

# PROJECT TITANIS

## Milestone 2

Anil Shanbhag 100050082  
Rahee Borade 100050028

# Repo Reader

This class adds the files stored in repository in a queue and gets the contents of the next file in the queue.

# Tokenizer

This class gets the contents of file from Repo Reader, converts it into a vector of tokens and returns plain text from it. It also returns a list of URLs present in the file.

# Doc Index

This class assigns a unique index to all the URLs obtained from tokenizer.

# Forward Index

This class maps the Doc Ids to the map of words present in the file with their occurrence count.

# Inverted Index

This class maps words to the list of URLs which contain them.

# Page Rank

This class computes the page rank of an URL and stores a map of Doc ID to its page rank.