Title: Catch arXiv papers Date: 2018-01-13 10:11 Category: Hacking Tags: arXiv, papers Author: Samdney

Have you ever wanted to catch the newest arXiv.org papers from your favorite research areas? I do.

At first, I was sure that arXiv offers a comfortable way to catch all the newest papers or all papers from a particular publication day. Sadly, I was wrong.

I didn't find such a possibility for an arbitrary day. You can only get an overview over the newest papers (E.g. like this https://arxiv.org/list/astro-ph/new). But already, only to get a list of publications longer ago than one week isn't possible. That's not nice.

The next idea was, to catch the papers by its arXiv-numbers. Nice idea, but it doesn't work if you only want papers from a particular area. The typical arXiv number consists of two parts. The first one gives you the publication date. The second part (the one after the dot) is a counter and tells you, this papers was the N'th paper of this day. Sadly, this counter goes over all areas. That means paper N belongs to area A, but paper N+1 can belong to area B. So we can't use the number for catching only papers from a particular research area.

Ok. I had to look for an other way. After some research, I found the possibility for subscribing on an arxiv mailing list (See also https://arxiv.org/help/subscribe), which sends you every working day an email about the newest papers of your favorite areas. Yeah, I subscribed on this about one year ago, so I have a nice overview for each day, now.

The content of this email looks like the following example from Fri, 12 Jan 2018

```
------------------------------------------------------------------------------
\\
arXiv:1801.03894
Date: Thu, 11 Jan 2018 17:41:20 GMT   (26kb)

Title: Stability in the homology of Deligne-Mumford compactifications
Authors: Philip Tosteson
Categories: math.AG math.AT math.GT
Comments: 15 pages, Comments welcome
\\
  Using the the theory of FS^op modules, we study the asymptotic behavior of
the homology of $\overline M_{g,n}$, the Deligne--Mumford compactification of
the moduli space of curves, for $n >> 0$. An FS^op module is a contravariant
functor from the category of finite sets and surjections to vector spaces. Via
maps that glue on marked P^1's, we give the homology of $\overline M_{g,n}$ the
structure of an FS^op module and bound its degree of generation. As a
consequence, we prove that the generating function $\sum_{n} \dim(H_i(\overline
M_{g,n})) t^n$ is rational, and its denominator has roots in the set $\{1, 1/2,
\dots, 1/p(g,i)\}$ where $p(g,i)$ is a polynomial of order $O(g^2 i^2)$. We
```

```
also obtain restrictions on the decomposition of the homology of $\overline
M_{g,n}$ into irreducible $S_n$ representations.
\\ ( https://arxiv.org/abs/1801.03894 ,  26kb)
------------------------------------------------------------------------------
```

You see, we receive a lot of information which we can use to catch our favorite papers, now.

In the following, I wrote a simple script to fetch all pdf-files based on the mailing list. You can also find it on https://github.com/Samdney.

```
#!/bin/sh

#*****************************
# Download of arXiv papers (pdf), based on mailing list information
# Author: Carolin Zöbelein
#*****************************

filename=$1 # Email file: arXiv mailing list
date=$2     # Date of paper submission

# Create and go to directory, move email to directory
mkdir $date
mv $filename $date
cd $date

# Read arXiv-ids from file and download belonging pdfs
readarray -t lines < "$filename"
for line in "${lines[@]}"; do
    if [[ $line == arXiv:* ]]
        then
        temp=$line; set - $temp; temp2=${*:1:1}; temp3=${temp2:6}; echo "$temp3"
        url=http://arxiv.org/pdf/$temp3.pdf
        echo "$url"
        wget --user-agent=Lynx $url
    fi
done

# Rename belonging email
mv $filename $date.txt
```

It's a very easy script. No hardcore hacking, but it makes your life a bit nicer :D. Enjoy!