
8.3.1	Fine-Tuning	87
8.3.2	Retrieval Augmentation	87
8.4	Inter-Annotator Agreement	87
8.5	<i>Mistral of Realms 7b</i>	87
8.5.1	Training Data Set	88
8.5.2	Fine-Tuning	88
8.5.3	Preliminary Results	89
8.6	<i>Mistral of Realms Instruct 7b</i>	91
8.6.1	Training Data Set	91
8.6.2	Fine-Tuning	93
8.6.3	Preliminary Results	93
8.7	Preliminary Results	98
8.7.1	<i>Mistral of Realms Instruct 7b</i> without RAG	98
8.7.2	<i>Mistral of Realms Instruct 7b</i> with RAG	102
8.7.3	<i>Mistral Instruct 7b</i> without RAG	103
8.7.4	<i>Mistral Instruct 7b</i> with RAG	105
8.8	Evaluation	106
8.8.1	Evaluation Method	106
8.8.2	Evaluation Results	107
9	Conclusion	111
9.1	Summary of the Work Done	111
9.2	Discussion	112
	References	115
	Appendix A Output - <i>Mistral of Realms Instruct 7b</i> with RAG	129
	Appendix B Output - <i>Mistral Instruct 7b</i> with RAG	139

Appendix C	Adventure Modules used for the Evaluation	155
C.1	Outputs from the <i>Mistral of Realms 7b</i> without RAG system	155
C.2	Outputs from the <i>Mistral Instruct 7b</i> without RAG system	173
C.3	Outputs from the <i>Mistral Instruct 7b</i> with RAG system	178
C.4	Outputs from the <i>Mistral of Realms Instruct 7b</i> with RAG system	186
Appendix D	Human Evaluator Details	191
Appendix E	Evaluation Results	193

LIST OF FIGURES

Figure	Description	Page
Figure 5.1	A screenshot of the Forgotten Realms Fandom Wiki article on <i>Wizard</i>	44
Figure 5.2	Different types of infoboxes in Forgotten Realms Fandom Wiki	50
Figure 5.3	Comparison of <i>Infoboxes</i> from 3 articles; <i>Elf</i> , <i>Human</i> , and <i>Tiefling</i> .	51
Figure 5.4	Poincare Embedding convergence after 50 epochs	52
Figure 5.5	First link traversal top 10 pages	57
Figure 6.1	<i>Lead Section</i> from the article on <i>Tiamat</i> in Forgotten Realms Fandom Wiki	63
Figure 6.2	Workflow on selecting the most appropriate label from the 3 different annotation levels in SHADE	65
Figure 6.3	Markdown text vs rendered version of a sample link found in the Forgotten Realms Wiki.	68
Figure 6.4	SHADE - First link list contains the correct label	68
Figure 6.5	SHADE - Noun phrase list contains the correct label	69
Figure 6.6	SHADE - Manually entering the correct label	70
Figure 6.7	Entity Relationship diagram for SHADE	70
Figure 6.8	Current annotation result breakdown by label source	71
Figure 7.1	Sample masked data item	75
Figure 7.2	An example for a Quest Graph	79
Figure 8.1	A graphical representation of the LLM agents and how their interactions	83

LIST OF TABLES

Table	Description	Page
Table 4.1	Different agents types in the multi-agent system	13
Table 4.2	Types of PCG strategies used for generating dungeons	19
Table 5.1	<i>FRW</i> data set	45
Table 5.2	<i>FRW-P</i> data set statistics	46
Table 5.3	<i>FRW-J</i> data set statistics	47
Table 5.4	<i>FRW-FJ</i> data set statistics	48
Table 5.5	<i>FRW-L</i> data set statistics	49
Table 5.6	<i>FRW-FL</i> data set statistics	49
Table 5.7	<i>FRW-CL</i> data set statistics	49
Table 5.8	<i>FRW-I</i> data set statistics	50
Table 5.9	Pearson Correlation for the pairwise text similarities across multiple similarity metrics	58
Table 7.1	Infilling data sets	74
Table 8.1	List of LLM Agents and their roles.	84
Table 8.2	Adventure module goodness evaluation questionnaire	108
Table 8.3	Evaluation result summary	110
Table 8.4	Inter-Annotator Agreement Scores	110
Table D.1	Human Evaluator Details	191
Table E.1	Evaluation results	193

LIST OF ABBREVIATIONS

Abbreviation	Description
<i>FRW</i>	Forgotten Realms Wiki
<i>FRW</i>	ForgottenRealmsWikidata set
D&D	Dungeons and Dragons
AI	Artificial Intelligence
BART	Bidirectional Auto-Regressive Transformers
BERT	Bidirectional Encoder Representations from Transformers
CBOW	Continuous Bag of Words
CNNs	Convolutional Neural Networks
DAG	Directed Acyclic Graphs
DM	Dungeon Master
GPT	Generative Pretrained Transformer
IAA	Inter-Annotator Agreement
JC	Jiang and Conrath
LLaMA	Large Language Model Meta AI
LLMs	Large Language Models
LoRA	Low Rank Adaptation
NLGen	Natural Language Generation
NLP	Natural Language Processing
NPC	Non Playable Characters
PCG	Procedural content generation
PV-DBOW	Paragraph Vector: Distributed Bag Of Words
PV-DM	Paragraph Vector: Distributed Memory
RAG	Retrieval augmented generation
RNNs	Recurrent Neural Networks
seq2seq	sequence to sequence models
WP	Wu and Palmer

LIST OF APPENDICES

Appendix	Description	Page
Appendix -A	Output - <i>Mistral of Realms Instruct 7b</i> with RAG	129
Appendix -B	Output - <i>Mistral Instruct 7b</i> with RAG	139
Appendix -C	Adventure Modules used for the Evaluation	155
Appendix -D	Human Evaluator Details	191
Appendix -E	Evaluation Results	193