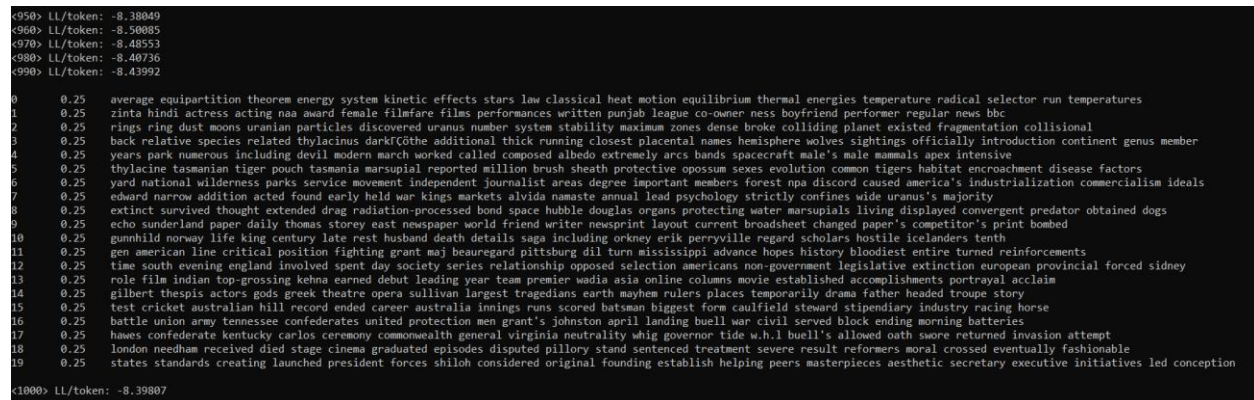# ANLP ASSIGNMENT 6

Team : Archish Ramesh Babu, Pravin Sundar, Prashanth Thirukkurungudi Sekar

& Barathwaaj Parthasarathy

I have used MALLET for the purpose of this assignment. MALLET which stands for MAchine Learning for LanguagE Toolkit is a java based package which can perform topic modelling based on input text documents. The tool imports multiple text documents at the same time and can create n number of topic's for the input data.

Below is a screenshot from the tool:

The text chosen for Topic Modelling exercise on the subject of illicit trading of Artifact. There are multiple reports on this topics and I have used 20 such documents to create the topic models. (Link → https://figshare.com/articles/looted_heritage_reports_txt.zip/91828)

Some of the questions I hope to answer based on the results from the Topic Modelling exercise are as follow,

1. Which are the countries that have a lot of illicit trading of stolen artifacts?
2. What are some of the recent historical artifacts that were stolen or recovered?
3. How is the government reacting to this?

5 TOPICS

```
0     0.1177     market china chinese antiquities artifacts court war high
golan peru items collection jewish jade documents malter fake art treasure
1     0.30133     sites archaeological looting site heritage cultural
archaeologists antiquities people archaeology national egypt university
local trade state history project international
```

2   0.24037   museum art antiquities statue objects getty museums looted university italy returned auction italian princeton sotheby york cultural potts cambodia
3   0.3646   museum police stolen greece antiquities ancient greek art theft found artifacts year culture ministry security thieves items government country
4   0.14963   antiquities egyptian egypt hawass syria british heritage iraq department cultural state aaa al country artifacts national museum syrian ancient

10 TOPICS

0   0.13225   statue auction sotheby art cambodia china looted market government united cultural chinese states cambodian sale private war relics years
1   0.13243   show artifacts sites american national metal treasure history archaeology tv archaeological st state malter augustine native wrote historic information
2   0.20554   sites site archaeological looting antiquities ancient egypt people city syria el area hibeh heritage department village ruins country syrian
3   0.05889   british peru yale prince image treasure gold university maqdala violin item collection return japan buyer paypal made loot peruvian
4   0.04104   mosaics aaa church bgsu learn university cathedral st de heart art letter stolen policy president mosaic turkey laurence christ
5   0.08553   egyptian antiquities egypt hawass objects committee state charges golan revolution red list cairo council minister sca zahi artifacts international
6   0.19981   museum art getty objects antiquities museums italy princeton returned italian university potts director looted dealer provenance collection york turkey
7   0.21535   cultural heritage antiquities looting national illicit archaeologists archaeological research project trade sites international museums university support local german community
8   0.20729   police antiquities coins found ancient artifacts year afghanistan greek coin market arrested artefacts century greece documents including made jewish
9   0.38299   museum stolen greece art theft ministry police culture archaeological ancient country security national government antiquities museums thieves items year

15 TOPICS

0   0.04798   getty museum potts statue art ancient policy true greek italian museums aphrodite kimbell goddess paul de roman director canavesi
1   0.11286   greece police greek museum ministry culture olympia ancient year gallery athens security antiquities coins crisis national museums cuts bronze
2   0.24225   sites archaeological site looting ancient egypt people city heritage el archaeologists digging damage antiquities hibeh security area ruins destroyed
3   0.04027   mosaics iraq university bgsu iraqi peru art yale ancient plates turkey mosaic hooper langin picchu machu ms united mosul

4   0.05378   show sites archaeologists german project nigeria archaeological nok archaeology gundu nigerian looting national university st tv american state augustine
5   0.0594   egyptian egypt hawass antiquities list revolution council red minister objects zahi sca museum foreign national state mubarak cairo emergency
6   0.05607   syria aaa syrian learn church policy letter cathedral heart army residents french gangs laurence funding citadel palmyra statement president
7   0.02887   golan ossuary jade henan fake pa wn court village jesus violin high trial fraud buyer image shabti shabtis paypal
8   0.53089   antiquities cultural country looted museums heritage international illicit trade year archaeological world years objects time director state government long
9   0.07515   statue sotheby cambodia auction cambodian illegally american malter mask federal sell khmer year york ker koh court coin law
10   0.04974   afghanistan mali jewish documents committee afghan sh jerusalem material mes aynak community israel covers resources development poverty region local
11   0.05921   british treasure coins gold japan loot return maqdala coin ebay king item spain silver family tewodros ethiopian emperor spanish
12   0.11227   museum art italy objects princeton university antiquities returned italian york turkey museums almagi dealer works acquired collections collection agreement
13   0.06453   china chinese metal art market heritage crime libya benghazi summer libyan palace theft buildings relics bank historic crimes gunter
14   0.49597   museum artifacts stolen found ancient items pieces made history art theft war years century return government ago including collection

20 TOPICS

0   0.09935   greece museum greek culture olympia ministry security police gallery athens archaeological national art crisis theft guard cuts museums guards
1   0.05343   sites syria iraq archaeological syrian iraqi heritage libya benghazi bank al department city libyan country looting plates museums antiquities
2   0.0262   art china chinese market jade paintings village henan summer palace relics bronze dynasty high gunter sales wei town goods
3   0.05921   show st sites artifacts american church national treasure archaeology history archaeological tv augustine state geographic cathedral diggers heart historic
4   0.08827   theft art metal thieves stolen heritage police de damage museum montreal buildings information reward official crimes crime arts thefts
5   0.08529   museum art objects italy princeton university italian antiquities returned dealer turkey museums york almagi met works collections investigation collecting

6   0.09244   police coins greek antiquities found artefacts arrested greece gang thessaloniki confiscated ring allegedly bronze suspected coin update year small
7   0.70974   museum antiquities artifacts ancient stolen years year back items made time country history pieces found government looted people century
8   0.05851   coin golan documents jerusalem court antiquities jewish ossuary trial israel accused jesus judge authority charges manuscripts weiss afghanistan james
9   0.03451   malter pa wn treasure spanish shabti shabtis american santo osiris spain park native national glyph miw dog court coins
10   0.05331   statue sotheby cambodia auction cambodian government art mask khmer states york koh ker case sell private united complaint stolen
11   0.05679   egyptian egypt hawass antiquities minister objects council red list state zahi revolution sca foreign cairo mubarak emergency al charges
12   0.02926   mosaics university bgsu images turkey mosaic image hooper langin frescoes zeugma qtip molholt purchase ms roman turkish analysis antioch
13   0.03833   british loot war gold maqdala japan garbage return tewodros ethiopian emperor sand sir britain treasure parthenon secret treasures family
14   0.02889   afghanistan item violin afghan tsirogiannis mes aynak buyer paypal ebay national prince playing shipping massoudi kabul scta bid cards
15   0.01696   nigeria german nok gundu nigerian project archaeologists mali looting university ncmm state dr germans archaeological sh aan africa local
16   0.27828   cultural heritage antiquities trade illicit sites archaeological museums looting excavation looted dealers objects future property international support past countries
17   0.01728   aaa committee peru learn yale policy research funding president letter picchu machu public inca peruvian report statement human recommendations
18   0.17759   site archaeological sites looting egypt people ancient hibeh digging el area antiquities security police archaeologists dr stop redmount tombs
19   0.04109   getty potts museum statue art policy greek aphrodite director museums true goddess kimbell paul canavesi exhibition los roman angeles

Result

We cannot make any statements based on the results from the Topic Modelling exercise to answer any of the questions. However, we get a fair idea on the countries, artifacts that were mentioned in the text. In addition to that, The Dirichlet parameter also provides us a way to compare the accuracy of topics relative to each other.

It could also be observed that as the number of topics increases the results also have improved, hence we can vary the number of topics based the available run time to get the best model for each case. Also on observing the result we can see that the tool is able to pick up the

words that occur frequency but fails to arrange them in an order that makes sense to the end user.

To be able to work on this tool the end user should be able to setup Java, install the Mallet tool at the appropriate location and be able to work on the command prompt. So, for a non-tech-savvy person, installing the software would be the first challenge and then running the code on the command prompt would be another challenge. Some of these issues could be avoided by setting up a user friendly Interface for the system.