

A Knowledge Graph Generated from An Indian Short Story

Akash Gupta

Department of Computer Science &
Engineering
JIS College of Engineering
West Bengal, India

Soma Giri

Department of Computer Science &
Engineering
JIS College of Engineering
West Bengal, India

Apurba Paul

Department of Computer Science &
Engineering
JIS College of Engineering
West Bengal, India

Abstract—Knowledge graph is a machine learning technique in which a graph is generated by scanning a set of documents detecting nodes (words) and relation within the nodes. In natural language processing Knowledge graph means a graph which provides a meaning knowledge. It can be considered as the process of obtaining required relation between bags of words.

In this project we will try to create a new knowledge graph which describes a text whole Indian short story.

Keywords— Sentence Segmentation, Dependency Parsing, Relation Extraction, POS Tagging, Knowledge Graph.

I. INTRODUCTION

The term “Knowledge Graph” is an encyclopedia is readable by machines. So, it is basically Knowledge organized in a manner that a machine can easily understand and extract information. Basically, we are looking at the graph theory - nodes, edges and attribute. More practically in our project we are looking at Knowledge graphs where we have got entities with attributes and relationship to other entities (also with attributes), which, we look from the Indian old short story “PANCHATANTRA”. In our project a knowledge graph is engineered manually, and the information are extracted from the source text (PANCHATANTRA.txt).

III. DATA PREPROCESSING

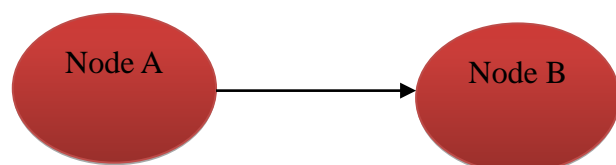
Panchatantra, which is considered as one of most famous prose written by Pandit Vishnu Sharma in Indian literature. The actual time span of the short story is unknown but according to the assumption the story was written near about 1300 to 300 BC. The word Pancha means Five and Tantra means stories. Basically, the short stories were based on older oral tradition with animals and bird's fable. These stories convey moral messages and the characters present in the stories are very interesting which attracts the children.

In our project we have created a knowledge graph with the help of Natural language processing, from an Indian short story as mentioned earlier. A particular story was hand-picked and refined by removing punctuations. The complex and compound sentences were reconstructed into simple sentences by removing brackets, commas etc.

According to our analysis in the short story there are 22 sentences and 204 words from the first story, 23 sentences and 173 words in second story, 35 sentences and 210 words from third story, 39 sentences and 175 words from fourth story.

IV. METHODOLOGY

To visualize the short story through a knowledge graph we used python libraries like **Pandas**, **regular expression**, **matcher**, **networkx** etc. which were constructed gradually in the knowledge graph. After the data preprocessing, spacy library helps us to perform sentence segmentation. The present simple sentences in our csv, are tokenized. Upon using dependency parsing the meaningful data from these sentences are obtained. Parts of speech from the sentences were identified using pos (Parts of Speech) tagging. Dependency parsing helps us to extract subject, object, main verb etc. and also highlights how each and every words are dependent on each other. To increase the rule primarily based strategies for relation /information extraction, we should have a proper knowledge of the dependency structure of the sentences at hand. With the help of spacy model (en_web_core_sm) we get an idea that how subject and object are connected with a main verb (root). Each and every subject, object, relations are stored into a list, in which we create the data frame with the column name subject, object, relation by their similarity of orders. The subject, object and root which were extracted from the text were stored in a list. These nodes (subject, object, relation) are called “Triples”. The csv file created by simple sentences is uploaded with the help of json. But initially the knowledge graph which was generated from the short story contained no relations between those nodes and the generated graph does not have any knowledge. It is necessary to understand however, info and Knowledge are embedded in these graphs. A node or an entity can have multiple relations for this sole reason graph viz was used to generate a graph at the end. Suppose there are two nodes A & B having totally different entities. These nodes are connected by an edge that represents the relationship or connection between the two nodes. Below is an example of the smallest knowledge graph we will build.



This is going to be a very fascinating aspect of this text. Our hypothesis concludes that the Predicate is truly the main verb in a sentence. Basically, this graph is a directed graph in which extracted subject, object and relations are used, subject

and object are nodes of the graph and the main verb is the actual relation between the nodes. After generating the directed graph, the story is visualized and accordingly the meaningful knowledge graph is obtained from which the short story could be properly comprehensible.

V. FIGURES & TABLES

The work proposed in this research paper helps to create a knowledge graph between different relationships and helps to establish relationships from different Story.

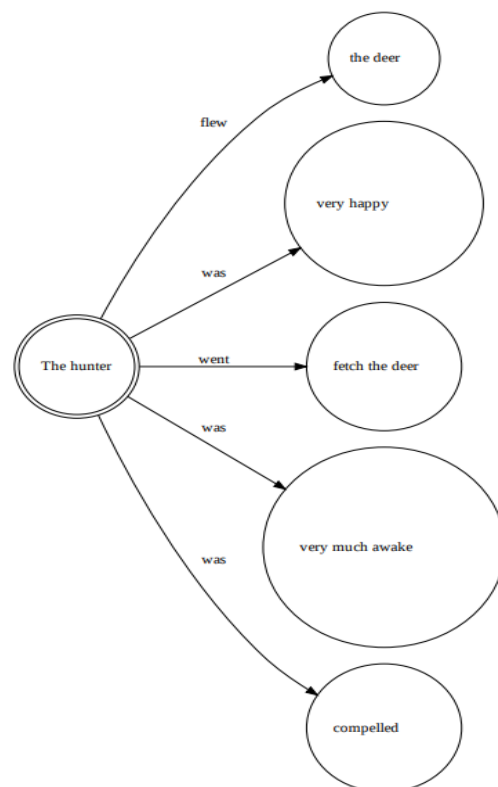
Here the following outputs have been created by extracting tokenization, dependence extraction.

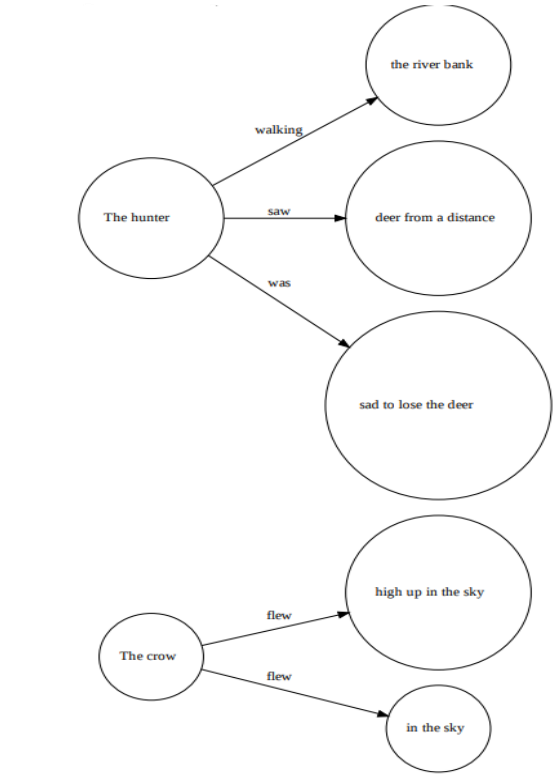
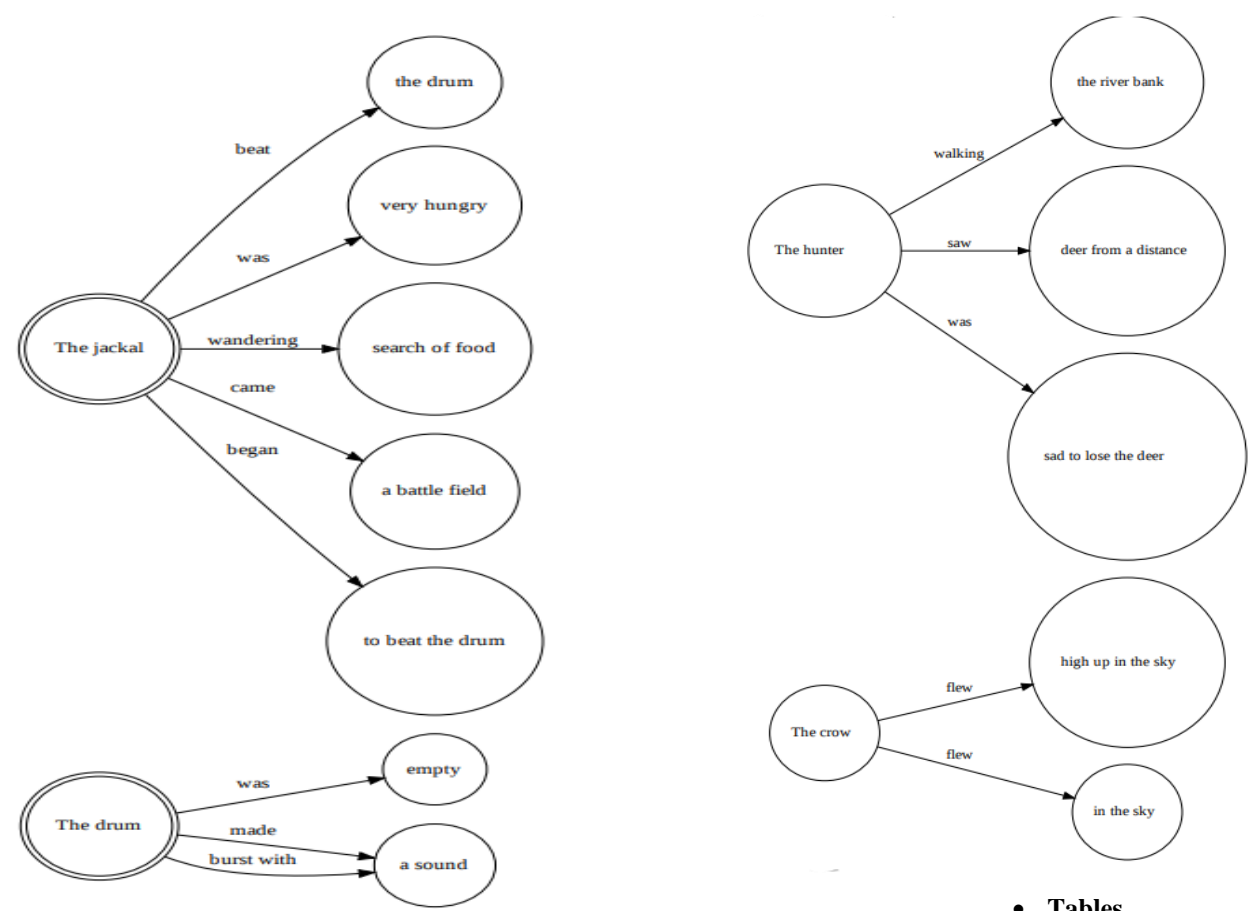
Here the sentence can be turned into a list. Each sentence can be presented as a knowledge graph.

	mysubject	myobject	myrelation
0	The jackal	some other small tasty animal inside the drum	thought
1	The drum	a sound	made
2	The jackal	the drum	beat
3	The jackal	very hungry	was
4	The jackal	the top of the drum too tough to tear off	found
5	The jackal	search of food	wandering
6	The jackal	a battle field	came
7	The jackal	to beat the drum	began
8	The sound of drumbeat	the whole jungle	filled
9	A leopard	towards the sound of the drum	was attracted
10	The jackal	there is some animal hiding inside the drum	said
11	The leopard	hungry	was
12	The drum	a sound	burst with

['The jackal', 'some other small tasty animal inside the drum', 'thought']
 ['The drum', 'a sound', 'made']
 ['The jackal', 'the drum', 'beat']
 ['The jackal', 'very hungry', 'was']
 ['The jackal', 'the top of the drum too tough to tear off', 'found']
 ['The jackal', 'search of food', 'wandering']
 ['The jackal', 'a battle field', 'came']
 ['The jackal', 'to beat the drum', 'began']
 ['The sound of drumbeat', 'the whole jungle', 'filled']
 ['A leopard', 'towards the sound of the drum', 'was attracted']
 ['The jackal', 'there is some animal hiding inside the drum', 'said']
 ['The leopard', 'hungry', 'was']
 ['The drum', 'a sound', 'burst with']
 ['The drum', 'empty', 'was']

Figures: - Image of the knowledge graphs of the IINDIAN short story





There lived three friends in a jungle.
They used to share their meals together.
I want to join your company.
You are most welcome.
There are many hunters around.
The hunters visit this jungle regularly.
I want to join your group.
A hunter appeared on the scene.
The crow then flew high up in the sky.
The hunter walking along the river bank.
The hunter saw the deer from a distance.
The crow flew in the sky.
The mouse ran into a hole.
The hunter was caught by the hunter.
The hunter was sad to lose the deer.
The hunter saw the deer from a distance.
The hunter was very happy.
The net and freed the turtle.
The hunter went to fetch the deer.
The hunter was very much awake.
The hunter was compelled.
That some miracle had taken place.
The four friends once again started living happily.

Table 2: Simple Sentences from second story

ACKNOWLEDGMENT

The analysis of the research work wishes to express our gratitude to Prof. Apurba Paul for allowing the degree attitude and providing effective guidance in development of this research paper. His conscription of the topic and all the helpful hints, he provided, contributed greatly to successful development of this work, without being pedagogic and overbearing influence. We also express our sincere gratitude to Dr. Dharmpal Singh, Head of the Department of Computer Science and Engineering of JIS College of Engineering and

all the respected faculty members of Department of CSE for giving the scope of successfully carrying out the research work. Finally, we take this opportunity to thank to Prof. (Dr.) Partha Sarkar, Principal of JIS College of Engineering for giving us the scope of carrying out the research work.

REFERENCES

- [1] Grigoris Antoniou and Frank Van Harmelen. A semantic web primer. MIT press, 2004.
- [2] V. Bryl et al. "What's in the proceedings? Combining publisher's and researcher's perspectives". In: Proceedings of the 4th Workshop on Semantic Publishing (SePub-lica). 2014.
- [3] Sarven Capadisli, Reinhard Riedl, and Sören Auer. "Enabling Accessible Knowledge". In: Conference for E-Democracy and Open Government. 2015, p. 257.
- [4] Matthew Horridge and Mark Musen. "Snap-SPARQL: A Java Framework for working with SPARQL and OWL". In: International Experiences and Directions Workshop on OWL. Springer. 2015, pp. 154–165.
- [5] M.A Musen. "The Protégé project: A look back and a look forward". In: AI Matters .Association of Computing Machinery Specific Interest Group in Artificial Intelligence 1.4 (2015).
- [6] Markus Nentwig et al. "A survey of current Link Discovery frameworks". In: Semantic Web 8.3 (2017), pp. 419–436.