Clustering

Performing clustering on a scatter output classes would give us much lower accuracies compared to data points spreaded in small cluster. The task is to create the cluster by breaking down our document files into words

This model of Hierarchical clustering is performed in multi dimensional space hence creating a lot of space complexity as well as time complexity in our model. Better to perform in chunks with the help of pickles(recommended)

Cluster number can be changed, ideal according to hierarchical clustering in this model it should be 2 as according to the max space distance methodology but cluster number can be vary

Dependencies

- **import** os
- **import** re
- **from** collections **import** Counter
- **import** numpy **as** np
- **from** nltk.tokenize **import** word_tokenize
- **from** nltk.tokenize **import** sent_tokenize
- **from** nltk.stem **import** WordNetLemmatizer
- **from** sklearn.cluster **import** KMeans
- **from** sklearn.externals **import** joblib
- **from** sklearn.feature_extraction.text **import** TfidfVectorizer
- **from** sklearn.decomposition **import** RandomizedPCA
- **import** matplotlib.pyplot **as** plt
- **from** nltk.corpus **import** stopwords
- **from** scipy.cluster.hierarchy **import** ward, dendrogram
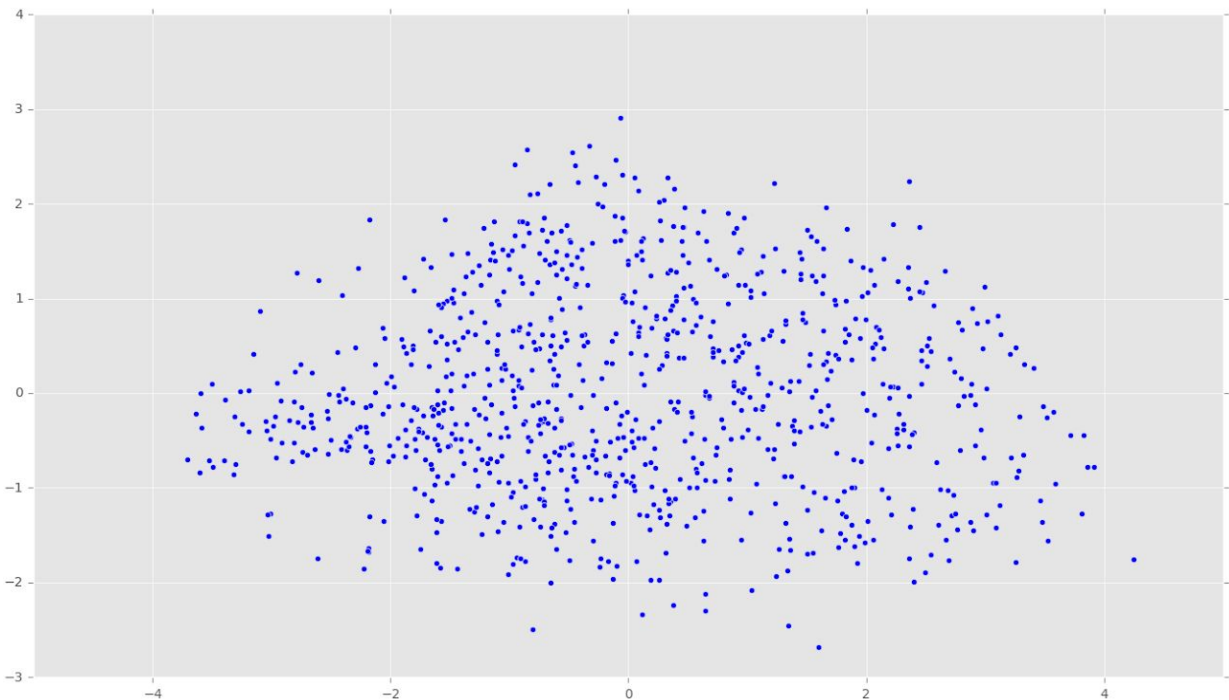
Features

- Trained model for hierarchical clustering
- Lemmatizer and tokenizer to get the proper dimension of the word in vector space
- Tfidf vectorizer used to project the words of the documents into the vector space
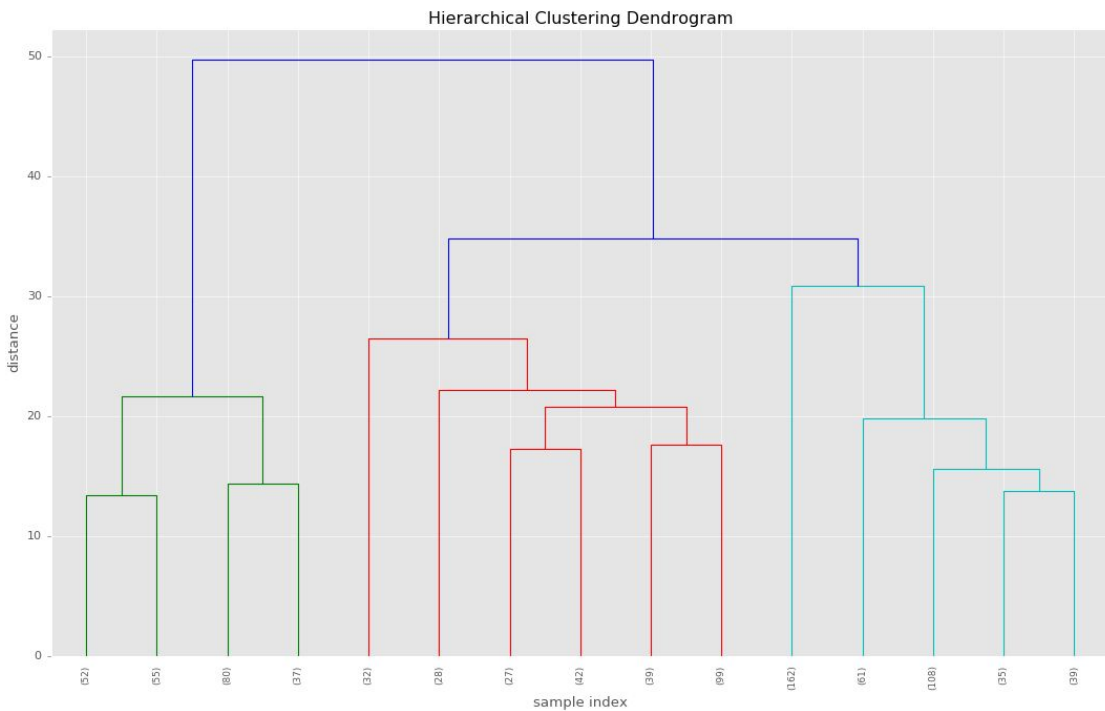- Counter used to get the most common element or tag in that particular cluster

Challenge

Technical assignment for the purpose to get select for machine learning internship at Gizbel.

Datasets

Provided by careers@gizbel.com through mail
(scatter datasets, drawn by reducing the dimension using PCA)

Results

Hierarchical Clustering Dendrogram

Results can vary as taking max space distance into consideration if we take distance as 30 we can see that in dendrogram ,horizontal line(at d=30) in graph cut-cross 4 vertical points hence 4 clusters, similarly various number of cluster can be taken depending upon which distance line chosen.

Taking threshold distance as 30 we got 4 clusters :

- 2: 267 (cluster no 2 contains 267 documents)

- 4: 243 (cluster no 4 contains 243 documents)

- 1: 224 (cluster no 1 contains 224 documents)

- 3: 162 (cluster no 3 contains 162 documents)

Tags / Common elements of each Clusters

- '4' : ['source', 'element', 'side', 'sound', 'emitting', 'surface', 'acoustic', 'de', 'slide', 'unit', 'drive', 'e', 'one', 'rod', 'r'],

- '3' : ['stress', 'effective', 'figure', 'pressure', 'pore', 'data', 'velocity', 'computer', 'environment', 'component', 'geomechanical', 'method', 'target', 'system', 'mean'], '

- 2' : ['de', 'moment', 'event', 'seismic', 'magnitude', 'sismique', 'la', 'evenements', 'formation', "d'un", 'sont', 'dans', 'un', 'le', 'partir']

- '1' : ['data', 'wavefield', 'survey', 'one', 'dip', 'least', 'seismic', 'structure', 'implementation', 'x', 'plane', 'using', 'according', 'subsurface', 'multicomponent']

Cluster 1 files :

['WO9729390A1.txt', 'WO16025030A1.txt', 'WO15167818A1.txt', 'WO16032566A1.txt', 'WO15158620A1.txt', 'WO15155595A1.txt', 'WO15085022A1.txt', 'WO15061098A1.txt', 'WO9820368A1.txt', 'WO15065952A1.txt', 'WO15138923A2.txt', 'WO9810312A1.txt', 'WO16001752A2.txt', 'WO15101832A2.txt', 'WO16075550A1.txt', 'WO16053326A1.txt', 'WO15077171A2.txt', 'WO16053945A1.txt', 'WO15120461A1.txt', 'WO15102754A1.txt', 'WO15104052A1.txt', 'WO16012826A1.txt', 'WO15061636A1.txt', 'WO16064845A1.txt', 'WO9522066A1.txt', 'WO15179130A1.txt', 'WO15143195A1.txt', 'WO9301508A1.txt', 'WO15153602A1.txt', 'WO15082421A1.txt', 'WO15102721A1.txt', 'WO9534037A1.txt', 'WO15118414A2.txt', 'WO9733184A1.txt', 'WO16094618A1.txt', 'WO16063124A1.txt', 'WO9825161A2.txt', 'WO9722077A1.txt', 'WO15134090A1.txt', 'WO15121755A2.txt', 'WO16011250A1.txt', 'WO15106065A1.txt', 'WO9917473A1.txt', 'WO16075538A1.txt', 'WO9819182A1.txt', 'WO15178789A1.txt', 'WO15123641A1.txt', 'WO16048852A1.txt', 'WO15159152A2.txt', 'WO16089878A1.txt', 'WO15157084A1.txt', 'WO16124878A1.txt', 'WO16027156A1.txt', 'WO15110912A2.txt', 'WO16042374A1.txt', 'WO9739367A1.txt', 'WO16076917A1.txt', 'WO9836292A2.txt', 'WO15121200A2.txt', 'WO15134910A1.txt', 'WO16005597A1.txt', 'WO15104639A1.txt', 'WO16083892A2.txt', 'WO16090031A1.txt', 'WO16038466A1.txt', 'WO9711390A2.txt', 'WO15104637A2.txt', 'WO16025028A1.txt', 'WO9708570A1.txt', 'WO15199757A1.txt', 'WO16133951A1.txt', 'WO9931528A1.txt', 'WO15127079A1.txt', 'WO15159151A2.txt', 'WO9820367A1.txt', 'WO15187136A1.txt',

'WO16005815A2.txt', 'WO15128732A2.txt', 'WO16079593A1.txt', 'WO15084429A1.txt',
'WO9508782A1.txt', 'WO15175053A2.txt', 'WO15102705A2.txt', 'WO15177631A2.txt',
'WO15118410A2.txt', 'WO15187150A1.txt', 'WO15136364A2.txt', 'WO9700485A1.txt',
'WO9635966A1.txt', 'WO15095248A1.txt', 'WO9733183A1.txt', 'WO9909434A1.txt',
'WO16060710A1.txt', 'WO15094458A1.txt', 'WO9323771A1.txt', 'WO15199800A1.txt',
'WO16001750A2.txt', 'WO9217798A2.txt', 'WO15120353A2.txt', 'WO15124961A2.txt',
'WO16089892A1.txt', 'WO16064462A1.txt', 'WO9632656A1.txt', 'WO9720235A1.txt',
'WO16037022A1.txt', 'WO16055826A1.txt', 'WO15077170A1.txt', 'WO9620417A1.txt',
'WO15104640A2.txt', 'WO15063595A1.txt', 'WO15108859A1.txt', 'WO16064607A1.txt',
'WO15112622A1.txt', 'WO9713166A1.txt', 'WO16014466A1.txt', 'WO16090033A1.txt',
'WO9738330A1.txt', 'WO9919749A1.txt', 'WO15082010A1.txt', 'WO15124960A1.txt',
'WO9726558A1.txt', 'WO15164388A1.txt', 'WO9744751A1.txt', 'WO15187628A2.txt',
'WO15168114A1.txt', 'WO15171192A1.txt', 'WO15132663A1.txt', 'WO15150728A1.txt',
'WO15073487A1.txt', 'WO9913357A1.txt', 'WO9931760A1.txt', 'WO16004157A1.txt',
'WO9742526A1.txt', 'WO16001041A1.txt', 'WO9821559A2.txt', 'WO16062710A1.txt',
'WO9633425A1.txt', 'WO9741456A1.txt', 'WO16011236A1.txt', 'WO9420864A1.txt',
'WO16065356A1.txt', 'WO15058177A1.txt', 'WO9607935A1.txt', 'WO16005598A1.txt',
'WO15118409A2.txt', 'WO16054008A1.txt', 'WO16028520A1.txt', 'WO15145256A2.txt',
'WO15089133A1.txt', 'WO16010588A1.txt', 'WO15102498A1.txt', 'WO15167646A1.txt',
'WO15136379A2.txt', 'WO15144453A1.txt', 'WO15193695A1.txt', 'WO15112746A1.txt',
'WO9843189A1.txt', 'WO16061293A1.txt', 'WO15092542A2.txt', 'WO15097556A2.txt',
'WO9903004A1.txt', 'WO15159149A2.txt', 'WO15063597A1.txt', 'WO15187396A2.txt',
'WO15078842A1.txt', 'WO16063123A2.txt', 'WO16063125A1.txt', 'WO15101831A1.txt',
'WO9200532A1.txt', 'WO16089884A1.txt', 'WO15125015A2.txt', 'WO15168130A1.txt',
'WO15191971A1.txt', 'WO15082419A1.txt', 'WO15185991A2.txt', 'WO16011581A1.txt',
'WO16090354A1.txt', 'WO16134210A1.txt', 'WO15171215A1.txt', 'WO16071728A1.txt',
'WO9928767A1.txt', 'WO15102722A1.txt', 'WO16046633A1.txt', 'WO15063444A1.txt',
'WO9818022A1.txt', 'WO15175780A1.txt', 'WO15065651A1.txt', 'WO16071769A1.txt',
'WO16036556A1.txt', 'WO15106879A1.txt', 'WO9815849A1.txt', 'WO9408256A1.txt',
'WO9713213A1.txt', 'WO16038458A2.txt', 'WO15160652A1.txt', 'WO9313434A1.txt',
'WO15065602A1.txt', 'WO15104638A2.txt', 'WO16005784A1.txt', 'WO15159150A2.txt',
'WO15143189A1.txt', 'WO15130441A1.txt', 'WO16030432A1.txt', 'WO15132662A1.txt',
'WO15145260A2.txt', 'WO9903054A1.txt', 'WO9837437A1.txt', 'WO16064483A1.txt',
'WO15092540A2.txt', 'WO15108862A1.txt', 'WO15153215A1.txt', 'WO15137821A1.txt',
'WO9618915A1.txt', 'WO9634301A1.txt', 'WO15187208A1.txt', 'WO16023598A1.txt',
'WO15097557A2.txt', 'WO16048194A1.txt', 'WO9705558A1.txt', 'WO15134379A1.txt',
'WO9926085A1.txt', 'WO16055867A2.txt', 'WO9906855A1.txt', 'WO15164478A1.txt']

Cluster 2 files :

['WO9013831A1.txt', 'WO16020500A1.txt', 'WO9714062A1.txt', 'WO15125019A3.txt',
'WO15170170A2.txt', 'WO9703370A3.txt', 'WO15136379A3.txt', 'WO9803885A1.txt',
'WO15145257A3.txt', 'WO9923350A1.txt', 'WO9304383A1.txt', 'WO15138923A3.txt',
'WO16003292A1.txt', 'WO15055904A1.txt', 'WO9804933A1.txt', 'WO9414086A1.txt',
'WO9815713A1.txt', 'WO9218377A1.txt', 'WO9114954A1.txt', 'WO15159000A2.txt',
'WO15097557A3.txt', 'WO9910620A1.txt', 'WO9609562A1.txt', 'WO15139706A1.txt',
'WO9812667A3.txt', 'WO15137916A1.txt', 'WO9218883A1.txt', 'WO15168114A9.txt',
'WO15161892A1.txt', 'WO9412896A2.txt', 'WO9401782A1.txt', 'WO9706452A3.txt',
'WO15108920A1.txt', 'WO9302371A1.txt', 'WO9750007A3.txt', 'WO15159152A3.txt',
'WO16064280A1.txt', 'WO9818293A1.txt', 'WO15167340A1.txt', 'WO15097556A3.txt',
'WO15161326A3.txt', 'WO9830942A1.txt', 'WO9503921A1.txt', 'WO9303402A1.txt',
'WO15102498A9.txt', 'WO15110912A3.txt', 'WO9738775A1.txt', 'WO9739366A1.txt',
'WO9812667A2.txt', 'WO15143480A1.txt', 'WO9748523A1.txt', 'WO15161326A9.txt',
'WO15076670A2.txt', 'WO9803886A1.txt', 'WO15155597A3.txt', 'WO9306311A1.txt',
'WO15185991A3.txt', 'WO15055905A1.txt', 'WO15142836A1.txt', 'WO15108922A1.txt',
'WO15173371A1.txt', 'WO15108865A3.txt', 'WO15077171A3.txt', 'WO15104638A3.txt',
'WO16011629A1.txt', 'WO16005815A3.txt', 'WO9620414A1.txt', 'WO9404944A3.txt',
'WO16028189A1.txt', 'WO15094582A3.txt', 'WO15123137A1.txt', 'WO9700451A1.txt',
'WO15177631A3.txt', 'WO15145260A3.txt', 'WO9603255A1.txt', 'WO16061422A1.txt',
'WO9219525A1.txt', 'WO15191087A1.txt', 'WO9503920A1.txt', 'WO9923351A1.txt',
'WO15138922A2.txt', 'WO15142815A1.txt', 'WO16055867A3.txt', 'WO9832034A1.txt',
'WO15136378A3.txt', 'WO15152521A1.txt', 'WO15177455A1.txt', 'WO15119781A9.txt',
'WO15138024A3.txt', 'WO15189415A2.txt', 'WO9104544A1.txt', 'WO9719370A3.txt',
'WO9923465A2.txt', 'WO16024026A1.txt', 'WO9524977A3.txt', 'WO15138325A1.txt',
'WO9621166A1.txt', 'WO15127939A1.txt', 'WO15193465A3.txt', 'WO9624861A1.txt',
'WO15092342A1.txt', 'WO15104633A3.txt', 'WO16026874A2.txt', 'WO9923465A3.txt',
'WO15063210A3.txt', 'WO9711394A2.txt', 'WO9602855A1.txt', 'WO15103608A1.txt',
'WO9524977A2.txt', 'WO9503914A1.txt', 'WO15125014A3.txt', 'WO15092540A3.txt',
'WO9740406A1.txt', 'WO15175053A3.txt', 'WO16054133A2.txt', 'WO9711391A1.txt',
'WO15138922A3.txt', 'WO15162479A1.txt', 'WO16046406A1.txt', 'WO9711392A1.txt',
'WO9626625A1.txt', 'WO9832033A1.txt', 'WO15200240A1.txt', 'WO9519241A1.txt',
'WO9218705A1.txt', 'WO9503915A1.txt', 'WO9212443A1.txt', 'WO9416471A1.txt',
'WO9312443A1.txt', 'WO15138485A3.txt', 'WO9301899A1.txt', 'WO9503913A1.txt',
'WO15142807A1.txt', 'WO9409385A1.txt', 'WO15124961A3.txt', 'WO16020540A1.txt',
'WO16066721A1.txt', 'WO9406033A1.txt', 'WO9212442A1.txt', 'WO16001750A3.txt',
'WO15174849A1.txt', 'WO9711390A3.txt', 'WO9744158A2.txt', 'WO15055377A1.txt',
'WO15083000A3.txt', 'WO9419707A1.txt', 'WO9506828A1.txt', 'WO16020554A1.txt',
'WO15125010A3.txt', 'WO15140644A3.txt', 'WO15159151A3.txt', 'WO9906267A1.txt',
'WO15159149A3.txt', 'WO16073483A1.txt', 'WO15159150A3.txt', 'WO9511110A1.txt',
'WO16093699A1.txt', 'WO16012191A1.txt', 'WO16012059A1.txt', 'WO9711393A1.txt',
'WO15187396A3.txt', 'WO15057607A1.txt', 'WO9711394A3.txt', 'WO15121755A3.txt',
'WO16051263A1.txt', 'WO15076670A3.txt', 'WO15169860A3.txt', 'WO9745756A1.txt',
'WO15128732A3.txt', 'WO9305411A1.txt', 'WO15104637A3.txt', 'WO15136364A3.txt',

'WO15189415A3.txt', 'WO9836292A3.txt', 'WO16026874A3.txt', 'WO16054133A3.txt',
'WO15159000A3.txt', 'WO16014883A3.txt', 'WO15108921A1.txt', 'WO15121614A1.txt',
'WO9730361A1.txt', 'WO15193322A1.txt', 'WO15101832A3.txt', 'WO16073720A1.txt',
'WO9832646A1.txt', 'WO15104636A3.txt', 'WO15126889A3.txt', 'WO9740434A1.txt',
'WO16063123A3.txt', 'WO9613788A1.txt', 'WO9819180A1.txt', 'WO15108919A1.txt',
'WO9921399A1.txt', 'WO9409386A1.txt', 'WO9106877A1.txt', 'WO16020554A8.txt',
'WO15189692A3.txt', 'WO15143546A1.txt', 'WO9709226A1.txt', 'WO15170170A3.txt',
'WO9738803A1.txt', 'WO15067864A1.txt', 'WO9503916A1.txt', 'WO9515506A1.txt',
'WO15138198A1.txt', 'WO9832309A1.txt', 'WO9524608A1.txt', 'WO15119647A1.txt',
'WO16001752A3.txt', 'WO9744158A3.txt', 'WO15092542A3.txt', 'WO9325919A1.txt',
'WO9710461A1.txt', 'WO9849049A1.txt', 'WO15145256A3.txt', 'WO9630782A1.txt',
'WO15104641A3.txt', 'WO15133908A1.txt', 'WO15125015A3.txt', 'WO9852072C2.txt',
'WO15145261A3.txt', 'WO15195532A3.txt', 'WO15148672A9.txt', 'WO15118410A3.txt',
'WO9524659A1.txt', 'WO9404944A2.txt', 'WO9501578A1.txt', 'WO15189483A1.txt',
'WO15189692A2.txt', 'WO16020542A1.txt', 'WO9217798A3.txt', 'WO15119781A3.txt',
'WO15118409A3.txt', 'WO16115437A1.txt', 'WO15104640A3.txt', 'WO15127237A1.txt',
'WO15092414A1.txt', 'WO9918443A1.txt', 'WO15100401A1.txt', 'WO9602854A1.txt',
'WO15118414A3.txt', 'WO16011670A1.txt', 'WO9707017A1.txt', 'WO15121200A3.txt',
'WO16028973A1.txt', 'WO9835244A1.txt', 'WO9414571A1.txt', 'WO9623234A1.txt',
'WO9921398A1.txt', 'WO16012857A3.txt', 'WO9200530A1.txt', 'WO9711395A3.txt',
'WO15168413A1.txt', 'WO9919750A1.txt', 'WO9827444A3.txt', 'WO9711395A2.txt',
'WO9904293A1.txt', 'WO9503912A1.txt', 'WO16074075A1.txt', 'WO15094582A2.txt',
'WO15171192A8.txt', 'WO9428439A1.txt', 'WO16010606A1.txt', 'WO15142841A1.txt',
'WO15102705A3.txt', 'WO15148672A3.txt', 'WO15120353A3.txt']

Cluster 3 Files :

['WO15177653A3.txt', 'WO15152902A1.txt', 'WO16115304A1.txt', 'WO15177653A2.txt',
'WO15102632A1.txt', 'WO15134515A1.txt', 'WO16122611A1.txt', 'WO16050942A1.txt',
'WO16051292A1.txt', 'WO16039773A1.txt', 'WO15063010A1.txt', 'WO15167553A1.txt',
'WO16041189A1.txt', 'WO15099773A1.txt', 'WO16060645A1.txt', 'WO16070073A1.txt',
'WO16076847A1.txt', 'WO9750007A2.txt', 'WO16007170A1.txt', 'WO15108865A2.txt',
'WO15127349A1.txt', 'WO15126889A2.txt', 'WO15117118A1.txt', 'WO16018426A1.txt',
'WO16126453A1.txt', 'WO15126423A1.txt', 'WO15168417A1.txt', 'WO15107079A1.txt',
'WO15195484A1.txt', 'WO15148666A1.txt', 'WO16137888A1.txt', 'WO15102508A1.txt',
'WO15157186A1.txt', 'WO16014358A1.txt', 'WO16057071A1.txt', 'WO15149237A1.txt',
'WO16080980A1.txt', 'WO16064476A1.txt', 'WO16085442A1.txt', 'WO16018229A1.txt',
'WO16069706A1.txt', 'WO15195129A1.txt', 'WO16032353A1.txt', 'WO15108641A1.txt',
'WO16001697A1.txt', 'WO15080739A1.txt', 'WO15059058A1.txt', 'WO15102791A1.txt',
'WO16076746A1.txt', 'WO16075231A1.txt', 'WO15167502A1.txt', 'WO15094306A1.txt',

'WO15103218A9.txt', 'WO16093794A1.txt', 'WO16044538A1.txt', 'WO15199759A1.txt',
'WO9700389A1.txt', 'WO16007808A1.txt', 'WO16036359A1.txt', 'WO16003786A1.txt',
'WO16115194A1.txt', 'WO16036411A1.txt', 'WO16011585A1.txt', 'WO16083861A1.txt',
'WO16085926A1.txt', 'WO16108872A1.txt', 'WO16111685A1.txt', 'WO15058015A1.txt',
'WO16025032A1.txt', 'WO16069322A1.txt', 'WO15138807A1.txt', 'WO15112529A1.txt',
'WO16115004A1.txt', 'WO16033435A1.txt', 'WO15102792A1.txt', 'WO16060641A1.txt',
'WO15138024A2.txt', 'WO16126762A1.txt', 'WO15099766A1.txt', 'WO16126759A1.txt',
'WO16028564A1.txt', 'WO15112449A1.txt', 'WO16022106A1.txt', 'WO15061266A1.txt',
'WO16025672A1.txt', 'WO15099779A1.txt', 'WO15076862A1.txt', 'WO15088878A1.txt',
'WO15152903A1.txt', 'WO15188951A1.txt', 'WO15102671A1.txt', 'WO16126761A1.txt',
'WO16099488A1.txt', 'WO16043980A1.txt', 'WO16115197A1.txt', 'WO15148672A2.txt',
'WO15142706A2.txt', 'WO15081103A1.txt', 'WO16018723A1.txt', 'WO15142711A1.txt',
'WO16114881A1.txt', 'WO15101829A1.txt', 'WO16014377A3.txt', 'WO15103494A1.txt',
'WO16069171A1.txt', 'WO16126233A1.txt', 'WO15134505A1.txt', 'WO15073483A1.txt',
'WO15178931A1.txt', 'WO15195801A1.txt', 'WO15134186A1.txt', 'WO16073198A1.txt',
'WO16028725A1.txt', 'WO15053876A1.txt', 'WO16089835A1.txt', 'WO15193465A2.txt',
'WO16043981A1.txt', 'WO16032489A1.txt', 'WO15199727A1.txt', 'WO16022237A1.txt',
'WO16043978A1.txt', 'WO16014377A2.txt', 'WO16137519A1.txt', 'WO16134443A1.txt',
'WO16099649A1.txt', 'WO15112876A1.txt', 'WO16018464A1.txt', 'WO15142706A3.txt',
'WO15126369A1.txt', 'WO16108909A1.txt', 'WO16099764A1.txt', 'WO15188396A1.txt',
'WO15061305A1.txt', 'WO15094307A1.txt', 'WO15160347A1.txt', 'WO15171669A1.txt',
'WO16011388A1.txt', 'WO15084655A1.txt', 'WO16037168A1.txt', 'WO16011385A1.txt',
'WO16010715A1.txt', 'WO15108980A1.txt', 'WO16007169A1.txt', 'WO15103229A1.txt',
'WO15119986A1.txt', 'WO16083893A1.txt', 'WO15104633A2.txt', 'WO15080890A1.txt',
'WO9827444A2.txt', 'WO16036809A1.txt', 'WO15099563A1.txt', 'WO16056936A1.txt',
'WO15103218A1.txt', 'WO16033054A1.txt', 'WO15117116A1.txt', 'WO16093793A1.txt',
'WO16043982A1.txt', 'WO16039813A1.txt', 'WO15105851A1.txt', 'WO16069170A1.txt',
'WO15080753A1.txt', 'WO16134376A1.txt']

Cluster 4 files :

['WO15133810A1.txt', 'WO9745006A1.txt', 'WO9001712A1.txt', 'WO15175646A1.txt',
'WO9814800A1.txt', 'WO9516212A1.txt', 'WO9410585A1.txt', 'WO15071491A1.txt',
'WO9317355A1.txt', 'WO16007505A1.txt', 'WO15162493A2.txt', 'WO15145261A2.txt',
'WO9713167A1.txt', 'WO9602856A1.txt', 'WO16008105A1.txt', 'WO9916129A1.txt',
'WO9429752A1.txt', 'WO15182608A1.txt', 'WO16064812A1.txt', 'WO9721115A1.txt',
'WO15088205A1.txt', 'WO9514246A1.txt', 'WO9530911A1.txt', 'WO16048772A1.txt',
'WO9504945A1.txt', 'WO9722892A1.txt', 'WO15088466A1.txt', 'WO9741457A1.txt',
'WO9810891A1.txt', 'WO15164863A1.txt', 'WO9809183A1.txt', 'WO16032999A1.txt',
'WO15063210A2.txt', 'WO16004348A1.txt', 'WO9506263A1.txt', 'WO16076731A1.txt',

'WO15169357A1.txt', 'WO9317354A1.txt', 'WO15072883A1.txt', 'WO15125010A2.txt',
'WO15195939A1.txt', 'WO15175766A1.txt', 'WO16011586A1.txt', 'WO16054046A1.txt',
'WO15192461A1.txt', 'WO9720234A1.txt', 'WO16077915A1.txt', 'WO15057640A1.txt',
'WO16044144A1.txt', 'WO9531735A1.txt', 'WO9701770A1.txt', 'WO9725632A1.txt',
'WO9819181A1.txt', 'WO16027055A1.txt', 'WO9825162A1.txt', 'WO15117149A1.txt',
'WO15069813A1.txt', 'WO9316930A1.txt', 'WO9812577A1.txt', 'WO9620451A1.txt',
'WO15083000A2.txt', 'WO9926179A1.txt', 'WO15089064A1.txt', 'WO15063209A1.txt',
'WO15161326A2.txt', 'WO16105765A1.txt', 'WO15128734A3.txt', 'WO9718488A1.txt',
'WO9222824A1.txt', 'WO15085155A1.txt', 'WO15109175A1.txt', 'WO15161216A1.txt',
'WO9923510A1.txt', 'WO15100544A1.txt', 'WO9305410A1.txt', 'WO15145195A1.txt',
'WO16051266A1.txt', 'WO16086304A1.txt', 'WO16093875A1.txt', 'WO9741453A1.txt',
'WO9530912A1.txt', 'WO16025249A1.txt', 'WO9713164A1.txt', 'WO16068715A1.txt',
'WO16057293A1.txt', 'WO9416344A1.txt', 'WO16114955A1.txt', 'WO15133903A1.txt',
'WO9843117A1.txt', 'WO9744685A1.txt', 'WO16067111A1.txt', 'WO9719370A2.txt',
'WO9806234A1.txt', 'WO9118302A1.txt', 'WO9510787A1.txt', 'WO16011502A1.txt',
'WO15104386A1.txt', 'WO9703370A2.txt', 'WO16008100A1.txt', 'WO9106879A1.txt',
'WO16032598A3.txt', 'WO15088997A1.txt', 'WO9106878A1.txt', 'WO16081579A1.txt',
'WO15147965A3.txt', 'WO16030228A1.txt', 'WO9618117A1.txt', 'WO9821559A3.txt',
'WO15125019A2.txt', 'WO16094332A1.txt', 'WO9219989A1.txt', 'WO15121749A3.txt',
'WO16093881A1.txt', 'WO15094303A1.txt', 'WO15074566A1.txt', 'WO16085509A1.txt',
'WO9812578A1.txt', 'WO15085426A1.txt', 'WO15137932A1.txt', 'WO16085511A1.txt',
'WO15128734A2.txt', 'WO9742525A1.txt', 'WO15187742A2.txt', 'WO9825161A3.txt',
'WO9631788A1.txt', 'WO15057642A1.txt', 'WO9737246A1.txt', 'WO16022192A1.txt',
'WO16014926A1.txt', 'WO15101643A1.txt', 'WO9323749A1.txt', 'WO15162493A3.txt',
'WO16087947A1.txt', 'WO16008103A1.txt', 'WO9809182A1.txt', 'WO15113031A1.txt',
'WO15069254A1.txt', 'WO9858336A1.txt', 'WO15138485A2.txt', 'WO15169860A2.txt',
'WO16060513A1.txt', 'WO15150813A1.txt', 'WO15136057A1.txt', 'WO15104636A2.txt',
'WO9635965A1.txt', 'WO9900295A1.txt', 'WO15174848A1.txt', 'WO16055331A1.txt',
'WO16011587A1.txt', 'WO15140644A2.txt', 'WO9923509A1.txt', 'WO16063126A1.txt',
'WO16014881A1.txt', 'WO9412896A3.txt', 'WO16041185A1.txt', 'WO15128221A1.txt',
'WO15121749A2.txt', 'WO15152758A1.txt', 'WO9410586A1.txt', 'WO15085104A1.txt',
'WO15181626A1.txt', 'WO15195532A2.txt', 'WO16079592A1.txt', 'WO16071775A1.txt',
'WO15066481A1.txt', 'WO15110255A8.txt', 'WO16012857A2.txt', 'WO9513548A1.txt',
'WO9824685A1.txt', 'WO9931615A1.txt', 'WO9834186A1.txt', 'WO9822835A1.txt',
'WO15187743A1.txt', 'WO16001753A1.txt', 'WO15199683A1.txt', 'WO16011164A1.txt',
'WO15104059A1.txt', 'WO15155597A2.txt', 'WO15136378A2.txt', 'WO15065604A1.txt',
'WO15102619A1.txt', 'WO16055565A1.txt', 'WO16014995A1.txt', 'WO16009270A1.txt',
'WO15145257A2.txt', 'WO15196779A1.txt', 'WO15199684A1.txt', 'WO9912055A1.txt',
'WO15127211A1.txt', 'WO16066719A1.txt', 'WO15187312A1.txt', 'WO9848301A1.txt',
'WO15089000A1.txt', 'WO9915913A1.txt', 'WO16030508A1.txt', 'WO9853344A1.txt',
'WO15101644A1.txt', 'WO16046648A2.txt', 'WO9706452A2.txt', 'WO15177637A1.txt',
'WO9535513A1.txt', 'WO16076729A1.txt', 'WO9718489A1.txt', 'WO8800711A2.txt',
'WO15089019A1.txt', 'WO9828636A1.txt', 'WO15125014A2.txt', 'WO9423313A1.txt',

'WO16038453A1.txt', 'WO15110255A1.txt', 'WO9736084A1.txt', 'WO9718491A1.txt',
'WO16076953A1.txt', 'WO8800711A3.txt', 'WO9857195A1.txt', 'WO9811455A1.txt',
'WO15104210A1.txt', 'WO15188829A1.txt', 'WO15063211A1.txt', 'WO16032598A2.txt',
'WO16108896A1.txt', 'WO16022194A1.txt', 'WO15057639A1.txt', 'WO9422036A1.txt',
'WO15119781A2.txt', 'WO9201358A1.txt', 'WO9852072A1.txt', 'WO9720233A1.txt',
'WO15065917A1.txt', 'WO9636888A1.txt', 'WO15173642A1.txt', 'WO15147965A2.txt',
'WO16014883A2.txt', 'WO9923511A1.txt', 'WO15065517A1.txt', 'WO9612200A1.txt',
'WO9317356A1.txt', 'WO16046648A3.txt', 'WO9513549A1.txt', 'WO9211546A1.txt',
'WO15104641A2.txt', 'WO15101645A1.txt', 'WO9741454A1.txt']