Recommendation using domain ontology

Bhagyashri B(181IT111)
C Sneha(181IT112)
Gayathri(181IT216)

Problem Statement

To find the relevant articles from a dataset of articles using domain ontology and giving the articles in the most relevant order as output.

Introduction

Recommendation systems are getting more and more popular with the increasing internet usage.
Many recommender systems are being developed, for movies, books, e-commerce, web page recommendations etc.
Domain Ontology is of great help when it comes to recommendation systems. It adds increased precision to the already existing recommender systems. They link related web pages, books etc based on the keywords. Based on the percentage of similarity they recommend the most relevant web pages, if it is a web page recommendation system, or books if it is a book recommender system.
In this project we are implementing article recommendation system based on

Literature survey

1)Web-Page Recommendation Based on Web Usage and Domain Knowledge

This paper shows that recommender systems are shown to work better with the inclusion of Domain Ontology. It gives out faster and more accurate results when compared to the existing recommender systems.

2)A Domain Ontology-based Information Retrieval Approach for Technique Preparation

In order to realize information sharing and information reuse of technique preparation process, domain ontology was applied in technique preparation process.

Latent Dirichlet Allocation (LDA)

LDA model algorithm is an algorithm that transforms document from Bag-of-Words model space into a topic space.

LDA is a probabilistic topic modeling technique. In topic modeling we assume that in any collection of interrelated documents there are some combinations of topics included in each document

LDA works in an unsupervised way. It is because, LDA use conditional probabilities to discover the hidden topic structure. It assumes that the topics are unevenly distributed throughout the collection of interrelated documents.

The main goal of probabilistic topic modeling is to discover the hidden topic structure for collection of interrelated documents. Following three things are generally included in a topic structure −

- Topics
- Statistical distribution of topics among the documents
- Words across a document comprising the topic

Example-

	Word1	word2	word3	word4	.....
Topic1	0.01	0.23	0.19	0.03	
Topic2	0.21	0.07	0.48	0.02	
Topic3	0.53	0.01	0.17	0.04	

Implementation

1)First we take in the dataset and read the words in each article in it individually. From this, we remove all the unnecessary words like is, the, were, etc. Removing words that occur way too frequently or too rarely helps us to get optimized results.

2)Then we create a Dictionary for words using gensim.

3)Further, using doc2bow we calculate the frequency of words.

4)Using gensim.models.ldamodel.LdaModel we converted Bag-of-Words model space into a topic space.

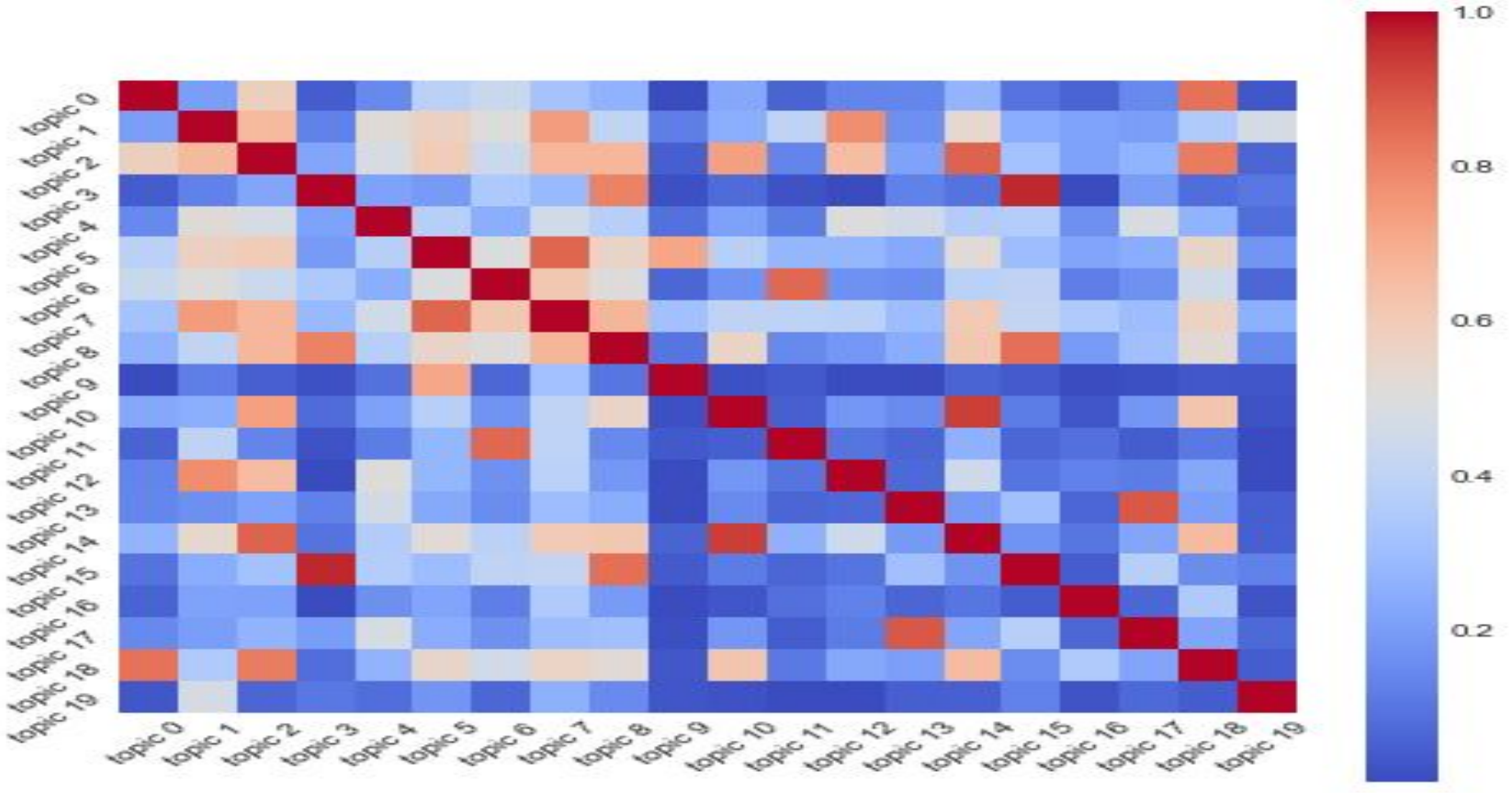
5)Using heatmap to visualize how closely one topic is related to another.

6)Then we take search input from user, which is then matched to the most relevant topics in the model.

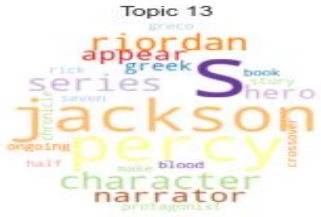
7)These relevant articles are then sorted from most relevant to least.

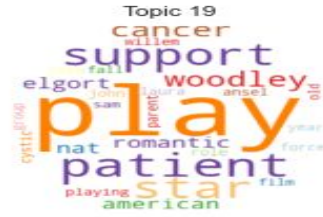
Results

Relation between the topics

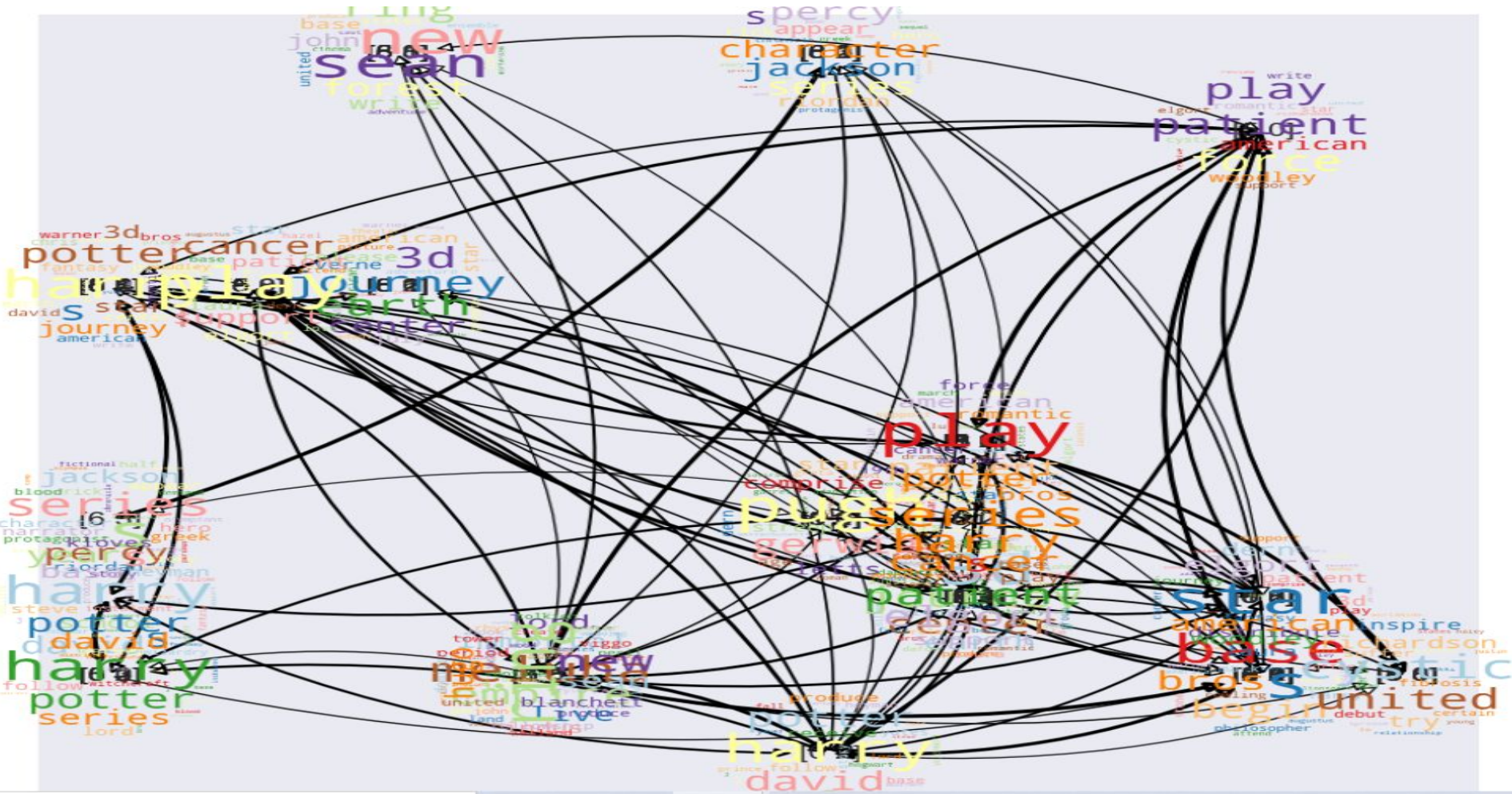


Words belong to each topic





Related Topics



Articles Related To Given Input (Final output)

```
Your question is: Harry Potter
['harry', 'potter']
TOP 0 RELEVANT (45, 0.9158288240432739): «Harry Potter and the Goblet of Fire is a 2005 fantasy film directed by Mike Newell and distributed by Warner Bros. Pictures, based on J. K. Rowling's 2000 novel of the same nam
Azkaban (2004) and the fourth instalment in the Harry Potter film series. The film stars Daniel Radcliffe as Harry Potter, with Rupert Grint as Ron Weasley, and Emma Watson as Hermione Granger. Its story follows Harry'

TOP 1 RELEVANT (36, 0.9095256328582764): «Harry Potter and the Goblet of Fire is a 2005 fantasy film directed by Mike Newell and distributed by Warner Bros. Pictures, based on J. K. Rowling's 2000 novel of the same nam
Azkaban (2004) and the fourth instalment in the Harry Potter film series. The film stars Daniel Radcliffe as Harry Potter, with Rupert Grint as Ron Weasley, and Emma Watson as Hermione Granger. Its story follows Harry'

TOP 2 RELEVANT (9, 0.8963267803192139): «Harry Potter and the Philosopher's Stone (released in the United States and India as Harry Potter and the Sorcerer's Stone) is a 2001 fantasy film directed by Chris Columbus and
Heyman and screenplay by Steve Kloves, it is the first instalment of the Harry Potter film series. The film stars Daniel Radcliffe as Harry Potter, with Rupert Grint as Ron Weasley, and Emma Watson as Hermione Granger.
t he is a famous wizard and begins his education. »

TOP 3 RELEVANT (27, 0.8732552528381348): «Harry Potter and the Prisoner of Azkaban is a 2004 fantasy film directed by Alfonso CuarÃ'n and distributed by Warner Bros. Pictures, based on J. K. Rowling's 1999 novel of the
e sequel to Harry Potter and the Chamber of Secrets (2002) and the third instalment in the Harry Potter film series. The film stars Daniel Radcliffe as Harry Potter, with Rupert Grint as Ron Weasley, and Emma Watson as
amed Sirius Black has escaped from Azkaban and intends to kill him. »

TOP 4 RELEVANT (18, 0.8556745648384094): «Harry Potter and the Chamber of Secrets is a 2002 fantasy film directed by Chris Columbus and distributed by Warner Bros. Pictures, based on J. K. Rowling's 1998 novel of the sa
sopher's Stone (2001) and the second instalment in the Harry Potter film series. The film stars Daniel Radcliffe as Harry Potter, with Rupert Grint as Ron Weasley, and Emma Watson as Hermione Granger. Its story follows
ns the Chamber of Secrets, unleashing a monster that petrifies the school's denizens. »

TOP 5 RELEVANT (90, 0.8554744124412537): «Harry Potter and the Half-Blood Prince is a 2009 fantasy film directed by David Yates and distributed by Warner Bros. Pictures.[3] It is based on J. K. Rowling's 2005 novel of t
loves and produced by David Heyman and David Barron.[4] The story follows Harry Potter's sixth year at Hogwarts as he receives a mysterious textbook, falls in love, and attempts to retrieve a memory that holds the key t

TOP 6 RELEVANT (81, 0.8490674495697021): «Harry Potter and the Half-Blood Prince is a 2009 fantasy film directed by David Yates and distributed by Warner Bros. Pictures.[3] It is based on J. K. Rowling's 2005 novel of t
loves and produced by David Heyman and David Barron.[4] The story follows Harry Potter's sixth year at Hogwarts as he receives a mysterious textbook, falls in love, and attempts to retrieve a memory that holds the key t

TOP 7 RELEVANT (54, 0.8397667407989502): «Harry Potter and the Order of the Phoenix is a 2007 fantasy film directed by David Yates and distributed by Warner Bros. Pictures.[5] It is based on J. K. Rowling's 2003 novel
(making this the only film in the series not to be scripted by Steve Kloves) and produced by David Heyman and David Barron. The story follows Harry Potter's fifth year at Hogwarts School of Witchcraft and Wizardry as th
alongside Rupert Grint and Emma Watson as Harry's best friends Ron Weasley and Hermione Granger. It is the sequel to Harry Potter and the Goblet of Fire and is followed by Harry Potter and the Half-Blood Prince.»

TOP 8 RELEVANT (72, 0.8245474696159363): «Harry Potter and the Order of the Phoenix is a 2007 fantasy film directed by David Yates and distributed by Warner Bros. Pictures.[5] It is based on J. K. Rowling's 2003 novel
(making this the only film in the series not to be scripted by Steve Kloves) and produced by David Heyman and David Barron. The story follows Harry Potter's fifth year at Hogwarts School of Witchcraft and Wizardry as th
alongside Rupert Grint and Emma Watson as Harry's best friends Ron Weasley and Hermione Granger. It is the sequel to Harry Potter and the Goblet of Fire and is followed by Harry Potter and the Half-Blood Prince.»

TOP 9 RELEVANT (53, 0.7276995182037354): «Little Women is a 2019 American coming-of-age period drama film written and directed by Greta Gerwig. It is the seventh film adaptation of the 1868 novel of the same name by Lou
etts, during the 19th century. It stars an ensemble cast comprising Saoirse Ronan, Emma Watson, Florence Pugh, Eliza Scanlen, Laura Dern, TimothÃ©e Chalamet, Meryl Streep, Tracy Letts, Bob Odenkirk, James Norton, Louis
```

Conclusion and future work

The articles which are most relevant to the search input are collected in order.

This can be used to give efficient web search results. Moreover, it can be used in different applications for getting relevant search results.

Technologies used

- Python
- Jupyter notebook

References

Web-Page Recommendation Based on Web Usage and Domain Knowledge https://ieeexplore.ieee.org/document/6514870

B. Liu, B. Mobasher, and O. Nasraoui, “Web usage mining,” in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed. Berlin, Germany: Springer-Verlag, 2011, pp. 527–603.

B. Mobasher, “Data mining for web personalization,” in The Adaptive Web, vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 90–135

G. Stumme, A. Hotho, and B. Berendt, “Usage mining for and on the Semantic Web,” in Data Mining: Next Generation Challenges and Future Directions. Menlo Park, CA, USA: AAAI/MIT Press, 2004, pp. 461–480.

David M. Blei, Andrew Y. Ng, Michael I. Jordan “Latent Dirichlet Allocation” Journal of Machine Learning Research 3 (2003) 993-1022