

**DEMO FILE**  
**INFORMATION RETRIEVAL AND WEB SEARCH**  
**(COMP 6791)**  
**CONCORDIA UNIVERSITY**

**PROJECT 4**

**SUBMITTED BY:**  
**MANJIT SINGH**  
**(40185580)**

## **UPPER BOUND**

At the beginning of the running code it's gonna ask the user that:-

*"How many pages you want to crawl? "*

Based on the answer that many pages will be crawled and downloaded.

## **CLUSTERING RESULTS FOR K=3**

### **Cluster number 0:**

Top 50 terms are:

[

quartier,conrad,café,cotutelle,streets,prizes,faq,visitors,tours,sector,units,spotlights,postdocs,governance,partnerships,shuttle,integrity,involved,bus,sustainability,orientation,abroad,history,operative,bipoc,offices,jobs,test,administration,global,facts,data,food,tuition,external,perspectives,researchers,fitness,athletics,continuing,language,engagement,chairs,black,fast,directions,aid,strategic,fees,residence

]

***Possible Cluster Name:- Transit Mediums and Administrative Services.***

### **Cluster number 1:**

Top 50 terms are:

[

efollettskip,content,main,november,december,williams,george,sir,loyola,blvd,maisonneuve,de,hg,page,canada,montreal,back,menu,skip,site,www,en,html,universityhttps,call,friends,street,campaign,directories,accessibility,centre,media,privacy,feedback,sgw,hb,map,expend,parents,study,academic,terms,search,news,new,campus,support,students,website,sherbrooke

]

***Possible Cluster Name:- Loyola Campus and its Student Support Services.***

### **Cluster number 2:**

Top 50 terms are:

[

sign,help,account,outlook,moodle,cu,schedule,find,hours,us,you,opportunities,ca,concordia,university,support,new,emergency,computer,students,terms,campus,study,search,contact,latest,events,maps,marketing,news,website,academic,general,share,online,linkedin,twitter,social,facebook,instagram,create,flickr,youtube,connect,myconcordia,cody,arts,schedules,colleges,services

]

***Possible Cluster Name:- Concordia University Latest events.***

## **AFINN SCORES FOR CLUSTERING RESULTS WITH K=3**

***Sentiment score of Cluster 0 using AFINN: 4.0***

***(Positive Sentiment)***

***Sentiment score of Cluster 1 using AFINN: 2.0***

***(Positive Sentiment)***

***Sentiment score of Cluster 2 using AFINN: 5.0***

***(Positive Sentiment)***

## **CLUSTERING RESULTS FOR K=6**

### **Cluster number 0:**

Top 50 terms are:

[

returns,collaborations,browzine,nod,annotated,citing,broken,scan,papers,scanner,edition,delivery,scanners,copiers,periodicals,spotlight,vanier,tablets,authors,wikipedian,booking,borrow,an,wi,laptops,printers,improvement,readings,interlibrary,communitydecember,microforms,intercampus,citation,citations,zotero,memiors,bibliometrics,thanks,journals,checkouts,sofia,deposit,retro,orcid,bibliographies,exhibitionnovember,spectrum,users,bound,reserves

]

***Possible Cluster Name:- Campus Services of Digital equipments.***

### **Cluster number 1:**

Top 50 terms are:

[

art,centres,human,interdisciplinary,exchange,technology,design,materials,biology,chain,supply,environment,physiology,biochemistry,kinesiology,applied,society,aerospace,cinema,dance,chairs,journalism,diversity,public,accountancy,environmental,contemporary,therapies,diverse,systems,humanities,summer,jewish,home,sciences,canadian,chemistry,mathematics,psychology,studio,music,institute,planning,finance,industrial,languages,modern,anthropology,hoppenheim,françaises

]

***Possible Cluster Name:- Courses Available.***

### **Cluster number 2:**

Top 50 terms are:

[

sign,main,content,page,help,support,new,policy,experience,necessary,analytical,technical,function,accept,properly,possible,cookies,uses,preference,best,contact,marketing,us,www,online,universityhttps,share,menu,en,linkedin,facebook,create,instagram,twitter,html,site,flickr,skip,social,youtube,academic,website,news,webmail,resourcescampus,ofengineering,relationsfaculties,serviceswhat,university,concordia

]

***Possible Cluster Name:- Social Media and Online Support.***

### **Cluster number 3:**

Top 50 terms are:

[

info,networks,different,like,value,features,generated,and,researchteachinglearning,rotate,or,located,ensembles,people\_places\_entities,club,ience,pt,general\_interest,ambassadors,active\_living,self\_care,zen\_den,mental\_health,healthy\_living,empathic\_listening,trained,zeynep,we,origin,arsel,keyword,campuses,concerts,audiences,know,let,little,screen,loy,markets,degrees,close,things,try,think,oscar,whole,peterson,countries,jump

]

***Possible Cluster Name:-Healthy Living Lifestyle.***

#### **Cluster number 4:**

Top 50 terms are:

[

ga,mu,td,va,ci,mi,fg,sb,er,pr,gm,fa,rr,mb,gn,lb,fb,ev,bike,rack,kiosk,cl,ls,vendôme,  
wheelchair,atm,legend,ld,cb,enforcement,gs,commuter,métro,panoramas,all,goo  
gle,parking,faqs,getting,west,rules,guy,accessible,buildings,using,bus,shuttle,direc  
tions,ps,vl

]

***Possible Cluster Name:- Transit mediums.***

#### **Cluster number 5:**

Top 50 terms are:

[

prizes,faq,quartier,cotutelle,café,conrad,streets,visitors,postdocs,spotlights,gover  
nance,partnerships,units,tours,bus,shuttle,involved,sector,integrity,operative,offi  
ces,test,jobs,bipoc,external,data,orientation,abroad,food,global,facts,tuition,sust  
ainability,history,engagement,athletics,administration,continuing,language,chairs,  
researchers,perspectives,residence,strategic,fees,fitness,aid,directions,fast,black

]

***Possible Cluster Name:- Administration Support and Services.***

## **AFINN SCORES FOR CLUSTERING RESULTS WITH K=6**

***Sentiment score of Cluster 0 using AFINN: 3.0***

***(Positive Sentiment)***

***Sentiment score of Cluster 1 using AFINN: 0.0***

***(Neutral Sentiment)***

***Sentiment score of Cluster 2 using AFINN: 9.0***

***(Positive Sentiment)***

***Sentiment score of Cluster 3 using AFINN: 2.0***

***(Positive Sentiment)***

***Sentiment score of Cluster 4 using AFINN: 1.0***

***(Positive Sentiment)***

***Sentiment score of Cluster 5 using AFINN: 4.0***

***(Positive Sentiment)***

## **CONCLUSION**

- Overall, for both k=3 and k=6 the sentiment score was inclined towards positive sentiment.
- Top 50 terms of each cluster are very relatable to each other hence one can say that optimized clusters are formed.



**NOTE**

**JUST RUN Clustering.py.**

**File Present in Web\_Crawler\WebCrawler\WebCrawler\spiders\**