INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD
R&D SHOWCASE 2020

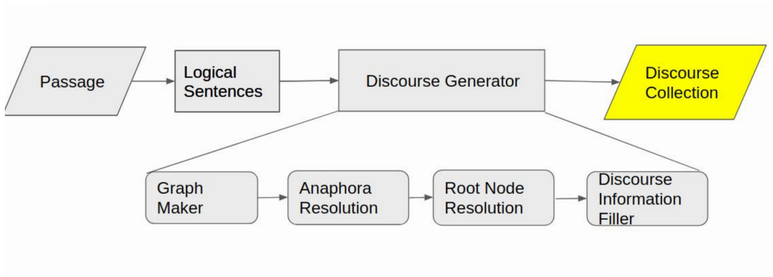
Samajh-Boojh: A Reading Comprehension System in Hindi

INTRODUCTION

Samajh-Boojh system presents an approach designed to answer questions on a reading comprehension passage on Panchatantra Stories.
It is an end-to-end system which focuses on
1) Comprehending the passage
2) Answer questions related to the passage

COMPREHENSION SYSTEM

The comprehension stage converts the unstructured passage into a structured representation called "**Discourse Collection**", which contains the relation shared amongst the sentences in the given passage along with the key characteristics of each sentence.



OUTPUT of COMPREHENSION

Comprehension Part:

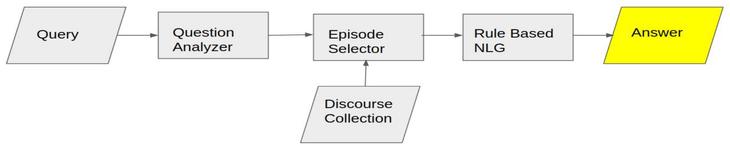
Passage:
rAma Eka acchA ladkA thA. vaha Eka dina pAThaSAlA jA rahA thA taba usE Eka billI dikhI. rAma usakE pAsa gayA para vaha bhAga gaI aura vO dukhI hO gayA

Discourse Collection:
{
"0": {
"OS": "rAma Eka acchA ladkA thA",
"karta": "rAma",
"kartaadj": ["rAma", "acchA"],
"ARS": "rAma Eka acchA ladkA thA",
"RNS": "rAma Eka acchA ladka thA"
}
"1": {
"OS": " vaha Eka dina pAThaSAlA jA rahA thA",
"time": "din",
"location": "pAThaSAlA",
"karta": "rAma",
"ARS": "rAma Eka dina pAThaSAlA jA rahA thA",
"RNS": "rAma Eka dina pAThaSAlA jA rahA thA"
}
"2": {
"OS": "usE Eka billI dikhI",
"time": "din",
"location": " pAThaSAlA",
"karta": "billI",
"given": " rAma Eka dina pAThaSAlA jA rahA thA",
"new": "rAma Eka billI dikhI",
"ARS": "rAma Eka billI dikhI",
"RNS": "rAma Eka billI dikha"
}
"3": {
"OS": " rAma billI pAsa gayA",
"time": "din",
"location": " billI pAsa",
"karta":"rAma",
"ARS": "rAma billI pAsa gayA",
"RNS": "rAma billI pAsa jA"
}
"4": {
"OS": "vaha bhAga gaI",
"time": "din",
"location": " billI pAsa",
"karta": "rAma",
"given": " rAma billI pAsa gayA",
"new": "billI bhAga gaI",
"ARS": "billI bhAga gaI",
"RNS": "billI bhAga jA"
}
"5": {
"OS": "vO dukhI hO gayA",
"time": "din",
"location": " billI pAsa",
"karta": "rAma",
"given": "billI bhAga gaI",
"new": "rAma dukhI hO gayA",
"ARS": "rAma dukhI hO gayA",
"RNS": "rAma dukhI hO jA"
}
}

QUESTION & ANSWERING SYSTEM

This module answers questions related to the passage using only "Discourse Collection".

With input as Discourse Collection and the query related to passage, this module chooses the best episode and selects the appropriate answer from that episode.



OUTPUT of QA SYSTEM

Question-Answering Part:

Discourse Collection:
{
"0": {
"OS": "rAma Eka acchA ladkA thA",
"karta": "rAma",
"kartaadj": ["rAma", "acchA"],
"ARS": "rAma Eka acchA ladkA thA",
"RNS": "rAma Eka acchA ladka thA"
}
"1": {
"OS": " vaha Eka dina pAThaSAlA jA rahA thA",
"time": "din",
"location": "pAThaSAlA",
"karta": "rAma",
"ARS": "rAma Eka dina pAThaSAlA jA rahA thA",
"RNS": "rAma Eka dina pAThaSAlA jA rahA thA"
}
"2": {
"OS": "usE Eka billI dikhI",
"time": "din",
"location": " pAThaSAlA",
"karta": "billI",
"given": " rAma Eka dina pAThaSAlA jA rahA thA",
"new": "rAma Eka billI dikhI",
"ARS": "rAma Eka billI dikhI",
"RNS": "rAma Eka billI dikha"
}
"3": {
"OS": " rAma billI pAsa gayA",
"time": "din",
"location": " billI pAsa",
"karta":"rAma",
"ARS": "rAma billI pAsa gayA",
"RNS": "rAma billI pAsa jA"
}
"4": {
"OS": "vaha bhAga gaI",
"time": "din",
"location": " billI pAsa",
"karta": "rAma",
"given": " rAma billI pAsa gayA",
"new": "billI bhAga gaI",
"ARS": "billI bhAga gaI",
"RNS": "billI bhAga jA"
}
"5": {
"OS": "vO dukhI hO gayA",
"time": "din",
"location": " billI pAsa",
"karta": "rAma",
"given": "billI bhAga gaI",
"new": "rAma dukhI hO gayA",
"ARS": "rAma dukhI hO gayA",
"RNS": "rAma dukhI hO jA"
}
}

+ rAma kaisA ladkA thA? → acchA

EPISODE SELECTION

Query: राम क्या खा रहा था?
Root Resolved Query: राम क्या खा?

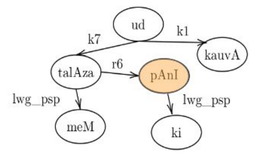
Original Sentence: उसने आम खाया.
Jaccard Score: 0

AR Sentence: राम आम खाया.
Jaccard Score: 3

Root Node Sentence: राम आम खा.
Jaccard Score: 7

EXAMPLE GRAPHS

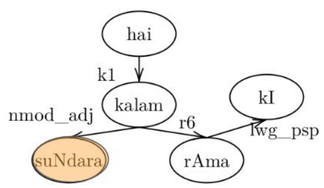
Sentence: kauvA pAni ki talAza meM uda



kauvA kiskE talAza meM udA? pAnI

Qtye: ['Kiske', 'talAza']
NLG Rule: r6 tag from node 'talAza'

Sentence: rAma kI kalam suNdara hai



kiskI kalama suNdara hai? rAma

rAma kI kalama kaisI hai? suNdara

Qtye: ['Adj_noun', 'kalama']
NLG Rule: nmod_adj tag from the node 'kalama'

RESULTS and ANALYSIS

Question Type	Questions	Accuracy
Karta	35	94.3%
Karma	7	100%
Time	7	100%
Loc	45	100%
Recipient	15	100%
Adj__Noun	15	100%
Intf	15	100%
Kya	179	71.6%
Kiske	13	84.62%
Kiska	5	100%
GivenNew	33	63.7%
Total	**440**	**75.45%**

Accuracy of the answers

Good accuracy for majority of the questions such as Karma and Time, Loc whose answers are straight forward from the tags.

Since the 'Kya' and 'GivenNew' format of the questions are versatile and the answers can be subjective.

More focus on testing versatile questions, more questions of type 'Kya' and 'GivenNew'.

Authors: Shalaka Vaidya , Hiranmai Sri Adibhatla, Radhika Mamidi ,Research Center Name: LTRC