

# Movie Recommendation System

Generated by Doxygen 1.8.13



# Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Namespace Index</b>                         | <b>1</b> |
| 1.1      | Namespace List . . . . .                       | 1        |
| <b>2</b> | <b>Namespace Documentation</b>                 | <b>3</b> |
| 2.1      | extended_dataset Namespace Reference . . . . . | 3        |
| 2.1.1    | Detailed Description . . . . .                 | 4        |
|          | <b>Index</b>                                   | <b>5</b> |



# Chapter 1

## Namespace Index

### 1.1 Namespace List

Here is a list of all documented namespaces with brief descriptions:

|  |   |
|--|---|
| <a href="#">extended_dataset</a> . . . . . | 3 |
|--|---|



## Chapter 2

# Namespace Documentation

### 2.1 extended\_dataset Namespace Reference

#### Variables

- `ps` = PorterStemmer()  
*using stemming in python library*
- `stop_words` = set(stopwords.words('english'))  
*stop word removal*
- `int url_column` = 4  
*url column in dataset\_with\_titleEng.csv*
- `final_dataset` = open('dataset.csv','w')
- `bool firstline` = True  
*skip the first line, as it is a header*
- `string comma` = ','
- `readCSV` = csv.reader(csvfile, delimiter=',')
- *dataset extended with English title*
- `res` = requests.get(row[url\_column])  
*fetch the url of dataset\_with\_titleEng.csv, one at a time*
- `soup` = bs4.BeautifulSoup(res.text,'lxml')
- `list act` = []
- `list ch` = []
- `attrs`  
*take storyline from the url*
- `href`
- `string actor_string` = ' | '.join(map(str,act))
- `string character_string` = ' | '.join(map(str,ch))
- `string old_row` = ' '.join(map(str,row))
- `summary` = story.text  
*make all characters small*
- `string sep` = 'written'  
*remove "written by" at the end of each storyline*
- `tokenizer` = RegexpTokenizer(r'\w+')  
*to remove punctuations*
- `li` = tokenizer.tokenize(summary)
- `list filtered_sentence` = [w for w in li if not w in stop\_words]  
*stop word removal*
- `list final_dataset_str` = []  
*stemming in python*
- `string story_clean` = ' '.join(final\_dataset\_str)

### 2.1.1 Detailed Description

This module takes input the title extended dataset, named dataset\_with\_titleEng.csv and gives the extended dataset with acto(s),character(s),storyline etc.(dataset.csv)  
For web scraping : BeautifulSoup was used  
For cleaning data and NLP : nltk library was used



# Index

`extended_dataset`, [3](#)