Clustering Wikipedia Articles

Lane Aasen

Department of Computer Science University of Washington Seattle, WA 98105 aaasen@cs.washington.edu

Abstract

Clustering Wikipedia articles using unsupervised learning techniques including K-Means and Latent Dirichlet Allocation (LDA). Exploration into K-Means including choice of K, K-Means++, the effect of random seeds, and TF-IDF thresholds.

1 Dataset

The provided dataset contains 15,903 Wikipedia articles in tf-idf format. Most of the articles are in English, but some are not. Sorting the documents by language would be an interesting project in itself! There are 10,574 unique words in this dataset. Each document is represented as a sparse vector with one dimension for each word. Stop words have been removed from the dataset, but uncommon words remain.

2 K-Means Clustering

My first goal was to implement a basic K-Means clustering algorithm from scratch. The code is available at https://github.com/aaasen/wiki-cluster.

2.1 Choosing K

2.1.1 Minimizing Distortion

Given K clusters $C_1, C_2, ..., C_K$ where each cluster is a set of document vectors and μ_i is the centroid of C_i , the total distortion is defined as follows:

$$\sum_{i=1}^{K} \sum_{d \in C_i} ||d - \mu_i||^2$$

To minimize the distortion, we could set K equal to the number of documents, but then the clusters would be meaningless. We want to choose a K with low distortion that also results in interpretable clusters. Figure 1 shows a plot of K versus total distortion. When $1 \le K \le 16$, adding additional clusters has a large impact on the distortion, but once K > 16, adding additional clusters has little impact on the distortion. From this alone, it makes sense to set K = 16 since it provides a good balance of distortion and interpretability.

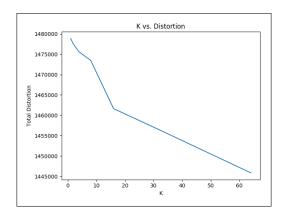


Figure 1: K versus total distortion for $K \in \{1, 2, 4, ..., 256\}$

2.2 K and Cluster Size

As K increases, the clusters become more sparse. Once K=256, over half of the clusters have only one document, and are essentially useless. When K=16, the median cluster size is 8.5, and the cluster sizes are as follows:

$$[10061, 3013, 1128, 909, 707, 30, 23, 13, 4, 4, 3, 2, 2, 2, 1, 1]$$

Over half of the clusters are very small, and one of the clusters is too large to be interpretable. This indicates that the data has significant outliers and may lack a structure conducive to clustering.

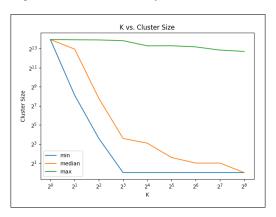


Figure 2: K versus minimum, median, and maximum cluster size for $K \in \{1, 2, 4, ..., 256\}$ with a log_2 scale on both axes.

2.3 Exploring Clusters

Table 1 shows the clusters with at least 10 documents for K-Means clustering with K=16. The words in each cluster are the dimensions of the centroid with the largest magnitude. The documents shown are those that are closest to the centroid of the cluster.

Overall, the generated clusters make sense, but there are some points of confusion:

- The words that make up cluster 0 have little relation to each other. This cluster contains the majority of the documents.
- Cluster 1 contains churches as well as colleges.
- Cluster 3 contains documents related to TV shows and sports because both contain the word "season."

2.4 K-Means++

K-Means++ is a method for selecting initial cluster centers for K-Means. In the normal version of K-Means, initial cluster centroids are selected by sampling random documents from the dataset. This can produce suboptimal clusters, increase time to convergence, and increase variability in clustering performance. K-Means solves this by selecting initial centroids one at a time to minimize distortion.

I implemented K-Means++ and expected it to increase clustering performance considerably. However, I found that it actually had a slightly negative impact on distortion and noticeably increased cluster sparsity.

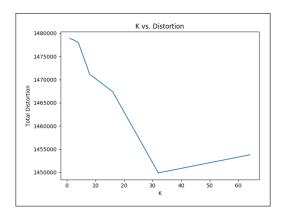


Figure 3: K versus total distortion for $K \in \{1, 2, 4, ..., 64\}$ using K-Means++.

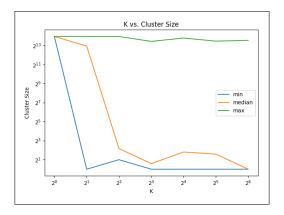


Figure 4: K versus minimum, median, and maximum cluster size for $K \in \{1, 2, 4, ..., 64\}$ with a log_2 scale on both axes. Using K-Means++.

2.5 Fighting Cluster Sparsity

After implementing K-Means++ with no success, it became apparent that there was a problem with the underlying data, and not with the clustering algorithm.

I began to investigate the provided dataset, which was in tf-idf format with no documentation. I found that stop words had been removed from the dataset, but rare words had been left in. Half of the words in the dictionary were used in 16 or fewer datasets (less than 0.1%). This was the obvious cause of cluster sparsity in the dataset.

2.5.1 Selecting a Threshold

Since the provided dataset does not contain stop words, I decided not to filter out common words. The most common words in the dataset appear to contain useful information for clustering.

Table 1: 10 least common words with the number and percentage of documents that they appear in.

Word	Documents	Percentage of Documents
frazioni	0	0.0%
threeletter	1	0.0062881%
budjovice	1	0.0062881%
gmina	1	0.0062881%
ortsgemeinden	1	0.0062881%
headwater	2	0.012576%
baronetage	3	0.018864%
breaststroke	3	0.018864%
voronezh	3	0.018864%
rosettes	3	0.018864%

Table 2: 10 most common words with the number and percentage of documents that they appear in.

Word	Documents	Percentage of Documents
well	4009	25.209%
second	3538	22.247%
high	2836	17.833%
family	2579	16.217%
group	2412	15.167%
north	2364	14.865%
major	2298	14.45%
large	2227	14.004%
general	2187	13.752%
long	2164	13.607%

Selecting a lower threshold involves finding the right balance between information loss and cluster sparsity. Words that are only used in one document should be ignored, since they provide no useful clustering information. The problem gets murkier with words that are used in just a few datasets. Table 3 lists ten words near the median, which is words that appear in 16 datasets. These words appear to have enough value for clustering that they should not be excluded. One of the large problems here is that we are working with only 15,000 of the 5 million English Wikipedia articles, which makes it difficult to produce fine-grained clusters.

2.6 Effect of Seed on Distortion

After observing some unexpected variability in my results, I decided to investigate the effect of random seeds on distortion.

I found that the choice of seed can have a large impact on cluster quality and distortion. Because of this, it is very important to run K-Means with at least a few different seeds and select the best one. I opted to not do this in the above experiments because it would have been too computationally expensive, but I will try many different seeds when generating the final clusters.

3 Conclusion

3.1 Further Research

One of the core assumptions of K-Means is that the data can be grouped into spherically symmetric clusters of the same size. This is quite a strong assumption, and it does not seem to be true in the case of Wikipedia articles. It would be interesting to experiment with other clustering techniques better suited to high dimensional text data.

Table 3: 10 words near the median.

Word	Documents Percentage of Documents	
multimillion	16	0.10061%
pisa	16	0.10061%
pranks	16	0.10061%
pesticides	16	0.10061%
hesitation	16	0.10061%
convection	16	0.10061%
ortiz	16	0.10061%
stagnation	16	0.10061%
gonzlez	16	0.10061%
cummins	16	0.10061%

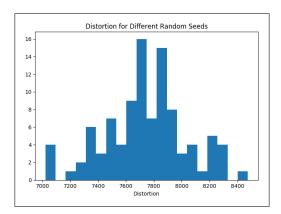


Figure 5: Distortion using K-Means++ on 100 training documents with different random seeds.

4 Next Steps

4.1 Recursive K-Means

One of the largest issues with applying K-Means to this dataset is that it produces clusters with huge size variations. Some clusters contain 10,000 documents, whereas others contain only one. Recursively applying K-means to large clusters could address this problem and even produce a sort of hierarchical clustering.

4.2 Latent Dirichlet Allocation

Some of the clusters contain documents that refer to difference meanings of the same word. For example, "season" could refer to a football season or a TV show season. I would like to explore Latent Dirichlet Allocation and whether or not it could help with this situation.

5 Conclusion

Table 4: K-Means clusters with K=16 and at least 10 documents.

Cluster	Size	Words	Documents
0	10061	females	mcgillpainquestionnaire
		station	historyofthefamily
		family	thetussaudsgroup
		located	nadiraactress
		north	mansfieldsummithighschool
1	3013	church	edmondscommunitycollege
		college	helderbergcollege
		students	oberlincongregationalchurch
		published	lundbyoldchurch
	1120	institute	dioceseoflimerickandkillaloe
2	1128	party	partyidentification
		served	labourfarmerparty
		general	democraticalliancesouthafrica
		member	liberaldemocratsitaly
		senate	christiancreditparty
3	909	season	dancingwiththestars
		club	davidmccracken
		playing	gilbertcurgenven
		seasons	bjsamsamericanfootball
		player	livingstonewalker
4	707	album	thegreatestdaytakethatalbum
		released	conflictingemotions
		songs	primalscream
		records	leftbacklp
5	30	rock	elisamartin
3	30	nba	kcjones
		basketball	hakeemolajuwon
		points	albertkingbasketball
		season	ballstatecardinalsmensbasketball
6	23	seasons riots	201011southfloridabullsmensbasketballteam sowetouprising
		police	1992losangelesriots
		murder	nikolaybogolepov
		captured	josephlamothe
		robbery	jenmi
7	13	congo	republicofcabinda
'	13	subtropical	brownrumpedbunting
		republic	copperbeltprovince
		zambia	leptopelisviridis
		zimbabwe	yellowthroatedpetronia
		ZIIIIUaUWE	успожиноанепрепоша

Table 5: 10 largest K-Means clusters with K=64. Using words that appear in at least 16 (0.1%) documents with K-Means++ for initializing cluster centroids.

Season S	Cluster	Size	Words	Documents
family north colimeads located mladeniiiubiofbribir students system helderbergeollege stargateschool miltonhighschoolmiltongeorgia technology govthazimuhammadmohsincollege molire oberlincongregationalchurch lundbyoldchurch stmaryschurchgrodno paris dioceseoflimerickandkillaloe party partyidentification labourfarmerparty served member served labourfarmerparty serbianliberalparty bronvenmaher democraticalliancesouthafrica bjsamsamericanfootball dancing withthestars playing seasons johnflanaganfootballer gilbertcurgenven player gilbertcurgenven general thegreatestdaytakethatalbum conflictingemotions records bornthisway rock sweetkisses frederickringer astorhousehotelshanghai 19221959 japan echinese pearl characters listofficlepisodes playbams on grock sweetkisses frederickringer atlentingemoney songs teenpop album britishpopmusic chart talkinginyoursleeperystalgaylesong bjahlahblahalbum doctor unitedstatesvlibby marsepper entomopter	0	8486	females	kotavamsa
north located mladenijiibiofbribir demondscommunitycollege system high program technology miltonhighschoolmiltongeorgia govthazimuhammadmohsincollege molire published daughter royal stmaryschurchgrodno paris dioceseoflimerickandkillaloe pariy served labourfarmerparty serolanliberalparty bronwenmaher senate democraticalliancesouthafrica bysamsamericanfootball dancingwiththestars dancidence player gilbertcurgenven piaper gilbertcurgenven primalscream records bromthisway records bromthisway rock songs album britishpopmusic chart talkinginyoursleepcrystalgaylesong system piune june investigation money astronomical imidothy polatile dancing marren party served in talkinginyoursleepcrystalgaylesong blabblablablabum marrenparty served in talkinginyoursleepcrystalgaylesong entomoper andynyman marrensvillestaticsexcandal johnlittehild unitedstates vibby marsamer marrenper astronomical			station	tornadoesof1993
located students edmondscommunitycollege helderbergcollege stargateschool miltonhighschoolmiltongeorgia govthazimuhammadmohsincollege molire oberlincongregationalchurch lundbyoldchurch stmaryschurchgrodno dioceseoflimerickandkillaloe paris dioceseoflimerickandkillaloe paris eserved labourfarmerparty servindiberalparty brownennaher senate democraticalliancesouthafrica bjsamsamericanfootball dancingwiththestars davidmccracken johnflanaganfootballer player gilbertcurgenven primalscream thegreatestdaytakethatalbum conflictingemotions bornthisway sweetkisses records papan chinese japan astorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce latkinginyoursleepcrystalgaylesong blahblahalbum doctor unitedstatesvlibby hostagefilm marsrapper entomopter			family	whitby
1 2226 students system helderbergcollege stargateschool mittonhighschoolmittongeorgia govthazimuhammadmohsincollege molifer royal gustargateschool mittonhighschoolmittongeorgia govthazimuhammadmohsincollege molifer royal stmaryschurchgrodno dioceseofimerickandkillaloe paris served labourfarmerparty member general bronvenmaher senate democraticalliancesouthafrica bisamsamericanfootball dancingwiththestars davidmcracken johnflanaganfootballer gilberteurgenven player gilberteurgenven thegreatestdaytakethatalbum conflictingemotions brothinses pearl characters listofflelepisodes popmusic teenpop album britishpopmusic talkinginyoursleepcrystalgaylesong blabblahblahalbum martensvillesatanicsexscandal june gustaren astronomical entomopter			north	colinmeads
system high stargateschool miltongeorgia govthazimuhammadmohsincollege miltonhighschoolmiltongeorgia govthazimuhammadmohsincollege moberincongregationalchurch daughter royal stmaryschurchgrodno paris dioceseoflimerickandkillaloe party served labourfarmerparty servianliberalparty servia				mladenjiiubiofbribir
high program technology govthazimuhammadmohsincollege miltonhighschoolmiltongeorgia govthazimuhammadmohsincollege molire published daughter royal stmaryschurchgrodno paris dioceseoflimerickandkillaloe strayschurchgrodno dioceseoflimerickandkillaloe paris served labourfarmerparty served labourfarmerparty serbianliberalparty bronvenmaher senate democraticalliancesouthafrica bronvenmaher season bjsamsamericanfootball dancingwiththestars davidmecracken johnflanaganfootballer gilbertcurgenven player gilbertcurgenven player gilbertcurgenven primalscream released thegreatestdaytakethatalbum conflictingemotions records bornthisway rock sweetkisses frederickringer japan chinese japana songs records listofficlepisodes 7 44 pop popmusic tenpop britishpopmusic talkinginyoursleepcrystalgaylesong blablahalabum money system martensvillesatanicsexscandal june doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter	1	2226		
program technology govthazimuhammadmohsincollege molire published daughter royal stmaryschurchgrodno paris dioceseoflimerickandkillaloe served labourfarmerparty served labourfarmerparty served bjsamsamericanfootball club dancingwithtestars davidmccracken johnflanaganfootballer player gilbertcurgenven songs records bornthisway rock sweetkisses parl imperialjapanesearmyairforce characters pince parl investigation money system june doctor intested in the pince part investigation marsapper astronomical investagelilm marsapper astronomical investagelilm marsapper astronopter intomopter intomolarity governational governational potential governational chart technological governational molecular molecular daughter moliron parts intomopter intomolarity and potential published intomoptic published intomoptic published intomoptic intomoptic intomoptic intomopter intomoptic intomoptic intomoptic intomoptic intomopter intomopter intomopter intomopter intomopter intomoptic intomopter intom			•	
technology church molire published daughter poyal stmaryschurchgrodno paris dioceseoflimerickandkillaloe party served member general bronwenmaher senate democraticalliancesouthafrica bjammamerous playing season playing season playing season player gilbertcurgenven released the grook songs records papan chinese paral characters album chart talkinginyoursleepcrystalgaylesong system minute doctor intestation intestates plum population digitalmonetarytrust andyhayman martensvillesatanicsexscandal johnlittlechild doctor unitedstatesvlibby istance and treated astronomical indicates in marsapper automorper entomopter intestations and process of the published daughter published daughter description of the published daughter description of discovered paral characters investigation marsapper astronomical intestates vibiby intended to the published daughter indicated and papan and published blandalbum digitalmonetarytrust andyhayman marsapper entomopter intestational characters astronomical intestates vibiby intended the published daughter indicated and published blandalbum digitalmonetarytrust andyhayman marsapper entomopter intended the published daughter indicated and particular daughter indicated and particular daughter indicated and particular daughter indicated and particular daughter daughte			_	
2 2085 church published daughter political daughter royal stmaryschurchgrodno dioceseoflimerickandkillaloe paris dioceseoflimerickandkillaloe paris served labourfarmerparty service democraticalliancesouthafrica bijsamsamericanfootball dancingwiththestars davidmccracken johnflanaganfootballer gilbertcurgenven primalscream thegreatestdaytakethatalbum conflictingemotions records japan chinese pearl characters listofficlepisodes palbum pritishpopmusic talkinginyoursleepcrystalgaylesong system june doctor astronomical daughter obertale daughter daughter date of the properties of the properties of the public daughter described in the properties of the proper				
published daughter royal stmaryschurchgrodno dioceseoflimerickandkillaloe 3 1001 party served labourfarmerparty serbianliberalparty bronwenmaher senate democraticalliancesouthafrica bjsamsamericanfootball dancingwiththestars playing season johnflanaganfootballer gilbertcurgenven player gilbertcurgenven prock sweetkisses records para astorhousehotelshanghai 19221959 japanese frederickringer japan astorhousehotelshanghai 19221959 jalabum britishpopmusic chart talkinginyoursleepcrystalgaylesong labum bjune johnlittlechild unitedstatesvlibby 8 27 investigation money andyhayman martensvillesatanicsexscandal june doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical				
daughter royal stmaryschurchgrodno dioceseoflimerickandkillaloe 3 1001 party partyidentification labourfarmerparty served labourfarmerparty served peneral bronwenmaher senate democraticalliancesouthafrica senate davidmccracken playing davidmccracken seasons johnflanaganfootballer player gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum songs conflictingemotions records bornthisway rock sweetkisses frederickringer japan astorhousehotelshanghai 19221959 japanesebadger pearl imperialjapanesearmyairforce characters listofficlepisodes 7 44 pop popmusic chart talkinginyoursleepcrystalgaylesong blahblahblahalbum system martensvillesatanicsexscandal june doctor unitedstatesvlibby 9 22 mars crater astronomical entomopter	2	2085		
royal paris dioceseoflimerickandkillaloe 3 1001 party served labourfarmerparty serbianliberalparty bronwenmaher senate democraticalliancesouthafrica bisamsamericanfootball dancingwiththestars davidmccracken player gilbertcurgenven player gilbertcurgenven prock sweetkisses papan astorhousehotelshanghai 19221959 japanese parl imperialjapanesearmyairforce listofficlepisodes 7 44 pop popmusic teenpop album chart talkinginyoursleepcrystalgaylesong blahblahblahalbum martensvillesatanicsexscandal june doctor unitedstatesylibby 9 22 mars crater astronomical services of the diocese per entomopter services of cost				
paris dioceseoflimerickandkillaloe party partyidentification labourfarmerparty served member serbianliberalparty bronwenmaher democraticalliancesouthafrica bjsamsamericanfootball dancingwiththestars playing seasons johnflanaganfootballer gilbertcurgenven player gilbertcurgenven player gilbertcurgenven primalscream released thegreatestdaytakethatalbum conflictingemotions songs records bornthisway rock sweetkisses frederickringer astorhousehotelshanghai19221959 japanese japane satorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce listofficlepisodes frederickringer satorhousehotelshanghai19221959 japanese japanese japanese japanese japanesearmyairforce listofficlepisodes frederickringer satorhousehotelshanghai19221959 japanese japanesearmyairforce listofficlepisodes japanesearmyairforce listofficlepis				
3 1001 party served labourfarmerparty served member general bronwenmaher senate democraticalliancesouthafrica bjsamsamericanfootball dancingwiththestars davidmccracken johnflanaganfootballer gilbertcurgenven player gilbertcurgenven primalscream released thegreatestdaytakethatalbum conflictingemotions records japanese japan astorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce listofflclepisodes listofflclepisodes lalbum britishpopmusic tenpop songs album britishpopmusic talkinginyoursleepcrystalgaylesong blablablahlalbum digitalmonetarytrust andyhayman martensvillesatanicsexscandal johnlittlechild unitedstatesvlibby hostagefilm marsrapper astronomical entomopter			1 7	
served member general serbianliberalparty bronwenmaher general bronwenmaher senate democraticalliancesouthafrica 4 932 season bjsamsamericanfootball dancingwiththestars davidmccracken playing seasons johnflanaganfootballer player gilbertcurgenven 5 640 album primalscream thegreatestdaytakethatalbum songs conflictingemotions records bornthisway rock sweetkisses 6 280 japanese frederickringer japan astorhousehotelshanghai19221959 chinese pearl imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic teenpop album britishpopmusic tentor chart talkinginyoursleepcrystalgaylesong blabblahblahalbum digitalmonetarytrust andyhayman system martensvillesatanicsexscandal june doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter			-	
member general serbianliberalparty bronwenmaher democraticalliancesouthafrica bysamsamericanfootball dancingwiththestars davidmccracken player gilbertcurgenven primalscream released thegreatestdaytakethatalbum conflictingemotions records bornthisway sweetkisses frederickringer japan astorhousehotelshanghai19221959 chinese japanese bornthisway rock sweetkisses frederickringer japan astorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce characters listofficlepisodes 7 44 pop songs teenpop album britishpopmusic chart talkinginyoursleepcrystalgaylesong blahblahblahalbum digitalmonetarytrust money andyhayman system martensvillesatanicsexscandal june doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter	3	1001		
general senate democraticalliancesouthafrica 4 932 season bjsamsamericanfootball dancingwiththestars davidmccracken seasons johnflanaganfootballer gilbertcurgenven 5 640 album primalscream thegreatestdaytakethatalbum conflictingemotions records bornthisway rock sweetkisses 6 280 japanese frederickringer astorhousehotelshanghai19221959 japanesebadger pearl imperialjapanesearmyairforce characters listofflelepisodes 7 44 pop popmusic chart talkinginyoursleepcrystalgaylesong blahblahalbum football digitalmonetarytrust andyhayman system martensvillesatanicsexscandal june johnlittlechild unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter				
senate democraticalliancesouthafrica season bjsamsamericanfootball club dancingwiththestars davidmccracken seasons johnflanaganfootballer player gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum songs conflictingemotions records bornthisway rock sweetkisses frederickringer japan astorhousehotelshanghai19221959 chinese japanesebadger pearl imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic chart talkinginyoursleepcrystalgaylesong rock blahblahblam digitalmonetarytrust money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 1 22 mars crater marsrapper astronomical entomopter				
4 932 season club dancingwiththestars davidmccracken johnflanaganfootballer player gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum conflictingemotions records bornthisway rock sweetkisses frederickringer japan astorhousehotelshanghai19221959 chinese japanesebadger pearl imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic teenpop britishpopmusic chart talkinginyoursleepcrystalgaylesong rock blahblahblahalbum money andyhayman system investigation digitalmonetarytrust andyhayman system june johnlittlechild unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter			general	
club playing seasons johnflanaganfootballer gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum conflictingemotions bornthisway rock sweetkisses frederickringer japan astorhousehotelshanghai19221959 japanesebadger pearl imperialjapanesearmyairforce listofficlepisodes 7 44 pop popmusic teenpop album britishpopmusic chart talkinginyoursleepcrystalgaylesong pine june johnlittlechild doctor unitedstatesvlibby 9 22 mars cater marsrapper entomopter	1	022		democraticalliancesouthafrica
playing seasons johnflanaganfootballer gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum conflictingemotions bornthisway rock sweetkisses 6 280 japanese japan astorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce listofflclepisodes 7 44 pop popmusic teenpop album britishpopmusic talkinginyoursleepcrystalgaylesong blahblahblahalbum digitalmonetarytrust andylayman system june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper entomopter	4	932		
seasons player gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum songs conflictingemotions bornthisway rock sweetkisses 6 280 japanese frederickringer japan astorhousehotelshanghai19221959 chinese japanesebadger imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic teenpop album britishpopmusic chart talkinginyoursleepcrystalgaylesong blahblahblahalbum rock listofflclepisodes 8 27 investigation money andyhayman system june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater astronomical entomopter				
player gilbertcurgenven 5 640 album primalscream released thegreatestdaytakethatalbum songs conflictingemotions records bornthisway rock sweetkisses 6 280 japanese frederickringer japan astorhousehotelshanghai19221959 chinese japanesebadger pearl imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic songs teenpop album britishpopmusic chart talkinginyoursleepcrystalgaylesong blahblahblahalbum rock blahblahblahalbum system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter			1 , 0	
5 640 album released thegreatestdaytakethatalbum conflictingemotions bornthisway rock sweetkisses 6 280 japanese frederickringer japan astorhousehotelshanghai19221959 japanesebadger pearl imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic songs album britishpopmusic chart talkinginyoursleepcrystalgaylesong rock blahblahblahalbum investigation money andyhayman system june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter				
released songs conflictingemotions bornthisway rock sweetkisses 6 280 japanese japan astorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce characters listofflclepisodes 7 44 pop popmusic songs album britishpopmusic chart talkinginyoursleepcrystalgaylesong blahblahalbum digitalmonetarytrust money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter		640		
songs records records bornthisway rock sweetkisses 6 280 japanese japan chinese pearl characters listofflclepisodes 7 44 pop songs album chart rock songs teenpop album britishpopmusic chart rock blahblahblahalbum 8 27 investigation money system june june doctor 9 22 mars rock songs records popmusic tenpop blahblahblah doctor unitedstatesvlibby 19 22 mars rock bornthisway sweetkisses frederickringer astronomical frederickringer satorhousehotelshanghai19221959 japanesebadger imperialjapanesearmyairforce listofflclepisodes popmusic teenpop britishpopmusic talkinginyoursleepcrystalgaylesong digitalmonetarytrust andyhayman system june johnlittlechild unitedstatesvlibby 9 22 mars hostagefilm marsrapper astronomical entomopter	3	040		
records rock sweetkisses frederickringer japan chinese pearl characters listofflclepisodes 7				
rock sweetkisses frederickringer japan astorhousehotelshanghai19221959 chinese japanesebadger pearl imperialjapanesearmyairforce characters listofflclepisodes 7				
6 280 japanese japan astorhousehotelshanghai19221959 chinese japanesebadger japanesearmyairforce listofflclepisodes 7 44 pop popmusic songs teenpop album chart talkinginyoursleepcrystalgaylesong rock blahblahblahalbum system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars crater astronomical entomopter				
japan astorhousehotelshanghai19221959 chinese japanesebadger pearl imperialjapanesearmyairforce characters listofflclepisodes 7	6	280		frederickringer
chinese japanesebadger imperialjapanesearmyairforce listofflclepisodes 7	Ü		l -	
pearl imperialjapanesearmyairforce listofflclepisodes 7			J 1	
characters listofflclepisodes 7			_	
7 44 pop popmusic songs teenpop album britishpopmusic chart talkinginyoursleepcrystalgaylesong rock blahblahblahalbum 8 27 investigation digitalmonetarytrust money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter			1	
songs teenpop album britishpopmusic chart talkinginyoursleepcrystalgaylesong blahblahblahalbum 8 27 investigation money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter	7	44		
album britishpopmusic chart talkinginyoursleepcrystalgaylesong rock blahblahalbum 8 27 investigation digitalmonetarytrust money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter	,			
chart talkinginyoursleepcrystalgaylesong rock blahblahalbum 8 27 investigation digitalmonetarytrust money andyhayman system june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter				
rock blahblahalbum 8 27 investigation digitalmonetarytrust money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter				
8 27 investigation money andyhayman system investigation june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater astronomical entomopter				blahblahalbum
money andyhayman system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter	8	27		digitalmonetarytrust
system martensvillesatanicsexscandal june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter				andyhayman
june johnlittlechild doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter			system	
doctor unitedstatesvlibby 9 22 mars hostagefilm crater marsrapper astronomical entomopter				johnlittlechild
9 22 mars hostagefilm crater marsrapper astronomical entomopter				
crater marsrapper astronomical entomopter	9	22		
astronomical entomopter			crater	
			astronomical	
expedition mars2			expedition	mars2
planets carbonatesonmars				