CS 6301.004

NIKITA VISPUTE NXV170005

ARPITA DUTTA AXD170025

PROJECT 2

REPORT

For this project we worked with the dataset of movie plot summaries. We built a search engine for the plot summaries that are available in the file "plot summaries.txt" and is uploaded on the UTD personal website.

We next created a corpus for the data and then using the tm package preprocessed the data by removing the stop words, punctuation, numbers, white spaces etc. We created the term document matrix and calculated the tf – idf values for each document pair.

The r code can be run from command line and takes in user input for the query. As an output it displays the top 10 most similar documents to the query.

Output of query and the results:

Query: Seventh-day Adventist Church

	Score	ID
Seventh-day Adventist Church pastor Michael Chamb	0.1775	595909
After faithfully serving a full time mission for his chu	0.0874	1405466
The film follows several inhabitants of the Italian tov	0.0745	10386740
This film is based on the true story of Carl Upchurch .	0.0713	35833714
A motley group of phony church leaders attempts to	0.07	24356631
As a group of bikers moves across the desert, they cc	0.0664	7279333
Giuseppe , an ex-convict meets with a village priest I	0.0577	16263909
A young lady about to get married realizes that she h	0.0573	13806284
Dissatisfied with the family architectural business, a	0.0549	27380492
Three brothers and a sister meet in Glasgow to prepa	0.0533	35658593

Query: Australia in 1939

	Score	ID
The film centers around Tom Dunn, an aspiring AFL star wl	0.4085	35724504
Elizabeth Paterson is the daughter of an important figure	0.3861	31992932
The plot of the movie is about newsreel cameramen and p	0.3794	6880509
The film has been shot in Chennai, Australia and New Zea	0.3682	30174841
The film tells the story of a man and wife whose marriage	0.3649	25553027
Based on the life of Dame Nellie Melba, the film traces th	0.3234	31658770
An American brother and sister, Bob and Kathy Prince, hav	0.3169	29629615
Joe Warr is a British sports writer who lives in Australia w	0.3025	22325279
Some aboriginals steal a child in rural Australia. Fifteen ye	0.3003	32987720
A number of mysterious accidents involving the deaths of	0.2938	21686513