

HTRC Data API Hands-on



Yiming Sun

Goal


- Show key pieces in writing code to request data from Data API
 - Request token from OAuth2 token endpoint
 - Request volumes from Data API
 - Processing ZIP
- Demo code:
 - Simple word count
 - Available in Java (v1.5+) and Python (v2.7)



Download code packages

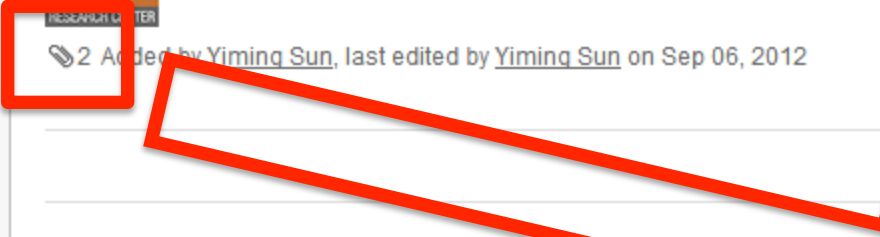
- <http://htrc.mine.nu/confluence/display/OUT/HTRC+Uncamp+Data+API+Demo>

Dashboard > Outreach > HTRC Uncamp Data API Demo




HTRC Uncamp Data API Demo

2 Added by [Yiming Sun](#), last edited by [Yiming Sun](#) on Sep 06, 2012





Dashboard > Outreach > HTRC Uncamp Data API Demo



HTRC Uncamp Data API Demo

Attached Files

Name	Size
 htrc-uncamp-python-demo.zip	8 kB
 htrc-uncamp-java-demo.zip	55 kB

Run the demo code

Java:

```
$ unzip htrc-uncamp-java-demo.zip
$ cd htrc-uncamp-java-demo
$ chmod 700 *.sh
$ ./volwc.sh 25vollist.txt <id> <scrt>
```

Python:

```
$ unzip htrc-uncamp-python-demo.zip
$ cd htrc-uncamp-python-demo
$ python VolumeWordCountClient.py 25vollist.txt <id> <scrt>
```



What about parallelization?

- At first level
 - Meandre workflows can leverage multiple cores
 - Use multi-threaded code when run on multi-core single node
 - Ideas on how to parallelize demo code
- More advanced level
 - MPI
 - Map/Reduce
- What algorithms do you use, and what are your basic needs?

