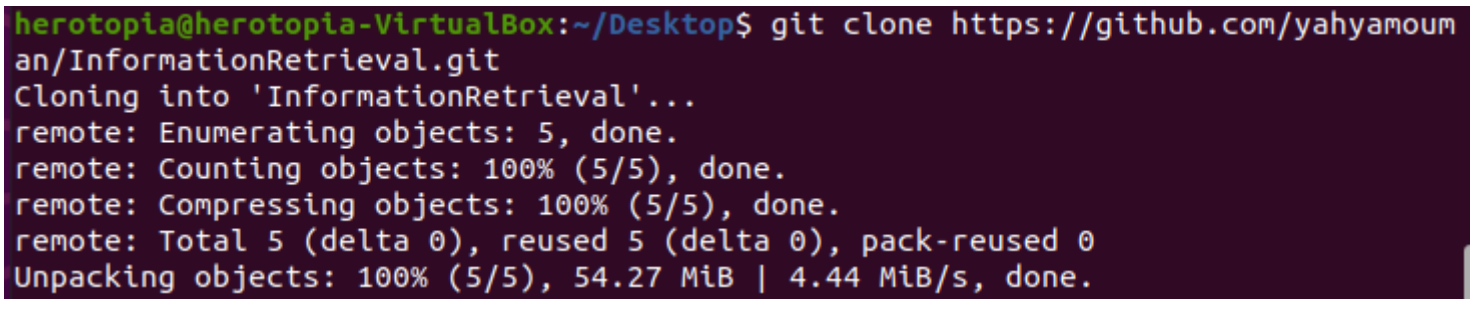
Information Retriaval (Setup)

Team	Yahya Mouman	Yasser Faleh	Ibtissam Slalmi	Younes Harir

Clone the git repository

```
git clone https://github.com/yahyamouman/InformationRetrieval.git
```

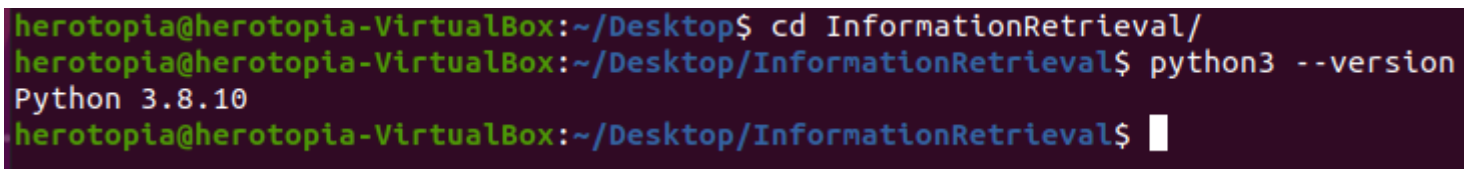


Access the folder

```
cd InformationRetrieval
```

Check python version (need 3.8.x)

```
python3 --version
```



Automatic

run setup.sh on sudo to unzip data and install dependencies (python libraries, etc)

OR Manual

Install unzip to extract XML data

```
sudo apt-get install unzip
unzip XML-Coll-withSem.zip -d XML-Coll-withSem
```

Install pip and dependencies

```
sudo apt-get install pip
pip install numpy
pip install nltk
pip install lxml
pip install beautifulsoup4
```

Run code

Run code to generate 50+ element grannularity runs with different configurations

(It takes 15 minutes to parse the entire 9840 XML documents in our machine but the time can differ on a virtual machine)

```
python3 invertedIndexGrannularityFinal.py
```

Alternative

The following link is a google collab cheat where you can run the code in a jupyter notebook format

https://drive.google.com/file/d/1jQQxq6HHHO8icIXArLYGG_MKcSenFErp/view?usp=sharing

Link to drive with runs

https://drive.google.com/drive/folders/1r5ekxYDB1kb8G6C1Au0trqAgcZM_IYp2?usp=sharing