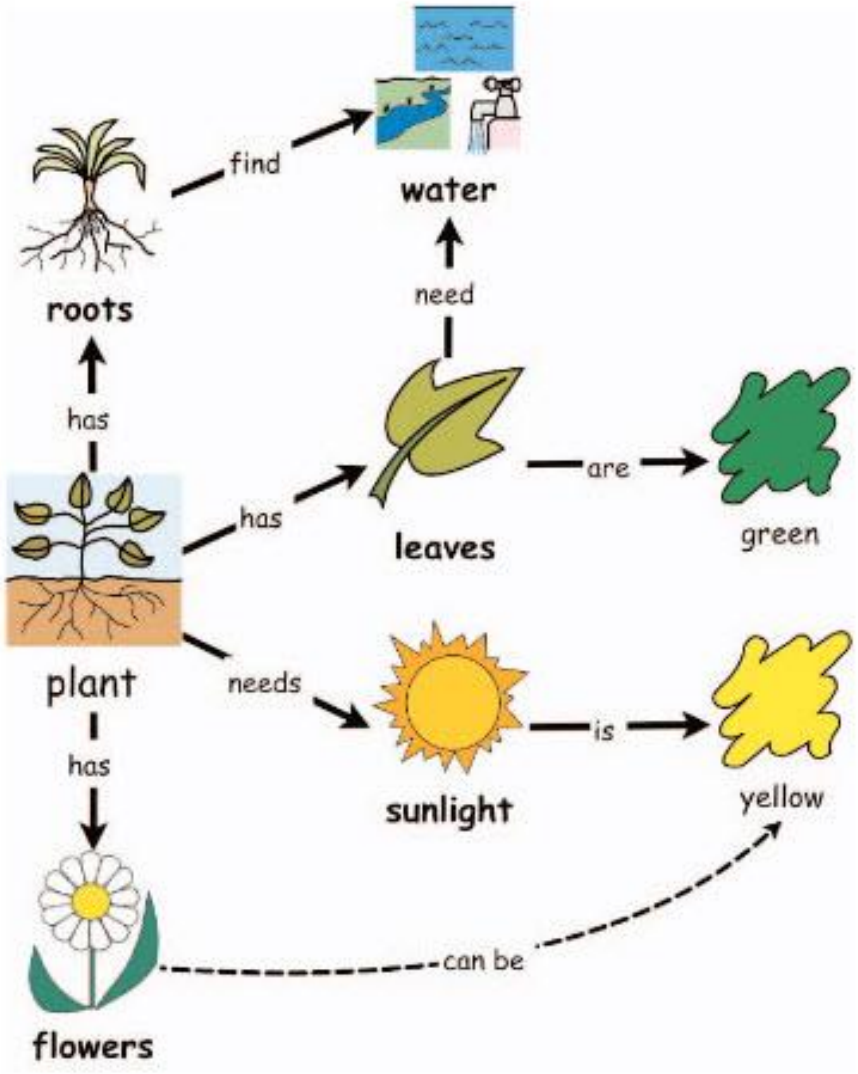
Automatic Concept Map Generation from Text-Based Learning Material



Guided By:

Prof Plaban Kumar Bhowmick

Group Members:

Smit Shah 17CS60R72

Ashish Malgawa 17CS60R81

Hussain Jagirdar 17CS60R83

Abhishek Prakash Singh 17CS60R48

Contents

Implementation

- Task: Summarizing the learning material.
- Step 1: Downloaded some learning materials.
- Step 2: For concept extraction, used Dbpedia spotlight python API.
- Step 3: Downloaded Predefined GloVe Vectors
- Step 4: For finding relationship between concepts, used OpenIE tool.
- Step 5: For visualization, used GePhi tool, Graphviz and NetworkX Module.

Dataset

- Learning Material 1
- NCERT Grade 6 social science book (Our Pasts-1)

  Chapter 2 – On the trial of the earliest people.
- NCERT Grade 7 social science book (Our Pasts-2)

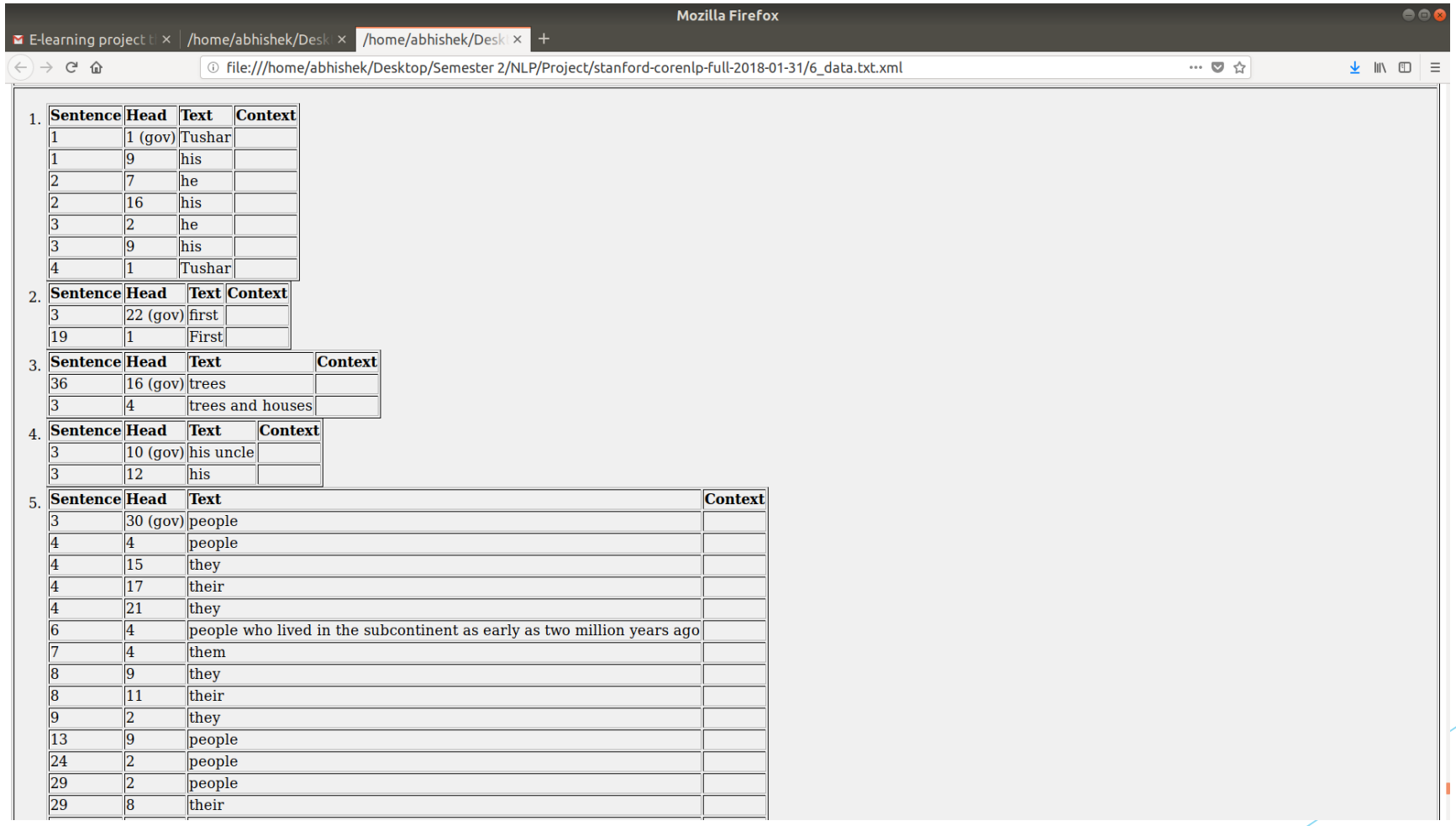
  Chapter 1 – Tracing changes through a thousand years
- Converted pdf into txt file.

  - http://pdftotext.com/
- Preprocessed the materials.

Dataset

- Preprocessing
  - Case folding
  - Coreference resolution
- Coreference resolution
  - Standford CoreNLP
  - Parsed XML file
  - Changed only pronouns.
- Index all 400000 Glove vectors through solr.

Co-Reference Resolution



1.

Sentence	Head	Text	Context
1	1 (gov)	Tushar	
1	9	his	
2	7	he	
2	16	his	
3	2	he	
3	9	his	
4	1	Tushar	

2.

Sentence	Head	Text	Context
3	22 (gov)	first	
19	1	First	

3.

Sentence	Head	Text	Context
36	16 (gov)	trees	
3	4	trees and houses	

4.

Sentence	Head	Text	Context
3	10 (gov)	his uncle	
3	12	his	

5.

Sentence	Head	Text	Context
3	30 (gov)	people	
4	4	people	
4	15	they	
4	17	their	
4	21	they	
6	4	people who lived in the subcontinent as early as two million years ago	
7	4	them	
8	9	they	
8	11	their	
9	2	they	
13	9	people	
24	2	people	
29	2	people	
29	8	their	

Co-Reference Resolution

6_data.txt

```
1 Tushar was going from Delhi to Chennai for his cousin's wedding.
2 They were travelling by train and he had managed to squeeze into the window seat his nose glued
3 As he watched trees and houses fly past his uncle tapped his shoulder and said Do you know that
  wondered when people couldn't travel quickly from one place to another did they spend their enti
4 Not quite.
5 We know about people who lived in the subcontinent as early as two million years ago.
6 Today we describe them as hunter-gatherers.
7 The name comes from the way in which they got their food.
8 Generally they hunted wild animals caught fish and birds gathered fruits roots nuts seeds leaves
```

7_data.txt

6_data.txt

```
1 Tushar was going from Delhi to Chennai for Tushar cousin's wedding.
2 They were travelling by train and Tushar had managed to squeeze into the window seat Tushar nose g
3 As Tushar watched trees and houses fly past Tushar uncle uncle tapped Tushar uncle shoulder and sa
  decades later. Tushar wondered when people couldn't travel quickly from one place to another did p
4 Not quite.
5 We know about people who lived in the subcontinent as early as two million years ago.
6 Today We describe people as hunter-gatherers.
7 The name comes from the way in which people got people food.
```

7_data.txt

Method 1.0

- Step 1 – Find relations between subject and object using OpenIE Tool for all sentences simultaneously.
- Step 2 - Extract relations corresponding to individual sentence through LCS.
- Step 3 – Find similarity between subject and objects (Cosine similarity using Glove Vectors)
- Step 4 – Apply the threshold to take the similar pair and add them in the Concept Map.
- Step 5 – Use networkx to create the graph and save it as gephi file.
- Step 6 – Change parameters in Gephi tool.

Method 2.0

- Step 1 – Find concepts using DBpedia corresponding to each sentence.
- Step 2 - Find relations between subject and objects using OpenIE Tool.
- Step 3 – Finding a pair of concept from a set of DBpedia concepts for each subject-object pair of relation through LCS.
- Step 4 – Find similarity between the concept pairs (Cosine similarity using Glove Vectors).
- Step 4 – Apply the threshold to take the similar pair and add them in the Concept Map.
- Step 5 – Use networkx to create the graph and save it as gephi file.
- Step 6 – Change parameters in Gephi tool.

Method 2.1

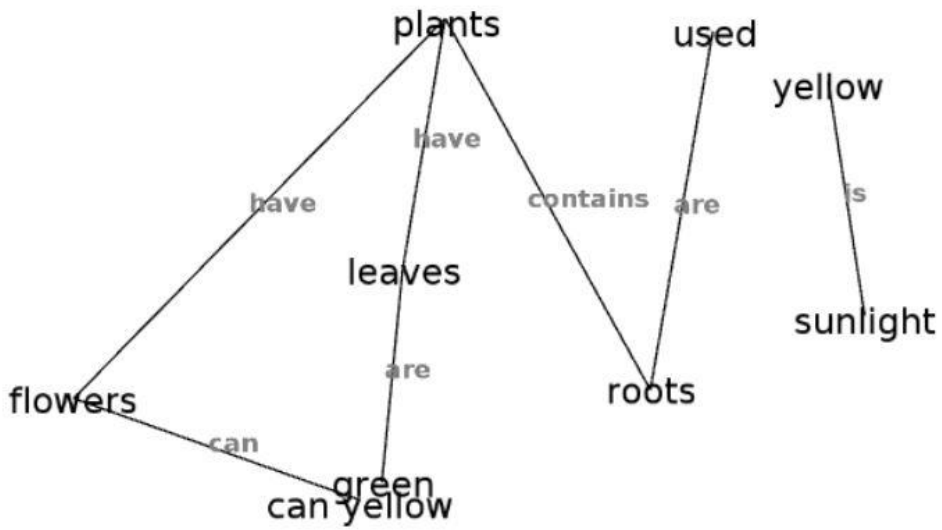
- Step 1 – Find concepts using DBpedia corresponding to each sentence.
- Step 2 - Find relations between subject and objects using OpenIE Tool.
- Step 3 – Finding a pair of concept from a set of DBpedia concepts for each subject-object pair of relation through cosine similarity.
- Step 4 – Find similarity between the concept pairs (Cosine similarity using Glove Vectors).
- Step 4 – Apply the threshold to take the similar pair and add them in the Concept Map.
- Step 5 – Use networkx to create the graph and save it as gephi file.
- Step 6 – Change parameters in Gephi tool.

Method 2.2

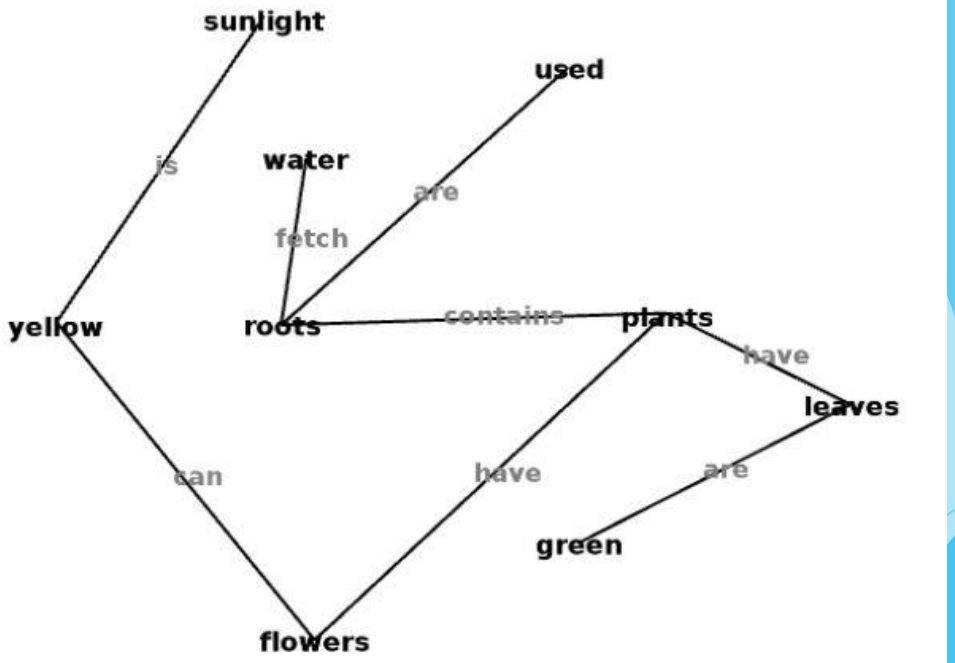
- Step 1 – Find concepts using DBpedia corresponding to each sentence.
- Step 2 - Find relations between subject and objects using OpenIE Tool.
- Step 3 – Finding a pair of concept from a set of DBpedia concepts for each subject-object pair of relation through cosine similarity as well as textual similarity.
- Step 4 – Find similarity between the concept pairs (Cosine similarity using Glove Vectors).
- Step 4 – Apply the threshold to take the similar pair and add them in the Concept Map.
- Step 5 – Use networkx to create the graph and save it as gephi file.
- Step 6 – Change parameters in Gephi tool.

Outputs

Method 1.0

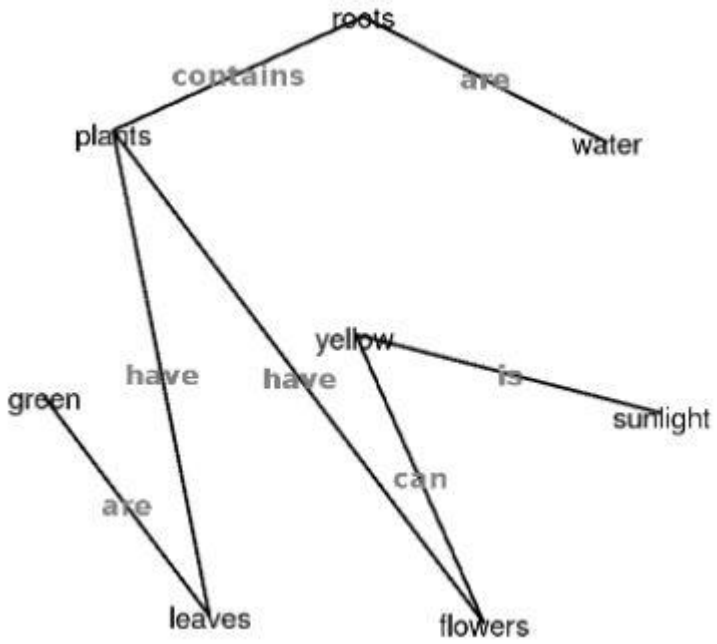


Method 2.0

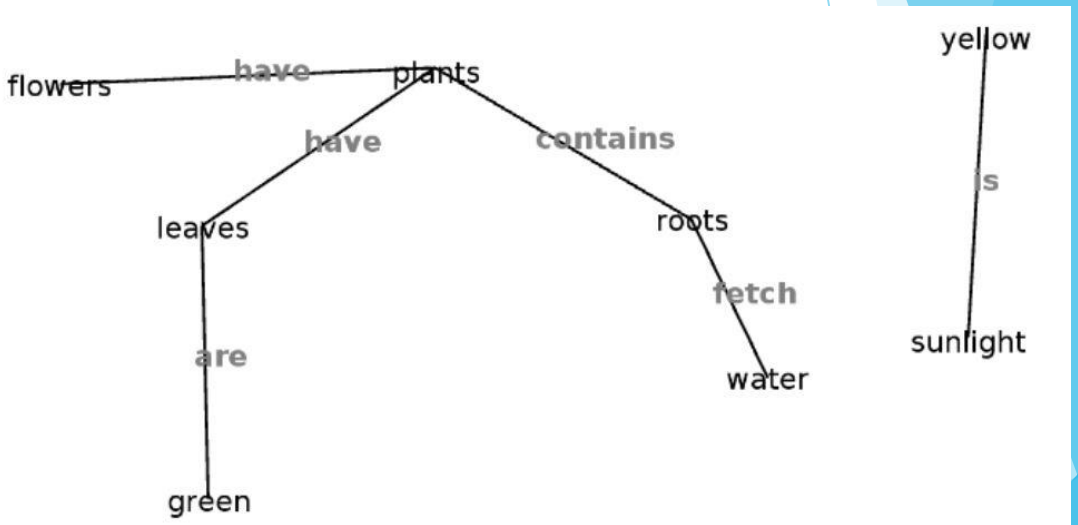


Outputs

Method 2.1



Method 2.2



Inference

Version 1.0				
File	#words	OpenIE time	Similarity time	Graph generation time
Hussain_txt	32	25.97	0.129	0.0045
grade 6	945	29.58	1.9529	0.0049
grade 7	2212	33.204	6.498	0.007428

Version 2.0				
File	#words	OpenIE time	DBpedia time	Graph generation time
Hussain_txt	32	25.08	4.65	0.0015
grade 6	945	28.528	33.916	0.00735
grade 7	2212	32.905	63.124	0.0361

Version 2.1				
File	#words	OpenIE time	DBpedia time	Graph generation time
Hussain_txt	32	26.71	5.344	0.0018
grade 6	945	33.04	35.5	0.0067
grade 7	2212	39.263	62.66	0.0433

Version 2.2				
File	#words	OpenIE time	DBpedia time	Graph generation time
Hussain_txt	32	24.244	4.677	0.0013
grade 6	945	30.24	33.313	0.0764
grade 7	2212	37.34	63.45	0.091

References

- https://nlp.stanford.edu/projects/glove/
- https://nlp.stanford.edu/software/openie.html
- https://pypi.python.org/pypi/pyspotlight/0.7.1
- http://demo.dbpedia-spotlight.org/
- https://gephi.org/
- http://ncertbooks.prashanthellina.com/
- http://pdftotext.com/