Story-Yarn: An Interactive story building application

Hryadyansh Saraswat[1], Snehal D. Shete[2], Vikas Dangi[1], Kushagra Agrawal[1], Anuj Aggarwal[1] and Aditya Nigam[1]
[1]IIT Mandi, [2]Independent Researcher

Abstract

Story building is an important part of language and overall development of a child. Developing an interactive and artificial intelligence (AI) based solution to create stories for children is an open and challenging problem. Methods combining large language models (LLMs) and knowledge graphs (KGs) - have further enabled high quality and coherent story generation. In this work, we present a platform, Story Yarn, developed for interactive story creation for children. We customise a KG, using children stories, which captures relationships between components of stories. This customised KG is then used along with LLM to collaboratively create a story. We have also built a simple app to facilitate user interaction. This platform can aid the creative development of children, and can be used at home or in schools.

1 Introduction

Storytelling is an important part of human society. Stories help in gaining and sharing knowledge in a memorable way and developing emotional connections. They are an important part of a child's development. While children are generally very creative, one can imagine ways to better channelize and support this creativity and at the same time help develop their linguistic skills. Play-and-Learn systems that help in interactive story-building can serve this purpose.

Interactive storytelling, a collaborative approach, encourages the co-creators of a story to make decisions and give a desired direction to the story. This engagement makes co-creators active participants rather than being passive observers. Apart from traditional interactive story building methods such as "story cubes/wheels" and "prompts in a jar", digital storytelling, using AI and non-AI methods, is gaining significance in language learning in children (Moradi and Chen, 2019).

The methods which are not based on AI, do not provide assistance in text creation part. Whereas, the contemporary LLMs (OpenAI, 2023; Touvron et al., 2023; Google, 2024), which come under the umbrella of generative AI, have facilitated the text creation component in story generation process. These models have humongous learnable parameters and are trained on huge corpus of data available on the internet, enabling them learn, and hence generate, semantically and grammatically correct text information.

LLMs generate text when they are given a context in the form of a prompt. Directly using LLMs for interactive generation is a cumbersome process, as a user has to give specific prompt for every interactive generation. There are many other shortcomings of directly using LLMs for story generation(Wang et al., 2023), lack of domain knowledge is one of them.

When evolving a story, in addition to causal connections in a story, the creative aspect of human mind looks for various new directions. Whereas, LLMs look backward at the story to provide directions in which stories can go. Hence, their choices have limited creativity and tend to get predictable. Also, given guidance in the form of specific external domain knowledge, the generation quality of language models have shown improvement in performance(Wang et al., 2023). This external knowledge is often stored in the form of structured graphs that follow a defined schema or topology (Hwang et al., 2020; Speer et al., 2017; Krishna et al., 2016).

Based on this, in our work we have created a KG using stories and have combined it with LLM Gemini (Google, 2024) for interactive story generation. Here, the role of our customised KG is to assist the LLM in building more creative and diverse stories. We have used various evaluation methods to verify the effectiveness of using KG and LLM for interactive story generation.

Many digital story building platforms are developed based on both AI (Akoury et al., 2020; Zhang et al., 2022; Golchha, 2023)and non-AI (Jumper, 2019; sourced digital stories, 2015; Storybird, 2019) technologies. Interaction with these tools is often challenging for young children. Hence, we developed a simple interactive platform, Story Yarn, which gives the users future path options for story progression.

Summary of contributions: Our proposed platform, Story Yarn, for interactive story creation involves:

- Constructing a KG of children stories.
- Using LLM along with our customised KG to facilitate generation of evolving story paths.
- A simple user interface for easier user-AI interaction.

The paper is organized as follows: Section 2, describes prior work related to story generation and more specifically to methods of interactive generation and usage of KGs in this area. The methods for creating KG and an interactive story generation framework proposed in this work are discussed in Section 3. The evaluation of our proposed framework and results are presented in Section 4, with relevant conclusions and future directions discussed in Section 5.

2 Related Work

In this section, we present a brief review of recent interactive story generation methods and existing story generation platforms for children.

2.1 Story generation and knowledge graphs:

Story generation has been a widely explored area in the domain of natural language generation. (Alabdulkarim et al., 2021) provides an excellent survey on automatic story generation methods and challenges in this area. Here, we explore a few interactive story generation works.

In (Yao et al., 2019), the authors propose a hierarchical story generation framework wherein a story plot is first planned and the story is generated based on this plot. There are two proposed models: static model, which takes complete story plot from user as input at the beginning of generation, and dynamic model which interacts with user and takes in plot points during the generation process. The work in (Brahman et al., 2020) use cue phrases given by a user at each sentence generation and previously generated sentences for interactively building a story.

The work in (Wang et al., 2023) captures review of story generation enhancements using structured knowledge. In (Ilievski et al., 2021), firstly commonsense axioms are derived for certain fixed story types, and they are used for querying a commonsense KG, based on the KG output story generation is done by filling templates defined for particular story types. They have explored 3 story types: unmet expectation, alternatives, object modification. The KG in this case is a combination of ConceptNet(Speer et al., 2017), ATOMIC(Hwang et al., 2020), and Visual Genome(Krishna et al., 2016).

In all of the above story generation works, there are no future path directions given to the user for evolving the story and also the user interaction is limited. These points are addressed in our work.

2.2 Story Generation platforms:

Here, we discuss a few of the well known existing non-AI and AI based story creating platforms. Notable non-AI platforms we studied are Story Weaver(sourced digital stories, 2015), StroyJumper(Jumper, 2019) and Storybird(Storybird, 2019). They offer illustrations and tools like templates, characters, props and scenes to their users for creating personalised storybooks. In both, StoryJumper and Storybird, the text part of a story is completely created by the users and no AI-assistance is given.

Since the advent of LLMs, many prompt based automatic story creating tools have been developed. More recent works in interactive generation that we studied are Storium(Akoury et al., 2020) and StoryBuddy(Zhang et al., 2022). (Akoury et al., 2020) uses digital story cards which represent characters, their strengths, scenes and other related information to provide addition information to language model for story construction. (Zhang et al., 2022) uses language model to form questions related to a story and offers an engaging way to create stories.

In (Holtz, 2024) along with prompt based interaction, one can add personal photos and choose a custom narrator style. (Golchha, 2023) creates characters in story based on given prompts and then add the details such as personality traits or fictional background related to characters. A user

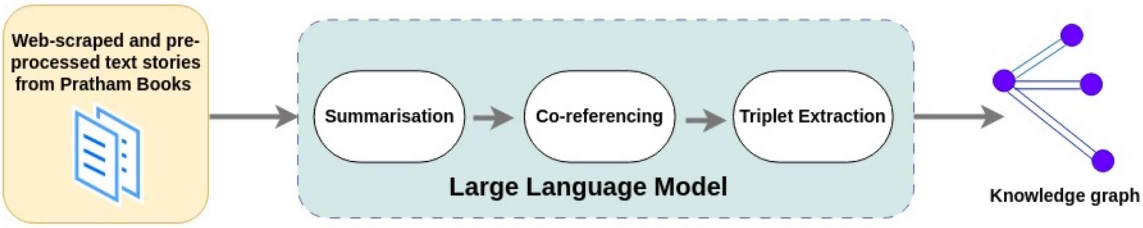


Figure 1: Knowledge Graph from Children stories

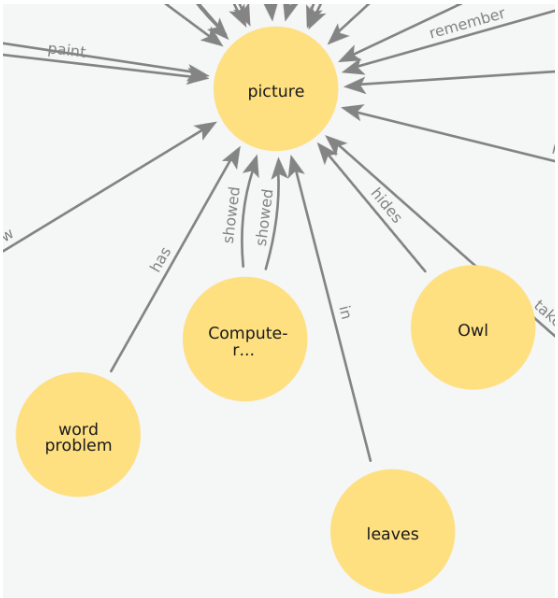


Figure 2: Snippet of created knowledge graph

can chat with the created characters and collaboratively build a story.

Not withstanding the above capabilities in these existing tools, they lack a fundamental creative aspect of providing multiple diverse but coherent possibilities for evolution of stories in a simplistic way suitable for young children. In this work, we address this important aspect.

3 Method

In this section, we discuss in detail the proposed story generation method, Story Yarn, which has interactive communication between a user and a framework consisting of LLM and our customized KG. We begin with the construction of our KG followed by elaborating on our framework.

3.1 Knowledge Graph construction:

A KG is a grid of interconnected ideas or concepts in a specific domain. Typically, in KGs data is arranged in the form of triplets of nodes and relations such as subject-relation-object. Here, subjects and objects are nodes that generally are entities such as people, places, concepts, events, fictional situations etc., and relations capture the context with which two entities are connected to each other.

In our case, the domain for which the KG is created is children stories. The dataset of stories used in our work is taken from an open digital books repository, Story Weaver, created by Pratham Books (sourced digital stories, 2015). Approximately 8.5k stories from this platform are pre-processed and KG triplets are extracted from these stories using LLM, Gemini 1.0-Pro. Figure 1 depicts block diagram of story KG construction. Following are the important points taken into account while extracting KG triplets.

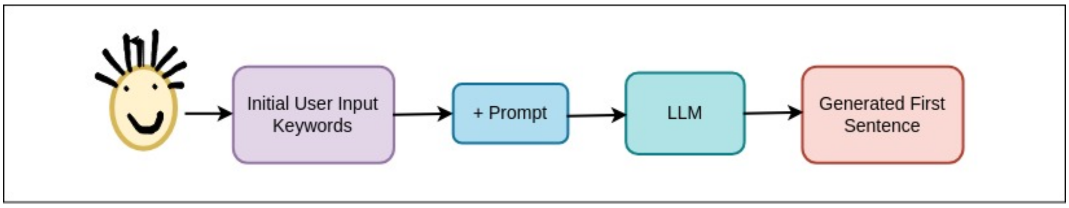
- Single word or a set of 2−3 words, that represent key concepts or entities in story sentence are stored as nodes.
- Relations are deduced by splitting the sentences and using the nodes extracted above.
- For example if a story sentence is: **"Fashion store has aisles of clothes."** then, **"Fashion store"** and **"clothes"** form nodes and **"has aisles of"** is extracted as relation.
- Co-referencing is taken into account while making nodes.
- A weight is assigned to each formed triplet as per Conceptnet(Speer et al., 2017) rules which are based on assertion of relation between nodes.

The extracted story triplets are then converted and stored in graph format using Neo4j graph database tool(Neo4j, 2012). This KG can easily be scaled, as one can extract triplets from new stories and add them to the graph using Neo4j tool. A snapshot of our KG is presented in Figure 2.

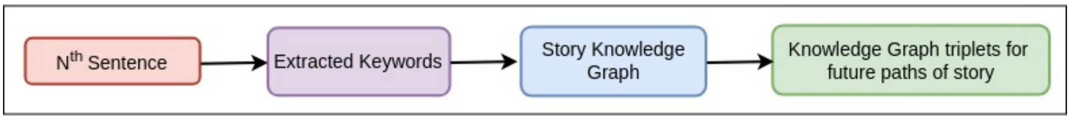
3.2 Story Yarn framework:

Figure 3 shows, our platform, Story Yarn's flow. Following are the steps using which a story is built by a user in collaboration with Story Yarn.

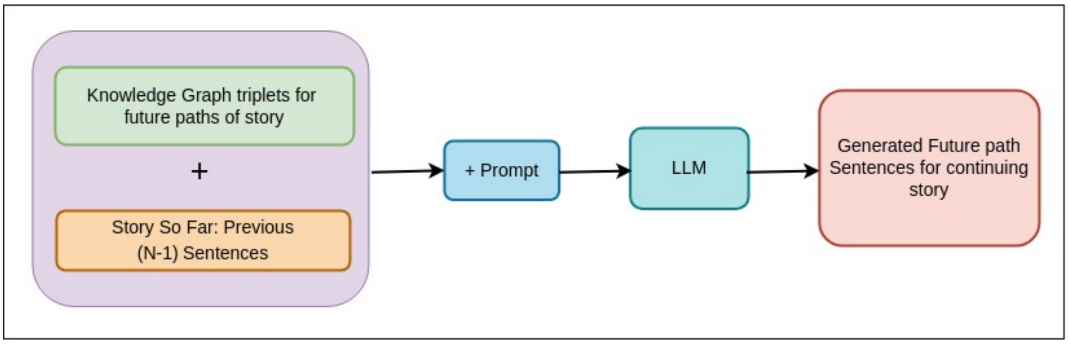
- At the beginning, a user is asked to provide input keywords based on which a story starts.
- These initial keywords and a designed prompt is given to LLM, to create first sentence (Figure 3(a)). The sentence is shown



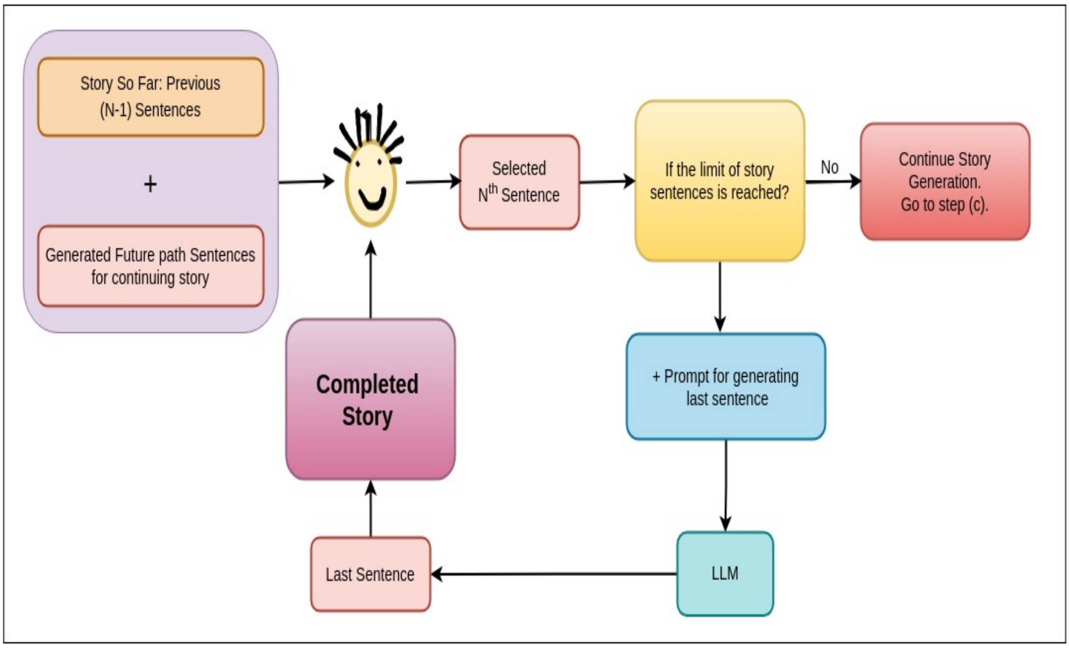
(a)



(b)



(c)



(d)

Figure 3: Story Yarn flow:(a) Processing user input keywords, (b) Obtaining KG triplets for future story paths, (c) Generating future story path sentences, (d) User input for story path selection and completion of story

to the user and is also appended in a buffer, 'Story so far', which stores story sentences.

- To continue the story generation, future story path options are given to the user. To create the future paths, keywords are extracted from the last generated $(N-1)^{th}$ sentence using basic NLP techniques like removal of stop-words, applying tokenization and lemmatization etc.

- The extracted keywords are used for querying our customised story KG. The output triplets of graph are first arranged according to the weight assigned to them, Figure 3(b).

- LLM is given prompt to use the keywords and the above output KG triplets to generate future path sentences. Figure 3(c) shows, this process of constructing future paths for continuing story.

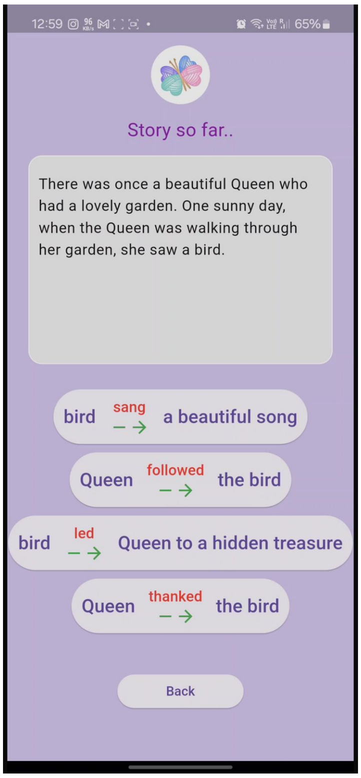


Figure 4: Story Yarn app screen

- The user is given the 'Story so far', that is (N- 1) previous sentences, and the future direction sentences in which story can progress. Figure 4, shows this step using a snapshot of our app interface.
- The user's choice of future sentence is then appended in the 'Story so far'. If the set limit of story length is not reached, the story generation continues in loop with extraction of keywords from N^{th} sentence and then on.
- If the set limit of story length is reached, LLM is given a prompt to generate last sentence for completing the story. This complete story is then presented to the user.

In our platform, at every step of a story the user is encouraged to think and create a continuation of story generated so far using given future paths. The user can then select a "creatively suitable" option, in her opinion, for progressing a story. This justifies the creativity and human touch aspect in using Story Yarn.

4 Experiments and Results

In this section, we present a detailed evaluation of our KG and story-creation platform, Story Yarn, including the dataset, the experimental settings, and their results.

4.1 Story Knowledge Graph evaluation:

A survey in (Wang et al., 2021), describes various metrics used to evaluate the performance and content of KG. The important relevant metrics that we considered during our KG construction are reliability and completeness of KG.

4.1.1 Reliability of KG:

Reliability of KG refers to the degree to which a KG accurately represents the information in the corpora from which it is created and how closely it is related to the domain. In our case the domain for which KG is created is children-stories.

1. **Parts of Speech (PoS) and Named Entity Recognition (NER)** : Since the KG triplets are used for creating paths for story completion, the conservation of information in story corpora and its language structure is of prime importance. This is verified by analysing the distribution of PoS tags and NER of our KG against that of our story corpora, consisting of approximately 8.5k stories taken from (sourced digital stories, 2015).

   PoS tags label each word in text data to its corresponding grammatical category, such as noun, verb, adjective, adverb etc. Our PoS analysis shows our children story corpus contains 54.1% of words as noun and verbs. Whereas, our story KG contains 52% of words as noun and verbs. This small difference between these numbers is because of resolution of co-references while making KG.

   NER classifies the entities in text data into predefined categories such as person, location, organization, quantities etc. The result of our NER analysis shows, the story corpus contains 47.6% of person, 21.1% organizations and 7.7% Geopolitical Entities. While our KG contains 47.3% of person, 26.2%of organizations and 9.1% of Geopolitical Entities. This analysis proves that our KG accurately captures and represents these essential named elements.

2. **Age-appropriate content**: It is a well-known fact that children-stories have great educational and entertainment values. But some of them also have violence or negative emotions or interactions in them. Since these events are depicted in a very subtle manner, which do not glorify the negative nature of

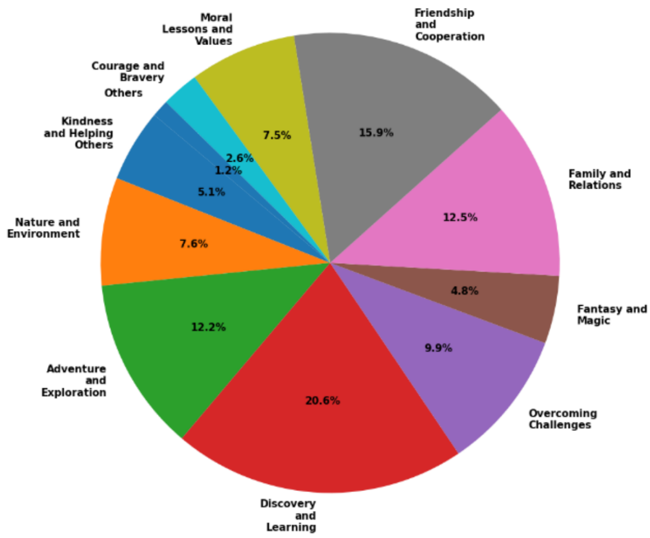


Figure 5: Distribution of Story Themes in Percentage

those events, such content is considered age-appropriate for children. Our KG is generated using crowd-sourced stories which are curated by experts and are suitable for children. Also we have added prompts to follow age safety measures while creating our KG using LLM.

4.1.2 Completeness of KG:

Completeness of KG, in our case, refers to the extent to which the knowledge graph covers a wide range of information needed to create engaging and diverse children's stories. The following analysis presents our KG completeness:

1. **Thematic variety of story corpus**: We find the thematic distribution of 8500 stories in our corpus. Figure 4 shows this distribution, proving our KG, which is constructed using these stories, contains data of multiple genres and themes of children stories.
2. **Comparison of KG vocabulary with children's lexical database**: In (Green et al., 2023), the authors have created an English vocabulary database,CBP-LEX, using childrens picture books, suitable for $0-8$ years age group. 98.05% of words in CBP-LEX data (Green et al., 2023) are present in our KG. This ensures the completeness of our KG.

4.2 Story Yarn platform evaluation:

Here, we evaluate of our interactive story creation platform, Story Yarn, using similarity, diversity and entropy metrics.

The test set-up for this uses a set of three words, randomly taken from "Creativity Words" list present in each of 47 stories in the mc500.dev.txt of MCTest dataset (Richardson et al., 2013), as initial input keywords. An initial story sentence is generated by LLM, and we continue to generate the future story path options for next 3 sentences (story parts) using our KG triplets, LLM-only(Gemini 1.0 Pro) option and our Story Yarn approach having combination of our KG + LLM output. These generated paths are then analysed and the results are as follows:

1. **Contextual Similarity**: This metric finds out how contextually similar the retrieved KG triplets, Story Yarn paths and the LLM-only paths are with previously generated story sentences. It provides valuable insight into parameters such as narrative continuation and contextual relevance. For this, vector distance between each generated path and the current story generation is calculated.
2. **Diversity**: It involves finding out how diverse the future story paths retrieved from each of the three methods are, indicating the variety of potential narrative directions and plot developments. We separately calculate the pairwise embedding distances amongst all the paths given by KG, Story Yarn and the LLM-only options separately.

$$d_{ij} = \|E(t_i) - E(t_j)\|$$

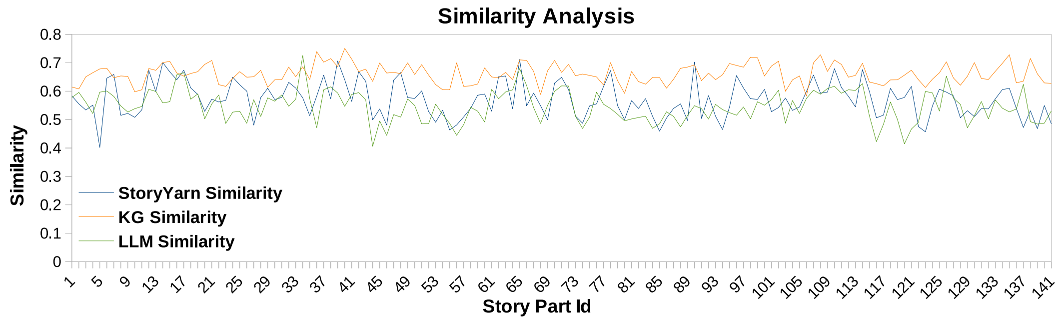
where, d_ij is the distance between i^{th} and j^{th} path embedding,$E(t_i)$, and $E(t_j)$.

We find an average embedding distance for each method and subtract average values from one to find diversity of paths for KG, Story Yarn and the LLM-only options.

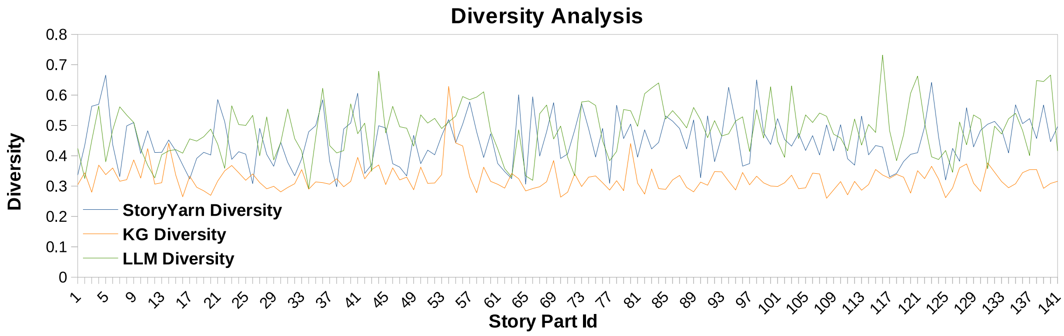
3. **Entropy of paths**: It is a measure of surprise elements that are present in the paths, contributing to the creativity of the paths and hence the story generated using them. The more the surprise elements and creativity, the more the children are encouraged to think out of the box. We use entropy for measuring the unpredictability or surprise elements of paths using below formula:

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

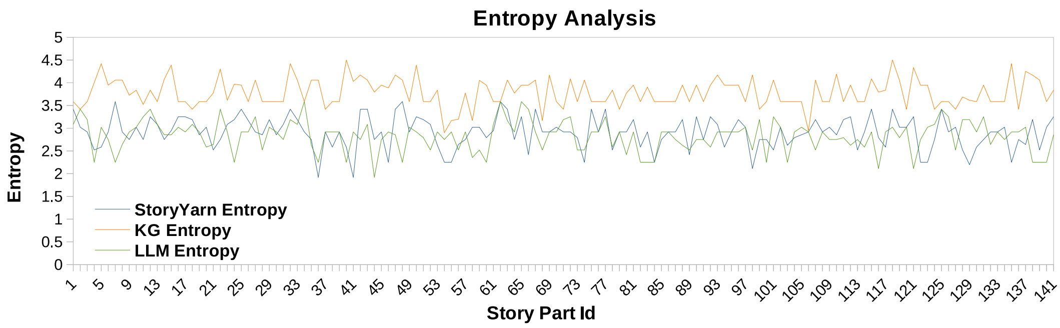
where p_i is the probability of occurrence of i^{th} word.



(a)



(b)



(c)

Figure 6: Story Yarn Evaluation: (a) Similarity Analysis,(b) Diversity Analysis, and (c) Entropy Analysis of story parts created using Story Yarn platform, KG output and LLM-only output. (Please refer to the soft copy for coloured figures)

	KG	LLM	Story Yarn
Similarity	0.66	0.55	0.57
Diversity	0.32	0.49	0.45
Entropy	3.78	2.83	2.90

Table 1: Story Yarn Platform Evaluation

Table 1 and Figure 6 show the trends of above analysis. In case of similarity, the KG triplets are more closely related to "the story generated so far" as compared to the paths generated by LLM-only. But the KG paths fall short in terms of diversity to the LLM paths. This is because of a very basic KG search method used in our approach. Entropy analysis shows that a standalone LLM system gives predictable paths with less surprise or creative elements when compared with paths generated by the KG. In all three analyses, the Story Yarn platform shows values that lie between the KG and LLM-only performances. This is because it uses KG induced information along with LLM and generates story paths combining the best of both KG and LLM-only options.

5 Conclusion and Future Work

In this paper, we have reported a complete platform for interactive story creation for young children using customised story knowledge graph and large language model. The effectiveness of our story knowledge graph further aids the LLM to generate coherent and diverse future story paths for evolving story.

In future, we plan to add voice and image assistance to the existing work. It would also be interesting to explore adding more stories and fusing the story graph with commonsense graphs.

References

Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyyer. 2020. STORIUM: A Dataset and Evaluation Platform for Machine-in-the-Loop Story Generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6470–6484, Online. Association for Computational Linguistics.

Amal Alabdulkarim, Siyan Li, and Xiangyu Peng. 2021. Automatic story generation: Challenges and attempts. In *Proceedings of the Third Workshop on Narrative Understanding*, pages 72–83, Virtual. Association for Computational Linguistics.

Faeze Brahman, Alexandru Petrusca, and Snigdha Chaturvedi. 2020. Cue me in: Content-inducing approaches to interactive story generation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 588–597, Suzhou, China. Association for Computational Linguistics.

Bharat Golchha. 2023. Story nest.

Gemini Team Google. 2024. Gemini: A family of highly capable multimodal models.

Clarence Green, Kathleen Keogh, He Sun, and Beth OBrien. 2023. The children's picture books lexicon (cpb-lex): A large-scale lexical database from children's picture books. *Behavior Research Methods*, 56.

Charlie Holtz. 2024. Once upon a bot.

Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2020. COMET-ATOMIC 2020: On symbolic and neural commonsense knowledge graphs. *CoRR*, abs/2010.05953.

Filip Ilievski, Jay Pujara, and Hanzhi Zhang. 2021. Story generation with commonsense knowledge graphs and axioms. In *Workshop on Commonsense Reasoning and Knowledge Bases*.

Story Jumper. 2019. Story jumper.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. 2016. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *CoRR*, abs/1602.07332.

Hamzeh Moradi and Hefang Chen. 2019. Digital storytelling in language education. *Behavioral Sciences*, 9(12).

Neo4j. 2012. Neo4j - the worlds leading graph database.

OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.

Matthew Richardson, Christopher J.C. Burges, and Erin Renshaw. 2013. MCTest: A challenge dataset for the open-domain machine comprehension of text. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 193–203, Seattle, Washington, USA. Association for Computational Linguistics.

Crowd sourced digital stories. 2015. Story weaver.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *AAAI*, pages 4444–4451. AAAI Press.

Storybird. 2019. Story bird.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Xiangyu Wang, Lyuzhou Chen, Taiyu Ban, Muhammad Usman, Yifeng Guan, Shikang Liu, Tianhao Wu, and Huanhuan Chen. 2021. Knowledge graph quality control: A survey. *Fundamental Research*, 1(5):607–626.

Yuxin Wang, Jieru Lin, Zhiwei Yu, Wei Hu, and Börje F. Karlsson. 2023. Open-world story generation with structured knowledge enhancement: A comprehensive survey. *Neurocomputing*, 559:126792.

Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019. Plan-and-write: towards better automatic storytelling. In *Proceedings of Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'19/IAAI'19/EAAI'19, Honolulu, Hawaii, USA. AAAI Press.

Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA. Association for Computing Machinery.