

[github.com / lmmx /](https://github.com/lmmx) scholaRdaemon



Bots for keeping up with the
flood of academic literature

Louis Maddox

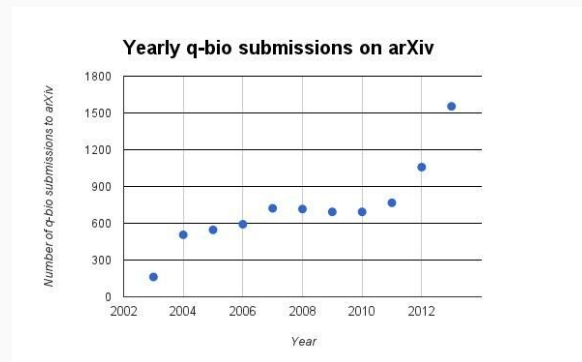
Bioinformatics and Systems Biology MSc

University of Manchester

 @permutans

Literature bots

- FlyPapers: Casey Bergman at UoM set off trend of Pubmed bots
 - 200+ now in his list: twitter.com/caseybergman/lists/literaturebots/members
- Pubmed: no direct links to papers
- ‘Preprints’ not indexed there, but can be found on Google Scholar (*arXiv*, *bioRxiv*, ...)



c(gmailr, RJSONIO, twitterR) + *cron*

- Quick and simple JSON storage, referring to local directories

```
louis ~ ~/Dropbox/Y3/Programming/scholarDaemon/config $ cat bot_registry.json
{"bots":{"idp":"IDP_papers", "mi-xpr":"miR_xpr_papers", "dna-topo":"DNAtopologyBot", "slims":"SLiM_papers",
"edgtx":"edgetics_papers","transxn-ns":"TransxnNoiseBot", "rna-ed":"RNA_edit_papers"}}
louis ~ ~/Dropbox/Y3/Programming/scholarDaemon/config $ ls
_citations  DNAtopologyBot  IDP_papers      RNA_edit_papers  TransxnNoiseBot
bot_registry.json  edgetics_papers  miR_xpr_papers  SLiM_papers
louis ~ ~/Dropbox/Y3/Programming/scholarDaemon/config $ ls IDP_papers/
gmail_authfile.json  gmail_id_log.json  sd_config.json  twitter_authfile.json
```

- 0 * * * * source /home/louis/.bashrc; date >> "\$scholarDaemon"sd.log; runsdaemon >> "\$scholarDaemon"sd.log

Google Developers Console: create project

- Your usual OAuth2 registration process to use a Google API, download JSON

The image displays three sequential screenshots from the Google Developers Console:

- New Project:** Shows the 'Create Project' dialog. The 'PROJECT NAME' is 'GmailReader' and the 'PROJECT ID' is 'gmail-tweets'. 'Create' and 'Cancel' buttons are visible.
- Create Client ID:** Shows the 'OAuth' section for the 'GmailReader' project. Under 'APPLICATION TYPE', 'Installed application' is selected. A warning message states: 'To create a Web Client ID or an Installed Application Client, you need to set a product name in the consent screen.' A 'Configure consent screen' button is highlighted.
- Consent screen:** Shows the 'Consent screen' configuration page. It includes fields for 'EMAIL ADDRESS', 'PRODUCT NAME' (set to 'GmailReader'), 'HOME PAGE URL', 'PRODUCT LOGO', 'PRIVACY POLICY URL', 'TERMS OF SERVICE URL', and 'GOOGLE+ PAGE'. A preview window on the right shows the user consent screen layout.

See wiki: <https://github.com/lmmx/scholarDaemon/wiki>

client_secret_... .apps.googleusercontent.com.json

Parsing papers from email

- Reads papers into XML directly from email HTML
- Pulls out direct links from Google redirect URLs

See wiki: <https://github.com/lmmx/scholarRdaemon/wiki>

```
scholar.alerts.mo <- messages('Scholar Alert microRNA oscillation')  
mo.msg.list <- scholar.alerts.mo[[1]][['messages']]  
eg.msg.id <- mo.msg.list[[1]][['id']]
```

Then proceed as above, but building a list of new papers by parsing the HTML with R's `XML` package:

```
eg.msg.resp.data <- message(eg.msg.id, format = "full")$payload$body$data  
eg.msg.html <- gsub('\r\n', '', Base64URL_Decode_To_Char(eg.msg.resp.data))
```

```
library('XML')  
mail.doc <- htmlParse(eg.msg.html, asText=TRUE, encoding = "UTF-8")  
mail.root <- xmlRoot(mail.doc)  
# See the XML package vignette for details: http://www.omegahat.org/R/XML/Tour.pdf  
  
article.list <- xmlElementsByTagName(mail.root[["body"]][["div"]], "h3")  
  
article.summaries <- lapply(article.list, GetPaper)  
names(article.summaries) <- NULL
```

```
> article.summaries  
[[1]]  
[[1]]$title  
[1] "A Dicer-miR-107 Interaction Regulates Biogenesis of Specific miRNAs Crucial for Neurogenesis"  
  
[[1]]$url  
[1] "http://www.sciencedirect.com/science/article/pii/S153458071400834X"  
  
[[2]]  
[[2]]$title  
[1] "Notch in memories: Points to remember"  
  
[[2]]$url  
[1] "http://onlinelibrary.wiley.com/doi/10.1002/hipo.22426/abstract"  
  
[[3]]  
[[3]]$title  
[1] "Glucocorticoids and 11 $\beta$ -hydroxysteroid dehydrogenases: mechanisms for hypertension"  
  
[[3]]$url  
[1] "http://www.sciencedirect.com/science/article/pii/S1471489215000065"
```

One app to rule them all

Application Management

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

See wiki: <https://github.com/lmmx/scholarDaemon/wiki>

messages

Search Twitter

- DNA topology papers** @DNAtopologyBot · Dec 5
Quantitative and Systems-Based Approaches for Deciphering Bacterial Membrane Interactome and Gene Function link.springer.com/chapter/10.100...
[View summary](#)
- DNA topology papers** @DNAtopologyBot · Dec 5
ESPS Manuscript NO: 20155 Manuscript Type: TOPIC HIGHLIGHT wjnet.com/esps/DownloadF...
- DNA topology papers** @DNAtopologyBot · Dec 5
Suppression of transcriptional drift extends C. elegans lifespan by postponing the onset of mortality elifesciences.org/content/4/e088...
- IDP Papers** @IDP_papers · Dec 5
Binding of disordered peptides to Kelch: Insights from enhanced sampling simulations pubs.acs.org/doi/pdfplus/10...
- IDP Papers** @IDP_papers · Dec 5
Structural biology: Structure in the cellular context nature.com/hmeth/journal/...
- IDP Papers** @IDP_papers · Dec 5
Effect of O-linked Glycosylation on the Equilibrium Structural Ensemble of Intrinsically Disordered Polypeptides pubs.acs.org/doi/pdf/10.102...
- SLiM papers** @SLiM_papers · Dec 3
In silico allergenicity prediction of several lipid transfer proteins sciencedirect.com/science/articl...
- Transcr'n noise bot** @TransxnNoiseBot · Dec 1
Characteristics and roles of lncRNAs in hepatocellular carcinoma: Novel insights into their mechanism wjnet.com/esps/DownloadF...
- Transcr'n noise bot** @TransxnNoiseBot · Dec 1
Computational Modeling of miRNA Biogenesis link.springer.com/chapter/10.100...
[View summary](#)