



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)

The screenshot displays the BMC Genome Biology journal homepage. The header includes the BMC logo, 'Part of Springer Nature', and navigation links: 'Explore Journals', 'Get Published', and 'About BMC'. A search bar and a 'Login' link are also present. The main banner features the journal title 'Genome Biology' in a yellow box, with a background image of a DNA double helix. Below the banner, there are links for 'Home', 'About', 'Articles', and 'Submission Guidelines'. A search bar with the placeholder 'Search articles within this journal' and a 'Submit a manuscript' button are located at the bottom of the header section. The main content area is divided into three columns. The left column features a colorful, abstract image of a microbial community and a 'Call for Papers: Microbiome Biology' link. The middle column displays a 'Article collection: Insights from Genome Editing' with a visualization of human chromosomes. The right column contains links for 'Editorial Board', 'Editorial Team', 'Sign up for article alerts and news from this journal', and 'Contact us'. Below these links, the '2017 Journal Metrics' section lists various impact factors and rankings, including a 2-year impact factor of 13.2, a 5-year impact factor of 16.5, a source normalized impact per paper (SNIP) of 3.1, and a SCImago Journal Rank (SJR) of 12.7.

BMC Part of Springer Nature Explore Journals Get Published About BMC Login Search

Genome Biology

Home About Articles Submission Guidelines

Search articles within this journal Submit a manuscript

Call for Papers: Microbiome Biology

Article collection: Insights from Genome Editing

2017 Journal Metrics

Citation Impact
13.2 - 2-year Impact Factor
16.5 - 5-year Impact Factor
3.1 - Source Normalized Impact per Paper (SNIP)
12.7 - SCImago Journal Rank (SJR)

Contribution of Group Members

Aakash Asli	main program: (loadAllArticles.R) Functions(util.R): analysisArticle extractAuthors extractAffiliation	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	<ul style="list-style-type: none">Carefully check the HTML of the article page and find the xpathUse a 2nd time extraction to get more detailed or related data



Thank you