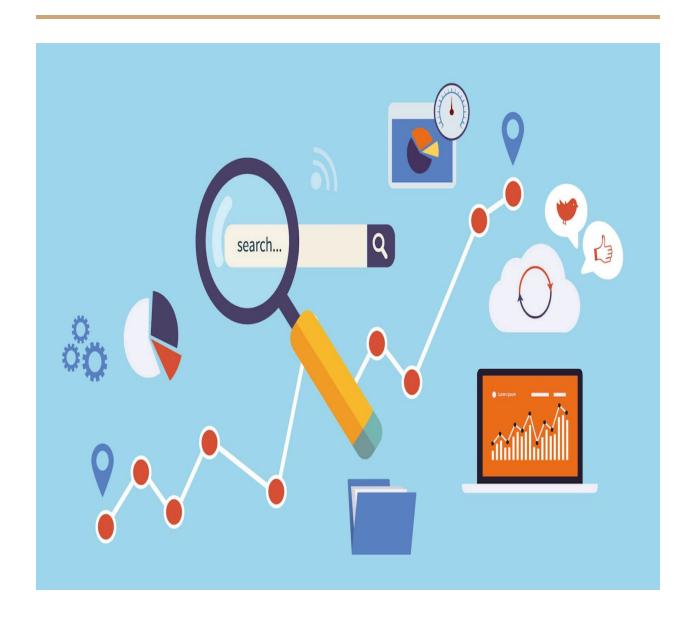
Search Engine Using R



Team Members

Omkar Nandkumar Dixit ond 170030

Dhwani Raval dsr170230 Karan Yashwant Kanani kyk170030 Introduction

A search engine is a software, typically accessed on the Internet, that searches a database

of information according to the user's query. The engine provides a list of results that best

match what the user is trying to find.

TF-IDF

TF-IDF, short for term frequency - inverse document frequency, is a numerical statistic that

is intended to reflect how important a word is a document in a collection or corpus. It is

often used as a weighting factor in searches of information retrieval, text mining, and user

modelling. The tf-idf value increases proportionally to the number of times a word appears

in the document and is offset by the number of documents in the corpus that contain the

word, which helps to adjust for the fact that some words appear more frequently in

general. TF-IDF is one of the most popular term-weighing schemes today, 83% of text-based

recommender systems in digital libraries use tf-idf.

Files

searchEngine.R

Implementation Details

Libraries: ™, dplyr, tidytext, wordcloud, SnowballC

Scripting Language: R

2

Screenshot

