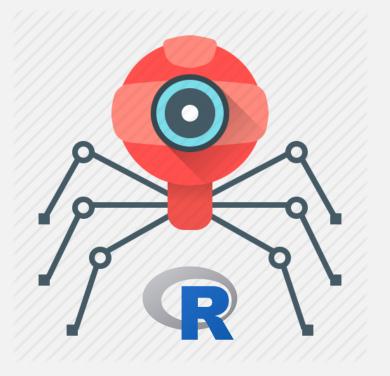
CS 636 Data Analytics with R Program Prof. Zhi Wei





Mid Term Project – Team 1

Aakash Asli | Shao Yu Chen | Tushar Asnani



Scope of Project

Use R program to extract required information from specified website

Main page:

https://genomebiology.biomedcentral.com

Total articles: 6300

Required information:

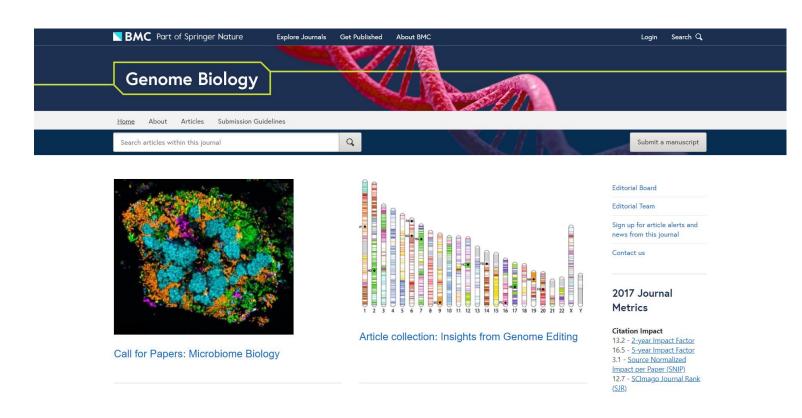
10 fields: DOI, Title, Authors, Author

Affiliations, Corresponding Author,

Corresponding Author's Email, Publication

Date, Abstract, Keywords, Full Text (Textual

format)



Contribution



Contribution of Group Members

Aakash Asli	main program: (loadAllArticles.R) Functions(util.R): analysisArticle extractAuthors extractAffliation	Testing and debugging Project report(PDF file)
Tushar Asnani	Function(util.R): • loadArticleList	Testing and debugging Readme.txt
Shao Yu Chen	Function(util.R): extract extracAttribute	Testing and debugging

Challenges

unfamiliar with R packages, like XML	Check the online document and samples
extract author, corresponding author, corresponding author's email	 Carefully check the HTML of the article page and find the xpath Use a 2nd time extraction to get more detailed or related data





Thank you